Investigating the Inhibition of Endoplasmic Reticulum α-Glucosidase II for the Development of Broad-Spectrum Antivirals

Fergus Ross Bremner

Lincoln College

Thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Submission Date

Supervisior: Prof Nicole Zitzmann

Department of Biochemistry

University of Oxford

Word Count: 49,937

Abstract

Enveloped viruses hijack the protein production machinery of their host in order to proliferate and their replication/infectivity relies on the correct folding of their surface glycoproteins. This dependence on glycoprotein quality control for viral proliferation makes the host endoplasmic reticulum quality control (ERQC) mechanisms an attractive target for the development of host-targeting antiviral agents, which may show broad-spectrum activity and a high genetic barrier to resistance. A key ERQC protein is α-glucosidase II (GluII), a heterodimeric enzyme bearing a catalytic α-subunit of the GH31 family and an accessory β-subunit that is necessary for full catalytic activity and for the localisation of the heterodimer in the ER. GluII mediates quality control of the folding of nascent glycoproteins by trimming glucose residues from its substrate glycan $Glc_{1\text{-}2}Man_9GlcNAc_2$. Inhibition of GluII can be achieved using iminosugars which mimic the glucose residue of the native substrate, causing the build-up of misfolded protein in the ER and prevents the secretion of correctly folded glycoproteins. In the case of enveloped virus infection, this inhibits virion secretion or reduces infectivity of secreted virions and has been shown to elicit antiviral effect against a range of enveloped viruses *in vitro* and *in vivo*. Iminosugar inhibitors of GluII are effective, but have been associated with significant drug-related adverse effects.

In this thesis, I present a novel approach to the inhibition of GluII that involves disruption of the heterodimeric protein-protein interface within the enzyme. Removal of the accessory β-subunit from the catalytic α-subunit has been shown to decrease catalytic activity and may induce the export of the catalytic domain out of the ER. Thiopyridones were discovered as small molecules that interfere with the α/β interface and I investigate this protein ligand interaction using techniques such as mass photometry, mass spectrometry and high performance liquid chromatography among others to confirm the capability of these molecules to disrupt the α/β interface and potentially inhibit GluII. Additionally, I outline the design and synthesis of novel iminosugar compounds with the aim of improving the selectivity for GluII relative to existing best-in-class iminosugars. I also use molecular dynamics for *in silico* examination of the full extent of the GluII α/β interface and the roles of peripheral domains of the β-subunit that are not fully understood based on existing structural data.

Acknowledgements

The work presented in this thesis would not have been possible without the calm and dedicated efforts of Prof Nicole Zitzmann, whose academic and pastoral support was invaluable to me. I would like to sincerely thank Nicole for continuing to promise that there was light at the end of the tunnel, and to thank her again for being right about it. Additionally, I would like to thank Prof Raymond Dwek for his encouragement, particularly towards the end of my degree, and for fostering an inquisitive attitude in many generations of glycobiologists.

It goes without saying that the work was not conducted in isolation, but I count myself particularly lucky to have been helped (sometimes dragged) along the way by a group of special colleagues. The collaborative environment of the Zitzmann group makes it a warm and welcoming place to work. I'd like to specifically thank Drs Mario Hensen and JL Kiappes for their direct supervision in the early stages of the project and for continuing to support my research even after they moved on to new opportunities. Their openness and patience allowed me to develop as an independent researcher. I am grateful to Drs Anu Chandran and Bevin Gangadharan for making the collection of much of the presented data possible and sharing their considerable knowledge with me. For keeping me sane and grounded, I want to thank Dr Juli Brun and Snezana Vasiljevic, who provided either honest and compassionate support or a stern telling-off, as appropriate, throughout my time in Oxford. I'm glad to count them among my friends. Other members and former members of the Zitzmann group who deserve my appreciation include Yohan Arman, Priyadarshini Chatterjee, Dr Dina Foteinou, Michelle Hill and Dr Dom Alonzi, for their contribution to a fulfilling working environment.

I owe my thanks to several collaborators – To Dr Weston Struwe for introducing me to mass photometry and for providing his time in the lab (and the pub). To Dr Sean Burnap for his assistance with glycan analysis. Dr Manish Kushwah and the rest of the Kukara group for their help acquiring MP data. To Wiktoria Sadowska for her help with native mass spectrometry. To Prof Syma Khalid for her time spent developing the *in silico* analysis of GluII. To Dr Pietro Roversi for his assistance with the structural biology that helped me finalise this thesis. To Prof Christina Redfield for her assistance in the NMR facility. I greatly appreciate the time and advice of my thesis examiners, Prof Justin Benesch and Dr Norica Nichita, who facilitated a greatly rewarding discussion of my research. Furthermore, the members of my thesis committee, Profs Christina Redfield and Mark Wormald, kept me facing in the right direction in the early stages of my DPhil.

A large portion of my gratitude goes to my family and friends. To Oscar, Ethan and James for providing me with a place to cry other than my own home, and for keeping me distracted alongside Bert and Louis, with our semi-regular Dungeons & Dragons sessions which provided exactly the escape I needed. I will again thank Juli Brun for her friendship, as she will be offended if she is not mentioned here at least twice. I am lucky that my list of friends is too long to complete here, but they know who they are and I am glad to know and love them. To Charlie, whom I consider to be comfortably in the top one people alive today. Your kindness kept me moving forward through the darkest times and no single person contributed as much to this as you, perhaps including even myself. You made it worth doing.

Finally, I would like to thank my immediate family who have been lifting me up from the very beginning. Mum, Dad, Hamish & Grace, thank you for all that you've done for me, but most of all for showing me what it means to be a decent person.

Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Definition
4-MUG	4-Methylumbelliferyl α-D-glucopyranoside
2-AA	2-Anthranilic Acid
aa	Amino Acid
ADPLD/ADPKD	Autosomal Dominant Polycystic Liver/Kidney Disease
AE	Adverse Effect
aLP	α-Lytic Protease
Ambic	Ammonium Bicarbonate
ANOVA	Analysis of Variance
At	*Arabidopsis thaliana*
BiP	Binding Immunoglobulin Protein
BnBr	Benzyl Bromide
Boc	tert-Butyloxycarbonyl
bp	Boiling Point
BSA	Bovine Serum Albumin
CADD	Computer-Aided Drug Design
CAST	Castanospermine
CC10	Cytotoxic Concentration required to decrease cell viability by 10%
CD	Circular Dichroism
CID	Collision Induced Dissociation
CNX	Calnexin
COVID-19	Coronavirus Disease-2019
CRT	Calreticulin
CryoEM	Cryogenic Electron Microscopy
cx	Compound x
DAA	Direct Acting Antivirals
DCM	Dichloromethane
DDA	Data Dependent Acquisition
DENV1-4	Dengue Virus Serotypes 1-4
DLS	Diamond Light Source
DMA	Dimethylamine
DMAP	4-Dimethylaminopyridine
DMEM	Dubesco's Modified Eagle Medium
DMF	Dimethylformamide
DMSO	Dimethylsulfoxide
DNJ	1-Deoxynojirimycin
DNJ-$(OBn)_4$	Tetrabenzyl-protected DNJ
dotp	Dot Product
DSF	Differential Scanning Fluorimetry
DTB	Desthiobiotin
DTT	Dithiothreitol
EBOV	Ebola Virus
EDEM1-3	ER degradation-enhancing α-mannosidase-like proteins 1–3 (EDEM1–3)

EPPIC	Evolutionary Protein-Protein Interface Classifier
ER	Endoplasmic Reticulum
ERAD	ER-Associated Degradation
ERD	Emerging Virus Disease
ERQC	Endoplasmic Reticulum Quality Control
ESI	Electrospray Ionisation
EtOAc	Ethyl Acetate
FA	Formic Acid
FBLD	Fragment-Based Lead Discovery
FGI	Functional Group Interconversion
FOS	Free Oligosaccharides
FRET	Fluorescence Resonance Energy Transfer
FWHM	Full Width at Half-Maximum
GalNAc	N-Acetylgalactosamine
GalNAc	Galactose
GH31	Glycosyl Hydrolase 31
GIIα	GluII Alpha Subunit
GIIβ	GluII Beta Subunit
Glc	Glucose
GlcNAc	N-Acetylglucosamine
GluI	ER α-glucosidase I
GluII	ER α-glucosidase II
GST	Glutathione
GXL	Glutaraldehyde Cross-Linking
HCD	High-energy Collision Dissociation
HCV	Hepatitis C Virus
HDX	Hydrogen-Deuterium Exchange
HexNAc	N-Acetylhexosamine
HexNAc	Hexose
HILIC	Hydrophilic Interaction Liquid Chromatography
His6	Hexahistidine
HIV-1/HIV-2	Human Immunodeficiency Virus 1 & 2
h*Mm*GluII	Humanised Murine GluII
HPLC	High Performance Liquid Chromatography
Hs	*Homo sapiens*
HSV1/2	Herpes Simplex Virus 1/2
HTA	Host-Targeting Antiviral
IAA	Iodoacetamide
IMP	Integral Membrane Protein
iSCAMS	Interferometric Scattering Mass Spectrometry
iSCAT	Interferometric Scattering Microscopy
Kif	Kifunensine
LDLRa	Low-Density Lipoprotein Receptor Class A
lDDT-Cα	Local Distance Difference Test
mAb	Monoclonal Antibody
MALDI-TOF	Matrix-Assisted Laser Desorption Ionisation-Time of Flight
MALS	Multi Angle Light Scattering

Man	Mannose
MANEA	α-1,2-Endomannosidase
MARV	Marburg Virus
Mm	*Mus musculus*
*Mm*GluII$_{Tryps}$	Trypsin-digested murine GluII
MMS	Membrane Mimetic System
M*O*N-DNJ	*N*-(9′-methoxynonyl)-DNJ
MOPS	(3-(N-morpholino)propanesulfonic acid)
MP	Mass Photometry
MRH	Mannose 6-Phosphate Receptor Homology
MS	Mass Spectrometry
MTS	(3-(4,5-dimethyl-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium)
MW	Molecular Weight
*N*B-DNJ	*N*-butyl-1-deoxynojirimycin
NEB	New England Biolabs
NeuAc	N-Acetyl-Neuraminic Acid
NP	Normal Phase
OS-9	Osteosarcoma protein 9
PBS	Phosphate Buffered Saline
PDI	Protein Disulfide Isomerase
PE	Petroleum Ether
PISA	Proteins, Interfaces, Structures and Assemblies
PMS	Phenazine Methosulfate
PNGase	Peptide:N-glycanase
pNPG	*p*-Nitrophenyl α-D-glucopyranoside
PPI	Protein-Protein Interface
PRM	Parallel Reaction Monitoring
RBV	Ribivarin
RFU	Relative Fluorescence Units
RSA	Retrosynthetic Analysis
S. pombe	*Schizosaccharomyces pombe*
SAH	Single α-Helix
SAR	Structure Activity Relationship
SAXS	Small Angle X-ray Scattering
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SEM	Standard Error of the Mean
SVR	Sustained Virologic Response
TCI	Targeted Covalent Inhibitor
TFA	Trifluoroacetic acid
TLC	Thin Layer Chromatography
T_m	Melting Temperature
ToP-DNJ	α-tocopherol DNJ
TR	Time Resolved
Ts	4-Toluenesulfonyl
UGGT	UDP-glucose:glycoprotein glucosyl transferase

UHMR	Ultra High Mass Range
UHPLC	Ultra High Performance Liquid Chromatography
UPR	Unfolded Protein Response
UV-4	*N*-(9′-methoxynonyl)-DNJ
UV-4B	*N*-(9′-methoxynonyl)-DNJ hydrochloride
VS	Virtual Screen
wt	wild type
XC50	Concentration required for half maximal interface disruption
YFV	Yellow Fever Virus
λ_{Em}	Emission Wavelength
λ_{Ex}	Excitation Wavelength

1. Introduction

1.1 Combatting Viral Infections

1.1.1. Emerging Virus Diseases

Emerging virus diseases (ERDs) are caused by a previously unidentified infectious agent and represent a significant public health threat, with around one new example occurring annually. Emerging and re-emerging diseases are commonly transmitted naturally from vertebrates to humans and viruses are overrepresented as the cause of such disease, with RNA viruses alone accounting for around one third of all emerging infections.[1] A major concern with diseases of this type is their ability to appear unexpectedly and spread rapidly, such that local and global healthcare services are quickly overwhelmed, with developing countries generally more vulnerable due to limited capacity to trace, treat and control the spread of infection, as demonstrated by the Zika outbreak in South America in 2015 and many outbreaks of Ebola virus (EBOV) across the African continent over the last 50 years.[2–4] In 2020, the necessity for preparedness against ERDs became acutely apparent, with the outbreak of coronavirus disease-2019 (COVID-19), caused by an enveloped coronavirus (SARS-CoV-2), bringing global human activity to an unprecedented halt, referred to as an 'anthropause'.[5] The initial outbreak quickly became a global pandemic and has seen confirmed cases exceed 676 million, with deaths in excess of 6.8 million worldwide, as of March 10th 2023.[6] This prompted a scientific scramble whereby researchers worldwide rushed to collect structural data for key viral proteins from this novel coronavirus strain and identify possible treatments by repurposing existing approved drugs.[7–13] Clearly, the identification of novel antiviral treatments is crucial for coping with both ERDs and known viral infections lacking suitable treatment options.

Introduction

The gold standard for antiviral therapy is the production of effective vaccines. Since the COVID-19 pandemic began, a concerted, global scientific effort has led to the expeditious development of several effective vaccines which were valuable tools in controlling the spread of the disease.[14–16] However, vaccine development is typically expensive and time consuming, generally taking 10-15 years from inception to approval, with 94% of vaccine candidates eventually proving unsuccessful.[17] Another significant disadvantage is that vaccines require specific knowledge of viral antigens and are therefore ineffective against novel viruses, those with multiple serotypes (*e.g.* dengue virus, DNV 1-4) and those that have a high rate of mutation, such as human immunodeficiency virus (HIV-1/HIV-2).[18,19] Indeed, the rapid increase in COVID-19 cases among highly vaccinated populations following the identification of the mutated omicron variant provided evidence that the recently developed vaccines offer significantly less protection from symptomatic disease following infection by mutant strains compared to earlier strains that predominated in 2021. While various combinations of booster doses significantly increase protection against omicron, the levels of protection decrease over time, meaning that regular boosters would be necessary to control the spread of mutant SARS-CoV-2 strains.[20] Despite their undeniable utility for disease control, emergency vaccines against COVID-19 and other ERDs by extension are far from the ideal solution when used in isolation. In the case of pandemic diseases, the use of emergency vaccines is reactionary by definition, aiming to prevent further spread of the disease in question at a time when large numbers of people are already infected. As such, they are ineffective at treating those who have already contracted the disease. With no approved antiviral treatment against COVID-19 at the start of the pandemic, medical professionals were forced to treat severe cases by alleviating patients' symptoms rather than dealing with the infection itself, leading to loss of life on a tragic scale. Optimal pandemic preparedness requires disease prevention by way of effective vaccines, supplemented with safe and effective antiviral drugs to treat existing infections.

1.1.2 Antiviral Drugs

In the absence of effective vaccines, treatment of a viral infection is typically achieved through one of two possible mechanisms. The first and most common method involves direct acting antivirals (DAAs) which directly target viral proteins. This effective strategy can be achieved in a number of ways, with commonly targeted processes including viral attachment and entry into host cells, as well as inhibition of viral proteases, polymerases and retroviral integrases.[21] Since they target essential viral proteins, DAAs are often effective against their target virus and exhibit sustained virologic response (SVR), meaning that the virus remains undetectable in the patient's blood for extended periods of time following treatment.[22] In theory, the targeting of viral proteins directly could lead to minimal side effects, as host proteins are not interfered with by the drug, however in practice this is often not the case, due to structural similarities between target viral proteins and their human analogues. Another possible benefit is the relatively short duration of treatment, though again this is entirely dependent on the nature of the infection, with chronic infections such as HIV requiring prolonged treatment, during which the risk of side effects as a result of exposure will be increased.[23,24] Key drawbacks of DAAs include their narrow range of applicability (they are often effective against only one or very few specific viruses) and their vulnerability to the development of resistance as a result of increased selection pressure on viral mutation, a common occurrence due to the low fidelity of viral DNA/RNA synthesis.[25] A key alternative method for treating viral infection is the use of host-targeting antivirals (HTAs), whereby drugs target host processes that are crucial for viral replication. This strategy is potentially beneficial as therapeutics of this type have a higher genetic barrier to drug resistance than DAAs, as the host's DNA replication is slower and less error-prone and therefore less likely to mutate and render the drug ineffective.[19,26] Moreover, HTAs have the potential for broad-spectrum antiviral activity, with many different viruses hijacking the same host processes and therefore vulnerable to modulation of the process in question.[27] Broad-spectrum activity is a hugely desirable property for novel antivirals as it potentially offers a first line of defence treatment against as yet untreatable viral infections,

including ERDs like COVID-19. It is possible that development of diverse, effective and safe HTAs could contribute to greater preparedness against currently unknown viruses with pandemic potential and reduce the impact of ERDs. Another useful application of HTAs is in the treatment of viral coinfections, which are commonplace and complicate infections to the detriment of patient health; HTAs that are active against all of the infecting viruses could alleviate these complications.[28] A key example of a possible HTA approved for treatment of viral infections is ribavirin (RBV), which features on the World Health Organisation's List of Essential Medicines for the treatment of viral haemorrhagic fevers and, when in combination with DAAs, hepatitis C virus (HCV).[29] Ribavirin is known to inhibit inosine-5′-monophosphate dehydrogenase, a crucial host enzyme for the biosynthesis of guanine nucleotides, which hinders viral replication by disrupting nucleic acid synthesis.[30] However, the precise mechanism of action of RBV is not fully understood and several other mechanisms have been proposed including inhibition of the viral protein RNA-dependent RNA polymerase. These alternative mechanisms of action suggest that RBV may act as an HTA, but also shows some DDA activity. Several other HTAs that have reached late stage clinical trials are summarised in Ji and Li's recent review.[27] Despite their apparent utility, development of HTAs is not without difficulty, with drug toxicity a frequent cause for concern. It is vital when developing HTAs that the modulation of the host process of interest is not significantly detrimental to the host cells. A host protein that is crucial for viral replication, but also for healthy cellular function is unlikely to be a suitable antiviral target for this reason. It is therefore important to identify targets that are either not vital for cellular function, perhaps as a result of functional redundancy between different proteins, or ones that are overexpressed upon infection to facilitate viral replication. In this case, it could be possible to partially inhibit the target such that viral replication is halted but normal cellular function can be preserved. It may be that some low-level toxicity is unavoidable, though if the HTA is effective the toxicity may be acceptable, and could be minimised by using the shortest possible duration of treatment. The potential benefit of using covalently binding molecules in this context should be noted, as they can provide extended

duration of action at low doses by irreversibly modifying the target protein, which may enhance their potency relative to similar non-covalent therapeutic agents.[31] Another barrier in HTA development is the often poor correlation between *in vitro* antiviral activity and *in vivo* clinical efficacy, leading to a high rate of attrition in this field which could only be improved by the expansion of reliable, predictive *in vitro* models. Despite these challenges, the rich and largely untapped potential of HTAs could inexorably reform the treatment of viral infections, and the in-tandem use of DAAs and HTAs is likely to improve our control over epidemic or potentially pandemic viruses.

1.2 Endoplasmic Reticulum Quality Control Mechanisms for Glycoprotein Folding

Many pathogenic viruses bear a host cell derived lipid bilayer containing virally encoded glycoproteins on their outer surface known as the viral envelope. High profile species of enveloped viruses include herpesviruses (e.g. herpes simplex virus, HSV1/2), flaviviruses (e.g. Zika, DNV1-4, HCV) and lentiviruses (e.g. HIV-1/HIV-2). The glycoproteins embedded in the viral envelope are crucial for membrane fusion of the virus with host cells during infection and can help virions to evade the host's immune system by shielding specific viral epitopes from antibody neutralisation.[32,33] The reliance of enveloped viruses on the correct folding of their envelope glycoproteins for proliferation and/or infectivity makes the host's endoplasmic reticulum (ER) protein-folding machinery an attractive target for novel HTAs. A validated, multi-protein mechanism for accurate folding of N-linked glycoproteins is the calnexin (CNX) cycle, a process fundamental to the ER quality control (ERQC) mechanism that can be targeted by potentially antiviral therapeutics.[34–38]

1.2.1 The Calnexin Cycle

During their biosynthesis, nascent glycoproteins are co-translationally modified at the ER's luminal face by the addition of precursor glycans with the formula $Glc_3Man_9GlcNAc_2$, which are added by the oligosaccharyltransferase (OST) enzyme.[39] It is these precursor glycans that facilitate the recognition of the nascent glycoproteins by several proteins in the ER that act as folding chaperones, mediate secretion and trigger ER-associated degradation (ERAD) where necessary.[40,41] The attachment of the $Glc_3Man_9GlcNAc_2$ N-glycans labels the glycoproteins as substrates for ERQC, which begins with the cleavage of the terminal glucose residue by ER α-glucosidase I (GluI), to give the $Glc_2Man_9GlcNAc_2$ glycan. The resulting diglucosylated glycan can then be recognised and bound specifically by the membrane bound lectin, malectin. While the precise function of malectin is not fully understood, it is believed to act as an early misfolding sensor by preferentially binding to non-native conformers of the nascent glycoproteins, preventing the secretion of aberrant gene products particularly under ER stress conditions such as the unfolded protein response (UPR), where malectin is significantly upregulated.[42–44] Recent studies suggest that malectin may play a role in innate immunity in several species of fish, with upregulation of the ER stress-induced protein correlating with downregulation of key viral genes during infection, indicating antiviral activity. Furthermore, these studies show that malectin is capable of recognising and binding several species of gram-positive bacteria, perhaps demonstrating the protein's role in both bacterial and viral infections.[45] Competing for binding of the $Glc_2Man_9GlcNAc_2$ N-linked glycoproteins is α-glucosidase II (GluII). Initially, this enzyme is responsible for the cleavage of the next terminal glucose residue to give the monoglucosylated glycan, $Glc_1Man_9GlcNAc_2$. Glycoproteins bearing this glycan are substrates for the membrane-bound folding chaperone CNX and its soluble analogue, calreticulin (CRT).[46] When bound to CNX/CRT, the nascent glycoprotein is retained in the ER and other folding chaperones and enzymes including binding immunoglobulin protein (BiP) and protein disulfide-isomerase (PDI) are recruited to assist in achieving native conformation.[47] Alternatively, the $Glc_1Man_9GlcNAc_2$ glycoprotein can again

be bound by GluII which cleaves the third glucose residue, yielding a fully deglucosylated $Man_9GlcNAc_2$ glycan, which has significantly reduced affinity for CNX/CRT but can interact with several different protein partners, depending on the quality of the glycoprotein fold. If correctly folded, the glycoprotein will be exported to the Golgi apparatus for further glycan processing. However, if misfolded, the glycoprotein can be recognised by UDP-glucose:glycoprotein glucosyl transferase (UGGT), an enzyme that reglucosylates the glycan to regenerate the $Glc_1Man_9GlcNAc_2$ species and restore its ability to be recognised by CNX/CRT, giving it another opportunity for correct folding. Alternatively, the misfolded glycoprotein can be bound by an ER degradation-enhancing α-mannosidase-like protein 2 (EDEM2) which removes the terminal mannose residue from the B-arm of the glycan, producing $Man_8GlcNAc_2$ glycans and targeting the glycoprotein for ER-associated degradation (ERAD) (Figure 1A). Subsequent activity by EDEM1/3 yields $Man_{5-7}GlcNAc_2$ (Figure 1B), which is recognised by an ER lectin, osteosarcoma protein 9 (OS-9), provided a Man-α(1,6)-Man linkage is exposed on the C-arm of the glycan. OS-9 is involved in retrotranslocation, delivering the misfolded glycoprotein to the HRD1 ubiquitination/export complex, trafficking it out of the ER.[48–50]

High concentrations of misfolded protein in the ER will initiate the UPR and subsequent ERAD, a complex, multi-mechanism process that involves the export of proteins to the cytosol, where they are bound to ubiquitin by a membrane-associated ubiquitin ligase (E3 ligase) at the cytosolic side of the ER membrane and subsequently hydrolysed by the proteasome. The precise mechanism of ERAD depends on the substrate and the nature of the incorrect fold and while originally assumed to be predominantly used as a quality control mechanism, there is evidence that it is implicated in regulated degradation of even correctly folded protein as part of normal cellular homeostasis.[40] A key feature of ERAD is the cleavage of the N-linked glycan from the rest of the protein by a peptide:N-glycanase (PNGase) before its degradation by the proteasome which releases free oligosaccharides (FOS) into the cytosol. These FOS

serve as biomarkers for ERAD, and have been harnessed in an assay in order to assess the relative inhibition of GluI and GluII, whereby the presence of glucosylated FOS, which are not able to enter the lysosome leading to cytosolic accumulation, provides insight into the precise point at which the calnexin cycle was interrupted, as determined by normal phase high performance liquid chromatography (NP-HPLC) (Figure 2). This FOS analysis has revealed that for small molecule inhibitors of ER-resident α-glucosidases with antiviral activity, such as iminosugars, it is arguably the inhibition of the second glucose cleavage by α-GluII that seems the most therapeutically relevant stage, as determined by the presence of $Glc_1Man_9GlcNAc_2$ glycans in the FOS assay at low inhibitor concentrations. It is only at high concentrations of the iminosugar inhibitor that triglucosylated FOS are observed, indicating inhibition of GluI. A possible complication in the interpretation of FOS HPLC data is the activity of α-1,2-endomannosidase (MANEA). MANEA is localised at the luminal membrane of the Golgi that catalyses the cleavage of di-, tri-, or tetrasaccharides (Glc1–3Man) from N-glycans that have escaped processing by GluI and/or GluII in the ER. For cell lines expressing MANEA, it is possible that glycoproteins that bypass ERQC and are erroneously exported to the Golgi may be deglucosylated upon arrival in the Golgi, indicating that MANEA may act as a backup to ER glucosidases to prevent secretion of glycoproteins that are improperly folded.[51] This ER glucosidase independent deglucosylation may prompt false positive interpretations of FOS assay data, as HPLC chromatograms may indicate the presence of fully deglucosylated high mannose glycans, even if ER glucosidase inhibition has been achieved.

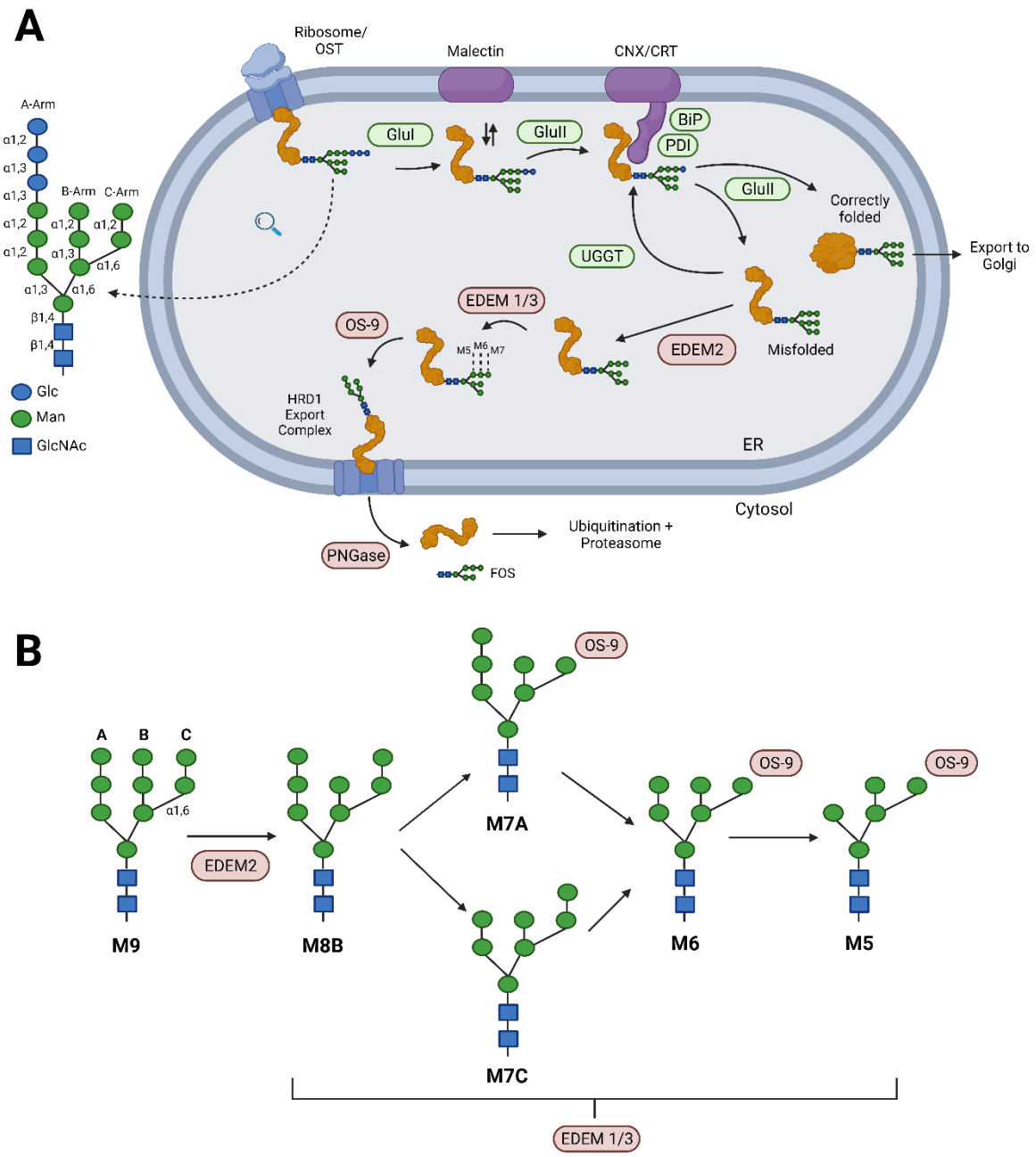


Figure 1 - The mammalian CNX cycle and ERAD. (A) The precursor glycan (left) is attached to the nascent glycoprotein (orange) co-translationally. Immediate cleavage of the terminal glucose residue by GluI produces a $Glc_2Man_9GlcNAc_2$ glycan, allowing either interaction with malectin or further glucose cleavage by GluII. The monoglucosylated glycoprotein is a substrate for CNX/CRT which can facilitate protein folding with the aid of other proteins including e.g. BiP and PDI. Following release from CNX/CRT, GluII cleaves the final glucose residue. Correctly folded glycoproteins are exported to the Golgi apparatus for further processing, while misfolded proteins are reglucosylated by UGGT to regenerate the CNX/CRT client. Terminally misfolded glycoproteins are targeted for ERAD by sequential removal of terminal mannose residues by EDEM 1-3 and OS-9. Glycoproteins export into the cytosol via the HRD1 complex is accompanied by removal of the glycan by PNGase before cytosolic degradation of both protein and FOS. (B) The activity of ER resident mannosidases, EDEM1-3. The substrate glycan is shown and the C-arm Man-α(1,6)-Man linkage necessary for recognition by OS-9 is shown.

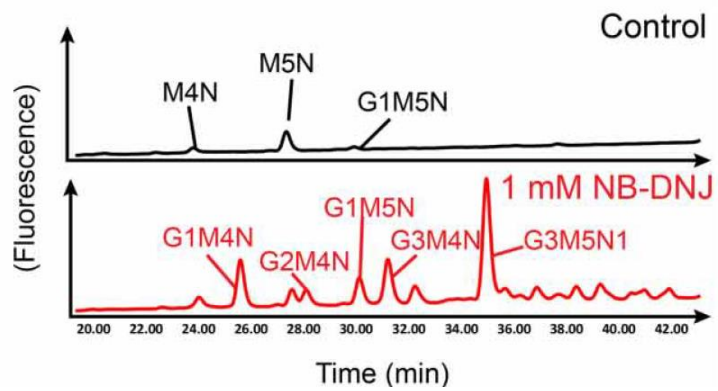


Figure 2 – Analysis of fluorescently labelled ER and cytosolic FOS by NP-HPLC. The presence of glucosylated FOS from samples where cells were treated with an iminosugar HTA, *N*-butyl-1-deoxynojirimycin (*N*B-DNJ), indicates inhibition of GluI and GluII. Degradation of FOS by cytosolic mannosidases results in low mannose FOS (≤ Man_5).[52] This figure was produced by Dr Dominic Alonzi and is presented here with the author's permission.

1.2.2 α-Glucosidase II

Viral proteins seem to rely more heavily on GluI and GluII than host proteins, and thus it is possible that there exists a therapeutic window where only partial inhibition is required for an antiviral effect.[53] With this in mind, the inhibition of GluII as a possible avenue toward novel antivirals has been pursued. Considering the apparent reliance of enveloped viruses on GluII (and the calnexin cycle more broadly) for their biosynthesis, it is important to identify key structural features of the enzyme that could be targeted by potential therapeutics. Previous work by Caputo *et al.* provides detailed insight into this structure, presenting small angle X-ray scattering (SAXS) data and a partial crystal structure of murine GluII (*Mm*GluII) complexed to various different ligands to clarify the mechanism of glucose cleavage.[54] Full length GluII is a soluble heterodimer consisting of a 108 kDa catalytic α-subunit (GIIα, encoded by the *GANAB* gene) belonging to the glycosyl hydrolase 31 (GH31) family and a 58 kDa accessory β-subunit (GIIβ, encoded by the *PRKCSH* gene) which is responsible for the localisation of the heterodimer in the ER through its C-terminal tetrapeptide ER retrieval signal, HDEL. Both subunits are highly conserved among eukaryotes, with human and murine analogues sharing 92% and 87% sequence identity for GIIα and GIIβ, respectively. Both GIIα and GIIβ are localised in the ER lumen, with no evidence of membrane association, either experimental or

inferred from sequence analysis, according to the relevant Uniprot entries and associated databases (Uniprot ID - *Mm*GIIα: Q8BHN3; *Mm*GIIβ: O08795)[55], though GIIβ appears to interact with the membrane protein polycystin-1 as part of a GluII independent function implicated in autosomal dominant polycystic liver/kidney disease (ADPLD/ADPKD). ADPLD/ADPKD is an inherited condition that has been associated with faulty variants of both *GANAB* and *PRKCSH*, with *GANAB* mutations causing decreased or complete loss of enzymatic activity of GluII by interfering with the architecture of the active site on GIIα, or with the protein-protein interface (PPI) between the two subunits.[56] In the case of *PRKCSH*, aberrant mutations lead to loss of GluII function, but may also induce ADPLD in a GluII-independent manner by interfering with interactions between GIIβ and other ER-resident proteins associated with this disease, such as polycystin-2.[57] Recently, GIIα expression was shown to correlate to disease severity in a small sample of multiple sclerosis patients, suggesting that it could act as a biomarker of neuroinflammation predictive for disease activity and treatment response.[58] Furthermore, GIIβ promotes tumorigenesis in various types of cancer by specifically binding to and promoting the activity of IRE1α, a key UPR signal activator, which helps protect tumor cells from ER stress.[59] This finding is supported by the fact that GIIβ knockout lung carcinoma cell lines demonstrate slower tumour growth and metastasis as well as altered sensitivity to certain chemotherapeutic drugs when compared with the same cell lines expressing GIIβ.[60] Despite these insights into GluII-associated pathologies, the full interactome of GluII and its subunits individually is beyond the scope of this thesis which will focus on its role in the ERQC mechanism.

Due to considerable difficulty in crystallising the full length heterodimer, GluII must be subjected to trypsin digestion to yield a crystallisable, but truncated version of *Mm*GluII (*Mm*GluII$_{Tryps}$). Here, GIIα is missing short sections of its N-terminal domain and GIIβ is significantly shortened, retaining only the N-terminal region which adopts two tandem low-density lipoprotein receptor class A (LDLRa) folds, each enclosing an octahedrally

coordinated calcium ion (Figure 3A). The *Mm*GluII$_{Tryps}$ crystal structures reveal that GIIα is itself a glycoprotein, with a single experimentally confirmed N-glycan attached at the N97 residue (PDB: 5F0E).[61] Despite the significant truncation, *Mm*GluII$_{Tryps}$ retains its affinity and catalytic activity for both natural substrates (*i.e.* $Glc_2Man_9GlcNAc_2$ and $Glc_1Man_9GlcNAc_2$ glycans) and synthetic substrates such as 4-methylumbelliferyl α-D-glucopyranoside (4-MUG), with 70% of the catalytic efficiency against 4-MUG retained, relative to the full length wild-type (wt) enzyme. Furthermore, the 1.74 Å resolution crystal structure contains both the active site and an appreciable amount of the α/β interface and has provided detailed insight into the catalytic activity of this enzyme alongside key interface interactions, with SAXS data illustrating the larger volume occupied by the full-length protein. Like other GH31 family enzymes, GIIα conducts its glycosidic bond hydrolysis via the Koshland double displacement mechanism, with the D564 and D640 residues acting as the catalytic nucleophile and general acid/base catalyst, respectively. Evidence for the role of these residues is provided by the crystal structure of *Mm*GluII$_{Tryps}$ soaked with various ligands which trap the pocket in key stages of the catalytic cycle (Figure 3B). The glycosidic bonds specifically cleaved by GluII are α(1,3) linkages and the specificity for this type of disaccharide is conferred by the side chains of the conserved and active site adjacent R617 and M565 residues, which would experience steric clashes with the β-anomers of these sugars. Demonstrated by the crystal structure of GluII with a bound α(1,3)-linked D-glucal pseudodisaccharide, the hexose ring occupying the -1 subsite of the catalytic pocket has three equatorial hydroxyl groups (C-2, C-3 and C-4) which form hydrogen bonds to conserved residues (R624, H698 and D451, respectively), explaining the preference for D-glucose over D-mannose or D-galactose in this position. While specific to α-anomers of the substrate glycans, GluII is capable of hydrolysing α(1,4)- and α(1,2)-linkages and their relative absence in the ER is thought to lift evolutionary pressure to lose activity against them. Crucial for the cleavage of both the second and third glucose residue from the N-linked precursor glycan of nascent glycoproteins is the ability to accommodate both Glc-α(1,3)-Glc and Glc-α(1,3)-Man in the catalytic pocket. The tolerance

for both glucose and mannose in the +1 subsite can be explained by the lack of close contacts between enzyme and the C-2/C-3 of the +1 hexose ring, which allows for some flexibility in the orientation of the stereochemical configuration at the C-2 position at which D-glucose and D-mannose differ. Interestingly, the equatorial C-2 hydroxyl group on a glucose residue in the +1 subsite does not appear to make any close contacts with α-GluII, whereas a mannose occupying this position can accommodate a hydrogen bond between the axial C-2 hydroxyl group and the protonated catalytic acid/base residue, D640. This results in stabilisation of the protonated D640 residue leading to less efficient hydrolysis of the Glc-α(1,3)-Man bond compared to the Glc-α(1,3)-Glc which is consistent with the earlier observation that the first glucose cleavage by GluII occurs more rapidly than the second at physiological pH.[62] This is further supported by the fact that this difference in rate of hydrolysis becomes negligible at higher pH, as the acidic proton from D640 is lost, thus nullifying the stabilising interaction.

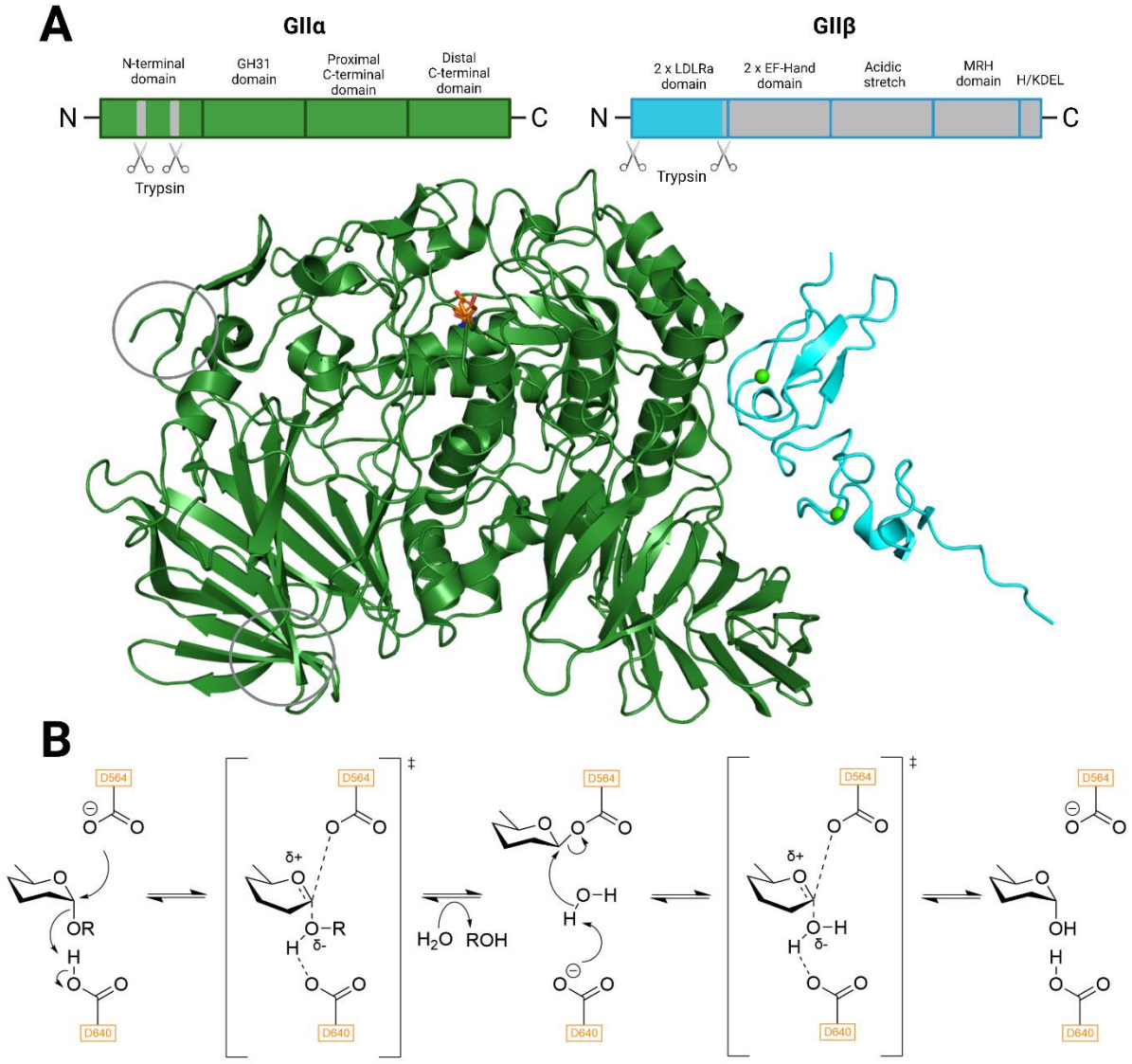


Figure 3 – The GluII heterodimer. (A) Crystal structure of *Mm*GluII$_{Tryp}$ (PDB: 5F0E). The simplified domain architecture is shown for GIIα (dark green) and GIIβ (cyan), with sections of the full length protein removed following tryptic digestion highlighted in grey. The catalytic nucleophile (D564) and acid/base residue (D640) are shown in orange and the octahedrally coordinated Ca^{2+} ions in GIIβ are shown as light green spheres. (B) The Koshland double displacement mechanism of action of glucose hydrolysis by GIIα, with catalytic residues highlighted in orange. For simplicity, hydroxyl groups of the substrate glucose residue are omitted and the rest of the glycan is denoted as a generic alkyl group, R.

While GluII shares its catalytic mechanism with other GH31 family members, it also bears key unique features. Firstly, two amino acids in the GIIα N-terminal domain (*Mm*GluII residues 305-317) form a loop that extends to the +1 and +2 subsites of the catalytic pocket. These residues form the “exclusion loop”, named for its apparent ability to preclude the binding of GluII to the non-cleavable α(1,4)-tetrasaccharide mimic, acarbose, a known intestinal α-glucosidase inhibitor (Figure 4). Caputo *et al*. investigated the significance of this unique

protrusion via mutagenesis and found that an *Mm*GluII F307G-Δ308 displays drastically reduced catalytic activity against the synthetic substrate 4-MUG, with only 17% of wt activity. Furthermore, the double deletion mutant Δ307-308 shows no activity against 4-MUG, suggesting that the exclusion loop is important for the native function of GluII and likely mediates access of ligands to the +1 subsite.

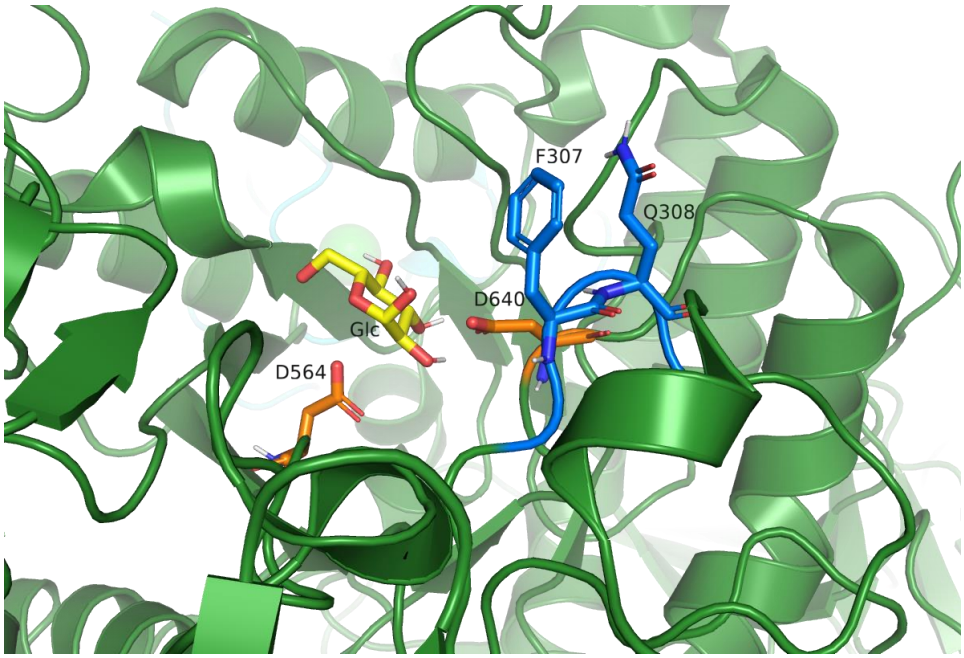


Figure 4 – *Mm*GluII$_{Tryp}$ with glucose (yellow) bound in the active site (PDB: 5H9O) with exclusion loop residues (F307, Q308) shown in blue and catalytic residues (D564, D640) in orange. The exclusion loop likely mediates the access of ligands to the +1 subsite of the active site. The double deletion mutant Δ307-308 shows no activity against the synthetic substrate 4-MUG.[54]

The second key feature of GluII is the functional importance of its heterodimeric nature. The close interaction between the catalytic pocket of GIIα and the N-terminal domain of GIIβ holds the pocket in a catalytically competent conformation.[54] Both Glc-α(1,3) cleavages of substrate glycans by GluII require an at least partially intact α/β interface, though a higher dependence exists for the cleavage of monoglucosylated species compared to diglucosylated since the latter can occur more rapidly, as outlined above.[54,63] In 2000, Arendt and Ostergaard identified the GIIβ domains that interact with GIIα by expressing recombinant GIIβ segments spanning its entire length as glutathione (GST) fusion proteins. The purified GST fusion proteins were

incubated with cell extracts and captured following introduction of a GST matrix and their association with GIIα was determined by measuring catalytic activity of GIIα against pNPG. This experiment identified the N-terminal region (GIIβ 1-118), which contains the LDLRa domains present in the *Mm*GluII$_{Tryps}$ crystal structure, and a region close to the C-terminus (GIIβ 273-400) which immediately precedes the mannose 6-phosphate receptor homology (MRH) domain on GIIβ (GIIβ 406-507).[64] Later, various publications by Stigliano *et al.* and Olson *et al.* show that in GIIβ knockouts, the catalytic activity of GluII against native $Glc_{1-2}Man_9$ substrates is reduced to 5-10% of wild-type activity *in vitro* and is slowed significantly *in vivo*; in both cases, the activity can be mostly restored by introduction of exogenous full length GIIβ.[63,65,66] Interestingly, introduction of GIIβ MRH domain to a lone GIIα provides virtually no restoration of catalytic activity by GIIα against a monoglucosylated glycan.[66] When taken with the observed reduction in catalytic activity of *Mm*GluII$_{Tryps}$ (in which GIIβ contains the LDLRa domain but not the MRH domain or the preceding *Mm*β273-400 region) relative to full length *Mm*GluII, this suggests that the N-terminal LDLRa domains of GIIβ are crucial for full catalytic activity of GluII, while the MRH domain may assist the catalytic GIIα in glucose hydrolysis by binding a mannose residue on the nascent glycoprotein.[54] This conclusion is supported by the observation that the accessory GIIβ is not necessary for the catalytic activity of GIIα against small synthetic substrates, such as *p*-nitrophenyl α-D-glucopyranoside (pNPG), suggesting that the efficient hydrolysis of glucose from larger N-glycans by GluII relies on enzyme-glycan interactions beyond those found in the active site.[63] These important allosteric interactions likely serve to orient the glucose residues into the active site as well as to avoid inhibitory interactions between glucose residues on the glycan's A-arm and neighbouring mannose residues from the B/C-arms (Figure 1).[65] Given the length of the N-linked $Glc_xMan_9GlcNAc_2$ glycan, it is plausible that the MRH domain could bind a mannose residue from the same glycan bound by the catalytic pocket of GIIα, provided that conformational rearrangement of GIIβ brought the two close enough together and this remains the dominant conceptual model in the literature.[66,67] Indeed, the treatment of the enzyme with a crowding agent (PEG 20,000)

in order to induce conformational change has been shown to increase the rate of the second glucose cleavage, perhaps suggesting that such a rearrangement would aid the enzymatic catalysis.[68] The feasibility and extent of this rearrangement is difficult to infer experimentally from the existing *Mm*GluII$_{Tryps}$ crystal structure, given the heavily truncated GIIβ, but the experimental evidence provided by Arendt and Ostergaard shows that both the N- and C-terminal regions of GIIβ interact with GIIα highlighting plausible proximity between the catalytic pocket of GIIα and the MRH domain.[64] Furthermore, the computational model of the full length murine GIIβ predicted with high confidence by AlphaFold (Uniprot: O08795) indicates several loops that could confer significant flexibility (Figure 5).[69,70] This tentatively supports the possibility that GIIβ may undergo conformational change to allow simultaneous binding of the substrate glycan to both the active site and the MRH domain. The apparent flexibility based on the AlphaFold model also provides an explanation of why GIIβ is notoriously difficult to crystallise. Despite this, more comprehensive structural and/or computational data would be required to assess the likelihood of this conformational change.

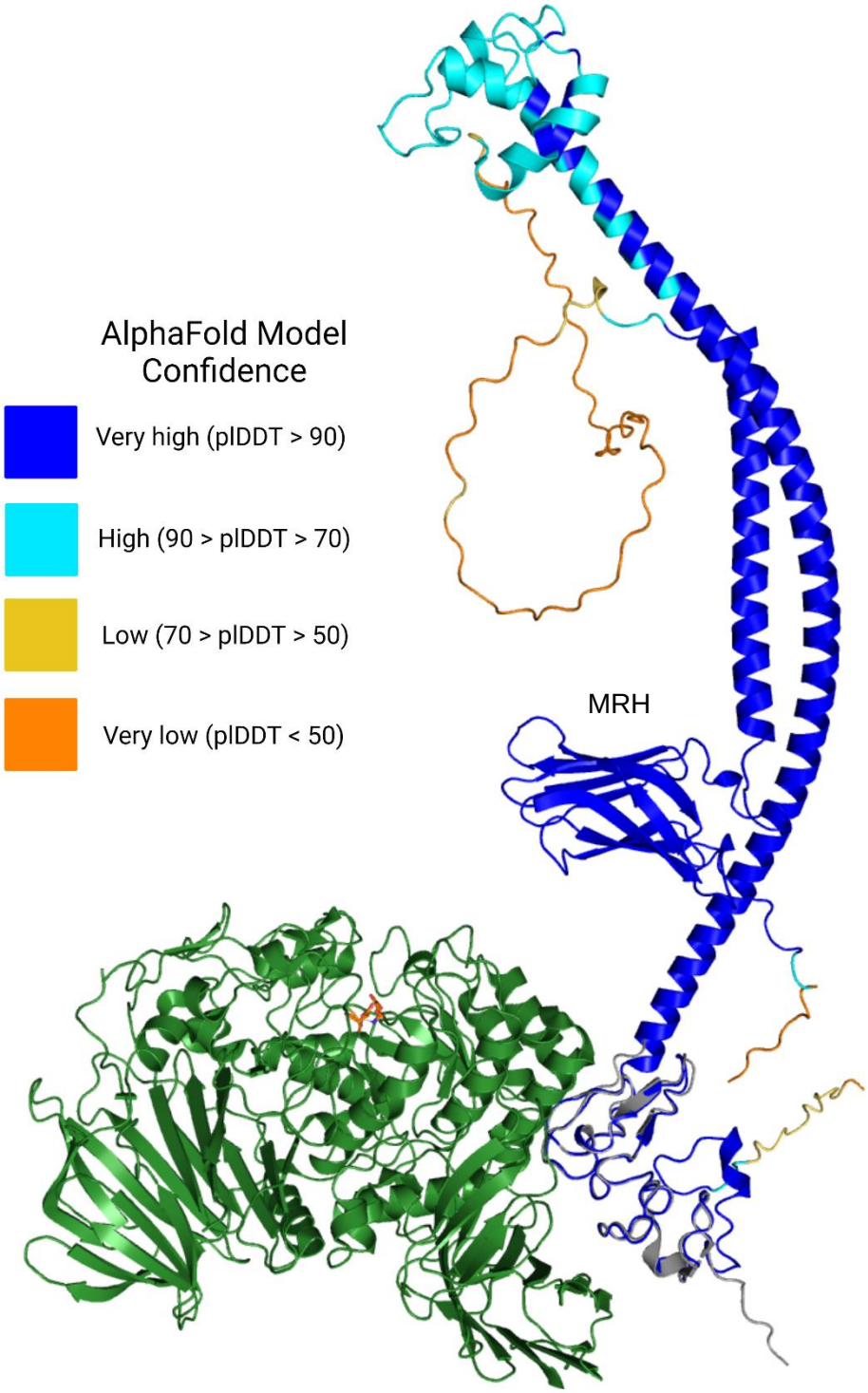


Figure 5 – The GluII heterodimer with the AlphaFold prediction of full length GIIβ superimposed onto GIIβ from $MmGluII_{Tryp}$ (grey). The AlphaFold GIIβ prediction is colour-coded based on the confidence of the prediction. The score, pLDDT, is ranked from 0-100 and corresponds to the predicted local distance difference test (lDDT-Cα) as a measure of local accuracy for the model.[69,70] The flattened β-barrel structure represents the MRH domain.

As mentioned above, the dominant conceptual model of GluII-glycan interactions involves the primary interaction between the GIIα active site and the Glc-Glc/Glc-Man residues of the

glycan's A-arm (which has been unambiguously defined using X-ray crystallography), as well as a secondary interaction between the MRH domain of GIIβ with the terminal mannose residue of the B and/or C arm on the glycan. The crystal structure of the *Schizosaccharomyces pombe* GluII MRH domain published by Olson *et al.* shows its structural homology with MRH domains from other ERQC-associated proteins such as OS-9 and Erlectin, both of which are involved in targeting misfolded proteins for ERAD.[67,71,72] The MRH domains in these proteins adopt a flattened β-barrel morphology and share conserved mannose binding residues (*Mm*GluII: Q428, R470, E489 and Y495). Removal of these mannose binding residues by mutagenesis reduces the ability of GIIβ to enhance glucose trimming from $Glc_{1\text{-}2}Man_9$ substrate glycans by GIIα *in vivo*, but does not influence its activity against pNPG, demonstrating that the MRH domain does not contribute to the α/β interaction but is required for efficient total deglucosylation of the Glc_2Man_9 substrate *in vivo*.[63] Another common feature of these MRH domains is a conserved tryptophan residue near the mannose binding site. Olson *et al.* demonstrated that mutagenesis of the W409 residue in the GIIβ of *S. pombe* (corresponding to W465 in *Mm*GluII) also inhibits GluII activity *in vitro* and *in vivo*, but to a lesser extent compared to the previously mentioned mannose binding residues.[67] The crystal structure of the MRH domain from the *S. pombe* GIIβ bound to mannose shows that W409 is oriented away from the mannose and likely modulates GluII activity by interacting with an adjacent hydrophobic pocket on the GIIα or GIIβ. Stigliano *et al.* provided more evidence for the importance of an interaction between MRH and a glycan mannose residue for GluII catalytic activity in an experiment involving removal of terminal mannose residues from the substrate glycan. In a cell-free assay, α-mannosidase treatment removed the terminal mannose residues to yield $Glc_2Man_{4\text{-}5}$ and $Glc_1Man_{4\text{-}5}$, which led to a decrease in glucose trimming rates by GluII to a degree comparable to the mutation of key mannose binding residues in the MRH domain.[63] This finding was later confirmed to be consistent *in vivo* with a similar experiment using mutant *S. pombe* which transferred low mannose truncated glycans

to nascent glycoproteins; the amount of glucosylated glycans detected increased with decreasing mannose content.[66]

1.3 α-Glucosidase II: Active Site Inhibition

To date, the inhibition of GluII has followed the traditional drug discovery strategy of targeting the active site on GIIα with small molecules. The well-established catalysis of glycan deglucosylation by GluII inevitably initiated attempts to competitively inhibit the enzyme's active site using non-cleavable glucose mimetic small molecules. The most commonly utilised glucose mimetic for this application is the iminosugar 1-deoxynojirimycin (DNJ) and derivatives thereof. Despite considerable success using DNJ iminosugars to inhibit GluII, an obvious disadvantage of their use lies in the lack of specificity of glucose mimetic compounds for GluII due to the widespread nature of glucose binding enzymes in the human proteome. Nevertheless, with appropriate medicinal chemistry optimisation to enhance specificity for GluII, DNJ derivatives remain a promising starting point for the development of broad spectrum, host-targeted antivirals.

1.3.1 Iminosugar Antivirals

Iminosugars are a class of sugar mimetic compounds in which the cyclic oxygen atom is replaced by nitrogen. Two naturally occurring iminosugars, DNJ and the bicyclic compound castanospermine (CAST), are the most promising candidates for antiviral drug development via competitive inhibition of ER glucosidases, GluI and GluII. Celgosivir is a prominent example of a CAST-derived antiviral agent and reached phase 2 clinical trials for treatment against DENV, HIV and hepatitis C (HCV) (Figure 6A).[73–75] Several DNJ derivatives are already approved for use in humans, albeit not for treatment of viral infections, including miglustat (Gaucher's disease) and miglitol (diabetes mellitus type 2); a galactose mimetic stereoisomer of DNJ, migalastat is also approved for treatment of Fabry disease (Figure 6B).[76]

Foremost among the DNJ-derived antivirals are *N*-(9′-methoxynonyl)-DNJ (M*O*N-DNJ), also known as UV-4, and the corresponding hydrochloride salt, UV-4B (Figure 6C). The latter completed phase 1 clinical trials for the treatment of DENV infections, which represent a significant health burden in tropical and sub-tropical regions and in severe cases can be fatal.[77,78,79] UV-4 has also shown *in vivo* antiviral activity against influenza A and B in mice.[76,80] Another key DNJ-derived antiviral candidate is IHVR19029, which has demonstrated antiviral efficacy against a range of haemorrhagic fever viruses including DENV, EBOV, Marburg virus (MARV) and yellow fever virus (YFV) *in vitro* and *in vivo* (Figure 6C).[81,82] Iminosugars are promising candidates for antiviral drug discovery and have demonstrated broad spectrum antiviral activity against multiple viral families, including the family *Coronaviridae*, to which the virus responsible for the COVID-19 pandemic, SARS-CoV-2, belongs. Indeed, CAST, celgosivir and UV-4 all show inhibition of viral replication in Vero E6 cells infected with SARS-CoV-2.[83] Furthermore, unpublished work by the Zitzmann group confirms that UV-4 is antiviral against all current SARS-CoV-2 variants of concern, highlighting the benefit of broad spectrum antiviral agents for improving pandemic preparedness. Despite their undeniable promise, iminosugars exhibit an unsurprising lack of selectivity, as the generic carbohydrate moiety that facilitates inhibition of ER GluI and GluII is also bound by numerous other glucosidases and glucosyltransferases. Intestinal disaccharidases (maltase, isomaltase, sucrase) are in fact more potently inhibited by UV-4 than GluII *in vitro*.[79] This off-target activity is thought to cause significant gastrointestinal distress, as observed in phase 1 clinical trials with miglustat, with osmotic diarrhoea being a common side effect with iminosugar treatment. In August 2022, results from a single oral dose phase 1 study with UV-4B was published and supports the continued investigation of the drug candidate in more advanced trials.[84] Interestingly, this trial showed that observation of severe gastrointestinal distress following UV-4B treatment is inconsistent between animal models, with dogs being particularly sensitive.[85,86] However, in humans, UV-4B is generally well tolerated and no maximum tolerated dose has yet been established, with 1000 mg the highest tested dose. Early clinical

studies are encouraging, but the customary limitations of a phase 1a study of course apply here and gastrointestinal side effects remain a distinct possibility, especially given that a 2x-3x higher dose is recommended for future studies to match efficacious doses in mouse models.[84]

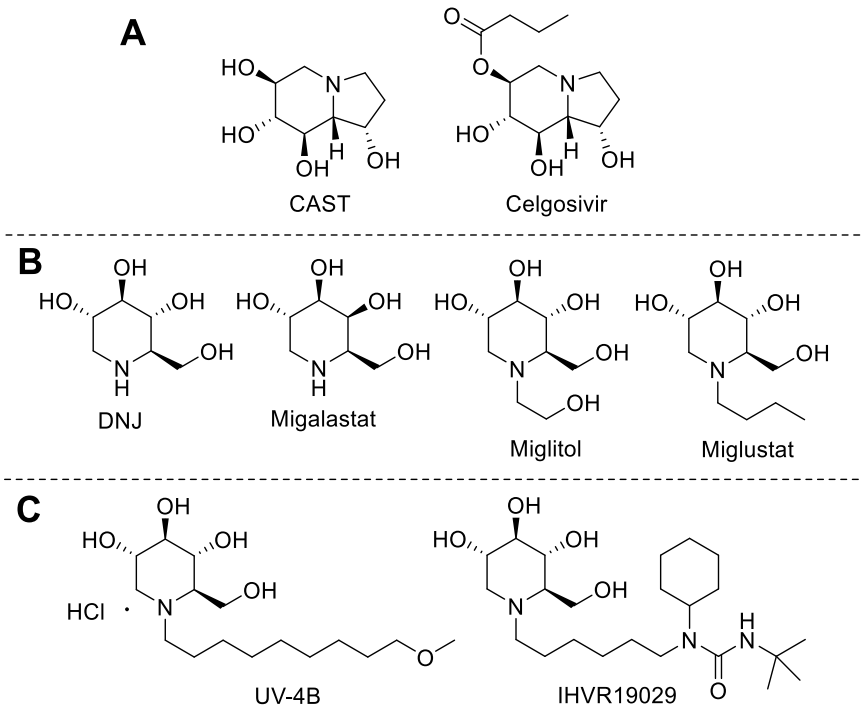


Figure 6 – Structures of iminosugars. (A) The bicyclic iminosugars, castanospermine and the ester prodrug thereof, celgosivir. Celgosivir were investigated for treatment of chronic HCV infection as part of a combination therapy that includes pegylated interferon alfa-2b and ribivarin, proceeding as far as phase 2 clinical studies (NCT00217139, NCT00292084).[87] Celgosivir also reached Phase 1b clinical trials for the treatment of DNV (NCT01619969).[88] (B) Structures of the monocyclic glucose mimetic, DNJ, and the stereoisomer and galactose mimic, migalastat, which is approved for use in treatment of Fabry disease.[89] Also shown are other approved DNJ-derived iminosugars, miglitol (diabetes mellitus type 2) and miglustat (Gaucher's disease).[90,91] (C) Promising DNJ-derived antiviral candidates, UV-4B and IHVR19029.[82,84]

1.3.2 Next Generation Iminosugar Antivirals

Iminosugars, such as CAST and DNJ, are demonstrably valid candidates for the development of broad spectrum antiviral drugs. That said, significant gastrointestinal distress, particularly osmotic diarrhea, is observed in animal and human models following treatment with leading

DNJ-derived candidates.[85,86,92] While early phase 1 clinical studies seem to suggest that the current best in class DNJ-derived candidate, UV-4B, may circumvent this issue in humans, it is sensible to assume that these severe side effects could persist, limiting the achievability of efficacious, antiviral doses in humans, particularly following oral administration.[84] Furthermore, the lack of specificity of this class of therapeutic raises concerns beyond the well documented gastrointestinal distress as off-target activity could lead to other unpredictable adverse effects. Additionally, inhibition of ceramide glucosyltransferase, the intended target of miglustat and migalastat, interferes with the biosynthesis of glycosphingolipids and could represent potential off-target activity for DNJ iminosugars used for antiviral purposes.[93,94] It is therefore appropriate to continue lead development of iminosugars to improve their specificity and identify novel next generation iminosugars that surpass those already under investigation.

In the case of DNJ-derived antiviral candidates, several approaches have been attempted to address the issues surrounding low specificity of these compounds. One such approach involves the synthesis of prodrugs of promising DNJ-derived compounds, such that they avoid interaction with intestinal glucosidases and are subsequently metabolised to their active form, which can inhibit ER glucosidases to exert antiviral activity. This strategy was exemplified by Ma *et al.*, who produced ester prodrugs of IHVR-19029 (Figure 7A).[95] Here, the hydroxyl groups on the DNJ moiety of IHVR-19029 were protected with ester groups that tolerated simulated gastric and intestinal conditions and were themselves not active against glucosidases, as shown by an *in vitro* enzymology assay using the synthetic substrate, 4-MUG. Despite low inhibitory activity towards ER GluI and GluII *in vitro*, the ester prodrugs demonstrate antiviral activity against DENV in infected Huh7.5 cells, albeit less potently than the parent compound. The conversion of the ester prodrugs into the active compound, IHVR-19029, showed inconsistent localisation *in vivo*, with the majority of conversion taking place in the liver in human models, while significant conversion was detected in plasma and the liver in mice. The most promising candidate presented in this study was the tetrabutyrate ester,

which demonstrated significantly improved pharmacokinetic properties compared to direct treatment with IHVR-19029, with a 2-4-fold increase in exposure following a 100 mg kg^{-1} oral dose.

An alternative approach to avoiding the dose-limiting side effects involves conjugating DNJ to a tissue-targeting moiety, as shown by Kiappes *et al*. in 2018.[96] By including a native metabolite into the chemical structure of novel iminosugars, it is possible to direct the compound towards clinically relevant cells and tissues. In this case, DNJ was conjugated to α-tocopherol, a non-toxic form of vitamin E that is directed to the liver following absorption in the gut and also accumulates in the membranes of immune cells (Figure 7B). The native metabolism of α-tocopherol makes this iminosugar, referred to as ToP-DNJ, an attractive candidate for treatment of HCV infection, for which the target organ is the liver, and DENV which targets immune cells and is believed to use the liver as a viral reservoir.[97,98] ToP-DNJ showed remarkable selectivity for ER GluII in *in vitro* enzyme activity assays compared to other glucosidases, including those located in the intestinal lumen that are responsible for the common gastrointestinal side effects. In fact, less than 50% inhibition of intestinal glucosidases was observed even with the highest tested concentration of ToP-DNJ (50 μM), whereas the IC_{50} value against GluII (9 μM – pNPG substrate) was comparable to that of DNJ (13 μM) and *N*B-DNJ (miglustat 16 μM). No explanation for the observed selectivity for GluII is offered in this article, but it is possible that the aromatic moiety in α-tocopherol is capable of interacting with the hydrophobic exclusion loop located near the GluII active site (see Section 1.1.2), which is unique to GluII compared with other GH31 family enzymes. The *in vitro* GluII inhibition by ToP-DNJ was also observed in cellular assays, though the inhibition is dependent on cell type. Only myeloid lineage cell lines (e.g. MDMΦ and HL60) showed GluII inhibition following treatment with ToP-DNJ, whereas Huh7.5, Jurkat and Raji cell lines were not affected, as seen by the presence (or absence) of mono- and/or di-glucosylated glycans in FOS assays. This observation was consistent with observed antiviral properties of ToP-DNJ

in infected MDMΦ cells, where DENV was inhibited with an IC_{50} value of 12.7 μM, measured by median tissue culture infectious dose ($TCID_{50}$) assays. ToP-DNJ shows poor oral bioavailability *in vivo* (maximum blood plasma concentration in mice = 46 nM after 8h), but biodistribution studies confirm that the compound is located in highest amounts in the liver following oral or intravenous delivery. Furthermore, ToP-DNJ demonstrates enhanced exposure in the liver and plasma compared to the approved iminosugars miglustat. Taken together, these observations demonstrate that the conjugation of DNJ to native metabolites like α-tocopherol is a valid approach to the production of selective, tissue targeting antivirals.

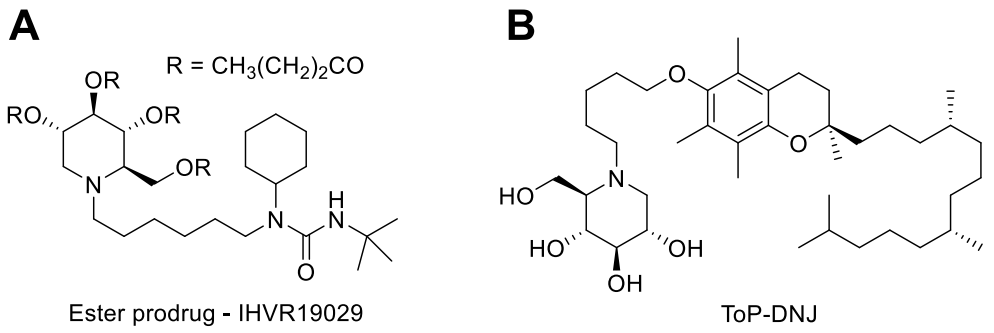


Figure 7 – Examples of next generation iminosugar HTAs designed to selectively inhibit GluII to avoid off-target activity. (A) Ester prodrug of IHVR19029.[95] (B) ToP-DNJ.[96]

Common to the chemical structure of many potent iminosugar inhibitors of GluII, including ToP-DNJ, is an N-linked aromatic/heteroaromatic moiety with some type of alkyl linker. Another established example is N-(6′-4″-azido-2″-nitrophenylamino) hexyl-1-deoxynojirimycin (NAP-DNJ) which is among the most potent known DNJ-derived inhibitors of GluI and GluII *in vitro*, but has not been pursued clinically owing to its unfavourable toxicity profile.[99] More recently, Karade *et al*. described a series of novel iminosugars that follow this common scaffold, many of which demonstrate low micromolar and even nanomolar IC_{50} values against GluI and GluII *in vitro*.[100] Here, valiolamine, the iminosugar scaffold for the type 2 diabetes drug, voglibose, was used as a starting point from which ten derivatives were synthesised. Two of these derivatives are structural analogues of *N*B-DNJ and UV-4 with valiolamine replacing DNJ and the rest contain a variety of N-linked nitrophenylamino groups, based on

NAP-DNJ. The nitrophenylamino moieties are further functionalised to contain azide (again, based on NAP-DNJ), triazole, tetrazole or pyrimidine and the linker length and/or composition is also varied. Some of these compounds are 10^4-10^5-fold more potent against GluII *in vitro* than the parent compound, valiolamine, and high resolution crystal structures of them in complex with *Mm*GluII capture the structural explanation for their potency. As a member of the GH31 family of enzymes, the active site of GIIα shares close structural similarity in its highly conserved -1 and +1 subsites with other related enzymes, such as the intestinal glucosidases sucrase-isomaltase and maltase-glucoamylase, hence the proclivity of iminosugars for off-target activity in the intestine. However, GluII is unique from its GH31 relatives in that the less conserved +2 and +3 subsites are effectively diminished by the close proximity of residues F307 and Q308 from the exclusion loop, which is absent in other GH31 family members and flanks the catalytic site of GIIα. Crystal structures of the GluII-valiolamine derivative complexes demonstrate that while smaller iminosugars, such as *N*B-DNJ, interact only with the -1 and +1 subsites, longer N-linked alkyl groups enable interaction with the +1, +2 and +3 subsites, hence the enhanced potency of UV-4 compared to *N*B-DNJ (Figure 8).[99] The same pattern is observed with the valiolamine analogues of these compounds (EB-0151 for UV-4, EB-0149 for *N*B-DNJ).[100] The +2 and +3 subsites are decorated with aromatic residues (F307, W523, W525, F571; residue numbering taken from *Mm*GluII) and this explains the increase in potency when aromatic groups are incorporated. Of the proposed compounds, the most potent are those that incorporate an additional heteroaromatic group such as triazole, tetrazole or pyrimidine. These iminosugars show clear interactions with all four subsites of the GIIα active site in the crystal structures and the most potent shows sub-nanomolar potency against GluII *in vitro.* The valiolamine derivatives are antiviral against DENV serotype 2 (DENV2) and SARS-CoV-2 in Vero and Calu-3 cells, respectively, as determined by an infectious virus yield reduction assay. This antiviral activity was inconsistent with the trends in inhibitory activity *in vitro*, likely due to poor pharmacokinetic parameters that prevent their localisation in the ER at effective concentrations. The antiviral activity is also inconsistent

between the two tested viruses, with some compounds being considerably more active against DENV2 than SARS-CoV-2 or vice versa, suggesting differences in the reliance of each virus on the CNX cycle. Despite these inconsistencies, the *in vitro* antiviral assays demonstrate the effectiveness of N-linked alkylation of iminosugars using aromatic/heteroaromatic fragments.

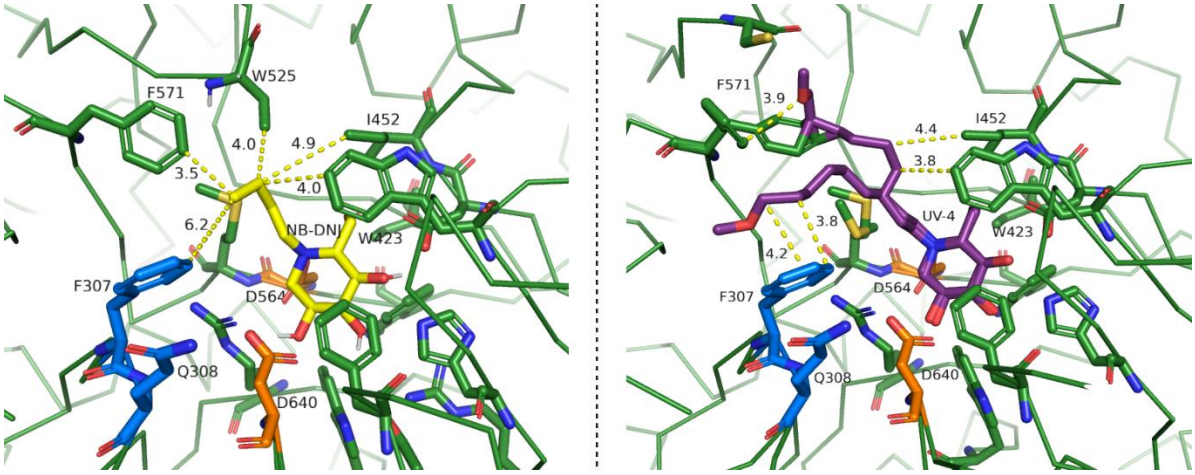


Figure 8 – Interactions between GluII and *N*B-DNJ (left, yellow, PDB: 5IEF) or UV-4 (right, purple, PDB: 5IEG) within the catalytic pocket of GIIα. Interatomic distances (Å) are highlighted as dashed yellow lines and key residues are labelled. The longer alkyl chain of UV-4 relative to *N*B-DNJ allows for closer contacts with the hydrophobic residues across the extended active site, with the interaction with exclusion loop residue, F307, of particular interest. Extended protein-ligand interactions may be responsible for the improved potency of UV-4 over *N*B-DNJ and may also enhance selectivity for GluII.

1.4 α-Glucosidase II: The α/β Interface

Treatment with iminosugars is a tried and tested method of inhibiting GluII to achieve antiviral effects against a range of viruses *in vitro* and *in vivo*.[53,100] It is therefore worthwhile to continue to seek iminosugars that are more potent and more selective against the desired target, in this case GluII, by building upon existing knowledge to design and produce compounds that exploit unique features of the target enzyme. As discussed in Section 1.2, functionalisation of DNJ iminosugars that interact with features of the GluII active site that are unique relative to other GH31 family enzymes is one avenue towards this goal, but off-target activity will likely be difficult to avoid given the ubiquity of glycan-binding enzymes throughout the body. By far the most striking feature that sets GluII apart from the otherwise monomeric GH31 family of enzymes is its heterodimeric structure. The catalytic GIIα subunit relies on its interactions with

the accessory GIIβ subunit for maximum activity against its native substrate, the $Glc_{1-2}Man_9GlcNAc_2$ glycan.[63,65–67] Herein, I explore the possibility of targeting the α/β interface of GluII as a novel approach towards its inhibition.

1.4.1 α/β Interface Architecture

Our current understanding of the α/β interface of GluII is mainly informed by the crystal structure of *Mm*$GluII_{Tryp}$, published by Caputo *et al.* in 2016.[54] This interface spans around 706 $Å^2$ and has a binding energy of -3.1 kcal mol^{-1}, as predicted by PISA and EPPIC servers which use classical thermodynamics or sequence entropy, respectively, to determine the biological relevance of PPIs from an input PDB file.[101,102] The binding energy can be largely attributed to the network of polar/ionic interactions, particularly salt bridges, that bind the two subunits together (Figure 9). These salt bridges are formed between the basic residues on GIIα (R837, R839, R840, R951, K952) and the acidic residues on GIIβ (D53, D54, D57, D94, D98), giving the interfacial regions of GIIα and GIIβ an overall positive and negative charge, respectively. The surface electrostatics of this crystal structure were calculated using the APBS-PDB2PQR server (Figure 10A).[103] In order to verify the evolutionary significance of the α/β interface, a surface conservation analysis was conducted using the ConSurf server.[104–106] This analysis revealed that the charged interface residues that line the α/β interface are conserved, with conservation of GIIβ interface residues comparable to that of the active site residues on GIIα, indicating the evolutionary significance of the interface (Figure 10B). This conclusion is supported by the fact that site directed mutagenesis studies confirm disruption of the α/β interface when GIIα residue R840 is replaced with glutamic acid (R840E). When co-expressed with the hexahistidine-tagged GIIα-R840E mutant, GIIβ does not co-purify on immobilised metal affinity chromatography, suggesting loss of the α/β interface. This finding is consistent *in planta*, as the equivalent *Arabidopsis thaliana* (*At*) GIIα mutant, R787E, also loses capacity for *At*GIIβ and shows reduced localisation in the ER and only 10% of the wt enzyme activity *in vitro*.[54] These experiments demonstrate the importance of the interfacial

salt bridges, but it is possible that by replacing the positively charged arginine residue with a negatively charged glutamic acid may overstate the impact the single mutation has on the stability of the interface by introducing electrostatic repulsion where there should be attraction. It may have been appropriate to include additional R840 mutants whereby the replacement residue is of similar size to arginine but carries no charge, perhaps glutamine or methionine, as polar and non-polar examples, respectively. This would help to demonstrate the significance of the salt bridges by providing a control in which the interaction is simply absent, rather than introducing its opposite.

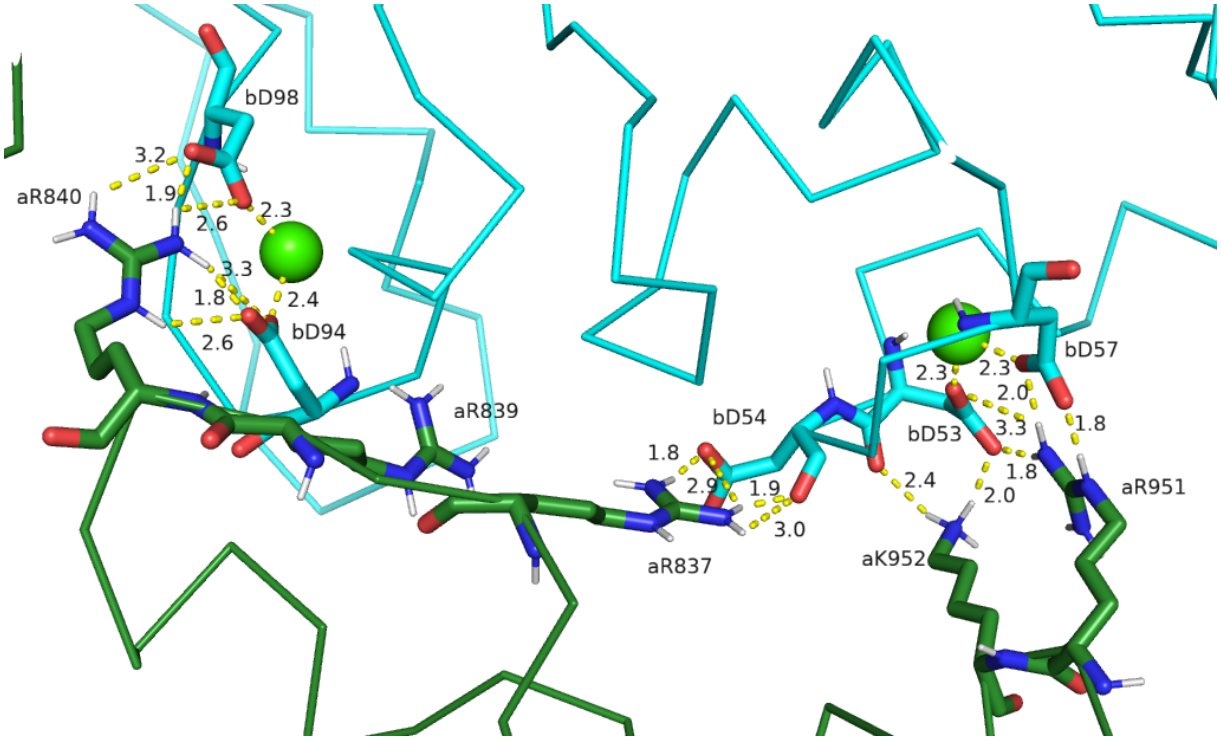


Figure 9 – A graphical representation of the key salt bridge interactions that dominate the binding energy of the α/β interface. Key residues are shown as sticks (GIIα – dark green: R837, R839, R840, R951, K952; GIIβ – cyan: D53, D54, D57, D94, D98) and hydrogen bond interactions are indicated as dashed yellow lines labelled with interatomic distances (Å).

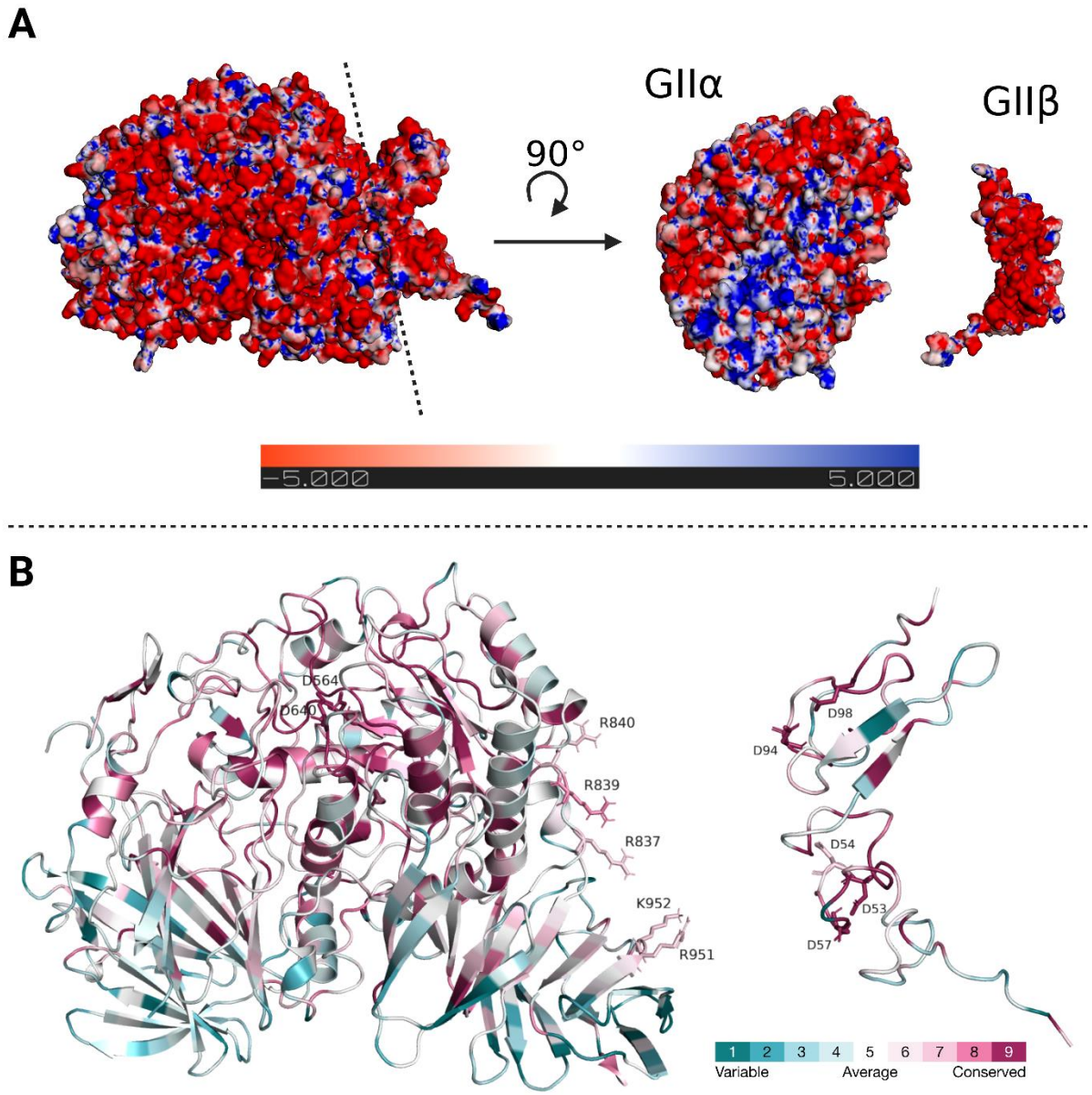


Figure 10 – Analysis of the partial α/β interface in the *Mm*GluII$_{Tryp}$ crystal structure. (A) Surface electrostatics calculated using the APBS-PDB2PQR server and visualised in Pymol.[103] The positive and negative charges of GIIα and GIIβ surfaces, respectively, indicate the ionic nature of the α/β interface. (B) Surface conservation analysis using the ConSurf server, visualised in Pymol.[104–106] Key interface and active site residues are shown as sticks and colour-coded on the basis of conservation. Residues involved in salt bridge interactions are moderately conserved in GIIα and stringently conserved in GIIβ, as the aspartic acid residues of GIIβ are also involved in coordination of the two Ca^{2+} ions.

While undeniably useful, the α/β interface described by the *Mm*GluII$_{Tryp}$ crystal structure tells us only part of the story. Trypsin digestion leaves GIIα mostly intact, but significantly truncates GIIβ, leaving only ~17% of the residues representing the LDLRa domains which remain bound to GIIα post-digestion. It is known that an important secondary interaction between the

substrate glycan(s) and the MRH domain of GIIβ contributes to full enzymatic activity, as discussed in Section 1.1.2.[63,65–67,72] It stands to reason that the true α/β interface spans a larger surface area than the 706 Å^2 observable in the *Mm*GluII$_{Tryp}$ crystal structure. Previous work conducted by Dr Alessandro T. Caputo, a former member of the Zitzmann group, and Dr Weston Struwe aimed to probe the full extent of the α/β interface for full length GluII by performing hydrogen-deuterium exchange (HDX) mass spectrometry (MS) experiments. By reasoning that residues that are significantly protected from HDX are potentially buried/not solvent accessible and could therefore be part of the full length α/β interface, Caputo, Struwe *et al*. constructed a map of both GluII subunits based on the relative HDX uptake of deuterium over the course of 60 minutes (data available in the supporting information of the referenced article).[54] An adapted version of this data is presented in Figure 11, which demonstrates the plausibility of an extended α/β interface relative to that shown in the *Mm*GluII$_{Tryp}$ crystal structure, given the high degree of protection from HDX shown by surface residues in close proximity to the LDLRa domains of GIIβ. Any PPI that uses these residues as points of contact between GIIα and GIIβ must be a transient one, as the LDLRa domains are the only sections of GIIβ that survive as a complex with GIIα following trypsin digestion.

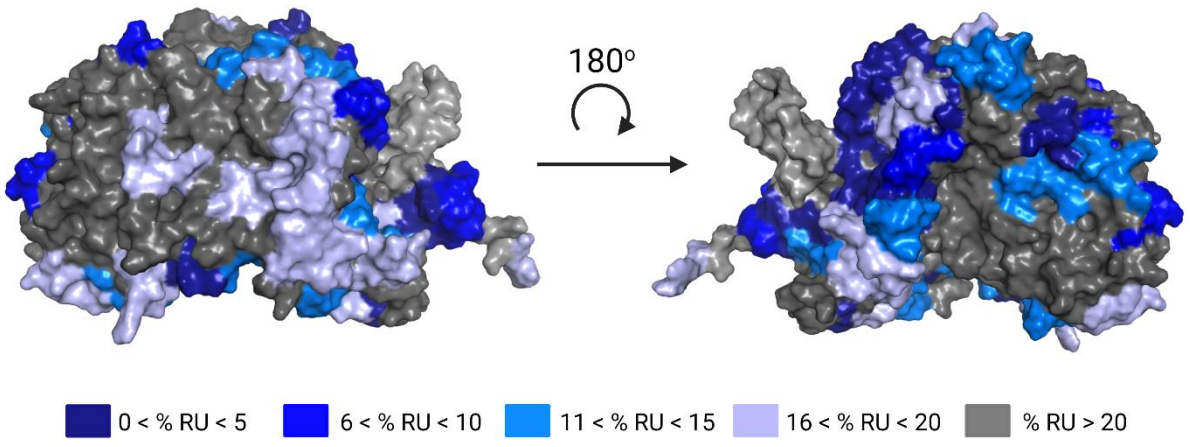


Figure 11 – Residues in GluII protected from HDX. The surface of GluII is colour coded according to the percentage relative uptake (% RU) of HDX in measured by MS. The regions most protected from HDX are located around the α/β interface and are likely to be buried in full length GluII, indicating potential sites of interaction between GIIα and GIIβ in full length protein that are not observed in the *Mm*GluII$_{Tryp}$ crystal structure.

1.4.2 Targeting the α/β Interface

Given the established importance of the GluII α/β interface for catalytic activity by GIIα, it is interesting to question the validity of targeting this structural feature with small molecules with the intention of disrupting the PPI to inhibit GluII and ultimately exert antiviral effect. In theory, disruption of the α/β interface could directly reduce the catalytic activity of GIIα by abrogating the putative interactions between the substrate glycan and the MRH domain of GIIβ, as shown *in vitro* by Stigliano *et al.* and Olson *et al.*.[63,65–67] A parallel mechanism for inhibition using this strategy lies in the separation of the ER/Golgi retrieval signal peptide, HDEL (located on the C-terminus of GIIβ), from the catalytic GIIα.[107] In theory, this would progress GIIα along the secretory pathway, causing it to abandon its post in the ER – thus removing it from CNX cycle. This may show ERQC inhibitory properties comparable to the inhibition of the enzyme by iminosugar active site inhibitors. The opportunity to explore this idea further presented itself somewhat serendipitously, when *Mm*GluII$_{Tryp}$ crystals were soaked with D-mannose and the crystal structure of the complex was determined (unpublished work conducted by Dr Pietro Roversi and Dr Alessandro T Caputo, former members of the Zitzmann group). This soaking experiment revealed two D-mannose molecules bound to GluII (Figure 12). One, unsurprisingly, was bound in the active site via hydrogen bonding to both catalytic aspartic acid residues, D564 and D640. The next and most intriguing of the D-mannose binding pockets is located at the α/β interface, with significant contacts made between the D-mannose ligand and both GIIα and GIIβ. Interestingly, this binding pocket is in close proximity with residues that form key salt bridges between the two subunits, particularly the interfacial arginine residues of GIIα, suggesting that it may be possible to exploit this binding pocket to identify compounds that interfere with the α/β interface.

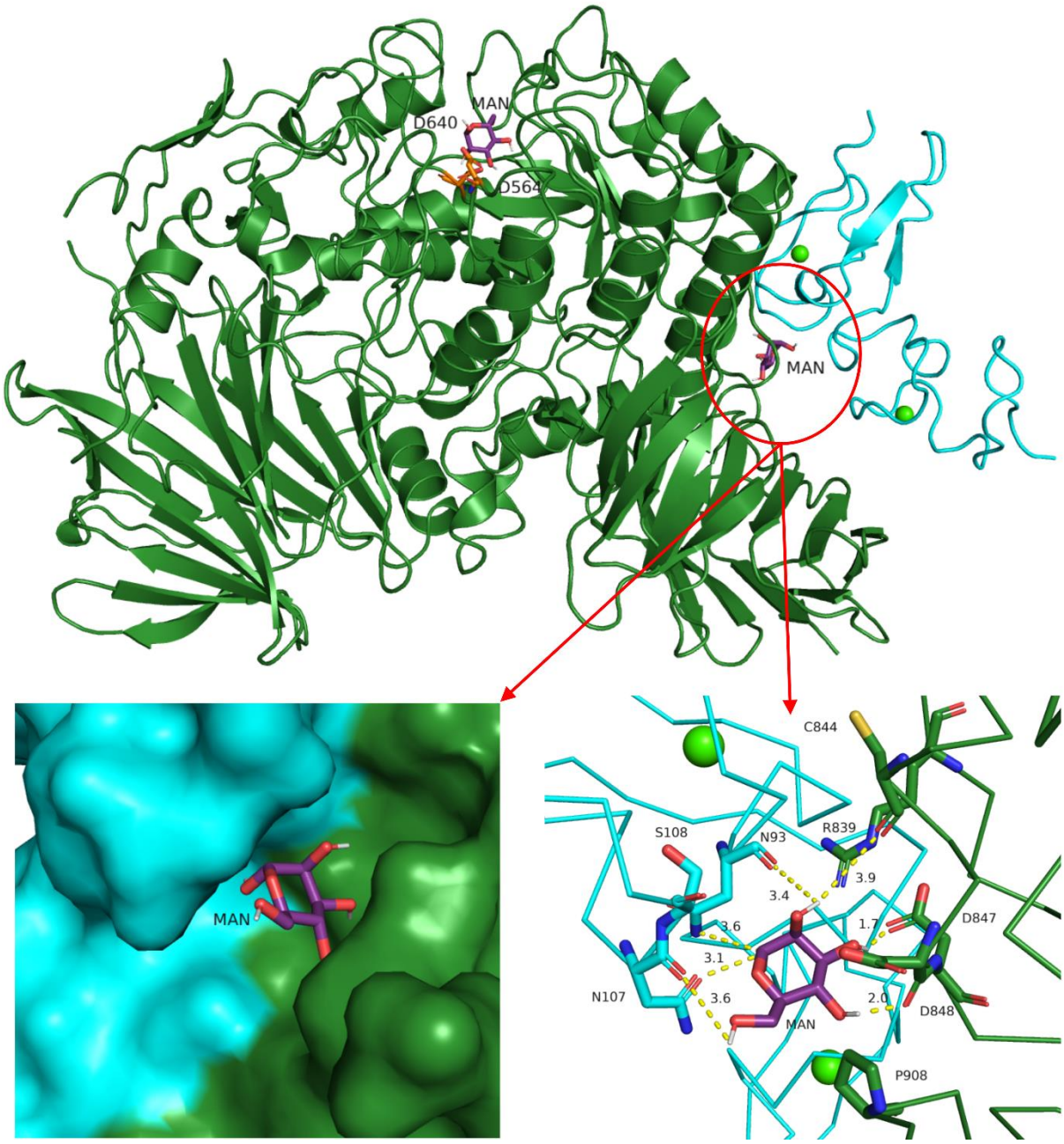


Figure 12 – Crystal structure of *Mm*GluII$_{Tryp}$ soaked with D-mannose. Two D-mannose ligands (purple sticks) can be seen; in the active site and in an interfacial pocket between GIIα and GIIβ. Zoomed in images of the interfacial D-mannose residue show the surface representation of the binding pocket (bottom left) and key protein-ligand interactions (bottom right). Residues within 4 Å of D-mannose are shown as sticks and likely hydrogen bonds are indicated as yellow dashed lines with interatomic distances labelled. Key salt bridge residues on GIIα are in close proximity to D-mannose, indicating the possibility of targeting this pocket with small molecules to disrupt the α/β interface.

Utilising the D-mannose pose seen at the α/β interface in the crystal structure of the GluII-D-mannose complex, the Zitzmann group, in collaboration with Evotec as part of a project funded by LAB282 (grant code: ALR02000), began to strategise about the type of

compound that could bind in the interfacial D-mannose binding pocket of GluII and ultimately disrupt the α/β interface. Two ligand binding strategies were considered:

1. The GIIα targeting strategy, in which the contacts observed between mannose and GIIα are maintained, but interactions with GIIβ are removed.
2. The GIIβ targeting strategy which aimed to maintain interactions with GIIβ but lose those with GIIα.

In both cases, the idea is to identify ligands that introduce electrostatic/steric clashes between GIIα and GIIβ, such that the α/β binding is weakened or lost. With this in mind, it would be desirable for these compounds to also form additional/alternative contacts with either GluII subunit that improve binding to the subunit to which they are anchored relative to that of the D-mannose ligand. Additionally, it was observed that the interfacial D-mannose residue is <4 Å away from the C844 residue on GIIα, which is not engaged in a disulfide bond to any other cysteine in the *Mm*GluII$_{Tryp}$ crystal structure. This prompted speculation over whether a weakly electrophilic moiety, a so-called 'covalent warhead', could be included in the target compound that allowed for covalent interaction between the protein and ligand via this free cysteine residue. The search for small molecules capable of binding covalently to a target protein in order to influence a PPI raised questions over the feasibility of such a task. These questions became the subject of an in depth literature review, summarised in Appendix 9.1, that I conducted in 2020.

To begin the search for potential PPI disrupting compounds, several types of compound were proposed. Firstly, to make use of the known interactions between mannose and the two GluII subunits, D-mannose analogues would be included in the search. While a useful starting point, this type of compound is potentially problematic owing to the presence of several D-mannose binding sites identified on GluII, let alone other proteins, which would somewhat defeat the point of seeking non-sugar mimetic GluII inhibitors. To identify alternative candidates with a

different molecular scaffold, a virtual screen (VS) of the mannose binding pocket was performed and computer-aided drug design (CADD) was conducted to optimise interactions with either of the two subunits of GluII, returning 82 compounds of interest to be considered for further investigation. Additionally, an Evotec library of 86 acrylamide compounds was considered to exploit the free C844 residue for covalent binding between the compound and GIIα. Inclusion of these covalent warheads was somewhat speculative as no information about the reactivity or solvent accessibility of C844 was known at the time and expectations of functional effect were understandably low. In addition to the D-mannose VS and acrylamide libraries, a suite of 107 fragments identified in an earlier crystallographic screen were also considered. Of these fragments, 41 were analogues of a fragment, NZ-IF3, known to bind GIIα near the interface. NZ-IF3 makes significant contacts with the GIIα residues F724, R834 and M836, placing it in a distinct binding pocket ~12-16 Å away from the D-mannose binding pocket and therefore not in the immediate proximity of the α/β interface. Soaking *Mm*GluII$_{Tryp}$ crystals with the remaining 66 fragments from the crystallographic screen yielded no complexes with *Mm*GluII$_{Tryp}$, but rather either destroyed the crystals during soaking experiments or induced loss of X-ray diffraction by the protein crystals. It was suggested that this may be indicative of α/β interface disruption which would in turn disrupt the formed crystal lattice and destroy the crystals. All told, 275 fragments were identified and considered as candidates for α/β interface disruption of GluII and the diversity of conceptual starting points (D-mannose VS, crystallographic hits, crystal disrupting compounds) translated to a wide variety of chemical scaffolds.

With this fragment library in hand, it became necessary to experimentally test their capacity for α/β interface disruption. A time resolved (TR) fluorescence resonance energy transfer (FRET) assay was developed and the compounds tested by Dr Mario Hensen, a former member of the Zitzmann group, again in collaboration with Evotec. Full details of the GluII TR-FRET assay can be found in Appendix 9.2. The FRET assay is a non-invasive technique

that monitors changes in proximity between two molecules by measuring the energy transfer from an excited donor fluorophore on the first to an acceptor fluorophore on the second, provided the acceptor is within 10 nm of the donor. To produce a FRET signal, the donor emission and acceptor excitation spectra must overlap (Figure 13A), which often complicates the readout by introducing artefacts into the FRET signal. Such artefacts must be minimised or removed by using experimental controls and/or data processing.[108,109] This technique requires the incorporation of fluorophores onto both protein binding partners in order to determine the proximity of their interaction. Commonly used fluorophores include fluorescent proteins, such as green, yellow or cyan fluorescent protein as well as organic dye compounds. Regardless of type, the properties of the donor and acceptor fluorophores must be tuned such that a sufficiently high percentage of the donor's emitted energy is transferred to the acceptor (the fraction of the energy transferred is known as FRET efficiency).[110] FRET can be applied to the identification and analysis of PPIs and their interactions with small molecules as the separation of two protein partners that occurs following PPI inhibition by some ligand will reduce/remove the FRET signal due to increased distance between the donor and acceptor fluorophores beyond the 10 nm limit for FRET (Figure 13B). *In vitro*, the energy transfer is measured using fluorescence microscopy, a technically challenging method that traditionally made HTS using FRET assays difficult. In recent years, many HTS-compatible FRET assays have been developed and described in the literature.[111,112] Combination of FRET with time-resolved fluorometry gives TR-FRET. This confers significant advantages over basic FRET, as the fluorescent lifetime of the donor fluorophore far exceeds that of background fluorescence that may interfere with the signal; a delay in acceptor fluorescence detection enhances reliability and signal-to-noise ratio.[111]

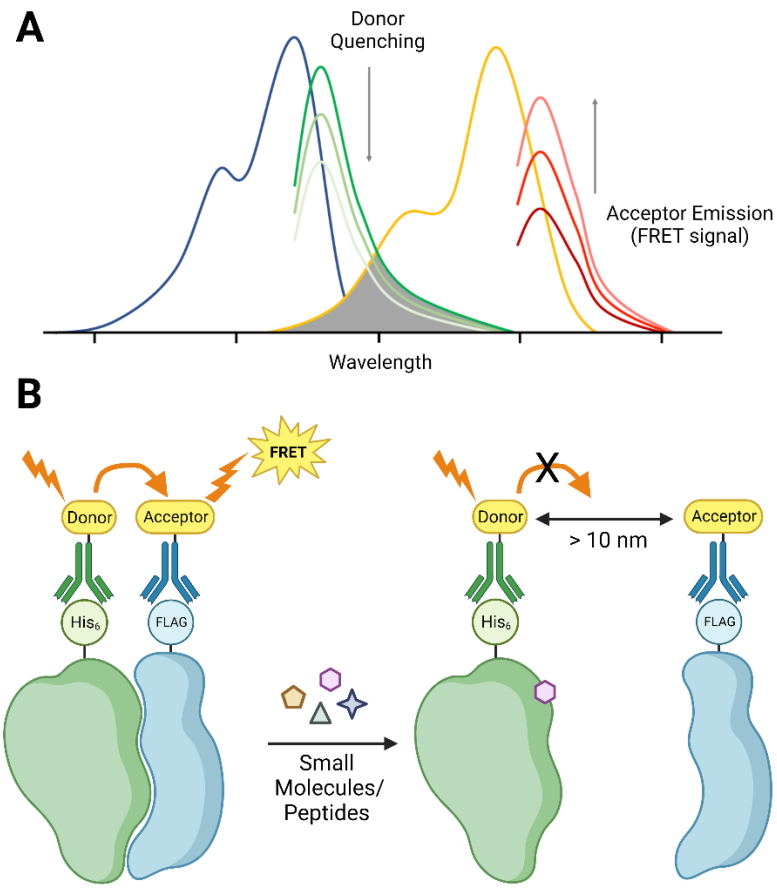


Figure 13 – Principles of the FRET assay. (A) The excitation of the donor fluorophore (blue trace) triggers fluorescence emission (green trace). Energy from this emission can be absorbed by the acceptor fluorophore (yellow trace) due to spectral overlap (grey area). The excited acceptor fluorophore emits energy by fluorescence (red trace), which is detected as FRET signal. Figure taken with permission from the DPhil thesis of Dr Mario Hensen. (B) High throughput screening using the TR-FRET assay. Protein binding partners are shown as green and blue shapes and are labelled with anti-His_6 and anti-FLAG monoclonal antibodies, as is the case for the GluII FRET assays. When the PPI is intact, FRET between the donor and acceptor fluorophores (yellow shapes) occurs. The PPI can be treated with a library of potential PPI disruptors. Successful PPI disruption increases the distance between the fluorophores beyond the FRET distance limit of 10 nm, so no FRET signal is observed.

An important consideration during TR-FRET assay development for measuring GluII α/β interface disruption was that the fragment screen should be as clinically relevant as possible, since the ultimate goal is to produce broad spectrum antiviral drugs for use in humans. Previous attempts by the Zitzmann group to express and purify human (*Hs*) GluII, had been unsuccessful, so a compromise was achieved by modifying the amino acid sequence of the recombinant *Mm*GluII by site directed mutagenesis such that the α/β interface more closely resembled that of the human protein. Sequence alignments of *Mm*GluII and *Hs*GluII revealed only three differences across the known area of the α/β interface, informing the production of

a *Mm*GIIα single mutant, F724G, and *Mm*GIIβ double mutant, L88P and S90N. These mutants were co-expressed to produce the so-called 'humanised' murine GluII (h*Mm*GluII) which was used in all subsequent experiments, including those presented for the first time in this thesis. From here, references to GluII, GIIα and GIIβ refer to the h*Mm*GluII mutant proteins unless specifically stated otherwise.

For the TR-FRET assays to assess the capability of small molecules to disrupt the GluII α/β interface, GIIα and GIIβ were tagged with hexahistidine (His_6) and FLAG tags, respectively. The tagged subunits were then treated with anti-His_6 and anti-FLAG monoclonal antibodies (mAbs) which were themselves tagged with donor/acceptor fluorophores (see Appendix 9.2). While the GluII heterodimer is intact, the donor and acceptor fluorophores will be within the distance limit for FRET (10 nm) and a signal will be measured. In the event of α/β interface disruption, the fluorophores will separate and the FRET efficiency will decrease, resulting in loss of signal. As a positive control for this assay, a version of GIIβ lacking the FLAG-tag was produced, allowing for a competition assay in which the untagged GIIβ can replace the FLAG-tagged GIIβ, resulting in a loss of FRET signal. This is also useful as it demonstrates that the kinetics governing the dissociation/association of the α/β interface allow it reform once lost, at least on the time scale of this TR-FRET assay (60 minutes). Additionally, this allowed for calculation of the K_d of the interaction between GIIα and GIIβ by performing a titration of 0.01 to 310 nM untagged GIIβ onto 3 nM GluII (GIIα-His_6, GIIβ-FLAG). This method yielded a K_d of 2.59 nM for the α/β interface under the tested conditions, illustrating a strong interaction between the two subunits. To further assist in the analysis of the TR-FRET primary assay, a counter assay was established in which GluII is replaced with a control peptide bearing both an N-terminal His_6-tag and C-terminal FLAG-tag (HHHHHHAQDYKDDDDK-CONH2). This peptide will bind both fluorophore tagged mAbs which cannot then be physically separated by interference with the peptide, resulting in a consistent FRET signal. This counter assay is useful as it allows for identification of fragments that may interact with the mAbs or

fluorophores in an unspecific manner, independent of GluII subunit interactions, which may give deceptive TR-FRET results.

Having gathered a library of compounds of interest, the Zitzmann group and Evotec began to screen the compound library using the TR-FRET assay described above to assess the ability of each fragment to disrupt the GluII α/β interface. Of the 275 compounds tested, 11 were found to have a stabilising effect on the α/β interface as illustrated by an increase in FRET signal with increasing fragment concentration. It is not clear why these fragments exerted such an effect and the TR-FRET assay provides little in the way of explanation given the lack of structural information provided. GluII α/β interface stabilisers may not be clinically relevant in the context of developing broad spectrum antiviral drugs, but may have biotechnical applications as tools to study the PPI that allow for more precise study of α/β interface kinetics in various experimental conditions. The α/β interface stabilising fragments are not explored further in this thesis. No FRET activity was observed for any of the 66 crystal breaking fragments. These compounds were not investigated further. More interestingly, 19 fragments showed positive TR-FRET results in the initial screen, indicating α/β interface disruption. Data from the TR-FRET assays were used to estimate the fragment concentration at which half maximal interface disruption, relative to the untreated protein (XC50) occurred, based on the activity of the control assay using the untagged GIIβ. The term IC50 is avoided here as it has not yet been established whether the α/β interface disruption would directly translate to enzyme inhibition. Of the 19 positive hits identified by TR-FRET, one was a structural analogue of the known crystallographic fragment NZ-IF3, which displayed modest FRET activity with a modest XC50 value of ~25 μM. The TR-FRET profile of this compound was unconvincing because of the similarity in response for both the primary and counter assays, suggesting that the positive FRET activity could be an artefact of the assay as opposed to genuine GluII interface disruption. To probe this further, 19 additional analogues of the fragment were acquired or synthesised and tested with TR-FRET. All but one were inactive, with only one

showing a 5-fold decrease in XC50 and slightly more convincing primary/counter assay comparison relative to the parent fragment. While this fragment seems to demonstrate genuine α/β interface disruption by TR-FRET, it was not investigated further in this thesis. The library of 86 acrylamide fragments yielded 15 modestly FRET active compounds. The most promising of these, EV-JJL001-052-001, showed an unexceptional XC50 value of ~80 μM, with most of the rest showing XC50 >500 μM. Following resynthesis of these acrylamide fragments by Evotec to improve their purity, none retained their weak FRET activity and they were subsequently disregarded for further investigation.

The most lucrative pool of fragments was the library derived from the mannose VS/CADD project. Three of the 82 fragments screened (with three unique chemical scaffolds) were FRET active; two were designed to follow the GIIα targeting strategy (i.e. binding to GIIα to induce steric/electronic clashes that outcompete GIIβ) while the third and least FRET active fragment was intended to follow the GIIβ targeting strategy. Of the two putative GIIα binding fragments, the most promising was a fragment with a thiopyridone scaffold, known as compound 60 (c60), which demonstrated an XC50 value of ~3.8 μM, roughly 10-fold lower than the next most active hit from this fragment series. Preliminary SAR studies investigated the effect of altering the substituents on the thiopyridone scaffold on FRET activity and revealed additional FRET active analogues. For one of these analogues, compound 61 (c61) the ester substituent was replaced with a carboxylic acid and this fragment demonstrated sub-micromolar XC50 (811 nM), demonstrating the value of continued SAR efforts and subsequent screening by TR-FRET. Eventually, a suite of thiopyridone compounds were evaluated. Many of these fragments were FRET active, with XC50 values ranging from 4.5-0.8 μM, as shown in Figure 14A. As the thiopyridone scaffold is predicted to bind in the interfacial D-mannose binding pocket and therefore in close proximity to the C844 residue on GIIα, researchers from the Zitzmann group and Evotec decided to probe the possibility of a disulfide bond between this free cysteine residue and the thiol on the thiopyridone ligands. Various analogues of c60 and

c61 were synthesised in which the sulfur atom is blocked, either directly via methylation or through steric clashes by methylation of neighbouring atoms, or replaced with oxygen. Interestingly, none of these compounds showed any meaningful FRET activity, indicating that the thiol is critical for the α/β interface disrupting activity of this chemotype (Figure 14B).

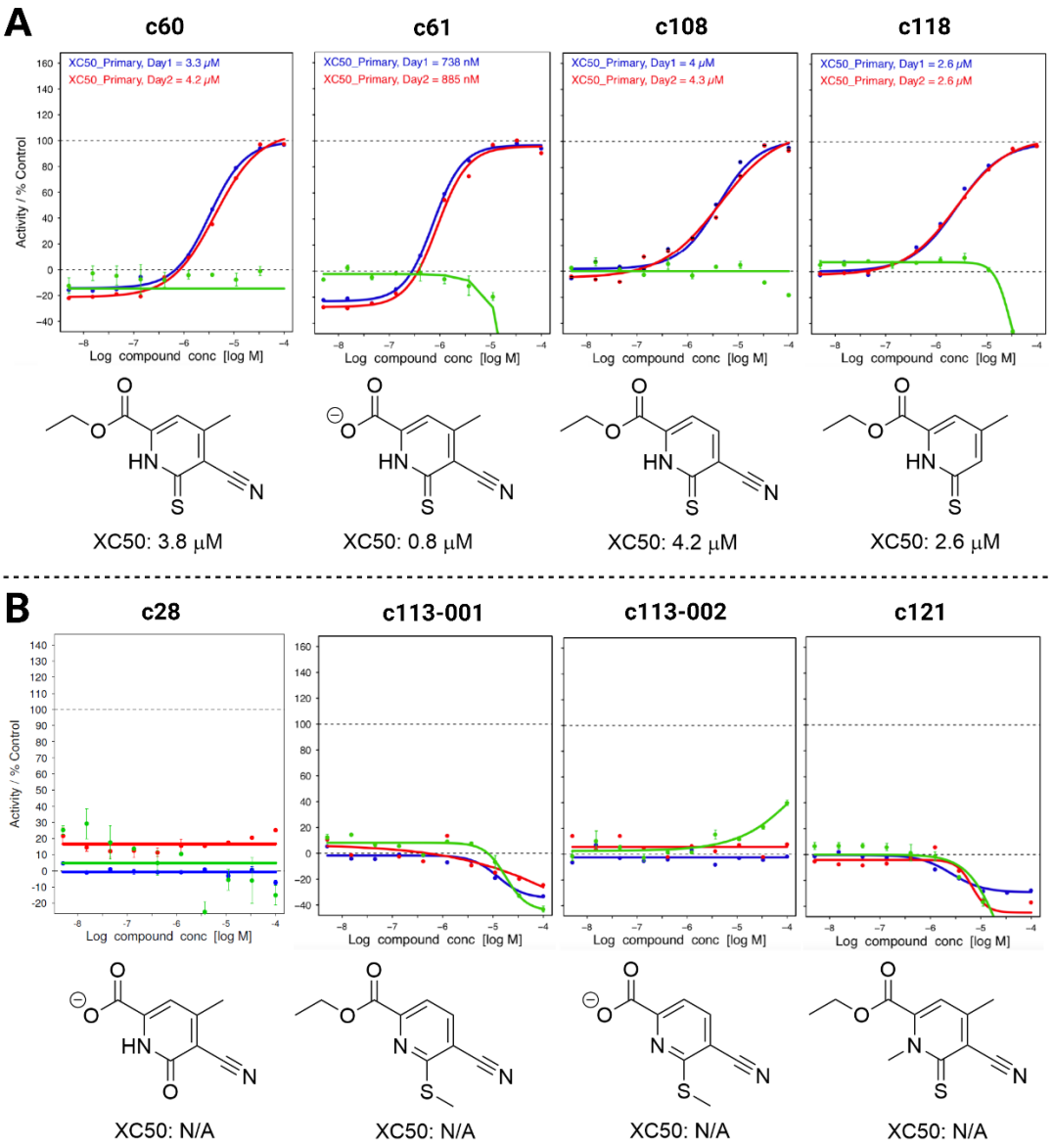


Figure 14 – TR-FRET measurements of GluII α/β interface disruption by thiopyridones. (A) Examples of FRET active thiopyridones with low micromolar potency against the GluII α/β interface. (B) Examples of FRET inactive thiopyridones in which the sulfur atom has been replaced by oxygen (c28), methylated (c113) or sterically blocked (c121). For both (A) and (B), TR-FRET activity was measured relative to a control where GluII is treated with increasing concentrations of an untagged GIIβ that competes off the GIIβ tagged with a fluorophore in a dose-dependent manner. The green trace indicates the results of the counter assay in which a peptide bearing N-terminal His_6 and C-terminal FLAG tags replaces GluII. Counter assay activity indicates compound interference with mAb binding. Blue and red traces indicate results of the primary assay collected on different days. Figure adapted from the DPhil thesis of Dr Mario Hensen with permission.

While the initial TR-FRET data were promising, several caveats limit the utility of these experiments. For example, several of the thiopyridones induce a negative response in the TR-FRET counter assay, indicating an increase in FRET efficiency for the control peptide. While it is unclear precisely why this happens, it is likely that the aromaticity of the thiopyridone scaffold may allow for absorbance by these compounds in the FRET range. Supporting this notion is the fact that many of the thiopyridone fragments are brightly coloured, typically red, orange or pink, indicating that they absorb visible light in the violet-blue range (400-470 nm). It is therefore plausible that the compounds could also absorb energy that contributes to donor quenching in TR-FRET and thereby interfere with both the primary and counter assays. For this reason, it was decided that these compounds would not be compatible with absorbance-based analytical techniques such as differential scanning fluorimetry. Furthermore, TR-FRET effectively informs us about GluII α/β interface disruption in a robust and reproducible manner, but gives little information about the binding mode or mechanism of action of thiopyridones against GluII.

Alternate techniques were attempted by the Zitzmann group to supplement the TR-FRET data. Firstly, the gold standard for identifying the ligand-protein binding mode would be to obtain high resolution X-ray crystallography data. Dr Mario Hensen performed several rounds of soaking experiments of *Mm*GluII$_{Tryp}$ and FRET active thiopyridones under various crystallography conditions, including those used to obtain the original *Mm*GluII$_{Tryp}$ crystal structure. Unfortunately, but perhaps unsurprisingly, these attempts were unsuccessful. In crystallographic terms, this is an inherently difficult system to work with as the compound of interest interferes with the crystal lattice by design as it disrupts the heterodimeric interface. Finding alternatives to the unsuccessful crystallography studies is still a valid avenue of research as they could categorically prove the binding mode of thiopyridones to GluII, thus streamlining all future medicinal chemistry optimisation of these lead compounds.

Many important questions around GluII α/β interface disruption as a novel method of GluII inhibition remain unanswered. TR-FRET provides convincing data to suggest that thiopyridones can disrupt the α/β interface *in vitro*, but limitations of the technique raise questions over whether the results are genuine or simply artefacts of an assay that involves molecules that potentially interfere with the energy transfer necessary for generating a FRET signal. It is therefore important to verify these results using alternative techniques. Furthermore, we do not yet know for certain that α/β interface disruption by thiopyridones actually translates to GluII inhibition *in* vitro, let alone *in vivo*, so it is necessary to establish GluII inhibition if these compounds can ever be considered as potential antiviral agents. Provided we can prove GluII inhibition by thiopyridones *in vitro*, it is then important to ascertain whether these compounds can cross the necessary cellular membranes to reach their target organelle, the ER, to exert GluII inhibition in living cells. Here, FOS assays developed by the Butters group are a useful tool, once the cytotoxicity of FRET active thiopyridones is measured to inform the choice of non-toxic concentrations to be used in the FOS assay; cytotoxicity of these compounds that has not yet been fully explored. The binding mode of thiopyridones also remains elusive at this stage and it is an important feature to identify if we hope to conduct medicinal chemistry optimisation of these fragments for eventual use in the clinic. All in all, it is still necessary to evaluate thiopyridones as an avenue towards GluII inhibition and ultimately their potential candidacy as broad spectrum, host targeting antiviral agents.

1.4.3 Covalent and Protein-Protein Interface-targeting Drugs

As discussed in Section 1.4.2, the Zitzmann group have identified a small molecule scaffold that appears to have a profound effect on the stability of a high affinity (K_d ~2.6 nM) PPI. Furthermore, it is a distinct possibility that this scaffold can engage in covalent interactions with the target protein, GluII, via a disulfide bond to a free cysteine residue. Understandably, this claim might raise concerns as it combines two traditionally avoided avenues of drug discovery: direct targeting of PPIs and incorporation of covalent protein-ligand interactions. To

address these concerns and assess the plausibility of this combination, I conducted an extensive literature review on the matter. The review is summarised in Appendix 9.1 and the full text can be provided in full on request.

2. Project Aims

2.1 Investigating Thiopyridones as α-Glucosidase II α/β Interface Disruptors

Previous work by the Zitzmann group used TR-FRET assays to assess the ability of various thiopyridones to disrupt the α/β interface of GluII. Several FRET active compounds were identified and the structural comparison of FRET active and inactive analogues suggest the possibility of covalent binding between these fragments and a free cysteine residue on GluII. The TR-FRET assays provide an excellent platform from which to further explore these compounds.

- Is the α/β interface disruption by thiopyridones legitimate, rather than an artefact of the FRET assay? Can we confirm disruption by using alternate techniques?
- Is the suggested covalent protein-ligand interaction legitimate?
- What is the most likely binding site of the covalent interaction?
- What is the effect of thiopyridone treatment on GluII activity *in vitro*?

2.2 Synthesis of Novel α-Glucosidase II Active Site Inhibitors

Thiopyridones represent an entirely novel class of GluII inhibitors that target the α/β interface. Even if they prove to be suitable for clinical study, which is far from guaranteed, several years of study are still required to reach this stage. In the meantime, it is pertinent to further explore

the tried and tested mechanism of inhibiting GluII to identify more potent and selective DNJ-derived iminosugars.

- Does the conjugation of DNJ to fragment 990, identified in crystallographic screening of *Mm*GluII$_{Tryp}$, yield iminosugars with inhibitory effects against GluII?
- Can I identify a practical synthetic strategy to achieve the synthesis of the target compounds?
- Do the linker length, nature of endocyclic heteroatoms or benzene substitution pattern of 990 influence the activity of the iminosugar?

2.3 Investigating the Role of the Mannose 6-Phosphate Receptor Homology (MRH) Domain in α-Glucosidase II Enzymatic Activity: An *In Silico* Analysis

Various publications have confirmed the contribution of the MRH domain of GIIβ to the full catalytic activity of GIIα. Current conceptual models suggest that the substrate glycan on the nascent glycoproteins will bind both the catalytic pocket of GIIα and the MRH domain of GIIβ, though there is currently no structural data to support this claim. The inherent flexibility of GIIβ and the transient nature of the enzyme:substrate complex make it a difficult system to gather structural data for, but *in silico* analysis may provide some insight.

- Is the current conceptual model of a glycan:GIIα:GIIβ complex sterically plausible when the glycan is attached to a nascent glycoprotein?
- What does the quaternary structure of full length GluII look like? Can we use computational modelling to estimate the general structure?
- Does the structural model suggest any additional binding pockets that could be targeted by small molecules to inhibit GluII?

3 General Materials & Methods - Protein Expression and Purification

Endogenous GluII is localised in the ER lumen as a result of the C-terminal HDEL peptide motif on GIIβ. For the sake of practical simplicity during protein expression, DNA constructs used here to produce GluII were designed to enable secretion of the recombinant enzyme and subsequent purification from the supernatant of the cell culture, as described in detail below. To this end, the pOPING vector was selected for expression as it carries a 31 residue secretion signal (MGILPSPGMPALLSLVSLLSVLLMGCVAETG) that was incorporated at the N-terminus of both the *GANAB* and *PRKCSH* inserts.[113] Furthermore, the HDEL sequence on GIIβ was removed to avoid the retention of the enzyme in the ER and a C-terminal hexahistidine (His_6) tag was incorporated on GIIα to enable purification by immobilised metal affinity chromatography (Section 3.3). DNA constructs described here were based on those previously employed by Caputo *et al*. and the modifications were validated in activity assays that confirmed the retention of catalytic activity of recombinant GluII.[54]

3.1 Transfection of Mammalian Cells

The GIIα/GIIβ in pHLsec/pOPING constructs were used to co-transfect FreeStyle™ HEK-293F cells (Thermo Fisher Scientific) for protein expression. A culture of HEK-293F cells with a cell density of $1.2x10^6$ cells/ml was transfected with a total of 1 μg/ml plasmid DNA. For co-transfection a GIIα:GIIβ molar ratio of 1:1.1 was used. In this project, the GIIα construct codes for a C-terminal His_6 tag. Cell transfection was carried out using a 1000:1 ratio of cell culture to transfection reagent (FreeStyle™ MAX - Thermo Fisher Scientific) and a total of 5% v/v (relative to cell culture volume) OptiPRO SFM medium (Thermo Fisher Scientific) following the manufacturer's protocol. Cells were incubated with shaking (110 rpm) for four days at 37°C and 8% CO_2 in polycarbonate vented Erlenmeyer flasks.

3.2 Harvesting Expressed Protein

Cells were removed from the culture by centrifugation at 3000 x g for 30 min, keeping the temperature <10°C to preserve protein stability, and the supernatant decanted. Phosphate buffered saline (PBS) was added to the stirring supernatant to a final concentration of 1X and the pH was adjusted to 7.4 using NaOH/HCl. To remove any residual cell debris, the supernatant was sequentially filtered through glass microfibre filters with 0.47 µm and 0.22 µm MF-Millipore™ Membrane Filters.

3.3 Immobilised Metal Affinity Chromatography (IMAC)

A 5 ml HisTrap™ excel column, prepacked with Nickel Sepharose excel affinity resins (General Electrics - GE Healthcare), was equilibrated with 1% buffer B (Table 1) and the filtered supernatant from Section 3.2 was applied at a flow rate of 5 ml/min. Next, the column was washed with 10 column volumes (CV) of 5% buffer B to remove any protein bound non-specifically to the column. To elute GluII bound to the nickel resin via the His_6-tag on GIIα, a step-elution with 70% buffer B over 10 CVs was performed. The eluate was collected in 1 ml fractions and the purification monitored using the UV absorbance at 280 nm. Peaks in the UV absorption profile were analysed by SDS-PAGE (see Section 3.6) to identify peaks containing GluII. Fractions positive for GluII by SDS-PAGE were combined in an Amicon® Ultra-15 centrifugal filter unit (10 kDa MWCO – Millipore) and concentrated by centrifugation at 4000 x g with simultaneous removal of imidazole by buffer exchange into 20 mM HEPES, 120 mM NaCl, pH 7.4. The final volume after concentration was dependent on the size of the sample loop used for size exclusion chromatography (Section 3.4).

Table 1 – Composition of buffers used during nickel affinity chromatography.

Buffer	Buffer Composition
A	1X PBS, 5% v/v Glycerol, pH 7.4

B 1X PBS, 5% v/v Glycerol, 500 mM Imidazole, pH 7.5

3.4 Size Exclusion Chromatography (SEC)

The concentrated protein sample from Section 3.3 was applied to either a Superdex™ 200 Increase GL 10/300 or HiLoad™ 16/600 Superdex™ 200 column, equilibrated with a 20 mM HEPES, 120 mM NaCl, pH 7.4. The choice of column was dictated by the size of the culture used to express the protein (Section 3.1). For larger cultures with higher protein yield, the HiLoad™ 16/600 Superdex™ 200 column was used, while smaller cultures called for the Superdex™ 200 Increase GL 10/300 column. The elution was monitored using the UV absorbance at 280 nm and 0.3 ml fractions were collected in a 96-well deep well plate. Peaks in the UV absorption profile were analysed by SDS-PAGE (see Section 3.6) to identify peaks containing GluII. Fractions positive for GluII by SDS-PAGE were combined and concentrated using a combination of Amicon® Ultra-15 and Ultra-0.5 centrifugal filter units until a recorded absorbance at 280 nm (A_{280}) > 4 was recorded on a Nanodrop®, as decribed in Section 3.5.

3.5 Determining Protein Concentration

A Nanodrop® 1000 (Thermo Fisher Scientific) was used for protein concentration determination. The device was set to detect absorbance at 280 nm (A_{280}) and measurements collected by applying 2 µl of sample to the objective. The protein concentration in mg/ml was determined by using the specific molar absorption coefficient (ε, M^{-1} cm^{-1}) and molecular weight (MW, Da) of the protein construct analysed (Table 2). Molar absorption coefficient and molecular weight were calculated by utilising the software tool ProtParam (ExPASy).[114]

Table 2 – Theoretical predictions of the molar absorption coefficients, ε, and molecular weights of GluII and its subunits according to the ProtParam software tool (ExPASy), The prediction for ε assumes that all paired cysteines form disulfide bonds.[114]

Protein	Molecular Weight / Da	ε / M^{-1} cm^{-1}
hMmGluII	166,600	242,490
hGIIα	108,044	184,650
hGIIβ	58,574	57,840

3.6 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was performed using NuPAGE® 4-12% Bis-Tris gels (Life Technologies) in NuPage® MOPS running buffer according to the manufacturer's protocol. Proteins samples were directly mixed with 4X NuPAGE® LDS Buffer (Life Technologies) and 10X NuPAGE® Reducing Agent. Samples were incubated for 5-10 minutes at 80°C to ensure full protein denaturation, then loaded onto the gel. Gels were run at 200 V for 40 – 50 minutes with Color Prestained Protein Standard (New England Biolabs) to estimate protein size. To visualise protein bands, gels were stained directly in QuickBlue Protein Stain (LubioScience) for at least 15 minutes.

Sample preparation for SDS-PAGE was performed as above, unless stated explicitly otherwise.

4. Investigating Thiopyridones as α-Glucosidase II α/β Interface Disruptors

4.1 Brief Introduction & Aims

The potential for thiopyridones to disrupt the GluII α/β interface has previously been identified by the Zitzmann group, as demonstrated by TR-FRET (see Section 1.4.2). The TR-FRET assays provide good evidence for α/β interface disruption and the initially identified FRET-active thiopyridone, c60, demonstrates low micromolar potency (XC50 = 3.8 µM) in this assay. TR-FRET assays were employed to examine the structure-activity relationships (SAR) of various thiopyridone analogues and identified one particularly potent interface disruptor, c61 (XC50 = 800 nM), which closely resembles c60, save for the presence of a carboxylic acid moiety in place of the ethyl ester on c60. The TR-FRET SAR studies also revealed disulfide bond formation as a possible mode of binding for this class of ligand, as interference with the thiol group on the thiopyridone scaffold invariably resulted in loss of FRET activity. Indeed, the scaffold was identified from a virtual fragment screen focused on an interfacial D-mannose binding site in close proximity to the C844 residue, which presents a free thiol in the *Mm*GluII$_{Tryp}$ crystal structure. Targeting this residue with a covalent interaction was an intended outcome of the virtual screen. The structures and TR-FRET profiles of key thiopyridones discussed in this thesis are presented in Figure 14.

The TR-FRET assays form a strong platform from which to further explore thiopyridones, but alone they are insufficient to describe and explain the GluII-ligand interaction. A key

disadvantage is that many of the FRET-active thiopyridones induce a FRET signal in the counter assay in which the fluorophore labelled GluII protein is replaced by a short synthetic peptide that can bind both the MAbs (at the N- and C-termini, respectively) that carry the donor and acceptor fluorophores. In this counter assay, the donor and acceptor fluorophores cannot be physically separated, so any FRET signal can be attributed either to interference with the peptide:MAb binding or donor quenching by the tested fragment. Thiopyridones often elicit a significant response in the counter assay (positive or negative), so it cannot be ruled out that the FRET signal in the primary assay is an artefact rather than legitimate GluII α/β interface disruption. Furthermore, TR-FRET does not provide information regarding the thiopyridone binding mode, so there is still no direct evidence of the protein-ligand disulfide connection predicted by the preliminary SAR studies (Zitzmann group/Evotec – unpublished).

In this thesis, I will investigate the effect of thiopyridone treatment on the GluII heterodimer with the intention of establishing the chemotype as a legitimate clinical lead for the inhibition of GluII to elicit an antiviral effect. Firstly, it is important to corroborate the findings of the TR-FRET assays using alternative techniques to confirm legitimate α/β interface disruption by thiopyridones *in vitro*. Here, I use size exclusion chromatography-multi-angle light scattering (SEC-MALS), MS and mass photometry (MP) to monitor GluII α/β interface disruption by FRET-active thiopyridones. It is important to demonstrate the binding mode of thiopyridones to inform any medicinal chemistry modification of the scaffold to optimise the pharmacodynamics and pharmacokinetics. MS is used in attempt to identify the covalent modification associated with the predicted protein-ligand disulfide bond. In parallel, I gathered indirect evidence for this interaction using MP and native MS in combination with site directed mutagenesis of GluII to identify which, if any, cysteine residue may be involved in this disulfide interaction (Sections 4.4.1-4.4.5).

I will also explore the effect of α/β interface disruption by thiopyridones on the enzymatic activity of GluII *in vitro*. It is not known if the capability of thiopyridones for interface disruption

translates to GluII inhibition, let alone antiviral efficacy. To examine this, I used HPLC to analyse the glycan content of the HIV envelope glycoprotein, gp120, following treatment with GluI and GluII in the presence and absence of c61, the most potent interface disrupting thiopyridone currently known. This experiment uses a native substrate of GluII, gp120, to assess whether treatment of GluII with thiopyridones and subsequent interface disruption influences the glycan content of the viral glycoprotein. I also used an MTS assay to assess the viability of Huh7 cells in the presence of thiopyridones at various concentrations and establish an estimate of cytotoxic concentrations. The results from the MTS assay informed the choice of thiopyridone concentration used in FOS assays used to determine whether the α/β interface disruption by thiopyridones observed *in vitro* is consistent in live cell assays (Section 4.4.6-4.4.7).

Currently, no structural data exists for the full length GluII protein and due to the putative α/β interface disruption by thiopyridones, gathering structural data of a GluII:thiopyridone complex to elucidate the binding mode using X-ray crystallography is challenging. Here, I present my efforts to obtain these structural data by cross-linking the two GluII subunits to prevent subunit separation and perhaps obtain full length GluII crystals, though time constraints limited my progress in this regard (Appendix 9.10).

4.2 Experimental Principles

4.2.1 Mass Photometry

Mass photometry (MP, Refeyn Ltd.) is a relatively novel technique that formed the basis of much of my study of the GluII α/β interface. MP, originally introduced as interferometric scattering mass spectrometry (iSCAMS) by Young *et al*. in 2018, is an optical technique that enables molecular counting on a single molecule level without the need for molecular labelling of any kind.[115] The technique makes use of the correlation between the polarisability of a molecule (proportional to the particle volume) and the extent to which it scatters incident light.

By applying the approximation that each amino acid can be treated as a single nano-object and the observation that amino acid volume and protein refractive indices show little variability (~1%), Young *et al.* demonstrated that the number of amino acids, and by extension the protein mass, are proportional to the scattering signal.

To detect the scattering signal from proteins in solution, MP builds upon methodology originally published as interferometric scattering microscopy (iSCAT) which utilises a modified confocal scanning microscopy experimental set up.[116] Here, a laser of instrument dependent wavelength (488 nm for the TwoMP mass photometer, the current state-of-the-art instrument from Refeyn Ltd.) is used as the light source. A glass coverslip is placed on a stage above the microscope objective and is loaded with a small volume (< 20 µl) of aqueous buffer. The position of the stage can then be adjusted to set and lock the focus of the microscope at the glass/water interface. A dilute protein sample (100 pM – 100 nM) is added to the buffer droplet and protein molecules will diffuse to the glass surface to bind non-specifically to the coverslip in what are known as 'landing events'. The landing events will result in elastic scattering of the incident light by the landing particles and the interference of the scattered and reflected light is detected and measured as ratiometric contrast which scales linearly with the mass of the landing particle.[115] Bespoke MP acquisition software (AcquireMP, Refeyn Ltd.) enables measurement of hundreds to thousands of landing events over a minute-long timescale (for measuring a sample of soluble protein) and data analysis software (DiscoverMP, Refeyn Ltd.) produces an interferometric contrast/molecular weight frequency distribution plot. Representative examples of MP output data are shown in Appendix 9.9.

Since its inception, MP has proven to be a particularly useful technique for studying proteins at a single molecule level. The technique provides high mass accuracy (< 5 kDa deviation between measured and sequence masses in a sample of bovine serum albumin (BSA)) that is independent of molecular shape.[115] Resolution, defined as the full width at half-maximum (FWHM) of the measured contrast/molecular weight, is sufficient to reliably distinguish

between protein molecules of markedly different size, but is limited by photon shot noise and influenced by molecular weight, with increased uncertainty for proteins at the lower mass range (TwoMP mass photometer, FWHM: 25 kDa at 66 kDa, 60 kDa at 660 kDa). A key benefit of MP is that it is an optical-only, non-destructive strategy that uses a relatively simple experimental set up and does not require any specific molecular properties or any type of analyte labelling which can complicate experimental workflow. This allows for reliable and practically straightforward imaging of a dilute sample of weakly scattering biomolecules whereby the stoichiometry, dynamics and energetics can be studied in heterogeneous mixtures.[117] Recently, MP has been shown to have a broad range of applicability beyond the realm of soluble proteins. Integral membrane proteins (IMPs) are notoriously difficult to study due to their aqueous insolubility which necessitates the use of various membrane-mimetic systems (MMS), including detergent micelles, amphipols, nanodiscs and lipid nanoparticles, to enable their study.[118] In 2021, Olerinyova *et al.* used MP to characterise membrane proteins and IMPs in solution and explore the subtleties of the inherently heterogeneous MMS in which they are solubilised at sub-chromatographic resolution.[119] Furthermore, Foley *et al*. recently used dynamic MP to not only measure the mass of single proteins associated with supported lipid bilayers, but also to track their diffusion across the membrane and monitor the formation of oligomeric biomolecular complexes.[120] The accurate mass of DNA has also been successfully identified using MP with up to 2 base pair accuracy, demonstrating the utility of the technique with biological macromolecules other than proteins.[121]

Perhaps the most obvious benefit of MP with respect to the GluII α/β interface is that it can rapidly reveal the binding affinities, dynamics and energetics of the PPIs involved in protein complexes. By distinguishing species based on their mass and using molecular counting to assess the relative abundances of complexes and their unbound component proteins, it is possible to extract kinetic and energetic information about the PPIs under investigation. This can be easily applied to simple stoichiometric systems, with examples in the literature

identifying the oligomerisation patterns of well-known proteins like BSA (Young *et al*., 2018), as well as quantifying monomer-dimer equilibrium of αβ-tubulin, which has implications in the biosynthesis of microtubules (Fineberg *et al*., 2020).[115,122] More complex multicomponent systems can also be analysed by MP, as shown by Soltermann *et al*. who investigated the immunoglobulin G:neonatal Fc receptor interaction, which consists of up to five different interacting species.[123]

Despite its considerable benefits, MP is not without drawbacks. For example, MP is limited to low protein concentrations (ideal range < 100 nM) that allow spatial separation of landing events within the field of view of the mass photometer; a requirement for well resolved contrast images. This limits measurable binding affinities of protein complexes to sub-micromolar values as reliable measurements of species populations require the bound complex to comprise at least 5% of the total protein concentration.[124] Furthermore, the lower mass limit of MP is currently around 30 kDa, but the lower relative resolution (FWHM) associated with smaller species makes distinguishing them in a heterogeneous sample less reliable. Nevertheless, the ability of MP to precisely monitor dynamic PPIs of varying complexity in real time on a single-molecule level makes it an exciting development in PPI analysis.

4.2.2 Circular Dichroism

Circular dichroism (CD) is an established technique for determination of protein secondary structure and is particularly useful for rapidly assessing the quality of protein folding for recombinant proteins and comparing their secondary structure with those bearing artificially introduced mutations.[125] A CD experiment involves exposing a protein sample to a beam of circularly polarised light, which occurs when the direction of the electric field vector rotates about its propagation direction while retaining constant magnitude. When viewed down the propagation axis, this can be visualised as rotation in a clockwise (E_R) or anticlockwise (E_L) direction. These two circularly polarised waves are out of phase with one another by 90

degrees and can be separated when passed through a certain medium, such as a prism. In the context of protein analysis, CD experiments exploit the asymmetry of protein secondary structure, which will differentially absorb E_R and E_L circularly polarised light with different indices of refraction for each wave. The addition of E_R and E_L vectors following their interaction with an asymmetric object results in an elliptically polarized vector and the degree of ellipticity can be measured. As a result, elements of protein secondary structure show characteristic CD spectra where proteins with α-helices show negative bands at 222 nm and 208 nm and a positive band at 193 nm, whereas well-defined antiparallel β-pleated sheets have negative bands at 218 nm and positive bands at 195 nm. Disordered proteins display low ellipticity at wavelengths >210 nm and negative bands near 195 nm.[126] These characteristic CD profiles allow for the rapid and qualitative assessment of a proteins secondary structure which can be used to evaluate the relative quality of fold for recombinant mutant proteins with respect to their wt parent protein. CD is not capable of determining the residue-specific detail of a protein fold and is less useful for analysis of polypeptide regions with variable structure, such as loops, as the measurement relies on the differential absorption of E_R and E_L by asymmetric objects like α-helices and β-sheets.

4.2.3 Differential Scanning Fluorimetry

Differential scanning fluorimetry (DSF) is a biophysical technique that measures the thermal stability of proteins with high throughput capability and is also referred to as the thermal shift assay.[127] Proteins in solution exist in a thermodynamic equilibrium between their folded and unfolded states. Incrementally increasing the temperature of the solution will shift the position of equilibrium to favour the unfolded state, allowing for quantification of the melting temperature (T_m) which is defined as the temperature at which the ratio of folded to unfolded protein is 1:1. By altering the protein or its environment (i.e. ionic strength of solution, pH, presence of certain ions), it is possible to measure any changes in stability of the protein fold as a difference in T_m, or thermal shift (ΔT_m), which represents a change in the Gibbs free

energy of the protein. To obtain a suitable thermal shift value, it is necessary to monitor the thermal unfolding of the protein over the designated temperature range. Typically, this is achieved by measurement of extrinsic fluorescence, whereby a fluorescent dye is added to the protein sample. In aqueous conditions and in the presence of a natively folded protein, the fluorescence of the dye is quenched, however as thermal unfolding progresses, the dye is capable of interacting with exposed hydrophobic residues and a fluorescence signal increases. To ensure an appropriate difference between the fluorescence background of the sample and the measured signal during the experiment, this method relies on the assumption that the protein of interest does not, in its native state, possess any significantly hydrophobic patches on its surface and is stable with respect to aggregation at the lowest tested temperature. Additionally, it is important that the native protein or other soluble components (e.g. ligands) do not engage in any significant interaction with the dye, which may increase the fluorescence background of the sample by shielding the quenching of the dye prior to protein unfolding.

DSF has proven incredibly useful since its inception in 1997, with a number of applications in protein biochemistry and drug discovery. In essence, the technique is designed to measure relative changes in stability of proteins in various conditions. By analysing the ΔT_m of proteins of interest in various buffers, it is possible to identify the conditions in which the protein is most stable (most positive ΔT_m with respect to a reference condition), enabling the optimisation of purification buffers for recombinantly expressed protein. Furthermore, the same application can be used to identify the most promising buffers for the purposes of crystallography, where a more stable protein will be more likely to crystallise and achieve high resolution structural data.[127] Indeed, this approach was employed by Caputo *et al*. to obtain crystal structures for *Mm*GluII$_{Tryp}$.[54] Another useful application is for identification of stabilising ligands in high-throughput screening.[128] Significant increases in protein T_m following addition of a small molecule suggests significant protein-ligand interactions that promote protein stability and

these small molecules may serve as a starting point in drug discovery. DSF can also be used to evaluate the impact of site directed mutagenesis on the tertiary structure of the protein as considerable stabilisation or destabilisation could be measured as significant ΔT_m, compared to the wt protein.

4.3 Materials & Methods

4.3.1 Size Exclusion Chromatography-Multi-Angle Light Scattering (SEC-MALS)

SEC-MALS experiments were run using GluII previously expressed and purified by immobilised metal affinity chromatography (Section 3.3) and SEC (Section 3.4). For initial SEC purification of h*Mm*GluII, a Superdex™ 200 Increase GL 10/300 column was used with PBS as the eluent.

For SEC-MALS, 8.3 µM GluII in PBS was incubated for 90 minutes at 37°C with either c61, c28, c113-002 (1.8% v/v from a 100 mM stock in DMSO) or pure DMSO (1.8% v/v). The protein/compound sample was applied to SEC on an analytical Superose™ 6 10/300 GL column (GE Healthcare) equilibrated with PBS at a flow rate of 0.5 ml/min. In this SEC experiment MALS was performed during elution, which was monitored via online static light-scattering (DAWN HELEOS 8+, Wyatt Technology), differential refractive index (Optilab T-rEX, Wyatt Technology) and UV (SPD-20A, Shimadzu) detectors. Data were analysed using the ASTRA software package v6 (Wyatt Technology).

4.3.2 Mass Spectrometry (MS)

Sample Preparation: To prepare samples for MS analysis, GluII was first subjected to SDS-PAGE (Section 3.6) to separate GIIα from GIIβ. Bands in the SDS-PAGE gel corresponding to either GluII subunit were excised and washed with Milli-Q water before removal of the QuickBlue Protein Stain (LubioScience) by sequential washes with 1:1 Milli-Q

water:acetonitrile and 1:1 100 mM ammonium bicarbonate (Ambic):acetonitrile. Destained gel pieces were dried on a Concentrator plus Centrifuge Concentrator (Eppendorf), then rehydrated with 10 mM dithiothreitol (DTT) in 100 mM Ambic and incubated for 45 minutes at 56°C with shaking at 650 rpm. After incubation, the mixture was cooled to room temperature and excess DTT solution was removed, then 55 mM iodoacetamide (IAA) in 100 mM Ambic was added and the sample was incubated in the dark at room temperature for 30 minutes. Excess IAA solution was then removed. For sample preparations using RapiGest SF (Waters), enough 0.1% RapiGest solution in 50 mM Ambic was added to cover the gel pieces and the sample was incubated at 37°C for 10 minutes. Excess solution was removed from the gel pieces which were then dried completely in a Concentrator plus Centrifuge Concentrator (Eppendorf). Dry gel pieces were rehydrated with a solution of digestion enzyme (trypsin, chymotrypsin or α-lytic protease – 12.5 ng/µl in 50 mM Ambic) and incubated on ice for 45 minutes. Excess solution was removed and replaced with enzyme-free 50 mM Ambic before overnight incubation at 37°C. The digestion buffer was then removed and placed into a clean Eppendorf tube and enough 25 mM Ambic to cover the gel pieces was added and the sample was incubated at room temperature with shaking (650 rpm) for 5 minutes, before addition of an equal volume of acetonitrile and a further 15 minutes of incubation with shaking. The supernatant was removed and pooled with the digestion buffer. The gel pieces were covered with a sufficient volume of 5% v/v formic acid in Milli-Q water and incubated with shaking for 5 minutes before another addition of an equal volume of acetonitrile and further 15 minutes of incubation with shaking. The formic acid/acetonitrile mixture was pooled with previous washes and the formic acid wash was repeated one more time. Combined sample washes were dried completely in a Concentrator plus Centrifuge Concentrator (Eppendorf) and resuspended in an appropriate volume of 0.05% v/v trifluoroacetic acid (TFA) in Milli-Q water followed by sonication for 5 minutes in a water bath sonicator. For RapiGest treated samples, higher TFA concentration (0.5% v/v) was used and an additional 45 minute incubation at 37°C was required to remove the bound RapiGest anionic surfactant. These samples were then

centrifuged at 13,000 rpm for 10 minutes to separate the water insoluble hydrolytic RapiGest products. The supernatant was transferred to a fresh HPLC vial and submitted for MS analysis.

Preparation of MS samples was carried out by myself and MS experiments themselves were kindly performed by Dr Bevin Gangadharan (Zitzmann group).

Liquid Chromatography: Peptides were separated on a Dionex Ultimate 3000 nano UHPLC system (Thermo Scientific). A nano analytical C18 reversed phase EasySpray column was used with dimensions 75 µm x 50 cm, 2 µm particle size (Thermo Scientific) with a flow rate of 250 nL/min at 55 °C. The mobile phase used was solvent A: 0.1% v/v formic acid in LC-MS grade water, and solvent B: 0.1% v/v formic acid in 80% v/v acetonitrile. The gradient used to separate peptides on the analytical column was: 2% B (0-6 min), 2-36% B (6-105 min), 36-60% B (105-125 min), 60-95% B (125-126 min), 95% B (126-131 min), 95-2% B (131-132 min) and 2% B (132-147 min) for column equilibration.

Mass spectrometry: Peptides from the nano UHPLC were analysed on a benchtop Q Exactive hybrid quadrupole-Orbitrap mass spectrometer (Thermo Scientific) using the EasySpray ion source. Prior to data acquisition, the mass spectrometer was calibrated for mass accuracy according to the manufacturer's recommendations using a positive ion calibration solution injected at 5 µl/min into a heated electrospray ionisation (HESI) probe (Thermo Scientific).

Data dependent acquisition: The conditions for data dependent acquisition (DDA) were as follows: chromatographic peak width was set at 20 s and the full MS conditions were with a resolution of 70,000, AGC target of 3e6, maximum IT (injection time) of 60 ms, scan range of 375 to 1500 m/z. The dd-MS2 conditions were with a resolution of 35,000, AGC target of 1e5, maximum IT of 60 ms, loop count of 10 (i.e. Top 10), isolation window of 2.0 m/z, fixed first mass of 120.0 m/z and normalised collision energy at 27 in a high-energy collision dissociation (HCD) cell. The data-dependent (dd) settings were with minimum AGC target: 2e3, minimum

intensity threshold of 3.3e4 ions, charge exclusion: unassigned, 1, 8, >8, peptide match: preferred, dynamic exclusion: 30 s.

MaxQuant: MaxQuant (version 2.0.3.0; Max Plank Institute of Biochemistry) was used to search the *.raw DDA files against a custom FASTA file for either humanised GIIα or GIIβ depending on the sample. Variable modifications were set as Oxidation (M), Carbamidomethyl (C) and Acetyl (Protein N-term). Additional custom variable modifications were also included for c61 native (C; C(8) H(3) N(2) O(2) S; +190.9915230409), c61 acidified (C; C(8) H(4) N(2) O(2) S; +191.999348073), c61 1x reduced (C; C(8) H(5) N(2) O(2) S; +193.0071731051), c61 2x reduced (C; C(8) H(7) N(2) O(2) S; +195.0228231693), c61 acidified 1x reduced (C; C(8) H(6) N(2) O(2) S; +194.0149981372) and c61 acidified 2x reduced (C; C(8) H(8) N(2) O(2) S; +196.0306482014). For digestion the protease was set as either trypsin (up to 2 missed cleavages), chymotrypsin (up to 4 missed cleavages) or alpha lytic protease (custom protease for TASVGIK with up to 4 missed cleavages). Instrument type was set to Orbitrap using the default settings except for the peptide tolerance at ±10 ppm. Minimum peptide length was set as 8 and maximum peptide length as 45. All other parameters were left as the default settings.

Sequence coverage: Protein Coverage Summarizer (version 1.3.6794.31818, Pacific Northwest National Laboratory) was used to compute the sequence coverage percentage based on the observed peptides for both GIIα and GIIβ. The peptide sequences identified for GIIα and GIIβ were copied from the peptides.txt file from MaxQuant. A *.txt file was created with these list of peptides and imported into the software along with either a custom FASTA file for the humanised GIIα or GIIβ depending on the sample.

Isolation list creation: Skyline software (version 20.2.0.343, 64-bit, Seattle, USA) was used to create *.csv isolation lists in Thermo Q Exactive format for GIIα tryptic peptides covering C822. *In silico* digestion for GIIα was achieved by using the following parameters in 'Peptide Settings':- Enzyme: trypsin; peptide length: 6 to 45 amino acids (aa) long; missed cleavages:

two. The peptides added to the isolation list were SSDCMKDDPITLFVALSPQGTAQGELFLDDGHTFNYQTR and RSSDCMKDDPITLFVALSPQGTAQGELFLDDGHTFNYQTR since these miscleaved unmodified peptides were the only peptides seen by DDA covering C822 and peptide SSDCMK was not observed. The isolation list only included these peptides with the methionine not oxidised since this was predominantly observed by DDA. Both unmodified and c61-modified peptides as well as +3, +4 and +5 charged peptides were included in the isolation list for native c61 ($C_8H_3N_2O_2S$; monoisotopic mass 190.991523) and acidified c61 ($C_8H_4N_2O_2S$; monoisotopic mass 191.999348). The Skyline file was saved as *.sky for data analysis.

Parallel reaction monitoring: Peptides were targeted by LC-MS using the Dionex Ultimate 3000 nano UHPLC system with similar settings as described above and the Q Exactive using parallel reaction monitoring (PRM). MS settings for PRM acquisition were: global settings – user role: advanced; lock mass: best; chromatographic peak width: 20 s; t-MS2 settings – polarity: positive; in-source collision induced dissociation (CID): 0.0 eV; default charge: 2; inclusion: on; microscan: 1; resolution: 70000; AGC target: 1e6; Maximum injection time: 240 ms; MSX count: 1; isolation window: 1.6 m/z; normalised collision energy: 27; spectrum type: profile. Full MS settings were as follows: scan range: 350-2000 m/z, fragmentation: none, resolution: 17,500, polarity: positive, microscan: 1, AGC target: 5e4, maximum IT: 20 ms. The isolation list from Skyline was imported into the inclusion list of the method. Various inclusion lists were used depending on the sample and contained the following precursor ions:

SSDCMKDDPITLFVALSPQGTAQGELFLDDGHTFNYQTR (m/z 1440.0071, +3; m/z 1080.2571, +4; m/z 864.4072, +5), SSDC[c61 native]MKDDPITLFVALSPQGTAQGELFLDDGHTFNYQTR (m/z 1503.6709, +3, m/z 1128.0050, +4; m/z 902.6055, +5), SSDC[61 acidified]MKDDPITLFVALSPQGTAQGELFLDDGHTFNYQTR (m/z 1504.0069, +3; m/z 1128.2570, +4; m/z 902.8070, +5), RSSDCMKDDPITLFVALSPQGTAQGELFLDDGHTFNYQTR (m/z 1492.0408, +3; m/z 1119.2824, +4; m/z 895.6274, +5), RSSDC[c61 native]MKDDPITLFVALSPQGTAQGELFLDDGHTFNYQTR (m/z 1555.7046, +3; m/z

1167.0303, +4; m/z 933.8257, +5), RSSDC[61 acidified]MKDDPITLFVALSPQGTAQGELFLDDGHTFNYQTR (m/z 1556.0406, +3; m/z 1167.2823, +4; m/z 934.0273, +5).

Skyline data analysis: The Skyline software was used to analyse PRM data by opening the *.sky files saved for isolation list creation. The acquisition method was changed to 'Targeted', the product mass analyser to 'Orbitrap' and the resolving power to 70,000. In transition settings the ion charges included were 1, 2, 3 and 4. The PRM *.raw files were imported and in most cases the automatic MS2 peak integration was used. In the case of any incorrect peak picking, the MS2 peaks were manually integrated. The unmodified peptides SSDCMKDDPITLFVALSPQGTAQGELFLDDGHTFNYQTR and RSSDCMKDDPITLFVALSPQGTAQGELFLDDGHTFNYQTR were observed in DDA and the msms.txt files for these peptides from MaxQuant were used to build spectral libraries in Skyline. In order to confirm the identity of the unmodified peptide in a PRM run and to rule out interference due to the matrix the chromatographic elution pattern of MS2 transition ions and the dot product (dotp) were also considered. A dotp value closer to 1 indicates better matching of the peptide peaks seen in PRM with the spectral library built from the DDA data. Peaks for the unmodified peptides were considered to be real if the dotp value was above 0.75, the retention time was within ± 1 minute as seen in the spectral library and there were at least 3 overlapping MS2 ions. Peaks for the c61 modified peptides were considered to be real if the retention time was within ± 1 minute between different samples / replicates, there were at least 3 overlapping MS2 ions and that the same peaks were absent in digests without c61.

Byonic data analysis: Glycopeptide data was analysed via Byonic™ v3.10.10 and Byologic™ v3.10-52x64 (Protein Metrics Inc.). Byonic search settings were as follows: Precursor mass tolerance = 10 ppm, Fragment mass tolerance = 20 ppm, FDR = 1% with a glycan library consisting of 182 N-glycan and 9 O-glycan structures. Additional permissible modifications were oxidation (Met/Trp), deamidation (Asn/Gln), Gln to pyro-Glu and Cys carbamidomethyl (fixed modification). Search criteria for glycopeptides was set for a minimum score ≥ 300 and

data was inspected manually in Byonic to ensure quality of MS/MS spectra and precursor monoisotopic peak assignments. Byologic was used to obtain the area of each extracted ion chromatogram, which were exported and used to quantify site-specific glycosylation at each site.

4.3.3 Mass Photometry (MP)

MP experiments were performed using the TwoMP mass photometer and the protocol described by Soltermann *et al*. in 2020.[123] Data acquisition was performed using AcquireMP software (Refeyn Ltd.). Firstly, borosilicate coverslips (24 x 50 mm, #1.5 thickness, Fisherbrand™) were cleaned by sequential 5 minute sonication (PS-40A Ultrasonic Cleaner, Cole Palmer) in Milli-Q water, isopropanol, Milli-Q water and dried under nitrogen gas flow. Silicone gaskets (CultureWell™ reusable gasket, 3mm diameter x 1 mm depth, Grace Bio-Labs) were applied to clean coverslips to enable sample application and the stage spatially adjusted to align the laser with the centre of the gasket. To these gaskets, 5-10 µl PBS (Gibco™ Dubesco PBS, no calcium, no magnesium, ThermoFisher) was added to enable focusing of the microscope to the glass/water interface. Meanwhile, 200 nM protein samples were diluted in PBS to a final concentration of 50 nM. The 50 nM protein sample was added to the gasket to achieve a final volume of 20 µl and mixed by pipetting. Measurements were started immediately (delay time ~2 seconds) and recorded for 60 seconds. The experimental mass photometry setup (TwoMP mass photometer, Refeyn Ltd.) involved a 488 nm laser diode used for illumination with the following instrument parameters: acquisition camera frame rate = 249 Hz, pixel binning = 6x6, 2-fold time-averaging.

The recorded movies were analysed using DiscoverMP software (Refeyn Ltd.). Mass calibration experiments were run as described above using the MS1000 protein standard in which various well resolved contrast peaks correspond to oligomeric states (typically 90, 180, 360, 540 and 720 kDa). Applying this mass calibration to recorded sample measurements

gave MW vs. counts histograms where each species was identified as a resolved peak. Total counts per species were obtained from Gaussian fitting to these peaks. Smaller species with molecular weights approaching the MP detection limit (ca. 30 kDa) were not quantified to ensure differentiation from background noise and were typically not treated as quantitative with respect to counts for larger species.

Herein, MP experiments were performed as above, unless stated explicitly otherwise. Sample preparation often differed between experiments and is outlined in detail in Section 4.4.4.

4.3.4 Native Mass Spectrometry

All experiments in this section were kindly carried out by Wiktoria Sadowska (Benesch group).

Purified GluII samples (Sections 3.2.3-4) were buffer-exchanged into 200-500mM ammonium acetate (pH 6.8) using Bio-Spin® centrifugal filter devices (Bio-Rad) according to the manufacturer's protocol. The concentration of protein was determined by UV-absorbance (NanoDrop spectrophotometer), and diluted into aliquots at appropriate concentrations (~10 µM). Nanoelectrospray was carried out in positive-ion mode on a Q Exactive UHMR mass spectrometer (ThermoFisher), using gold-coated capillaries prepared in house and the application of a modest backing pressure (~0.5 mbar). Nitrogen gas was introduced into the HCD collision cell to improve transmission. The instrument was operated at a resolution of 12500 (at 200 m/z), with "Low Detector Optimization", and a trapping pressure of 7.5 in the HCD cell. Experimental parameters were optimised on a sample-by-sample basis, but approximate values are as follows: Scan range: 1000-12000 m/z; capillary voltage: 1.1-1.5 kV; capillary temperature: 50°C; in-source trapping from -10 to -70 V; injection times: < 100 ms; microscans: 10; automatic gain control target: 1e6. Mass spectra were deconvolved using UniDec software.[129]

4.3.5 Circular Dichrosim (CD)

Proteins were purified according to the protocol in Sections 3.2.3-4. To remove HEPES and chloride ions, the protein was buffer exchanged into phosphate buffer (75 mM sodium phosphate dibasic heptahydrate, 25 mM sodium phosphate monobasic monohydrate) using Bio-Spin® centrifugal filter devices (Bio-Rad) according to the manufacturer's protocol. Buffer exchanged protein samples were 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 260-190 nm with a 0.5 nm data pitch, 1.0 s DIT, 1.0 nm bandwidth, 20 nm/min scanning speed, and 3 accumulations. Signal contribution from the buffer was subtracted by collecting a blank under the above conditions first. Results were plotted as wavelength (nm) versus optical rotation (mdeg).

4.3.6 Differential Scanning Fluorimetry (DSF)

Differential scanning fluorimetry (DSF) experiments were set up by mixing the protein of interest (0.1 mg/ml) with SYPRO Orange (Molecular Probes) at a concentration of 5X in a final volume of 20 µl in white, polypropylene, non-skirted PCR plates (Starlab) sealed with optically-clear ThermalSeal RT2 film (Alpha Laboratories). All measurements were recorded in triplicate. A 25-95°C thermal ramp with 1 degree per minute was performed on a MX3005P real time PCR machine (Stratagene) measuring fluorescence with the ROX filter which corresponds approximately to λ_{ex} of 494 nm and λ_{em} of 602 nm.

4.3.7 Glycomics

Enzymology: For analysis of glycans to determine GluII activity, the HIV envelope glycoprotein, gp120, was expressed according to the protocol described in Section 3.2, using a Bg505 gp120 construct in pPI4 vector provided by Snezana Vasiljevic (Zitzmann group).

Prior to transfection of FreeStyle™ HEK-293F cells with the gp120-pPI4 construct, cells were treated with the mannosidase inhibitor, kifunensine (Kif) (4.3 μM final concentration) and glucosidase inhibitor, *N*B-DNJ (1 mM final concentration). This ensures that gp120 is expressed with predominantly Man_9Glc_3 glycans, thus making it a suitable substrate for GluI and GluII.

To assess GluII activity, 1.2 mg of gp120 was simultaneously treated with GluI and GluII in a 10:1:1 (gp120:GluI:GluII) mass ratio. The mixture was incubated at 37°C overnight and gp120 was then isolated by SDS-PAGE, performed according to the protocol in Section 3.6. The gp120 band was excised from the gel and destained using sequential washes with MilliQ water before addition of PNGase F (New England BioLabs) and overnight incubation at 37°C to release the N-glycans from the protein. The following day, the reaction mixture containing the N-glycans was transferred to a fresh tube and the gel pieces washed twice with 75 μl Milli-Q water, with the supernatant combined with the reaction mixture. The mixture was dried for 2-3 hours using the Concentrator plus Centrifuge Concentrator (Eppendorf) and the residue resuspended in 30 μl Milli-Q water.

2-AA Labelling: Extracted glycans were fluorescently labelled with 2-aminoanthranilic acid (2-AA; Sigma-Aldrich). Here, the 30 μl glycan solution was combined with 80 μl labelling mixture (30 mg/mL 2-AA and 45 mg/mL sodium cyanoborohydride in a solution of sodium acetate trihydrate (4% w/v) and boric acid (2% w/v) in methanol). The mixture was incubated at 80°C for 60 minutes, before being combined with 1 ml 97% acetonitrile in Milli-Q water and applied to Spe-ed Amide-2 columns (Applied Separations). The columns were washed with 95% acetonitrile in Milli-Q water and eluted with 2 x 750 μl Milli-Q water. The eluate was frozen with liquid nitrogen and lyophilised to obtain dry, 2-AA labelled glycans.

HPLC: Fluorescently labelled glycans were analysed by Hydrophilic Interaction Liquid Chromatography (HILIC)-UHPLC using a 2.1 mm × 10 mm Acquity BEH Amide Column

(particle size 1.7 µm, Waters). The mobile phase consisted of two solvents: A - 50 mM ammonium formate, pH 4.4; B - acetonitrile. 2-AA labelled glycans were separated using the elution system described in Table 3. Fluorescence was measured using λ_{Ex} = 360 nm and λ_{Em} = 425 nm.

Table 3 – Elution system used for analysing 2-AA labelled glycans by HILIC-UHPLC using a 2.1 mm × 10 mm Acquity BEH Amide Column (particle size 1.7 µm, Waters). Solvent A = 50 mM ammonium formate, pH 4.4; Solvent B = Acetonitrile.

Time / min	Solvent Ratio A:B	Flow Rate / ml/min
0	22:78	0.5
38.5	44:56	0.5
39.5	100:0	0.25
46.5	22:78	0.5
48	22:78	0.5

4.3.8 MTS Assay

Cytotoxicity of thiopyridones was assessed by measuring the mitochondrial metabolic activity using the CellTiter 96® Aqueous One Solution Cell Proliferation Assay (Promega) following the manufacturer's protocol. Briefly, Huh7 cells were seeded in a 96-well plate, then treated with varying concentrations of thiopyridones in Dulbecco's Modified Eagle Medium (DMEM, ThermoFisher). After three days when cells had achieved 90% confluency, 20 µL of a solution containing MTS (3-(4,5-dimethyl-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium) and the electron coupling reagent, phenazine ethosulfate, was added to seeded cells to a total of 100 µL DMEM. Samples were incubated for 2 hours at 37°C, 5% CO_2 and the absorbance measured at 490 nm on the CLARIOstar microplate reader (BMG Labtech). Absorbance of blank readings (DMSO) was subtracted and samples were normalised to untreated controls.

4.3.9 Free Oligosaccharide (FOS) Assay

To determine the composition of free oligosaccharides accumulating inside cells, Huh7 were seeded in a 12-well plate treated with either α/β interface disrupting thiopyridones at their 90% cell survival concentrations (CC10: c61 = 220 µM; c108 = 91 µM), *N*B-DNJ (250 µM) or DMSO (0.2% v/v). After 48 hours when the cells had achieved 90% confluency, the supernatants were removed and cells washed thrice with PBS. The cells were then resuspended in 1 ml PBS and lysed by sequential freeze-thaw cycles before 20 minute centrifugation at 13,000 rpm. The homogenate was placed in a fresh 1.5 ml Eppendorf tube (~1 mg protein) and was desalted and deproteinated by passage through a mixed-bed ion-exchange column [0.2 ml of AG50W-X12 (H+, 100-200 mesh) over 0.4 ml of AG3-X4 (OH−, 100-200 mesh)], pre-equilibrated with water (2 × 2 ml). The homogenate was added to the column which was washed with 2 × 2 ml of water, and the eluate collected. The extracted and purified FOS were then dried by lyophilizing.

Normal Phase (NP)-HPLC was used to separate the purified 2-AA labelled FOS using a 4.6×250 mm TSK Gel Amide-80 column (Sigma-Aldrich) in a Waters Alliance 2695 separations module. An in-line Waters 474 fluorescence detector set at λ_{Ex} = 360 nm and λ_{Em} = 425 nm. Data were processed using Empower software (Waters) and oligosaccharides identified based on their retention time, with peak identity previously confirmed by Alonzi *et al.* using glycosidase digests and MALDI-TOF MS.[52]

4.4 Results & Discussion

4.4.1 Size Exclusion Chromatography-Multi-Angle Light Scattering (SEC-MALS)

To test the legitimacy of FRET results, SEC-MALS was performed to assess the capacity of thiopyridones to disrupt the GluII α/β interface. Here, GluII was incubated with a selection of FRET active and inactive thiopyridones, as well as with an equivalent volume of pure DMSO

as a negative control, before being applied to a Superose™ 6 10/300 GL column (GE Healthcare). The column served to separate species in the mixture based on their molecular size and multi-angle light scattering was employed simultaneously to determine the molecular weight of the species resolved by SEC based on the extent to which they scatter incident light. Details of the experimental setup can be found in Section 4.3.1.

The thiopyridones used in this experiment were chosen based on their potency as α/β interface disruptors determined by TR-FRET assays (Figure 14). The most potent thiopyridone, c61 (XC50 = 800 nM, TR-FRET), was chosen to maximise the observed response while minimising the volume of DMSO added to the protein sample, as higher proportions of DMSO promote protein degradation. For this reason, c61 is the only FRET active thiopyridone used in the majority of experiments in this thesis. Two FRET inactive thiopyridones were used: c28 (the pyridone analogue of c61 where the sulfur atom is replaced by oxygen) and c113-002 (in which the thiol group on c61 is replaced with a thioether). In both c28 and c113-002, the sulfur atom postulated to bind to GluII covalently via a disulfide bond to a free cysteine residue is either removed (c28) or blocked (c113-002). It is proposed that the removal of their capacity to form disulfide bonds may be the explanation for the complete absence of FRET activity compared to their close structural analogue, c61. The rationale behind the choice of thiopyridones to test was extended to all subsequent experiments.

If α/β interface separation was occurring under these experimental conditions, it would be easily observable as the heterodimer, GIIα and GIIβ have distinct molecular masses (166.6, 108.0 and 58.6 kDa, respectively). Indeed, well resolved peaks based on the UV absorbance trace were visible in the SEC profile, as shown in Figure 15, where UV absorbance traces have been normalised relative to the GluII-DMSO control sample to correct for minor differences in protein sample concentration. The SEC profile for GluII-DMSO control sample (Figure 15, black trace) shows the only significant peak at an elution time of around 22 minutes and MALS predicts this peak to correspond to the GluII heterodimer with acceptable accuracy

(predicted mass = 155.4 kDa, sequence mass = 166.6 kDa). Given the strong affinity of the GluII α/β interface (K_d = 2.59 nM), the absence of any significant peaks for either GIIα or GIIβ in this trace is consistent with expectations. The SEC profiles of the samples treated with c28 and c113-002 (Figure 15, green and blue traces, respectively) are very similar to the DMSO control, with the only real difference being the presence of a minor peak at an elution time of ~16 minutes. The molecular weight cannot be accurately determined by MALS for this peak as it exceeds the upper limit of mass resolution for the SEC resin used. This is typical of protein aggregation and species eluted here can be said to have a molecular weight >600 kDa. The similarity of the profiles of c28 and c113-002 with the DMSO control suggests little to no GluII interface disruption under these experimental conditions, in good agreement with their TR-FRET profiles. Conversely, the SEC profile of the GluII-c61 sample is strikingly different. Here we see three significant peaks at elution times of 16, 24 and 28 minutes. The largest peak elutes at 16 minutes and corresponds to the protein aggregate peak previously described for c28 and c113-002. For the next significant peak (24 mins), MALS predicts a molecular weight of ~75.9 kDa. This most closely matches the sequence mass of GIIβ, though its identity cannot be confirmed by SEC-MALS alone. MALS could not accurately predict the corresponding mass for the final peak (28 mins), so it has been omitted due to inaccuracy. The near absence of any SEC peak corresponding to the GluII heterodimer, as seen for the DMSO, c28 and c113-002 traces, could suggest α/β interface disruption, but it is not possible to draw this conclusion from SEC-MALS data alone.

To identify the species responsible for the observed SEC peaks for the GluII-c61 sample, I collected the fractions corresponding to each peak and ran SDS-PAGE using Broad Range Color Prestained Protein Standard (10-250 kDa, New England BioLabs), as shown in Appendix 9.6. The protein aggregate peak (~16 mins) ran as three distinct bands (approximately 110, 90 and 20 kDa), suggesting that perhaps GIIα is responsible for the putative protein aggregation. The smaller SEC peak (~24 mins) produced one clear band on

SDS-PAGE with a molecular weight of ~70 kDa, in good agreement with the MALS prediction. The SEC peak at 28 minutes gave no observable bands on SDS-PAGE. The gel provided some evidence of α/β interface disruption as the combination of bands matches those seen in the untreated GluII control lane. Subsequent MS experiments (in-gel trypsin digestion, Section 4.3.2) confirmed that the peak at elution time 16 minutes corresponds to aggregated GIIα and that the additional bands seen for this species on SDS-PAGE are proteolytic fragments of GIIα. Protein identification by MS also confirmed the ~70 kDa species predicted by MALS and SDS-PAGE was indeed GIIβ (see Appendix 9.6 for MS sequence coverage). Taken together, these data provide evidence for α/β interface disruption by c61.

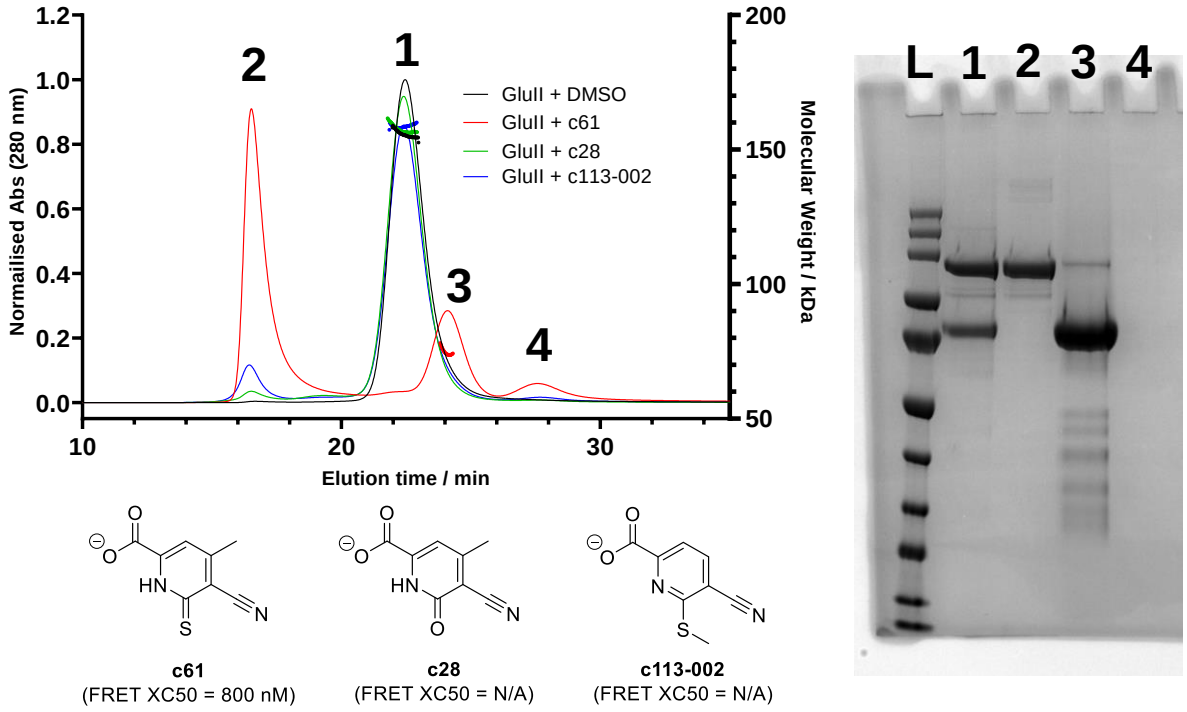


Figure 15 - SEC-MALS profile of GluII treated with FRET active (c61) and inactive (c28, c113-002) thiopyridones. The structure and FRET activity of each thiopyridone is shown. The area under the curve in the UV absorbance traces from SEC are normalised relative to the DMSO control (black). SEC UV profiles (left-hand Y-axis) show the similarity between the FRET inactive compounds, c28 (green) and c113-002 (blue) to the DMSO control (black). The FRET active c61 (red) shows striking difference in elution pattern to the other profiles, including loss of the GluII heterodimer peak (~22 mins), suggesting that the compound disrupts the α/β interface. MALS predictions of molecular weight are shown as dotted lines (right-hand Y-axis). SDS-PAGE of the significant SEC peaks confirmed interface disruption.

4.4.2 Proteomics - Mass Spectrometry

4.4.2.1 Evotec – LAB282

In theory, mass spectrometry is the most straightforward way to identify covalent modification of peptides by small molecules. Indeed, it was the first technique employed by the Zitzmann group/Evotec during LAB282 funded research (grant code: ALR02000). Due to limited availability of the data, MS experiments performed by Evotec are not described in detail in this thesis, though some of their key experiments that informed my own work are summarised here and figures provided by Evotec are shown in Appendix 9.7.

Firstly, we sought to use MS to prove in principle that the thiopyridone scaffold is capable of forming covalent bonds via the thiol moiety. As discussed in Section 1.4.2, the interfacial D-mannose binding pocket to which thiopyridones were predicted to bind in a virtual screen is in close proximity to the C822 residue on GIIα, prompting the suggestion that this residue may be targetable by covalent warheads. A 2017 publication by Chen *et al.* demonstrated the identification of a covalent PPI disruptor affecting the subunits of human cytomegalovirus DNA polymerase, which inspired the workflow to establish thiopyridones as covalent binders.[130] Here, the PPI inhibitor was first incubated with small synthetic peptides containing nucleophilic residues to establish compound reactivity and this idea was adopted for application to the GluII-thiopyridone system.

To verify the capacity for disulfide bond formation without the complications of protein secondary or tertiary structure, Evotec synthesised several nucleophilic peptides: the generic peptide LWXAAGR (where X = T, S, C, K or A) and SSDCMK, which matches the GIIα sequence at C822. The peptides (25 µM) were incubated for five minutes with either c60, c61 or probe 37 (50 µM), with the latter being analogous to c60/c61 but with an alkyne reporter tag to enable analysis by in-gel fluorescence. The peptide solutions were diluted to 1.75 pmol/µl and 5.25 pmol of each was injected into a Q Exactive HF-X hybrid quadrupole-Orbitrap mass

spectrometer (Thermo Fisher) and analysed by tandem MS (MS/MS). For the LWXAAGR peptide, Evotec detect covalent modification by the thiopyridones only when the variable residue is cysteine with the other nucleophilic residues showing no covalent binding to the tested compounds. The detected covalent modification corresponds to disulfide bonding with acidified thiopyridones. Even with LWCAAGR, unmodified peptide is detected with every thiopyridone, but with c61 (and to a lesser extent, c60) the intensity of the modified peptide in MS1 far exceeds that of the unmodified peptide. With probe 37, intensities of the modified and unmodified peptide are roughly equal. For the second synthetic peptide, SSDCMK, Evotec MS/MS data show broadly similar results as LWXAAGR, with c61 giving the highest ratio of modified to unmodified peptide. These MS data (not shown) provide evidence for the possibility of disulfide bond formation between thiopyridones and cysteine.

Evotec attempted to determine covalent binding of GluII-thiopyridones using conventional protein MS experiments, whereby native protein was incubated with the ligand prior to in-solution digest with trypsin and/or LysC digestion enzymes. A key drawback of this approach was the use of dithiothreitol (DTT) and iodoacetamide (IAA) reducing and alkylating agents, respectively. This is common practice for MS analysis of proteins as it ensures full denaturation by breaking disulfide bridges and alkylating the resulting thiols to prevent refolding of the protein, allowing unrestricted access of the digestion enzymes to their digestion sites. However, when searching for a peptide modified by a disulfide bonding ligand, the stated conditions are likely to remove the ligand entirely. Consequently, it is unsurprising that Evotec were unable to identify thiopyridone modifications of GluII peptides in their initial MS/MS experiments.

In an attempt to indirectly identify GluII-thiopyridone binding while continuing to use their established MS sample preparation workflow, Evotec developed a cysteine proteomic profiling experiment. Here, cysteines potentially modified by thiopyridones are identified by looking at the competition for covalent interaction with solvent exposed cysteine residues between

thiopyridones and a desthiobiotin-labelled IAA alkylating agent (DTB-IAA). A figure provided by Evotec summarising the DTB-IAA experiment is provided in Appendix 9.7. Briefly, GluII (50 µg) was incubated with c60, c61 or probe 37 (50 µM). In theory, the thiopyridones would form their covalent interaction with the solvent exposed cysteine(s) in these non-denaturing conditions. Following this incubation, 100 µM of DTB-IAA was added to the mixture, resulting in the DTB-labelled alkylation of any free cysteines that had not already been covalently bound by thiopyridones. After a washing step to remove excess DTB-IAA and unbound thiopyridones, GluII was subject to the standard reducing (DTT) and alkylating (IAA – unlabelled) conditions used in previous MS experiments, which likely removes any bound thiopyridones by breaking disulfide bonds. Next, GluII was digested using trypsin and the resulting peptides analysed by LC-MS/MS using EASY-nLC 1000 (Thermo Scientific) coupled to Q Exactive HF, 40 cm C18 column. Theoretically, cysteine residues that are modified with a DTB-IAA alkylation are solvent exposed and likely not bound by thiopyridones in the initial incubation. Conversely, unlabelled IAA modifications highlight cysteines that may have been initially bound by thiopyridones (later removed during the DTT reduction) or that are not solvent exposed and therefore not available for disulfide bonding with ligands. Comparing the MS spectra from these experiments (data not shown) with the *Mm*GluII$_{Tryp}$ crystal structure, which easily identifies buried core residues, reveals several solvent exposed cysteines that may be bound covalently by thiopyridones. This experiment informed my site directed mutagenesis experiments which attempted to identify which of these cysteines, if any, are involved in the proposed covalent interaction (Section 4.4.3).

4.4.2.2 Zitzmann Group MS

Independently of the work performed by Evotec, I attempted a search for GluII peptides covalently modified by thiopyridones using MS with all experiments performed by Dr Bevin Gangadharan (Zitzmann group). During the initial screen in which thiopyridones were discovered, it was demonstrated that these compounds follow the GIIα-targeting strategy in

which significant protein-ligand interactions are made to GIIα to sterically interfere with GIIβ (see Section 1.4.2). The prediction that GIIα and not GIIβ is the target of thiopyridones has been validated by several experiments, discussed in more detail in Sections 4.4.3 and 4.4.4, so GIIβ was omitted from the MS experiments to identify covalent binding of thiopyridones to GluII. The bulk of MS searches for peptide modification by thiopyridones focused on the C822 residue on GIIα as it is close to the putative thiopyridone binding pocket. Furthermore, mass photometry (MP) experiments provide evidence that of the solvent exposed free cysteines on GIIα identified in the Evotec DTB-IAA MS screen (see 4.4.2.1), C822 is the most likely to influence thiopyridone binding, as discussed in Section 4.4.4. Sequence coverage for GluII digests discussed here are presented in Appendix 9.8.

To detect GIIα peptide modification by thiopyridones it is important to preserve the presumed disulfide bond during sample preparation by avoiding reducing conditions. I first attempted the generic in-gel sample prep of c61-treated GIIα as described in Section 4.3.2 and simply skipped the reduction and alkylation steps prior to the enzymatic digestion by trypsin. Unsurprisingly, this resulted in poor sequence coverage (29.56%) in data-dependent acquisition (DDA) MS/MS relative to the data gathered when identifying the species recovered in the earlier SEC-MALS experiment (70.65%) due to restricted access of the digestion enzymes to their cleavage sites. Importantly, this method of sample preparation did not result in sufficient GIIα cleavage to produce MS compatible peptides at the residue of interest, C822. To counteract the limited activity of digestion enzymes against native proteins, I incubated GIIα with 0.1% RapiGest SF (Waters) in 50 mM aqueous ammonium bicarbonate. RapiGest is a mild anionic surfactant that denatures proteins without modifying the peptides or inhibiting the digestion enzyme.[131–133] In theory, this would expose the cleavage sites and lead to more comprehensive sequence coverage of GIIα by MS. Indeed, for GIIα digests following treatment with RapiGest, we see improved coverage (79% for trypsin, 73% for chymotrypsin) in DDA-MS/MS experiments. The data collected following the RapiGest-assisted in-gel GIIα

digestion were used for downstream analysis and this workflow was repeated for all subsequent experiments.

Despite good sequence coverage for the trypsin/chymotrypsin digests of GIIα, the DDA-MS experiments usually did not detect the presence of any peptides covering the C822 residue and invariably did not identify any modifications corresponding to a cysteine-thiopyridone disulfide bond. Based on the presence of cleavage sites in the GIIα sequence and assuming no missed cleavages, both trypsin and chymotrypsin digestion should give MS compatible peptides containing C822 (trypsin – RSSD**C**MK; chymotrypsin - RVRRSSD**C**M), based on predictions made by the Expasy PeptideCutter tool and setting the lowest cleavage probablility to 45%.[114] In the event of missed cleavages, larger peptides containing C822 may also be seen, though DDA-MS/MS often does not detect any even when factoring in the variable methionine oxidation states possible for these peptides. An abundance of the peptide(s) of interest below the detection limit in DDA-MS/MS may explain their absence in the data, but another possibility is that C822 could be involved in a disulfide bond with another GIIα cysteine. This suggestion is supported by the fact that when the protein is subject to reducing conditions during sample preparation, C822 often is observed though often in longer peptides with missed cleavages. To test this theory, raw MS data from a RapiGest/Chymotrypsin-treated GIIα sample with and without c61 was analysed using SEQUEST and MaxQuant to search for likely disulfide bonds.[134–136] Here, it was revealed that in the absence of c61 there is a possible disulfide bond at the C822 residue. Interestingly, this disulfide bond is not predicted for the equivalent c61-treated sample. This could suggest that the treatment of GluII with c61 interferes with a native disulfide bond within GIIα. That said, it is also possible that the C822 disulfide bond simply does not exist and the absence of an MS compatible C822 peptide corresponds to incomplete/excessive chymotrypsin digestion. Indeed, according to the *Mm*GluII$_{Tryp}$ crystal structure, C822 exists as a free thiol, however C423 (also a free thiol) may be in close enough proximity to form a disulfide bond to C822

without drastic conformational change from the crystal structure. It is therefore possible that such a disulfide bridge exists in GIIα in solution, but the evidence we currently have is insufficient to confidently conclude either way.

The treatment of GluII with c61 and subsequent RapiGest-assisted enzymatic digestion and MS analysis was repeated several times in attempt to identify an MS compatible C822 peptide. In disagreement with the disulfide mapping described above, DDA-MS/MS detected surprisingly long C822 peptides spanning 39 and 40 amino acids in at least two separate GIIα tryptic digestion sample preparations. These long peptides differ only by the presence/absence of the N-terminal R818 residue and would typically be considered too large for facile detection by conventional MS analysis, but were detected here with reasonable MaxQuant scores and ion intensities (Table 4). The long peptides are visible for both c61-treated and untreated GluII samples following tryptic digestion.

Table 4 - Long C822 peptides detected following tryptic digestion of c61-treated GIIα. C822 is indicated in bold and apostrophes show missed trypsin cleavages.

Digestion enzyme	Peptide Charge	Sequence	MaxQuant Score	Ion Intensity
Trypsin	+4	R'SSDCMK'DDPITLFVALSPQGTAQGELFLDDGHTFNYQTR	98.63	3.71×10^7
		SSDCMK'DDPITLFVALSPQGTAQGELFLDDGHTFNYQTR	34.74	1.87×10^6

When cross-referencing the presence of the long 39 amino acid C822 peptide to the Generalized Proteoform Meta-analysis Database (gpmDB), we found that it actually appears nine times more than the shorter peptide (SSDCMK).[137] Having identified C822 peptides, we examined the c61-treated sample further using parallel PRM-MS/MS in which the peptides of interest were specifically targeted for detailed analysis. Confusingly, the peptides of interest (Table 7) were not observed in MaxQuant searches of the PRM-MS/MS data and manual searching of the MS1 spectra confirmed the absence of the precursor ions. However, analysis

of the raw data using Skyline software revealed the presence of the fragmented, unmodified C822 peptide with high confidence (dotp 0.8), with 32 unique daughter ions detectable at a similar retention time as seen in the DDA data. It is possible the long peptide is not detected well by the Orbitrap of the Q Exactive used, hence its relative and total absence in the DDA and PRM MS1 spectra, respectively. However, in PRM the quadrupole successfully isolates the peptides of interest and transfers them to the higher-energy C-trap dissociation (HCD) collision cell where it is comprehensively fragmented such that the smaller daughter ions are easily detected in MS2. The peptides of interest confidently detected in the MS2 spectrum of the PRM-MS/MS experiment were not modified by c61. We also manually searched MS1 spectra for the same C822 peptide with a mass adduct corresponding to the disulfide bond to c61 and it was not observed. In MS2 a broad peak with the correct mass-to-charge ratio (m/z) was present when analysed on the Skyline software with at least 5 overlapping MS2 transitions confirming that this peak is real, but the ion intensity is much smaller than for the unmodified peptide. Additionally, the retention time shifts from 94 minutes for the unmodified peptide to 112 minutes for the supposed c61 modified peptide. Since c61 is a reasonably hydrophobic molecule (LogP = 0.88, predicted by ChemDraw - PerkinElmer Informatics), it is possible that its addition to a peptide would increase the affinity for the C18 resin on the column and increase the retention time. Although the absence of the modified peptide in the MS1 spectrum and poor quality of data in MS2 make it difficult to confidently assign the peak at later retention time to the c61-modified peptide, the probability of detecting an unrelated peak of precisely the correct m/z in MS1 and at least 5 MS2 ions is very low. If the weaker signal truly does correspond to the modified C822 peptide, then it is certainly far less abundant in the PRM MS2 spectrum, suggesting a much lower quantity of modified peptide compared to unmodified. Based on MS2 peak area, the c61-modified peptide is observed with 6-fold lower abundance on average than the unmodified peptide, although we cannot confirm this relative abundance with certainty since something in the sample matrix could supress/enhance ions at the two different retention times for the modified and unmodified peptides. Ideally, a synthetic peptide

with a matching sequence to Entry 1 in Table 8 could have been incubated with c61 and then both the modified and unmodified synthetic peptide run in DDA to confirm that c61 binds to the cysteine. This DDA data would have helped to establish a spectral library for c61 bound to the peptide and well as retention times. This means that when samples are run in PRM, we would have known with even greater confidence and with dotp values if the peaks we are seeing are real.

The DDA experiment identified the long C822 peptide (Table 4) in the +4 charge state and therefore the +4 charge state was targeted specifically by PRM. To ensure a thorough analysis, we acquired PRM data on a sample of c61-treated GIIα to search the MS2 spectrum for daughter ions of the peptide at +3 and +5 charge states. The lower charge state (+3) peptide was not identified, but the relatively high m/z of the modified peptide is above the upper limit of a typical peptide search using Skyline (m/z 1500).[138] However, the search for the peptide in the +5 charge state revealed peaks for the expected m/z value with several MS2 ions visible. The retention time matches that of the putative c61-modified C822 peptide with +4 charge (112 minutes) seen in the previous PRM search, increasing the confidence that the peaks for the modified peptide are true. While the ion intensity is modest, the peaks are more well-defined than for the equivalent +4 charge state peptide. Searches were performed for peptide modification by native and acidified c61 and both returned similar MS2 peaks. To ascertain whether these signals originate from a legitimate c61 modification or are simply artefacts of the MS spectra, PRM data were acquired for the untreated GIIα sample targeting the peptide modified by native and acidified c61. Unsurprisingly, there was no sign of these peaks in the MS2 spectrum for untreated GIIα and at the expected retention time (112 minutes) the spectrum displays only background noise for this sample. This proves that the MS2 peaks seen in the c61 treated GIIα sample correspond to the c61-modified C822 peptide suggesting that covalent interaction between thiopyridones and C822 is possible (Figure 16B + C). Nevertheless, the intensity of these ions is still considerably lower than for the unmodified

peptide (Figure 16A) meaning that this cysteine residue is usually not modified by c61 and the disulfide bond is therefore unlikely to contribute significantly to the mechanism of action of thiopyridones.

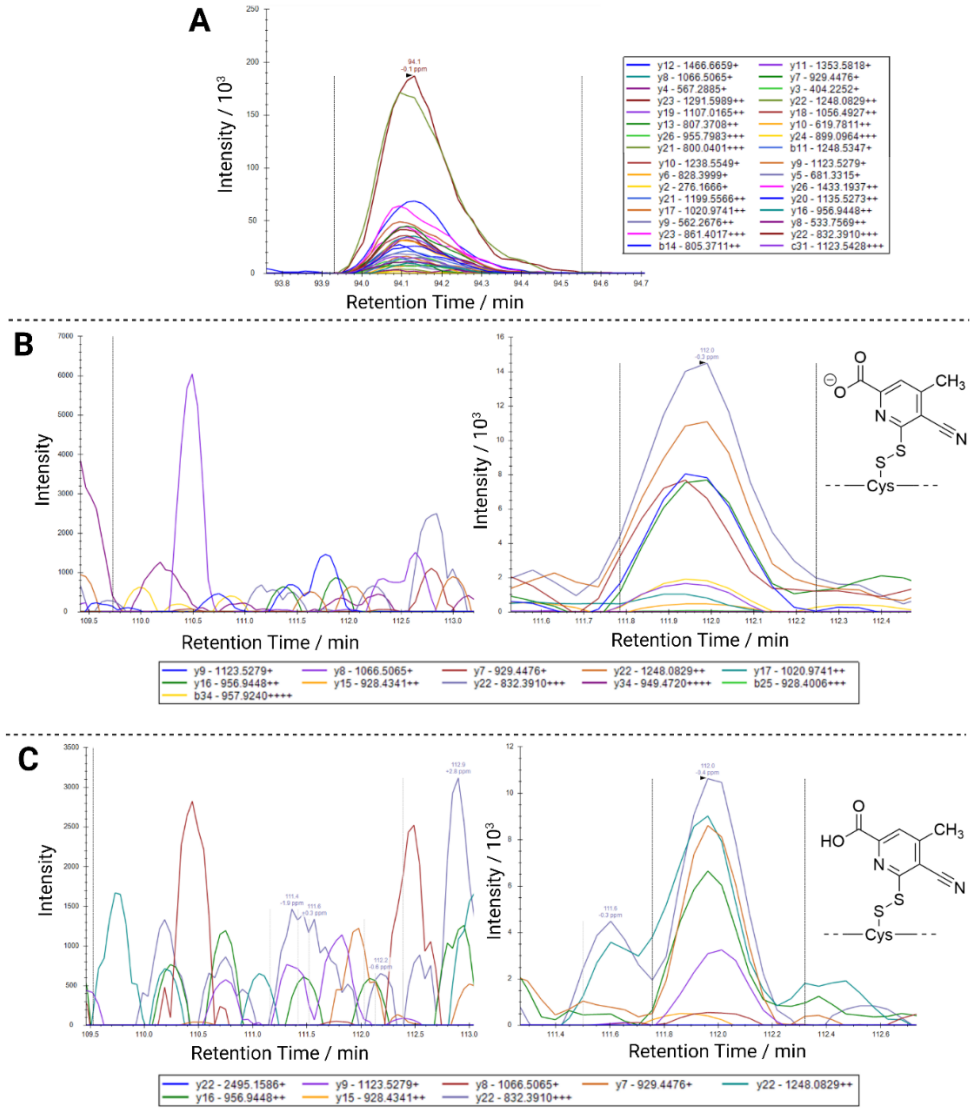


Figure 16 – Extracted ion chromatograms of PRM MS2 ions of peptide RSSDCMKDDPITLFVALSPQGTAQGELFLDDGHTFNYQTR in the +5 charge state, with MS2 ions and their m/z values shown in the key. (A) The unmodified peptide detected in c61-treated GIIα. The ion intensity is highest among the MS2 spectra presented here, indicating that this species is the most abundant in the sample. (B + C) A PRM search for the long peptide modified with native (B) or acidified (C) c61, with the modifications of interest shown on the right. On the left are the PRM searches of the untreated GIIα sample; no modified peptide is detected here. The right-hand spectra shows the same PRM search for the c61-treated GIIα sample and confirms the presence of the c61 modifications.

In addition to the search for thiopyridone binding by MS, I also set out to identify the glycosylation profile of the recombinant GluII heterodimer itself, which is trafficked along the

secretory pathway during expression as a result of artificial modifications of the gene introduced to allow facile protein expression. Secretion of GluII from the ER does not occur endogenously, so any post-translational modifications (PTMs) corresponding to more advanced stages of the secretory pathway, such as O-linked glycosylation in the Golgi, should be interpreted with care as they are likely to be functionally irrelevant with respect to wt GluII. In-gel tryptic digestion of both GIIα and GIIβ was performed, this time using the full reduction (DTT) and alkylation (IAA) protocol with RapiGest for maximum sequence coverage. To inform the search, I used the NetNGlyc and NetOGlyc servers from DTU Health Tech to predict the potential N- and O-linked glycosylation sites on GIIα and GIIβ, the results of which are shown in Table 5.[139,140] NetOGlyc predicted one O-linked glycosylation for GIIα and several for GIIβ and NetNGlyc predicted N-linked glycosylation sites are in agreement with the potential N-glycosylation sites shown in the UniProt entries for murine GIIα and GIIβ (Uniprot Entry Numbers: GIIα – Q8BHN3; GIIβ – O08795).[55]

Table 5 - Potential GluII glycosylation sites predicted by NetNGlyc/NetOGlyc servers from DTU Health Tech. NetNGlyc/NetOGlyc scores >0.5 are considered potential glycosylation sites.

GluII Subunit	N- or O-Linked?	Predicted Glycosylation Site	NetNGlyc/NetOGlyc Score
GIIα	N	N75	0.665
	O	S31	0.597
GIIβ	N	N76	0.627
	N	N473	0.545
	O	S28	0.686
	O	S112	0.545
	O	S130	0.620
	O	T138	0.526
	O	T240	0.530
	O	T246	0.623
	O	T269	0.621
	O	S271	0.513
	O	S286	0.681
	O	T291	0.847
	O	T299	0.649
	O	S346	0.737
	O	T348	0.657

To test the predictions in Table 5, DDA-MS/MS data from alpha-lytic protease (aLP) and trypsin digests of GluII subunits using RapiGest without reduction or alkylation were analysed using Byonic™ (Protein Metrics by Dotmatics). The Byonic search of the DDA data from GIIα confirmed the presence of the predicted N75 N-linked glycosylation on peptide LELQGLQKNMT with $Hex_6HexNAc_2$ as the predominantly observed glycan. In the context of mammalian ER resident proteins, the $Hex_6HexNAc_2$ can be confidently presumed to be $Man_6GlcNAc_2$. The N75 glycan was also observed with more or fewer mannose residues, indicating varying efficiency of mannosidases during protein production (Figure 17). The peptide covering N75 was only seen by Byonic and not MaxQuant suggesting that this peptide is either always occupied or the level of unoccupied peptide is below the limit of detection. The gpmDB shows that tryptic peptide VLLVLELQGLQKNMTR has been observed before but the number of observations are approximately 1000-fold lower than that of the most observed GIIα peptides, suggesting that N75 is predominantly occupied. Indeed, a $Man_2GlcNAc_2$ glycan is observed in this position in crystal structures of *Mm*$GluII_{Tryp}$ (PDB: 5HJO), confirming N75 as a genuine glycosylation site. Table 5 shows the NetOGlyc prediction of an O-linked glycan on residue S31, however this was not observed in the Byonic search. Instead, O-linked Core 1 glycans with the formula $HexNAc_1Hex_1NeuAc_1$ and $HexNAc_1Hex_1NeuAc_2$ were identified on peptide EGNGAQPEA**T**PGDGDKPEE**T** containing T205 and T215, with the most likely structure being $GalNAc_1Gal_1NeuAc_{1\text{-}2}$. Glycosylation of these residues was not predicted by NetOGlyc, though T205 had a NetOGlyc score just below the threshold for glycan assignment while T215 has a much lower score (NetOGlyc score: T205 = 0.468; T215 = 0.089). Analysis using Byonic suggests that the O-linked glycan is most frequently detected on T205, based on analysis of MS2 daughter ions of the peptide of interest. This finding is in agreement with the relative probability of T205 and T215 glycosylation predicted by NetOGlyc suggesting that T205 is the more likely glycosylation site, however a more thorough protein digest that can separate the

two threonine residues, perhaps GluC, would be necessary to categorically confirm this assignment. Alternatively, electron-transfer dissociation (ETD) could help to confirm the exact location of this O-glycan by fragmenting the peptide of interest along the backbone without interfering with the glycosylation.[141]

For GIIβ, the two tryptic DGSDEPGTAACPNGSFHCTNTGYK and YEQGTGCWQGPNRSTTVR peptides containing the predicted N-linked glycosylation sites at N76 and N473, respectively (Table 5) were identified, but did not carry any glycans. To further support that these sites are not occupied, gpmDB shows that both these peptides are the most highly observed peptides of GIIβ.[137] Several possible O-linked glycans were identified, with the most confidently assigned on residue T291 on peptide SEVPPTDIPVPEETEPK with $HexNAc_1Hex_1NeuAc_2$, $HexNAc_1Hex_1NeuAc_1$ and $HexNAc_1Hex_1$ in a roughly 4:3:1 ratio. This prediction is in good agreement with the NetOGlyc predictions in Table 5 as T291 has the highest score of all the predicted O-glycosylation sites, however the remaining predictions were not confidently observed in these data. We did not see peptide SEVPPTDIPVPEETEPK in MaxQuant. Therefore, this peptide could be always occupied with an O-glycan, or the abundance of the unoccupied peptide could be below the limit of detection. The aLP digest of GIIβ yielded large peptides spanning the region R285-S372 that contain multiple possible glycosylation sites. As a result, most of the O-linked glycans on GIIβ identified by Byonic cannot be assigned confidently to any particular residue. The exception to this conclusion is the previously mentioned T291 O-glycan which is unambiguously identified on a shorter peptide (SEVPPTDIPVPEETEPK). To improve the confidence in the Byonic O-glycosylation assignments, alternative digestion enzyme could be used to enable better coverage of the amino acid sequence. The observed glycosylation of GluII, confirmed experimentally by MS, is listed below in Table 6 and Figure 17. As previously stated, the identified O-glycans are likely not functionally relevant as the endogenously expressed GluII is retained in the ER and is unlikely to exist in the Golgi for long enough to carry any significant Golgi-associated PTMs. Consequently, the glycomics data

presented here indicate that endogenous GluII is likely to carry a single high mannose N-glycan on GIIα, as previously identified by X-ray crystallography (PDB: 5F0E).

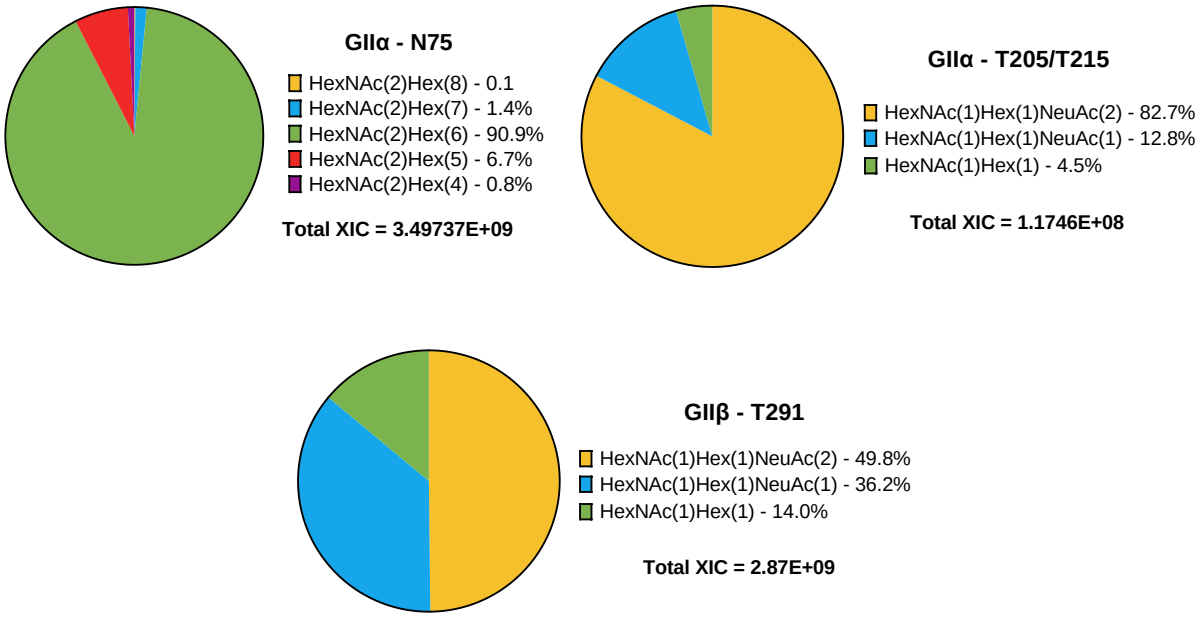


Figure 17 – Pie charts showing the distribution of glycans found at each of the confirmed glycosylation sites on GluII. In the case of the GIIα O-glycan, the precise site is not known and is stated as either T205 or T215. The generic glycan structures predicted by Byonic are stated and the most likely specific structures are listed below in Table 6.

Table 6 - Confirmed GluII glycosylation sites identified by DDA-MS/MS of aLP digested GluII subunits and analysed using Byonic. The GIIα N75 and GIIβ T291 glycans are in agreement with those predicted by NetNGlyc/NetOGlyc (Table 5). Many of the predicted O-glycosylation sites on GIIβ are not identified by MS. These additional O-linked glycans may or may not be present. O-Glycosylation of recombinant GluII is an artefact of its non-native progression along the secretory pathway so is unlikely to be functionally relevant in the wt system.

GluII Subunit	N- or O-Linked?	Confirmed Glycosylation Site	NetNGlyc/ NetOGlyc Score	Glycan Formula
GIIα	N	N75	0.665	$Man_{4-7}GlcNAc_2$
	O	T205	0.468	$GalNAc_1Gal_1NeuAc_{1-2}$
GIIβ	O	T291	0.847	$GalNAc_1Gal_1$, $GalNAc_1Gal_1NeuAc_{1-2}$

MS experiments undertaken by Evotec on synthetic peptide SSDCMK suggest that thiopyridones can bind covalently, though the additional PRM data from the Zitzmann group suggest that GIIα is predominantly unmodified when treated with c61. However, PRM-MS/MS data of c61-treated GIIα did highlight a C822-c61 disulfide interaction, albeit in the form of low

abundance MS2 ions whose parent peptide is not visible in MS1 data (see Section 4.4.2.2). The only observable C822 peptide is much longer than would typically be considered compatible with MS analysis, so it is possible that the difficulty in observing the relevant peptide is responsible for the relative absence of evidence for the target interaction.

4.4.3 Site Directed Mutagenesis of α-Glucosidase II

With limited direct evidence of the c61-GluII disulfide bond from MS, I set out to probe the interaction using alternate techniques, the first of which involved using site directed mutagenesis (SDM) to investigate whether GluII mutants lacking key surface exposed cysteines showed any attenuated α/β interface disruption following treatment with FRET-active thiopyridones. If the protein-ligand interaction is indeed a covalent disulfide bond with cysteine, then removing the relevant cysteine should have an observable effect on α/β interface disruption. In MS experiments on synthetic nucleophilic peptides performed by Evotec, it was shown that solvent exposed cysteines are capable in theory of forming disulfide bonds to FRET active thiopyridones. Another Evotec MS experiment investigating the competition for cysteine modification between thiopyridones and DTB-IAA (Section 4.4.2.1) highlighted several potentially thiopyridone-modified cysteine residues which were considered for SDM: C3, C19, C25, C423, C502 and C822 from GIIα and C491 from GIIβ. Interestingly, C491 on GIIβ is located within the MRH domain, which is known to be necessary for full catalytic activity of GIIα against native substrates.[63,65–67]

During their search for thiopyridone modified GluII peptides, Evotec made use of probe 37, an alkyne-labelled analogue of c60/c61. Following treatment with the red-fluorescent tetramethylrhodamine azide (N_3-TAMRA, Thermo Fisher), the terminal alkyne on probe 37 reacts with the azide of N_3-TAMRA in a copper-catalysed click reaction which allows detection of probe 37 by fluorescence imaging. To identify the subunit to which thiopyridones bind, GluII was first incubated with 1 µM probe 37 for 30 minutes then treated with N_3-TAMRA and applied

to SDS-PAGE. By measuring fluorescence intensity of the gel bands, Evotec quantitatively assessed the incorporation of probe 37 predominantly into the GIIα SDS-PAGE band, with a much lower fluorescence intensity measured for the GIIβ band. This provides good evidence that the main GluII-thiopyridone interaction responsible for interface disruption is mediated by GIIα. Evotec then performed competition assays by treating the GluII-probe 37 sample with either c60 or c61 at increasing concentrations form 2.5-50 μM. Interestingly, they observed lower fluorescent intensities at higher concentrations of the unlabelled thiopyridones, suggesting that c60 and c61 are able to outcompete probe 37 in a dose-dependent manner (Appendix 9.7). This could be evidence that the GIIα-thiopyridone interaction is in fact non-covalent as a disulfide bond between cysteine and probe 37 would likely be unaffected by increasing concentrations of the other thiopyridones. Nevertheless, this N_3-TAMRA SDS-PAGE experiment along with MP data that is discussed in Section 4.4.4 informed the omission of the GIIβ cysteine residue (C491) from my SDM study due to low likelihood of thiopyridone modification of GIIβ. The GIIα C3 residue was also omitted as it is within the secretion signal that was artificially introduced onto the N-terminus of both GIIα and GIIβ to ensure efficient harvesting of the recombinant protein during protein expression in mammalian cells (Section 3.1) and is not physiologically relevant to native GluII. With this in mind, I identified five solvent exposed cysteines on GIIα for replacement by SDM: C19, C25, C423, C502 and C822. In the *Mm*GluII$_{Tryp}$ crystal structure, C19 and C25 are engaged in a disulfide bond with one another so are unlikely to be bound by thiopyridones, but their detection in the Evotec DTB-IAA MS/MS assay qualified them for further investigation by SDM and I decided to simultaneously replace them with alanine in SDM to produce a double mutant of GIIα. This resulted in four proposed GIIα mutant constructs (C19/25A, C423A, C502A and C822A) to be produced using the protocols described in Section 3. Figure 18 shows the location of the cysteine residues of interest on the *Mm*GluII$_{Tryp}$ crystal structure.

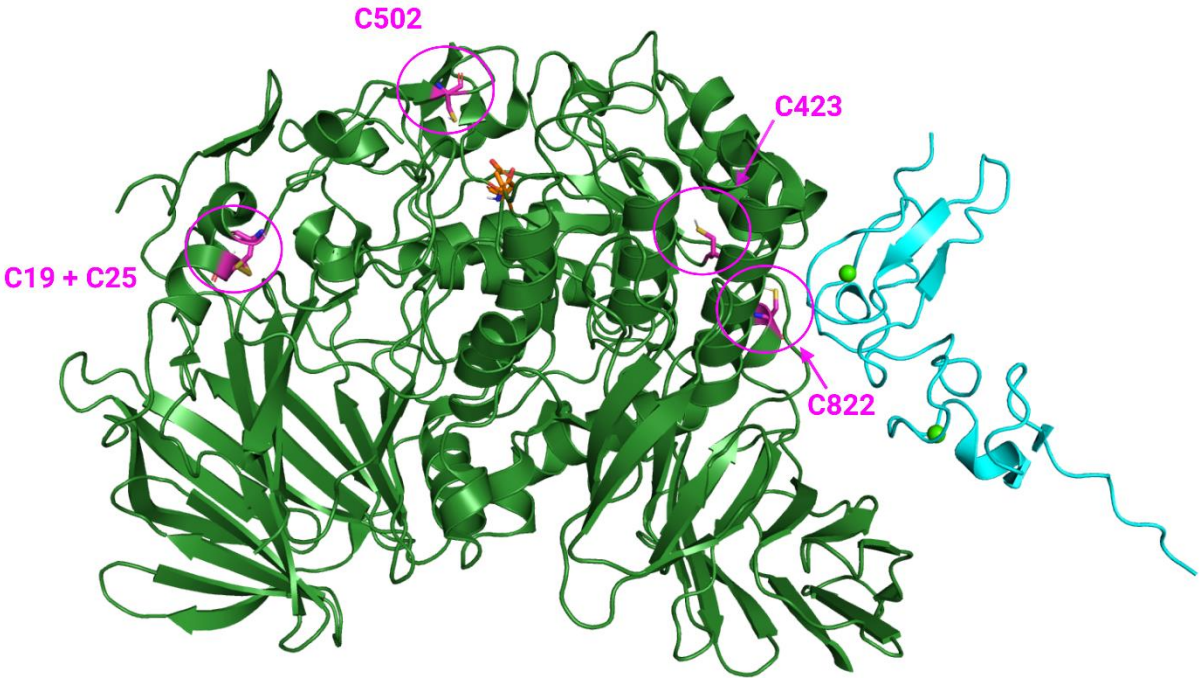


Figure 18 – Cysteine residues of GIIα suggested to be potentially modified by thiopyridones based on the Evotec DTB-IAA MS/MS experiment. In the *Mm*GluII$_{Tryp}$ crystal structure, C423, C502 and C822 exist as free thiols, while C19 and C25 are engaged in a disulfide bond to one another. The highlighted cysteine residues were targeted for SDM.

After several attempts to produce the cysteine mutants of GIIα through cloning experiments (outlined in Appendix 9.3 and 9.4), the desired *hGANAB*-pOPING mutant constructs were purchased directly from Twist Bioscience and these were expressed, purified (Section 3) and used for all downstream analysis. The commercial *hGANAB*-pOPING constructs proved very effective for the expression of the GluII mutants, achieving high yields from a single 200 ml cell culture (HEK293F) for C423A (0.43 mg), C502A (0.34 mg) and C822A (0.33 mg), relative to those achieved using the h*GANAB*-pHLsec construct provided by Dr Mario Hensen in equivalent expressions (approx. 0.22 mg). The expression of the C19/25A double mutant provided a lower yield (0.06 mg), warranting a second expression, again using 200 ml of HEK293F cells, which achieved similar yield. It is possible that interference with the disulfide bond between C19 and C25 disrupts the GIIα fold sufficiently to target the recombinant protein for ERAD and decrease the secretion of GluII from the ER. Nevertheless, all of the expressed mutants were easily purified by IMAC and SEC, with SEC profiles consistent with that of the h*Mm*GluII parent protein.

Having obtained the four cysteine mutant proteins, I performed CD experiments to assess the quality of the protein fold and DSF to determine the relative stability of the mutant proteins compared to the humanised GluII parent molecule.[127] CD (Section 4.3.5) is a commonly used technique to estimate the secondary structure of proteins based on the extent to which they absorb right- and left-handed circularly polarised light.[125] The CD spectra shown in Figure 19 were generated by first normalising the traces relative to that of C502A to correct for discrepancies in protein concentration of the measured samples. The traces for all cysteine mutants of GluII (C19/25A, C423A, C502A and C822A) closely match that of h*Mm*GluII, suggesting that the overall protein fold is not dramatically affected by artificial point mutations. The relative stability of the recombinant mutant proteins was compared using DSF (Section 4.3.6) and results are shown in Figure 20A. In Figure 20B, the measured melting temperature (T_m) of h*Mm*GluII is used as a reference against which the cysteine mutants are measured, with significant increases or decreases in T_m representing relative stabilisation or destabilisation, respectively. Of the cysteine mutants, C822A is the most similar to h*Mm*GluII, with a negligible increase in T_m. When taken with the results from the CD experiment, we can say with reasonable confidence that the C822A mutation has no significant impact on the secondary structure or stability of the 3D fold of the GluII protein, so any downstream experiments comparing the two are directly comparable. In the case of C19/25A, and to a lesser extent C423, the T_m decreases with respect to h*Mm*GluII suggesting the mutations are destabilising. In the *Mm*GluII$_{Tryp}$ crystal structure, C19 and C25 are engaged in a disulfide bond, so loss of this interaction would understandably destabilise the tertiary structure. In the same crystal structure, C423 exists as a free thiol but is somewhat buried and less solvent exposed than C822, suggesting that steric or electronic interference with neighbouring core residues could be responsible for the slight destabilisation when this cysteine is replaced by alanine. Unique among the cysteine mutants is C502A which has a T_m around 4°C greater than h*Mm*GluII, indicating significant stabilisation. This effect is not so easily explained by its position in the crystal structure where it sits in close proximity to the catalytic pocket. The

significant differences in stability for C19/25A, C423A and C502A should be taken into consideration when analysing the results of any downstream analysis.

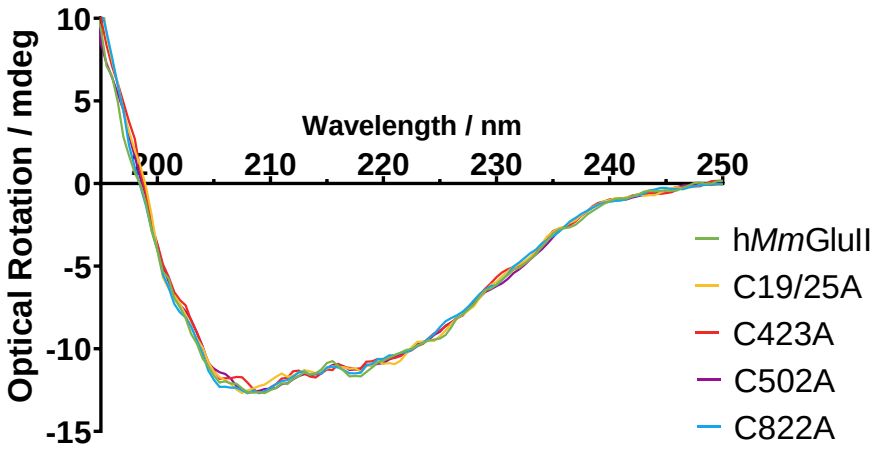


Figure 19 – Circular dichroism (CD) data for humanised GluII and the cysteine mutants thereof. The data were normalised relative to C502A to correct for differences in protein concentration between samples. The similarity in shape of all the traces can be used to deduce that the secondary structure and by extension the overall protein fold is consistent for all tested GluII mutants.

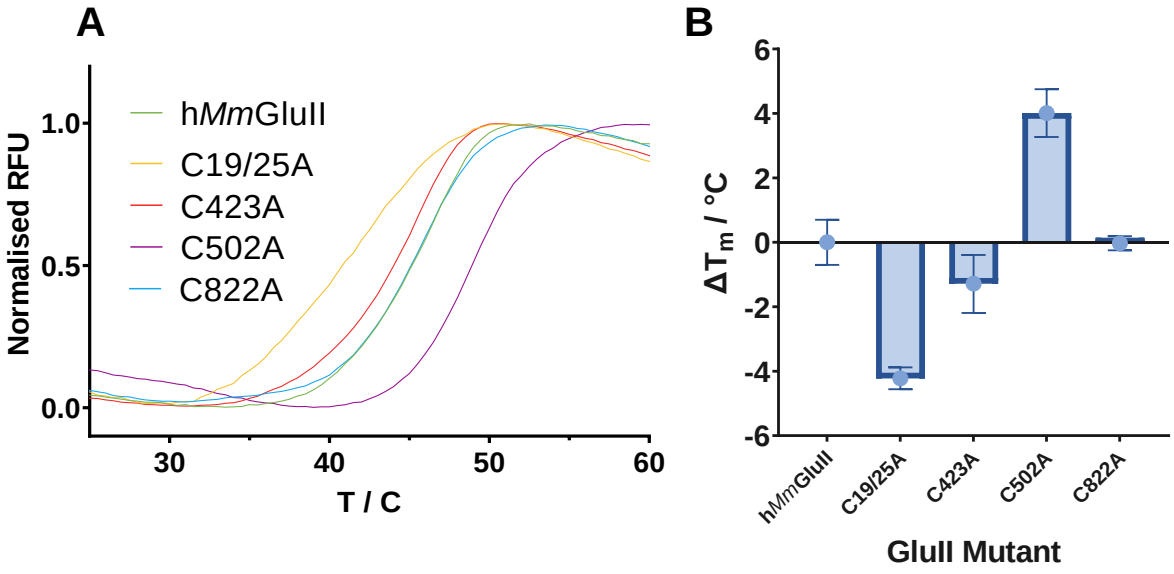


Figure 20 – Differential scanning fluorimetry (DSF) data for humanised GluII and the cysteine mutants thereof. **(A)** Spectra showing the degree of unfolding of each protein, measured as the increase in relative fluorescence units (RFU), as the temperature is incrementally increased. DSF experiments were run as technical quadruplicates and the mean trace is shown for each protein. Fluorescence of the ROX reporter dye was measured and normalised to a maximum RFU value of 1 in each trace. The melting temperature (T_m) is measured at the steepest point of each curve. **(B)** A comparison of the changes in mean T_m relative to that of humanised GluII, with standard error of the mean (SEM) shown as error bars. The C19/25A and C423A mutants show a lower T_m indicating destabilising mutations, while the opposite is true for C502A. The T_m of the C822A mutant is not significantly different from the parent protein, humanised GluII.

4.4.4 Mass Photometry

Mass photometry (MP) is a relatively novel technique capable of measuring the relative distribution of proteins in solution based on their molecular weight and is explained in more detail in Section 4.2. MP has recently been used to investigate PPIs and monomer-dimer equilibria, making it a particularly attractive method for analysing the dynamics of the GluII α/β interface.[122,124] The technique can easily distinguish between the large heterodimer (166 kDa) and the smaller GIIα and GIIβ (108 and 58 kDa, respectively) and the strong affinity of the α/β interface (K_d = 2.59 nM according to TR-FRET) is well suited for MP analysis, as the heterodimer will remain intact even at the low measurement concentrations used in MP (< 100 nM). GluII samples were measured by MP to produce a distribution frequency histogram whereby individual landing events are counted and the molecular weight measured as ratiometric contrast, allowing measurement of the distribution of the heterodimer and separated subunits. To process the data, I chose to determine the degree of α/β interface separation by measuring the ratio of counts of GIIα relative to the GluII heterodimer (GluII:GIIα) as this information is easily gathered by fitting Gaussian curves to the observed peaks in the frequency histogram. I decided to omit the peaks corresponding to GIIβ as the smaller subunit was not reliably detected between measurements and the number of counts was variable if the peak was observed at all. The resolution of the TwoMP mass photometer (Refeyn) is poorer at the lower end of the mass range (FWHM = 25 kDa for a 66 kDa species) so the measured mass of the GIIβ was often inaccurate. It is possible to observe smaller species by lowering the threshold of detection in DiscoverMP (Refeyn) such that lower ratiometric contrasts are considered true landing events, however in my experience, detection of GIIβ was still significantly variable between measurements. Given that the degree of α/β interface disruption is easily measured using the ratio GluII:GIIα and both species are reliably detected with good mass accuracy and reproducibility between measurements, GIIβ was neglected for all MP experiments described here.

My first objective was to use MP to confirm GluII α/β interface disrupting capabilities of thiopyridones observed in both TR-FRET (Section 1.4.2) and SEC-MALS (Section 4.4.1). To this end, I designed an experiment whereby GluII was treated with increasing concentrations of c61, c108 (FRET active) and c28 (FRET inactive) in attempt to observe dose-dependent α/β interface disruption. Due to limited availability of thiopyridone stocks, c108 was used in place of c60 as they are close structural analogues with comparable potencies based on TR-FRET (XC50: c60 = 3.8 μM; c108 = 4.2 μM). Here, aliquots of a GluII stock (200 nM) were diluted to 50 nM and thiopyridones (100 mM stocks in DMSO) were added to final concentrations of 0.05, 0.5, 1.58, 5, 15.8, 50 and 158 μM. The maximum DMSO content for these samples was 0.32% v/v, comfortably within the tolerable range for the GluII heterodimer (< 2%). The samples were incubated at room temperature for 60 minutes prior to MP measurements. The use of a more concentrated stock from which GluII is diluted prior to MP measurements was intended to counteract the decrease in number of counts in MP data observed after prolonged incubation. It is believed this is caused by non-specific binding of the protein to the Eppendorf tube in which it is contained. The results in Figure 21 clearly demonstrate dose-dependent disruption of the GluII α/β interface for FRET active thiopyridones c61 and c108 as the GluII:GIIα ratio approaches zero at high concentration, indicating near complete separation of the two subunits. In contrast, no appreciable interface disruption is observed for c28 even at high thiopyridone concentration. These results are consistent with both TR-FRET and SEC-MALS data, confirming beyond reasonable doubt that thiopyridones are capable of disrupting the GluII α/β interface *in vitro*. In order to estimate an XC50 value for the tested thiopyridones, it was necessary to normalise the GluII:GIIα ratios relative to that of the untreated sample, so measurements at higher thiopyridone concentrations could be taken as a fraction of the maximum GluII:GIIα ratio. The normalisation allowed estimates of XC50 for each compound, shown in Table 7, which shows good agreement with the TR-FRET predictions.

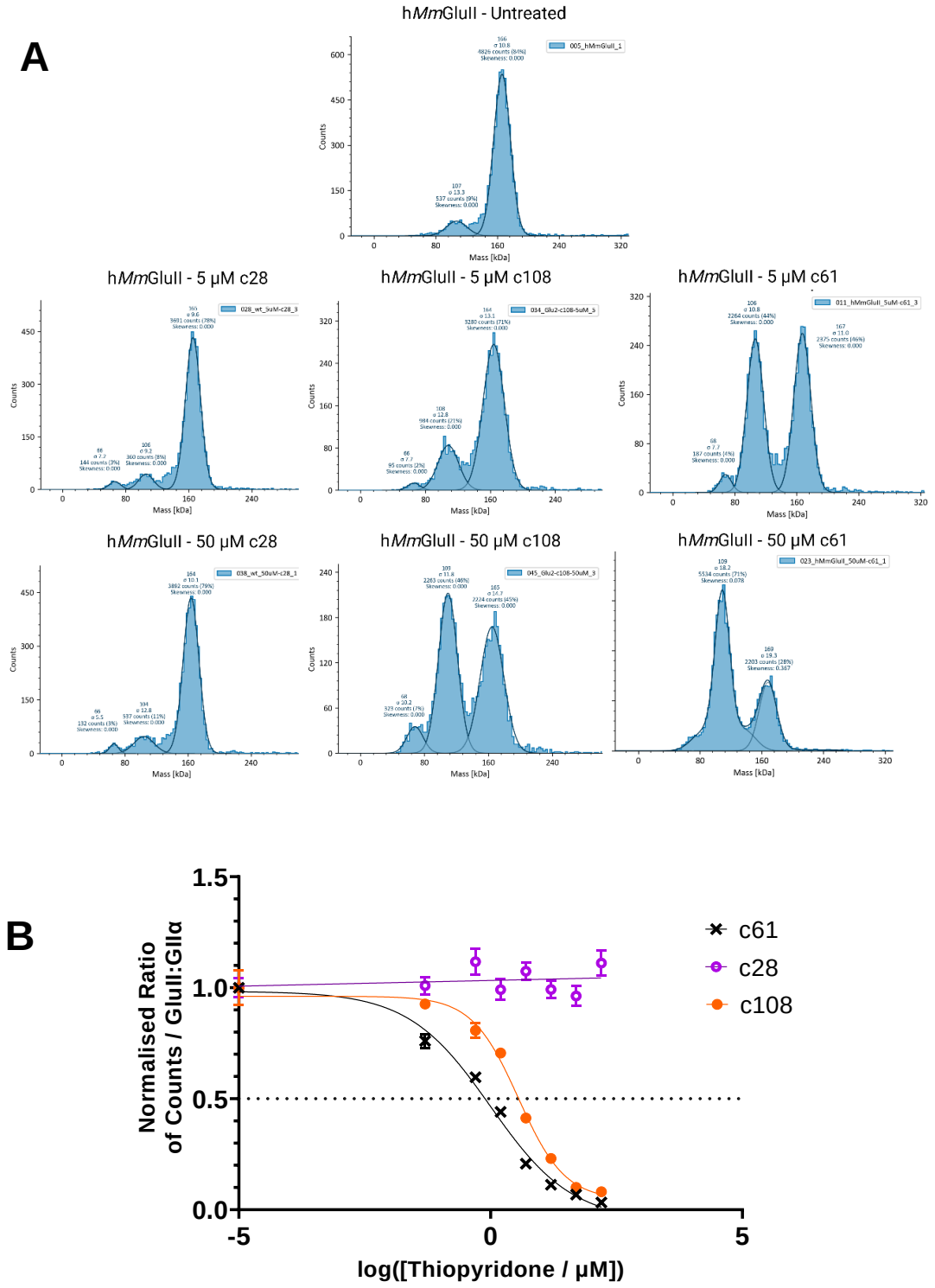


Figure 21 – MP data showing the effect of increasing thiopyridone concentration on the ratio of GluII:GIIα in solution. The final protein concentration in each sample was 50 nM. (A) Frequency histograms of a representative sample of MP measurements of h*Mm*GluII treated with various thiopyridones. Six technical replicates were measured for each sample. (B) The mean GluII:GIIα ratio, normalised to that of the untreated protein, extracted from the frequency histograms above. Error bars measure standard error of the mean (SEM) and points with no error bars represent samples where SEM was too low to be displayed on the graph. A non-linear regression analysis (GraphPad Prism, v9.5.0) fitted Boltzmann sigmoidal curves to the data for c61 and c108, while a straight line was fitted to c28 indicating no α/β interface disrupting activity. For c61 and c108, the data show clear dose-dependent separation of the GluII subunits.

Table 7 – A comparison of the predicted XC50 values for the tested thiopyridones by MP and TR-FRET. The XC50 values from MP were predicted using the Boltzmann sigmoidal fit in Figure 21 by interpolating the mean thiopyridone concentration when the normalised GluII:GIIα ratio is 0.5, indicating 50% α/β interface disruption. Predictions of XC50 values by TR-FRET were previously determined by Dr Mario Hensen (Zitzmann group) and are shown in Figure 14 in Section 1.4.2. The predictions for thiopyridone potency between the two analytical techniques are in good agreement.

Thiopyridone	Predicted XC50 from MP / µM	Upper Limit of predicted XC50 from MP / µM	Lower Limit of predicted XC50 from MP / µM	Prediected XC50 from TR-FRET / µM
c61	0.776	0.928	0.643	0.800
c108	3.613	4.297	3.058	4.150
c28	N/A	N/A	N/A	N/A

To rule out the possibility that the pronounced α/β interface disruption by thiopyridones could be explained by the hour-long incubation of GluII at room temperature, I performed timecourse experiments with untreated GluII, as well as samples with c61 at final concentrations of 1 and 100 µM. Samples were prepared as described above. Given the time-sensitive nature of these experiments, I measured each sample only twice by MP because increasing the number of technical replicates would have introduced inaccuracies in the time stamps for each measurement. The results, shown in Figure 22, prove that in the absence of interface disrupting thiopyridones, the GluII heterodimer remains intact for the duration of the experiment. The higher c61 concentration (100 µM) achieves almost total α/β interface disruption during this experiment as the normalised GluII:GIIα ratio approaches zero. The less concentrated c61 treatment (1 µM) results in a roughly 1:1 ratio of heterodimer to GIIα, supporting the the XC50 value of 780 nM predicted by MP (Table 7). Equilibrium for α/β interface disruption is reached at around 25 minutes for both tested concentrations. These data seem to support a non-covalent binding mode for thiopyridones as the lower c61 concentration, which is still a 20-fold molar excess of the compound relative to the protein, achieves equilibrium at a 1:1 GluII:GIIα ratio after around 25 minutes. If GluII-thiopyridone binding was covalent and effectively irreversible, one would expect a significant excess of thiopyridone to still achieve maximum interface disruption, even if it took longer to do so. The fact that equilibrium is achieved at 20-fold molar excess of the small molecule would suggest that the interface disruption is reversible under these experimental conditions.

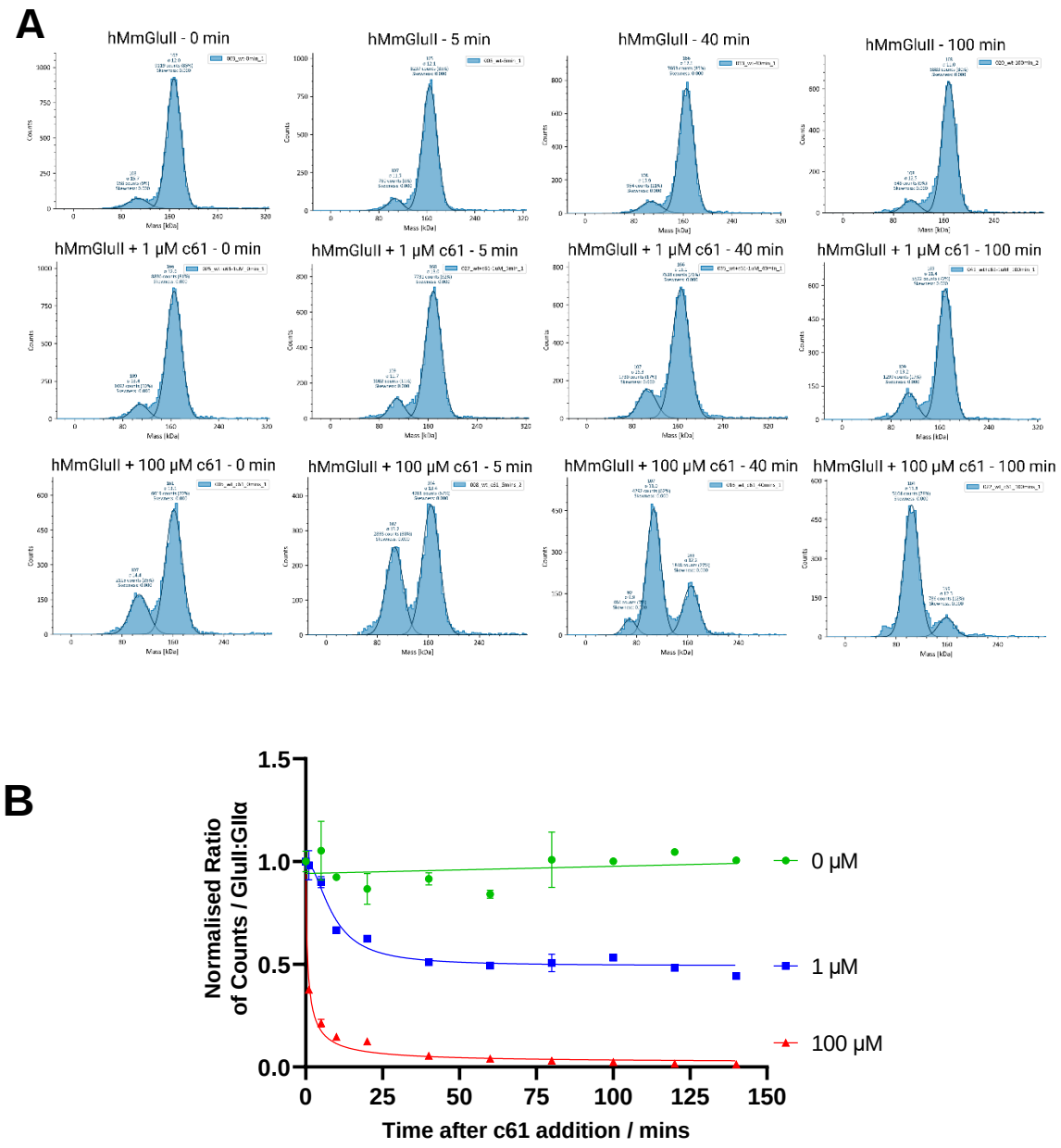


Figure 22 – Timecourse MP experiment of GluII treatment with c61. (A) Frequency histograms of a representative sample of MP measurements of h*Mm*GluII treated with various thiopyridones. Two technical replicates were measured for each sample. (B) Points are plotted as mean values from technical duplicates (N = 2) with SEM shown as error bars. Due to the narrow spread in the measured GluII:GIIα ratio, the SEM is too small to be displayed for most measurements and is omitted. These data show that the GluII heterodimer remains intact after incubation at room temperature, with no discernible α/β interface disruption up to 140 minutes. When treated with 1 µM c61, GluII experiences half the maximum subunit separation which is observed for the 100 µM c61 treatment after roughly 25 minutes.

By fitting a one phase exponential decay model to the data in Figure 22 (GraphPad Prism, v9.5.0) and comparing the slope for the c61 experiments, it is possible to determine relative

dissociation rates at the different c61 concentrations. The relevant formulae for the exponential decay fitted by GraphPad Prism and the derivative thereof are shown in equations 1 and 2, respectively, where the value 'k' is the calculated rate constant for the fitted model and 'y_0' is the initial GluII:GIIα ratio.

$$y = y_0 \cdot e^{-k \cdot x} \quad \text{(Equation 1)}$$

$$\frac{dy}{dx} = -k \cdot y \quad \text{(Equation 2)}$$

Using Equation 2 and taking 'y' as the midpoint between the initial and minimum GluII:GIIα ratios, I calculated a predicted rate of decay for both tested concentrations of c61, with results shown in Table 8. These calculations indicate that the rate of α/β interface disruption is dependent on thiopyridone concentration as a roughly 10-fold increase in the initial rate of GluII subunit separation is observed with 100 µM c61, relative to 1 µM c61.

Table 8 – Calculations of the relative rate of GluII subunit separation based on a one phase exponential decay model fitted to the data in Figure 22. The model was fitted and rate constant values extracted by GraphPad Prism (v9.5.0) and the rate of decay was calculated using Equation 2, where 'y' is the midpoint between the initial and minimum GluII:GIIα ratios from the fitted model.

[c61] / µM	Rate constant (k) / min^{-1}	Midpoint / $(y_{max}+y_{min})/2$	Rate of Exponential Decay / min^{-1}
1	0.0808	0.750	-0.0606
100	1.0860	0.538	-0.584

Having established that the observed behaviour of the GluII heterodimer in the presence of thiopyridones can be attributed to the compound itself rather than to the experimental conditions, I decided to explore further the possibility of a covalent interaction between the protein and c61. I opted to make use of the redox chemistry of disulfide bonds by investigating whether reducing conditions affect the GluII α/β interface disruption by thiopyridones. In

theory, if the protein ligand interaction is covalent then reducing conditions would disfavour the formation of disulfide bonds and would reduce the impact of thiopyridone treatment on the GluII heterodimer. To ensure that GluII would tolerate the proposed conditions, I incubated GluII (50 nM) and DTT (5 mM) for one hour at room temperature before taking MP measurements (Figure 23, Entry 2). I observed no appreciable difference relative to the sample of GluII alone (50 nM), suggesting that under these conditions the heterodimer tolerates the presence of reducing agent at high molar excess. I then repeated the described incubation of GluII but added 50 µM c61 (Figure 23, Entry 3) in order to compare the effect of reducing conditions in two ways (final concentrations in the measured sample are shown in parentheses):

1. Can reducing conditions *prevent* α/β interface disruption by c61? This experiment involved incubation of GluII (50 nM) with c61 (50 µM) and DTT (5 mM) together (Figure 23, Entry 4).
2. Can reducing conditions *reverse* α/β interface disruption by c61? Here, GluII (50 nM) was incubated with c61 (50 µM) for an hour, then DTT (5 mM) was added and the sample incubated for a further 15 minutes at room temperature (Figure 23, Entry 5).

The results of MP experiments with DTT are displayed in Figure 23, which shows the expected GluII:GIIα ratio for both the untreated and c61-treated GluII sample. Interestingly, performing a Šídák's multiple comparisons test (GraphPad Prism v9.5.0) reveals that when incubated with c61 and DTT simultaneously, GluII experiences significantly less α/β interface disruption than when treated with c61 alone. Similarly, when GluII is incubated with c61 alone and the sample is later treated with DTT, a significant increase in the GluII:GIIα ratio relative to the GluII/c61 sample is seen. These data suggest that reducing conditions significantly reduce the capacity of c61 to disrupt the GluII α/β interface without interfering with the protein directly and can reverse the disruption when it does occur. The measured differences in GluII:GIIα ratio for the described experiments, while significant, are modest and perhaps insufficient to

conclude that the changes observed are a result of reduction of a GIIα-c61 disulfide bond. Indeed, if this reduction were responsible for the prevention and/or reversal of α/β interface disruption by c61 then one would expect such harshly reducing conditions (10^6-fold excess of DTT relative to GluII) to have a more pronounced effect on the GluII:GIIα ratio. Nevertheless, the data in Figure 23 could suggest that the observed differences are mediated by some interaction between DTT and c61, since these MP data suggest DTT has no appreciable effect on the GluII heterodimer under these conditions. However, the data should be interpreted with caution, as MP is capable of measuring molecular weight independently of molecular shape so while the technique convincingly indicates the preservation of the α/β interface in the presence of DTT, it is unlikely to reveal any unfolding of the heterodimer that may contribute to the results shown in Lanes 4 and 5 of Figure 23.

To further probe the mechanism of GluII-thiopyridone binding, I first investigated the role of GIIβ on the behaviour of the GluII heterodimer following treatment with thiopyridones to validate the choice to omit the surface exposed cysteine on GIIβ (identified by Evotec during the DTB-IAA assay, Appendix 9.7) from the SDM experiments. The potential GIIβ cysteine of interest, C491, is located on the C-terminus of the protein and is therefore not present in the *Mm*GluII$_{Tryp}$ crystal structure. I decided to use MP to analyse a recombinant *Mm*GluII$_{Tryp}$ mimic, provided by Dr Mario Hensen, in which only the LDLRa domain of GIIβ that is present in the crystal structure is expressed. The GIIα-LDLRa complex will therefore will not contain C491. If C491 on GIIβ is key to the GluII-thiopyridone interaction, then I would expect to see less or no interface disruption when using the GIIα-LDLRa complex in place of the full length GluII heterodimer. To test this theory, the GIIα-LDLRa complex was incubated at room temperature for one hour both with and without 50 μM c61, then measured by MP with the results shown in Figure 24. The molecular weight of the LDLRa domain of GIIβ is only around 8.2 kDa, so is too small to be detected alone by MP, however the difference in size between the GIIα-LDLRa complex and lone GIIα is readily detectable. Indeed, in the absence of c61 MP detects a single

peak at 116 kDa, representing the GIIα-LDLRa complex, and following c61 treatment the MP peak shifts to 104 kDa corresponding to GIIα. These data suggest that C491 on GIIβ is not significantly involved in the protein-ligand interaction, which validates the decision to omit this cysteine from the SDM study.

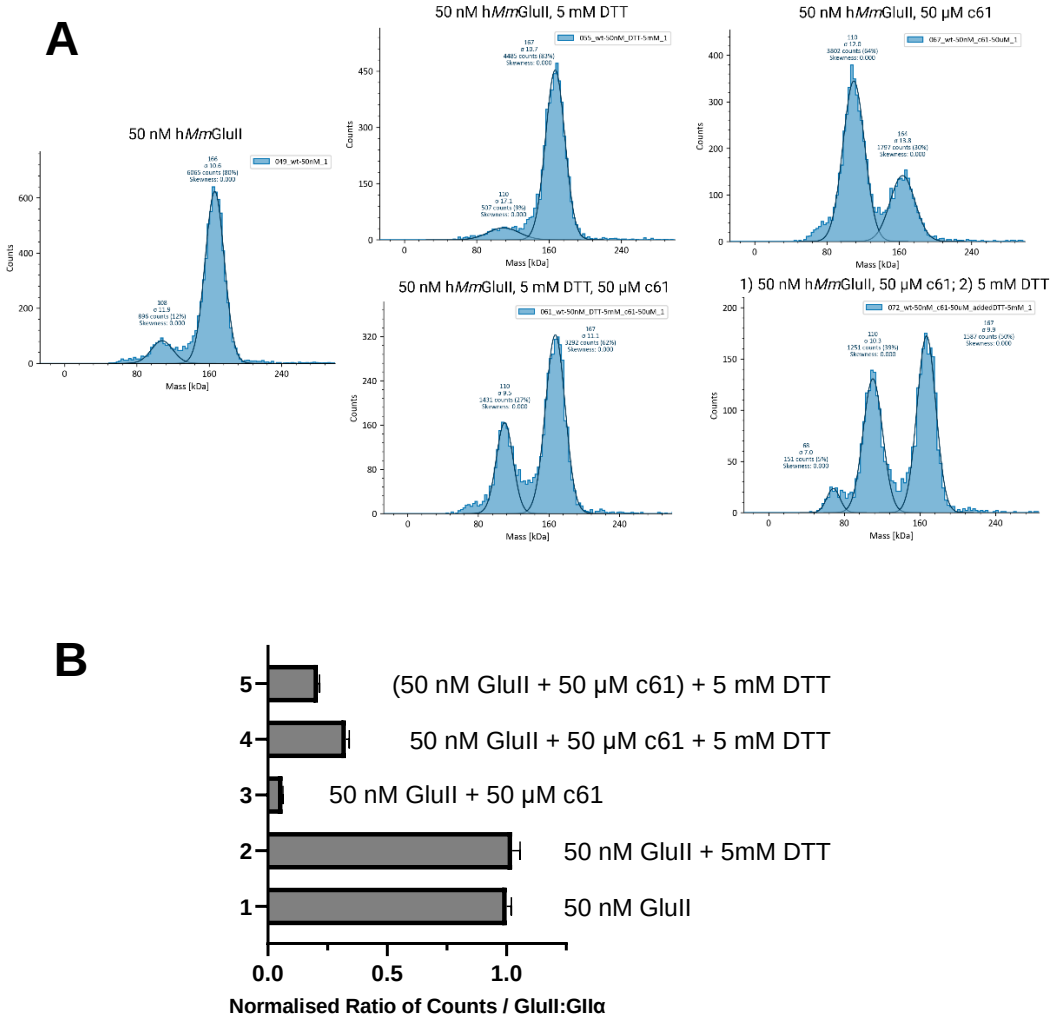


Figure 23 - Measuring the effect of DTT treatment on GluII α/β interface disruption by c61 using MP. (A) Frequency histograms of a representative sample of MP measurements of h*Mm*GluII treated with various thiopyridones. Six technical replicates were measured for each sample. (B) Interface disruption is measured as the ratio of counts, GluII:GIIα, normalised to the mean ratio in the untreated GluII sample. The mean value is shown as bars with SEM error bars. A Šídák's multiple comparisons test (GraphPad Prism v9.5.0) revealed no significant difference in GluII:GIIα between entries 1 and 2 suggesting that DTT alone does not affect the GluII heterodimer. The same test shows significant, albeit modest, differences in GluII:GIIα between entry 3 and entries 4 and 5, indicating that reducing conditions are somewhat capable of preventing (Entry 4) and/or reversing (Entry 5) GluII α/β interface disruption by c61.

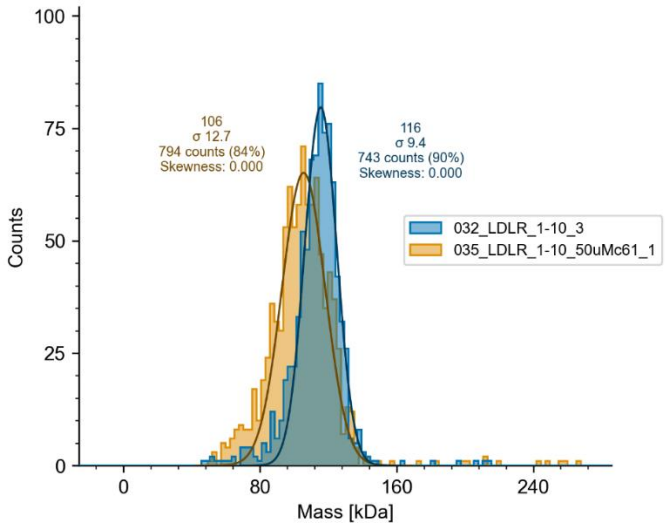


Figure 24 – MP measurements of the GIIα-LDLRa complex in the presence (blue) and absence (orange) of 50 µM c61. The data are shown as frequency distribution plots and the peaks are fitted with a Gaussian curve. The quoted counts and standard deviation (σ) are based on the Gaussian plot, with the centre of the peak corresponding to the molecular weight of the species. The data show a shift in molecular weight from 116 kDa (GIIα-LDLRa) to 106 kDa (GIIα) following treatment with 50 µM c61, suggesting that solvent exposed C491 on GIIβ, which is not present in the LDLRa domain, is not significantly involved in the GluII-thiopyridone interaction. The LDLRa domain is below the mass detection limit of MP, so no peak for this species is seen.

Next, I explored the effect of increasing concentrations of c61 on the four cysteine mutants described in Section 4.4.3. The theory was that if any of the solvent exposed cysteine residues identified during the DTB-IAA MS assay (Section 4.4.2.1) were involved in interactions with α/β interface disrupting thiopyridones, then their removal from the GluII primary sequence should reduce the impact of thiopyridone treatment on the α/β interface. The data in Figure 25 show that for C19/25A, C423A and C502A the response to c61 treatment is similar to that of h*Mm*GluII and indeed a two-way analysis of variance (ANOVA) performed using GraphPad Prism shows that there is no significant difference in the mean GluII:GIIα ratio between these mutants and the parent protein at c61 concentrations over 5 µM. For lower concentrations, C19/25A and C502A show modestly significant differences in mean GluII:GIIα ratio compared with h*Mm*GluII. In the case of C822A, representing the cysteine residue closest to the putative thiopyridone binding pocket discovered in the D-mannose VS, the data show a reduced sensitivity to α/β interface disruption by c61. Indeed, the two-way ANOVA test shows a significant difference in mean GluII:GIIα ratio relative to h*Mm*GluII at all tested concentrations.

By interpolating the mean value for c61 concentration from the Boltzmann sigmoidal fit modelled to the data of each mutant, I extracted the predicted XC50 value for the thiopyridone against each protein, with the results shown in Table 9. This table shows that c61 is modestly more potent against C19/25A and C502A, while the difference in potency against C423 is not significant. In contrast, c61 potency is markedly reduced against C822A, with a predicted 13-fold increase in the mean XC50. These data provide convincing evidence that C822 is involved in the GluII-thiopyridone interaction.

While the data in Figure 25 and Table 9 allude to a possible role of C822 in the GluII-c61 interaction, they are not sufficient to prove that the interaction in question is covalent. If the interaction was a disulfide bond then the MP profile for C822A in the presence of c61 should be similar to that of h*Mm*GluII with c28, because in both cases one of the sulfur atoms required for a disulfide bond has been removed. In reality, c61 is able to almost totally disrupt the α/β interface in C822A, though higher concentrations are required to induce the same effect seen for equivalent experiments with h*Mm*GluII. In Figure 26, I show the comparison in MP data for h*Mm*GluII and C822A with c61 and c28, which demonstrates that my initial predictions for this experiment were proven incorrect. In combination, the presented MP data suggest that the predicted disulfide interaction between thiopyridones and GluII is unlikely, though MP measurements of C822A with c61 confirm experimentally the computationally predicted D-mannose binding pocket as the likely site of interaction for thiopyridones. While GIIα-thiopyridone binding may be non-covalent, questions still remain about the stark difference in SAR between c28 and c61 given their close structural similarity and isoelectronic nature.

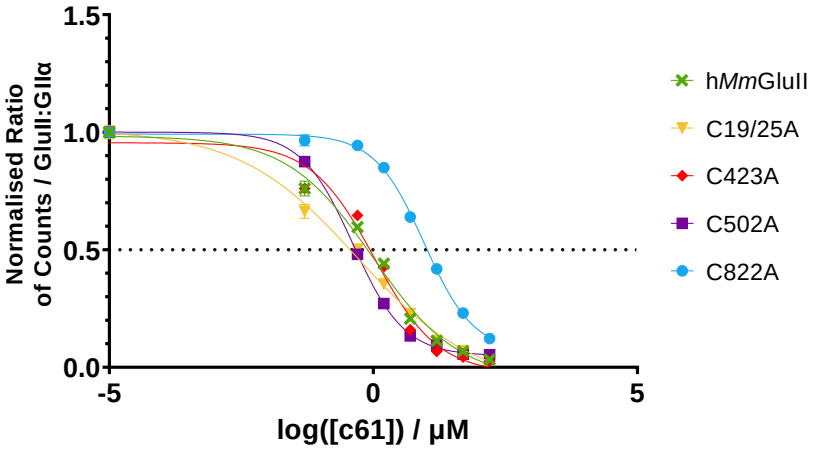


Figure 25 – An MP experiment comparing the impact of increasing concentrations of c61 on h*Mm*GluII and several GluII cysteine mutants described in Section 4.4.3. The final protein concentration in each sample was 50 nM. The points on the graph represent the mean GluII:GIIα ratio, normalised to that of the untreated protein, and error bars measure SEM. Six technical replicates were measured for each sample and points with no error bars represent samples where SEM was too low to be displayed on the graph. A non-linear regression analysis (GraphPad Prism, v9.5.0) fitted Boltzmann sigmoidal curves to the data for each protein. A two-way ANOVA test (GraphPad Prism, v9.5.0) was used to compare the mean GluII:GIIα ratios for each cysteine mutant relative to h*Mm*GluII (green) at each concentration of c61. This revealed that C822A (blue) is the only mutant for which the mean GluII:GIIα ratio deviates significantly from that of h*Mm*GluII at every tested concentration. For C19/25A (yellow) and C502 (purple), differences in mean GluII:GIIα ratio relative to h*Mm*GluII were modestly significant at c61 concentrations below 5 µM, but insignificant at higher concentrations. C423A (red) showed no significant difference from h*Mm*GluII at any c61 concentration. Examples of representative MP histograms are shown in Appendix 9.9.

Table 9 – Predicted XC50 values for c61 against various GluII mutants. Predictions were made by using the Boltzmann sigmoidal fit in Figure 25 by interpolating the mean thiopyridone concentration when the normalised GluII:GIIα ratio is 0.5, indicating 50% α/β interface disruption. The data show comparable potency of c61 for h*Mm*GluII, C19/25A, C423A and C502A but a roughly 10-fold decrease in potency against C822A, suggesting involvement of C822 in the GluII-thiopyridone interaction. This experiment also legitimatises the interfacial D-mannose binding pocket as the likely binding site for thiopyridones, as predicted computationally during the initial mannose VS.

GluII Mutant	Predicted XC50 / µM	Upper Limit of predicted XC50 / µM	Lower Limit of predicted XC50 / µM
h*Mm*GluII	0.776	0.928	0.643
C19/25A	0.363	0.442	0.293
C423A	0.861	1.031	0.708
C502A	0.461	0.497	0.427
C822A	10.469	11.639	9.457

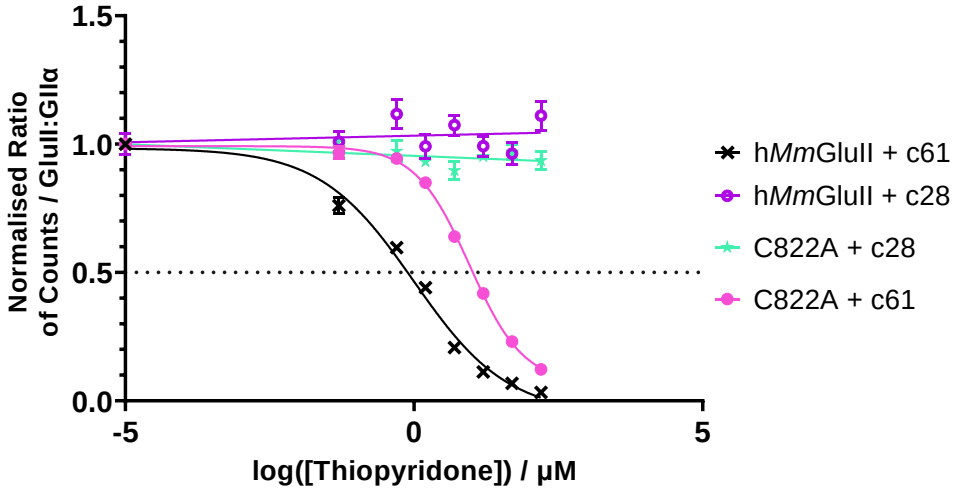


Figure 26 – MP data comparing the effect of treating h*Mm*GluII and C822A with active (c61) and inactive (c28) thiopyridones. The final protein concentration in each sample was 50 nM. The points on the graph represent the mean GluII:GIIα ratio, normalised to that of the untreated protein, and error bars measure SEM. Six technical replicates were measured for each sample and points with no error bars represent samples where SEM was too low to be displayed on the graph. A non-linear regression analysis (GraphPad Prism, v9.5.0) fitted Boltzmann sigmoidal curves to both data sets for c61 treatment, while a straight line was fitted to the data for c28 treatment. These data suggest that active thiopyridones are not likely to disrupt the α/β interface by forming disulfide bonds to GIIα, as the h*Mm*GluII/c28 and C822A/c61 traces are not comparable.

4.4.5 Native Mass Spectrometry

Proteomics had so far provided limited evidence of thiopyridones interacting specifically with GIIα. The PRM MS/MS experiments of c61-treated GluII outlined in Section 4.4.2.2 reveal the presence of a C822 peptide bound to c61 in MS2 spectra, but at markedly lower abundance than the unmodified peptide such that no peak is visible in MS1 or indeed DDA MS/MS with the same sample. This suggests that while thiopyridones potentially can bind free cysteine residues via a disulfide bond, usually they do not, so this is unlikely to be the main mechanism of action of the compounds. To supplement these MS data, I chose to perform native MS experiments to discover more about the GluII-thiopyridone interaction without needing to destroy the protein during sample preparation. Native MS experiments were carried out by Wiktoria Sadowska (Benesch group) according to the protocol in Section 4.3.4.

To determine the suitability of GluII for analysis by native MS, the untreated protein in 500 mM ammonium acetate was measured and the spectrum is shown in Figure 27. Here, GIIα is

readily detected with good ion intensity with an experimental mass of 108.92 kDa. However, the GluII and GIIβ are detected at much lower abundances heterodimer (Expasy ProtParam theoretical mass: GIIα = 108.044 kDa; GIIβ = 58.574 kDa; GluII = 166.600 kDa)[114]. The variable glycosylation described in Section 4.4.2.2 may partially explain the observed difference in experimental and theoretical mass for GIIα (876 Da), as the glycan masses are not included in the ProtParam prediction, however the combined mass of the most commonly detected GIIα glycans (Table 6) is 2.3 kDa which exceeds the observed mass difference. When focusing on the GIIα peaks (m/z ~4950-5750) a low abundance peak with a molecular weight of 109.9 kDa is observed, representing a 1 kDa increase relative to the experimental mass of GIIα and a 1.8 kDa increase relative to the theoretical mass. It is possible that the low abundance species at 109.9 kDa indicates the presence of glycans identified in Table 6, with the difference in mass relative to the main GIIα peak closely matching that off the O-glycan on GIIα (MW $HexNAc_1Hex_1NeuAc_2$ = 947 Da; experimental native MS mass difference = 951 Da). If this hypothesis is correct, it could indicate a common cleavage during MS whereby GIIα most commonly does not carry the O-glycan leaving a less abundant subset that does. Another explanation could be that the low abundance species indicates a glycosylation site that is not usually occupied. Variable truncation of N- and O-linked GluII glycans may explain the discrepancy between the theoretical and experimental masses of GIIα by native MS.

Simultaneous quality control analysis by MP (data not shown) proved that the GluII sample exists predominantly as the heterodimer prior to injection into the mass spectrometer, suggesting that the relative difficulty in observing the GluII complex is an artefact of the native MS experiment.[142] It is possible that during sample ionisation and/or desolvation in native MS, the largely ionic α/β interface is disrupted by neutralisation of key negative charges such that GIIα, already significantly positively charged at the α/β interface, is more readily detectable than either GIIβ or GluII. Since GIIβ carries significant negative charge at the interface, it may be that positive ion mode native MS is not suitable for its detection; repeating this experiment

in negative ion mode could provide deeper insight into the reason for the absence of GIIβ in the data presented here. The overrepresentation of GIIα by native MS for a sample of untreated GluII is unusual, but does not disqualify the technique as a method of investigating the GluII-thiopyridone interaction, since I have collected ample evidence to suggest that GIIα and not GIIβ is key for this association. In some instances, the native MS spectrum also shows the presence of a high molecular weight species (279.12 kDa) that could correspond to a heterotrimer of GluII subunits with a 2:1 ratio of GIIα:GIIβ (data not shown). This species appears to have very low abundance and is likely not biologically relevant. The experimental mass of the native GluII heterodimer is detected as two distinct charge series with high (m/z 6300-7100) and low (m/z 3700-4000) m/z. This splitting of the GluII population suggests that the heterodimer is unfolding under these experimental conditions which may partly explain the low abundance of the native heterodimer in the data.

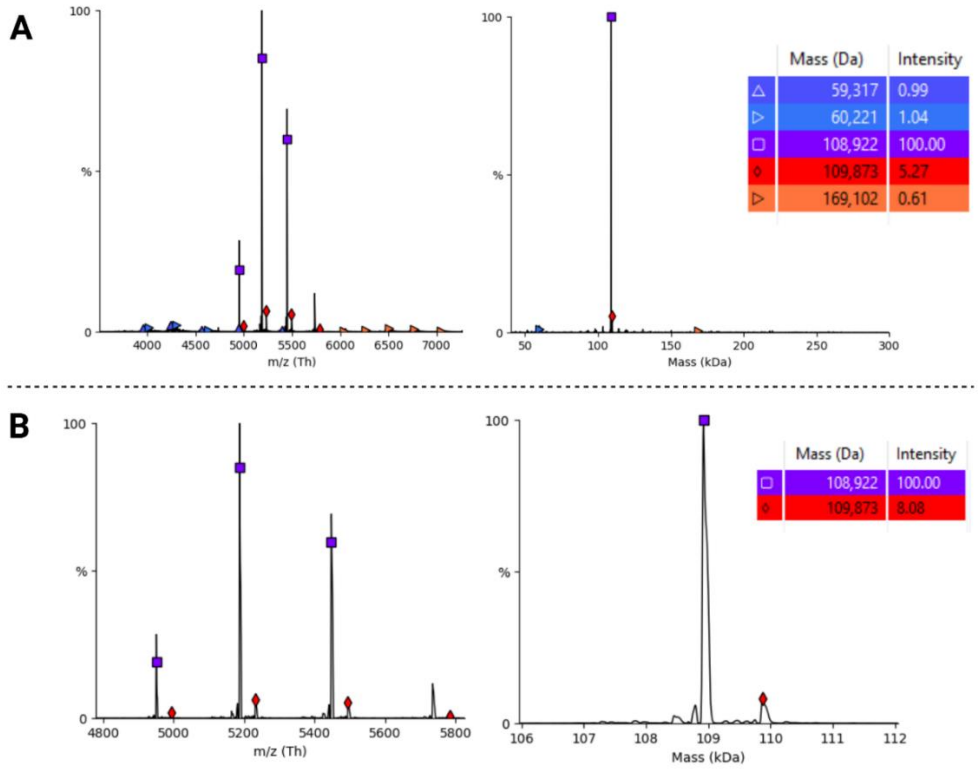
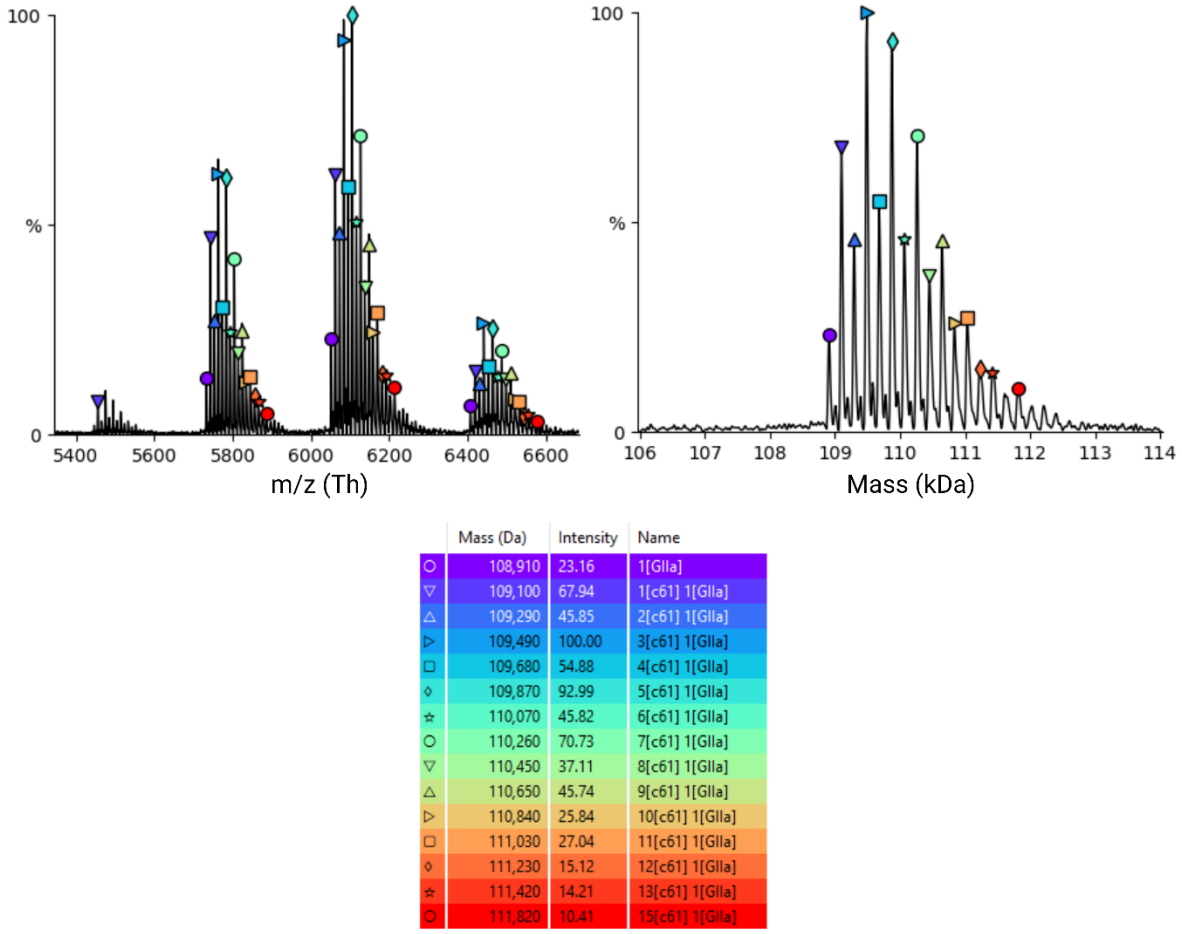


Figure 27 – Native MS spectrum of untreated GluII. (A) The full native MS spectrum. Here, GIIα is reliably detected while GIIβ and the GluII heterodimer appear at much lower ion intensity. (B) A zoomed spectrum showing the signals corresponding to GIIα only.

Having established facile detection of GIIα by native MS, I then examined the protein in the presence of c61. I opted to use 1.5% v/v of a 100 mM c61 stock in DMSO to achieve a final c61 concentration that was sufficiently high to maximise protein-ligand binding while keeping incubation times relatively short. This decision necessitated a control experiment using an equivalent volume of DMSO to rule out the possibility that the organic solvent alone would interfere with the GluII sample and/or its injection into the mass spectrometer (not shown). The data for the DMSO control was largely the same as that of untreated GluII, with the caveat that the charge state is altered, with DMSO-treated GluII appearing at higher m/z than untreated GluII. With short-term DMSO tolerance of GIIα verified, I treated a sample of GluII with 100-fold molar excess of c61 (1.5% v/v, 100 mM stock) relative to GluII and incubated the sample for 15 minutes at room temperature. The deconvolved native MS spectrum (UniDec[129]) of this sample is shown in Figure 28 and demonstrates the interaction of c61 with GIIα as periodic additions of roughly 200 Da (MW c61 = 193 Da) relative to the GIIα peak (108.92 kDa). The predominant peaks in each of the GIIα charge series correspond to GIIα bound to three units of c61 which shows roughly three-fold greater intensity than the GIIα alone. In a similar experiment using a lower concentration of c61 (20-fold molar excess compared to GluII), the predominant peak in the native MS spectrum was a 1:1 GIIα:c61 complex, suggesting that protein-ligand complex formation displays some dependence on ligand concentration. In Figure 28, several peaks are visible representing protein-ligand binding at higher c61 stoichiometry but interestingly, the peaks for GIIα bound to even numbers of c61 molecules (GIIα:c61 = 1:2, 1:4, 1:6) are lower in intensity than for odd numbers of bound c61 (GIIα:c61 = 1:1, 1:3, 1:5). This prompted the question, is it possible that c61 itself is dimerising? To test this theory, I performed ^{1}H and ^{13}C NMR and small molecule LCMS experiments on c61. These MS data identify molecular weights corresponding to both the c61 monomer ([M+H] = 194.079 Da) and dimer ([2M+H] = 387.021 Da, $C_{16}H_{11}N_4O_4S_2$) suggesting that the thiopyridone may indeed be capable of dimerisation, though the corresponding NMR data were inconclusive. The possibility of c61 dimerisation could explain the relatively low

abundance of GIIα seemingly bound to even numbers of c61 molecules, but it is unclear how relevant the c61 dimer is in the data obtained from the GluII-c61 sample.

Over time, the c61-treated GluII sample became more difficult to measure due to inability to establish stable spray of the sample likely caused by blockage of the injection needle. In SEC-MALS experiments confirmed with MS (Section 4.4.1), I observed the aggregation of GIIα following c61 treatment at the GluII concentration used (8.3 µM). Given that native MS used similar protein concentrations (7-14 µM), it stands to reason that GIIα aggregation is responsible for the limited unstable sample spray after prolonged incubation. To rule out DMSO as a cause for the poor data quality after hour-long incubation, I prepared a sample using 0.2% v/v DMSO using a 1 mM c61 stock (1% DMSO in PBS) and incubated for an hour at room temperature. The data collected from this sample were comparable to equivalent measurements using a higher c61 concentration, however the prolonged incubation led to the near disappearance the peak for GIIα without c61 in favour of the 1:1 GIIα:c61 species. The previously described stoichiometric relationship indicating the presence of the c61 dimer was observed again in this sample. Unfortunately, the lower concentration of DMSO had a meagre impact on the deterioration of data and sample quality over time, suggesting that GIIα aggregation is indeed responsible for the limited timeframe for dependable measurement by native MS.



	Mass (Da)	Intensity	Name
○	108,910	23.16	1[GIIa]
▽	109,100	67.94	1[c61] 1[GIIa]
△	109,290	45.85	2[c61] 1[GIIa]
▷	109,490	100.00	3[c61] 1[GIIa]
□	109,680	54.88	4[c61] 1[GIIa]
◇	109,870	92.99	5[c61] 1[GIIa]
☆	110,070	45.82	6[c61] 1[GIIa]
○	110,260	70.73	7[c61] 1[GIIa]
▽	110,450	37.11	8[c61] 1[GIIa]
△	110,650	45.74	9[c61] 1[GIIa]
▷	110,840	25.84	10[c61] 1[GIIa]
□	111,030	27.04	11[c61] 1[GIIa]
◇	111,230	15.12	12[c61] 1[GIIa]
☆	111,420	14.21	13[c61] 1[GIIa]
○	111,820	10.41	15[c61] 1[GIIa]

Figure 28 – Native MS spectra of GluII treated with 100-fold molar excess of c61. The charge series are shown on the left and the deconvolved mass spectrum on the right. These data demonstrate the binding of many equivalents of c61 to GIIα.

Native MS data of c61-treated GluII provide further evidence for the interaction between GIIα and thiopyridones. However, higher order c61 stoichiometries observed in these MS spectra may not be biologically relevant, especially given that c61 is capable of hydrogen bonding and is therefore likely to engage with GIIα in a non-specific manner. To test the specificity of the GluII-c61 interaction, we tested the effect of increasing HCD energy on the protein-ligand complex and the results are shown in Figure 29A. For c61-treated GluII, isocratic increase of collision energy in the HCD cell (0-180) resulted in conspicuous reduction in intensity of peaks corresponding to higher order c61 stoichiometries, while the peak representative of the 1:1 GIIα:c61 species became the most intense in each charge state. These data suggest that the 1:1 GIIα:c61 interaction is specific and characterised by a relatively high binding energy as it

is better able to withstand the harsh conditions of the HCD180, while other c61 molecules bound to GIIα are mostly lost. By performing the same experiment with c28 in place of c61 (Figure 29B) it is clear that the inactive thiopyridone binds less readily to GIIα even when no additional collision energy is applied, as only three significant peaks are visible, corresponding to unbound GIIα as well as GIIα bound to one to two units of c28. When HCD is increased to 180, all but the unbound GIIα peaks disappear suggesting that the limited binding of c28 is unspecific, in good agreement with data gathered from other techniques.

The final native MS experiment aimed to corroborate the MP data suggesting that reducing conditions can prevent and reverse some α/β interface disruption by c61 (Figure 23). To do so, I prepared a sample of 8 µM GluII with 0.8 M DTT in attempt to maintain the molar ratios used in MP. Perhaps unsurprisingly, the native MS data for this sample were poor and showed considerably lower signal-to-noise ratio compared to samples without DTT. Furthermore, the data showed a broadening of the charge series in the presence of DTT, suggesting that at such high concentrations DTT is inducing unfolding of the protein (Figure 30). A further complication for data analysis is that native MS shows peaks representing GIIα bound to DTT, with a mass difference from the unbound GIIα of roughly 160 Da. Due to the similarity in molecular weight between DTT and c61, it could become difficult to accurately distinguish between the two ligands by native MS when they are added as a mixture, especially if the quality of the data is sub-optimal. In hindsight, use of such high DTT concentrations was a misstep and it would be prudent to repeat this experiment with lower DTT concentrations while still using molar excess and perhaps longer incubation time. This conclusion is also pertinent to the MP experiment in reducing conditions outlined in Figure 23, as the native MS data confirm the unfolding of at least GIIα in the presence of large excess of DTT. The extent of α/β interface disruption by c61 may well be influenced by the unfolding of the two subunits in reducing conditions which could modify the access of the ligand to its binding pocket.

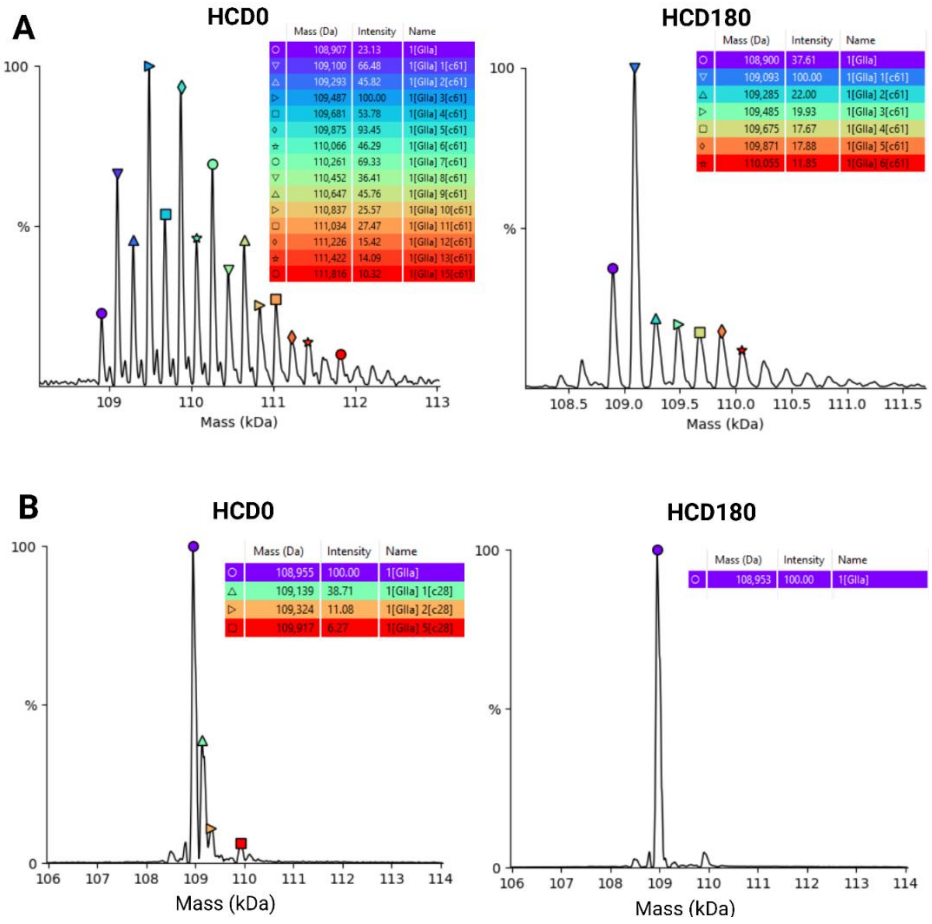


Figure 29 – The native MS2 spectra (m/z 5400-6700) of thiopyridone-treated GluII at low and high HCD energy (0-180). (A) GluII treated with 100-fold molar excess of c61. At HCD0 (left) many equivalents of c61 bind to GIIα, most of which are lost when HCD is increased to 180. The presence of the 1:1 GIIα:c61 peak at HCD180 indicates that the interaction is specific as it tolerates the high energy environment without dissociation. (B) GluII treated with 100-fold molar excess of c28. Fewer equivalents of c28 associate to GIIα and all are removed by increasing HCD. There are no specific interactions between GIIα and c28, consistent with MP and TR-FRET data.

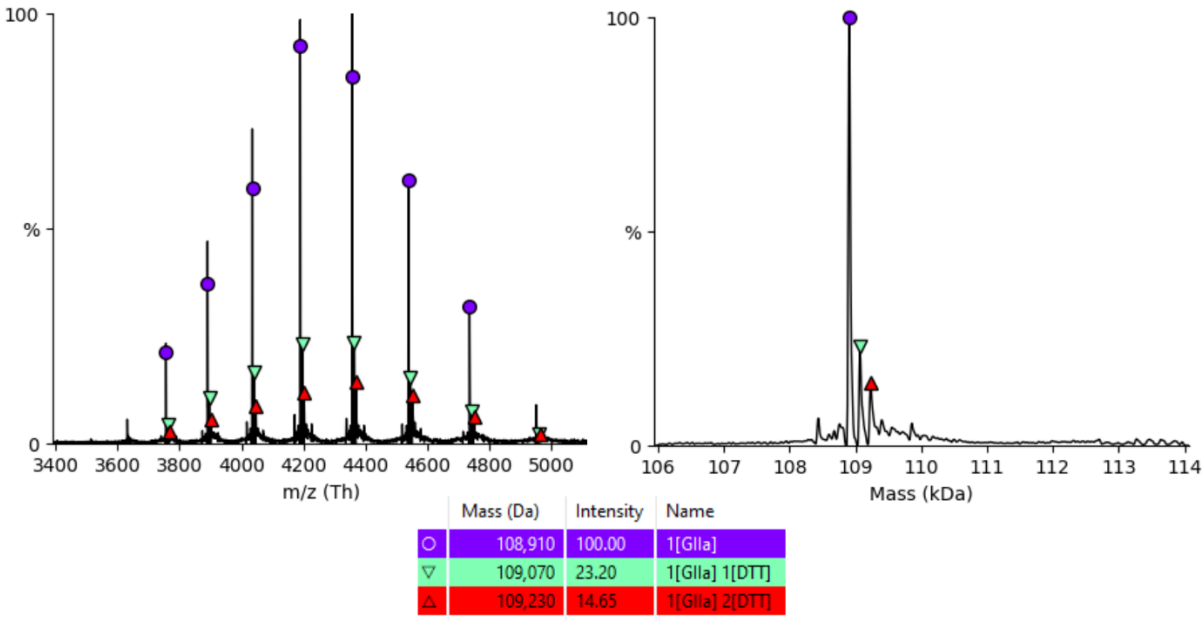


Figure 30 – Native MS spectra of GluII treated with 100,000-fold molar excess of DTT. GIIα is readily detectable but the broadening of the charge series with respect to the data for untreated GluII indicate that the high molar excess of DTT is inducing denaturation of the protein.

4.4.6 Impact of Thiopyridone Treatment on *In Vitro* α-Glucosidase II Activity

Until now, I have discussed thiopyridones in terms of their ability to disrupt the GluII α/β interface. This is quantified from MP and TR-FRET data as XC50, defined as the compound concentration that gives 50% α/β interface disruption relative to an untreated GluII sample. While a useful parameter for the described enzymology assay, it is not demonstrative of whether this α/β interface disruption corresponds to GluII inhibition. Subunit separation of GluII is likely to decrease the catalytic activity of GIIα, as shown by mutagenesis studies in which key arginine residues forming salt bridges in the α/β interface are replaced by glutamic acid (R837/840E), thereby abrogating the main interaction between GIIα and GIIβ.[54] These mutations result in reduction of GluII activity to around 10% of wt levels, but it is possible that the replacement of a positively charged residue with a negatively charged one may exaggerate this effect. In the case of α/β interface disruption by thiopyridones, it is likely that steric, rather

than electronic, effects dominate the mechanism of action meaning that thiopyridone treatment and mutagenesis of key salt bridges are not directly comparable in terms of their influence over GluII activity. Furthermore, the data presented here indicate that thiopyridones likely bind GIIα non-covalently and achieve equilibrium between the GluII heterodimer and the separated subunits. These factors suggest that thiopyridones may have a much less permanent effect on the α/β interface than introducing R837/840E and the associated electrostatic repulsion. Consequently, it is necessary to establish that thiopyridone treatment not only separates the GluII heterodimer but also inhibits the catalytic activity of GIIα.

Previous GluII activity assays make use of synthetic substrate mimics such as 4-MUG or pNPG, or free glycans with the same composition as the native substrate, i.e. Glc_1Man_9.[54,65] However, it has been shown that in the absence of GIIβ, GIIα retains more of its activity against synthetic substrates than for native substrate glycans, probably due to the action of the MRH domain of GIIβ.[63,65] I decided that synthetic substrates would not be suitable for the purpose of assessing the inhibitory properties of thiopyridones. Native substrate glycans are the most obvious choice for this analysis, however obtaining them involves either costly commercial purchase or their extraction and purification from recombinant glycoproteins. Given that native substrates of GluII are glycans on developing glycoproteins, I decided to use the heavily glycosylated HIV envelope glycoprotein, gp120, as the substrate in the activity assay as this more closely mimics the native substrate than free oligosaccharides. To simplify the analysis of the assay results, it was necessary to ensure homogeneity of gp120 glycans by treating cells with the mannosidase inhibitor, Kif, and glucosidase inhibitor, *N*B-DNJ during gp120 protein production. Use of these inhibitors yields gp120 with predominantly Glc_3Man_9 glycans. In theory, treatment with *N*B-DNJ and Kif would cause inhibition of the CNX cycle and ERAD, respectively (Figure 1), leading to the secretion of misfolded gp120. By comparing CD data for gp120 and Kif/*N*B-DNJ treated gp120, I demonstrate experimentally that treatment with glycosidase inhibitors has little effect on the secondary structure of recombinant gp120

(Appendix 9.13). In the context of the enzymology assay discussed below, the quality of the gp120 fold likely has no bearing on GluI/GluII activity, as the interactions of these glycoside hydrolases with their N-glycan substrates is independent of glycoprotein fold.[54,143]

The presence of Glc_3Man_9 glycans necessitates the use of GluI to perform the first glucose cleavage and for practical simplicity, I opted to perform GluI and GluII treatment of gp120 simultaneously. I performed the assay as described in Section 4.3.7 using the gp120/GluI/GluII mixture treated with high (1.5 mM) and low (0.78 µM) concentrations of c61 as well as untreated and DMSO-treated control samples. The lower concentration was chosen on the basis of XC50 values predicted by TR-FRET and MP, while the higher concentration used the maximum DMSO concentration tolerated by GluII (1.5% v/v, 100 mM c61 stock). Following GluI/GluII treatment, the gp120 glycans were analysed by HPLC and the results are shown in Figure 31. Using a $Man_9GlcNAc_2$ standard (Ludger) confirms the identity of the peak eluting at around 28 minutes as this glycan, with roughly equally spaced peaks (approx. 2 min) representing the addition or removal of single hexose residues. The data in Figure 31 indicate that the treatment of gp120 with Kif and *N*B-DNJ was mostly successful in generating predominantly $Glc_3Man_9GlcNAc_2$ N-glycans, with all Glc residues successfully removed following simultaneous treatment with GluI and GluII. At low concentrations of c61 (0.8 µM), no considerable inhibition of either ER glucosidase is observed, however using roughly 500-fold molar excess of c61 relative to the proteins (1.5 mM c61) yields glucosylated glycans in the HPLC profile (approx. elution times = 29.6, 31.1, 33.2 min). Interestingly, the elution of $Glc_3Man_9GlcNAc_2$ species at 33.2 min indicates the inhibition of GluI. In combination with the incomplete mannosidase inhibition by Kif to give considerable $Man_8GlcNAc_2$ and $Man_7GlcNAc_2$, the ostensible GluI inhibition somewhat confounds interpretation of the presented data, as the assumed $Glc_2Man_9GlcNAc_2$ (Figure 31, Lane 5) and $Glc_1Man_9GlcNAc_2$ (Figure 31, Lane 4) peaks could in fact correspond to $Glc_3Man_8GlcNAc_2$ and $Glc_3Man_7GlcNAc_2$, respectively.

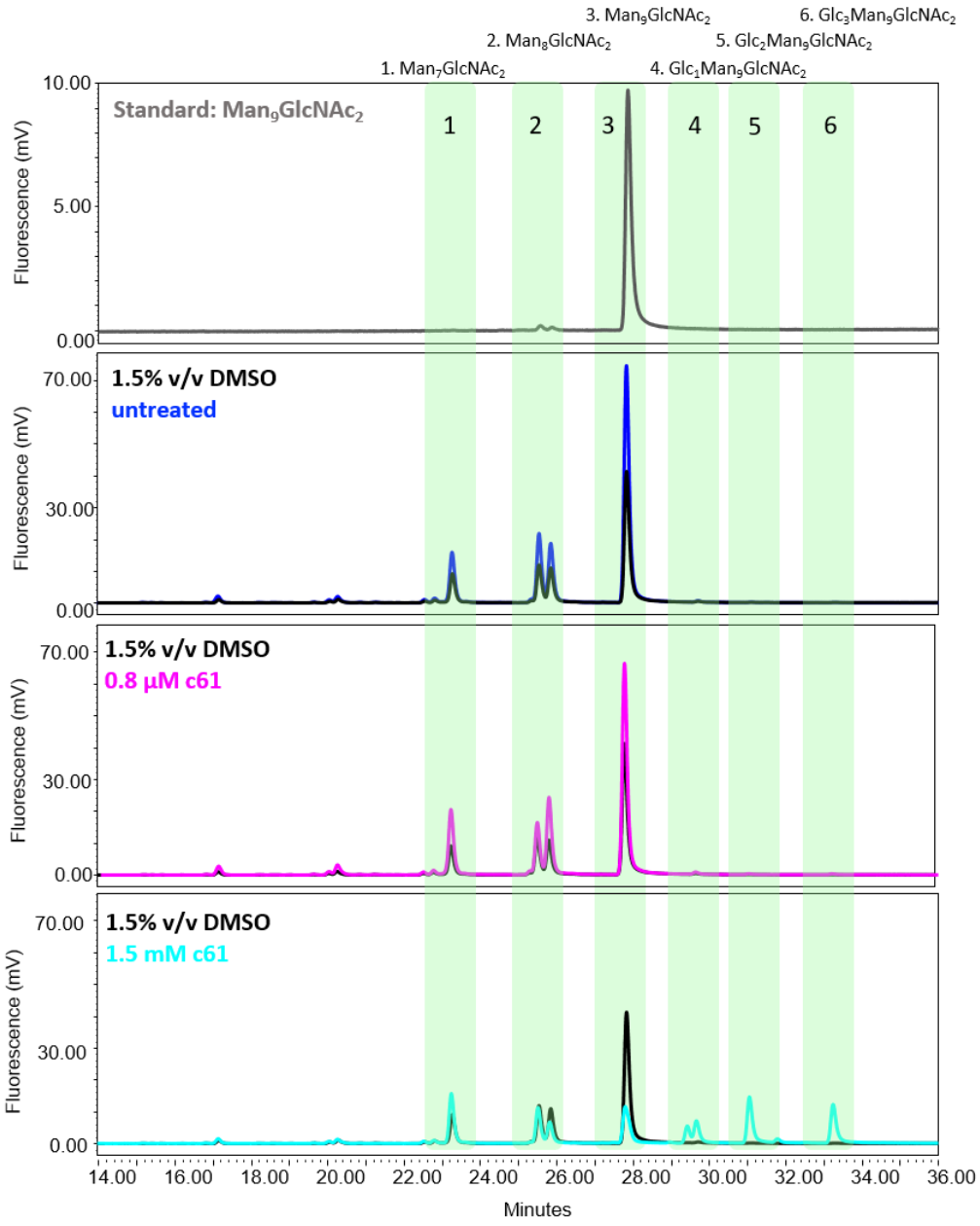


Figure 31 – HPLC chromatogram for 2-AA labelled glycans released from gp120 treated with GluI and GluII in the presence and absence of c61. The presence of glucosylated glycans (lanes 4-6) indicates ER glucosidase inhibition.

To confirm the off-target inhibition of GluI by c61, I repeated the *in vitro* enzymology assay using GluI alone with the data presented in Figure 32. The observation of a peak in HPLC corresponding to the $Glc_3Man_9GlcNAc_2$ species confirmed the inhibition of GluI when treated with 1.5 mM c61. The experiment in which GluI was not treated with c61 shows successful

removal of the terminal Glc residue from the substrate glycan, as observed in lanes 4 and 5 ($Glc_2Man_8GlcNAc_2$ and $Glc_2Man_9GlcNAc_2$, respectively). It is therefore likely that the species in lane 5 in the chromatogram for c61-treated GluI corresponds to the same $Glc_2Man_9GlcNAc_2$ species, suggesting that GluI is only partially inhibited by the thiopyridone which in turn suggests that in Figure 31, lanes 4 and 5 contain contributions from both glucosylated and deglucosylated glycans. If this hypothesis is correct, then we can conclude that c61 inhibits both ER glucosidases. However, it remains possible that in Figure 31, the species in lane 5 for the c61-treated chromatogram is $Glc_3Man_8GlcNAc_2$ again showing the complication introduced by the incomplete mannosidase inhibition by Kif during expression of the gp120 substrate glycoprotein. Further characterisation of the species from each peak is necessary to determine the extent to which c61 inhibits GluI.

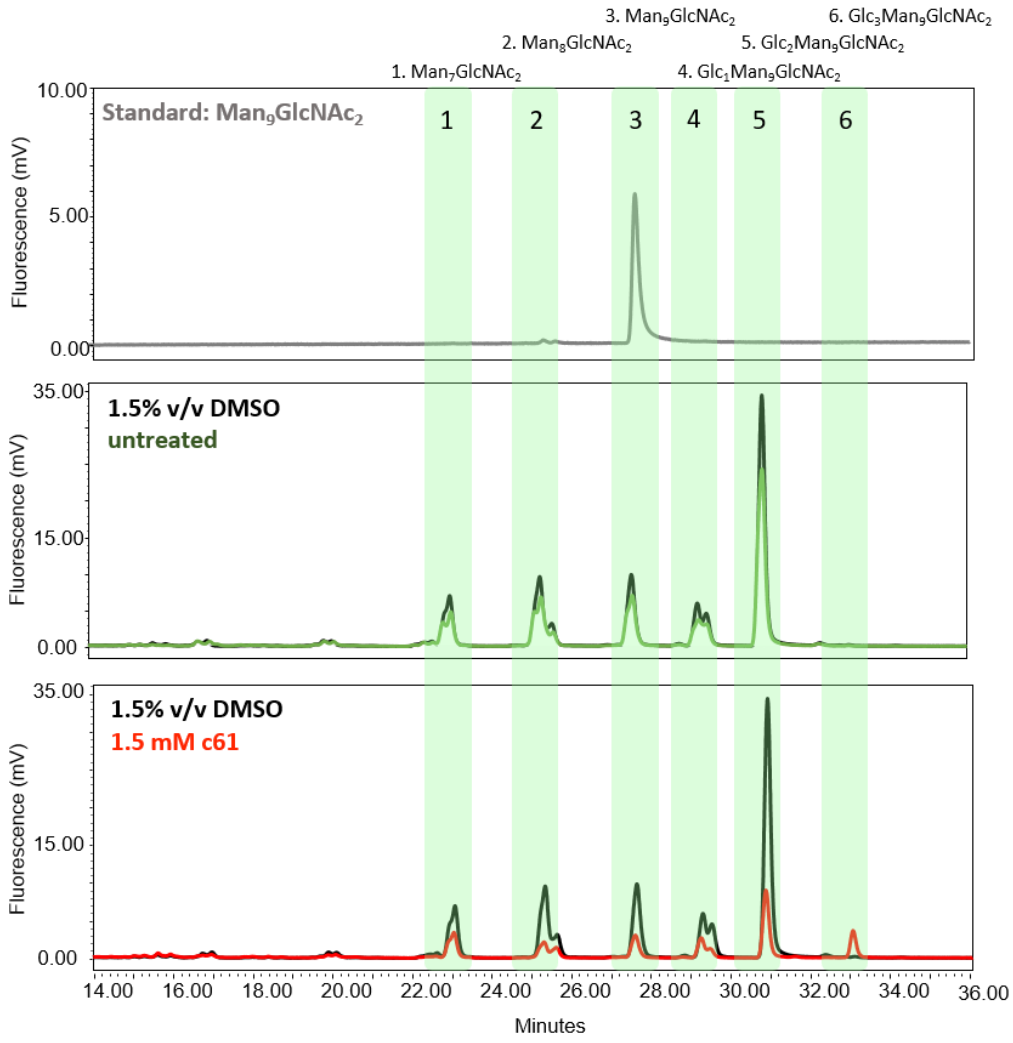


Figure 32 – HPLC chromatogram for 2-AA labelled glycans released from gp120 treated with GluI in the presence and absence of c61. The presence of triglucosylated glycans (Lane 6) indicates at least partial inhibition of GluI.

4.4.7 Evaluation of Thiopyridone Activity in Live Cells

The data presented in this thesis demonstrate that though the precise mechanism of action remains unclear, thiopyridones are capable of disrupting the GluII α/β interface *in vitro* and this disruption may lead to loss of enzymatic activity. However, GluII along with other ER resident proteins are often difficult to target with small molecules in live cells on account of compounds having to cross both the cell membrane and that of the ER itself. Drug delivery to the ER can be achieved using biologically inert liposomes that mimic and target the ER membrane before releasing encapsulated compounds.[144] More recently, several types of

ER-targeting nanoparticles have been reported in the literature.[145] Liposomes and nanoparticles, while effective at delivering their payload to the ER, would ideally be unnecessary in the event that drugs can reach their target protein unaided as is the case with iminosugar antivirals.[53] With this in mind, I decided to assess the capability of thiopyridones to reach GluII in the ER in live cells.

I first assessed the cytotoxicity of the GluII α/β interface disrupting thiopyridone compounds, c61 and c108, using an MTS assay with Huh7 cells (Section 4.3.8). This assay involves the metabolic reduction of the tetrazolium salt, MTS (in combination with phenazine methosulfate (PMS)) to give an insoluble formazan product which absorbs light at a wavelength of 490 nm. Since this reaction relies on cellular metabolic activity, the relative absorbance at 490 nm correlates to cell viability, where cells with healthy metabolism reduce more of the dye to give a higher absorbance reading.[146] By treating cells with active thiopyridones (c61 and c108), it is possible to assess the extent to which cell viability is reduced to extract a cytotoxicity value for the tested compounds. The results of the assay are shown in Figure 33, where thiopyridone concentration is plotted against the absorbance at 490 nm normalised to that of the untreated Huh7 cells. The assay shows that DMSO does not significantly affect cell viability at the tested concentrations and that the thiopyridones are non-toxic in Huh7 cells up to concentrations markedly higher than their *in vitro* α/β interface disrupting potency (XC50, Table 7) predicted by MP and TR-FRET. This is demonstrated by the predicted 90% cell survival concentrations (CC10), defined as the cytotoxic concentration of each compound that decreased cell viability by 10%, which are around 220 μM and 91 μM for c61 and c108, respectively. The tested thiopyridone concentrations revealed no totally cytotoxic concentration of either compound, with the highest tested concentration of 1 mM resulting in 46% cell viability for the c61 treated cells and 22% for c108, relative to the untreated cells. The data indicate that c108 is the more toxic of these two compounds, but limited availability of the thiopyridones restricted the use of higher concentrations in this assay. The observation of low cytotoxicity for these compounds

could indicate that thiopyridones do not effectively cross cell membranes, but this assay provides a platform from which to further explore the effect of the compounds on live cells.

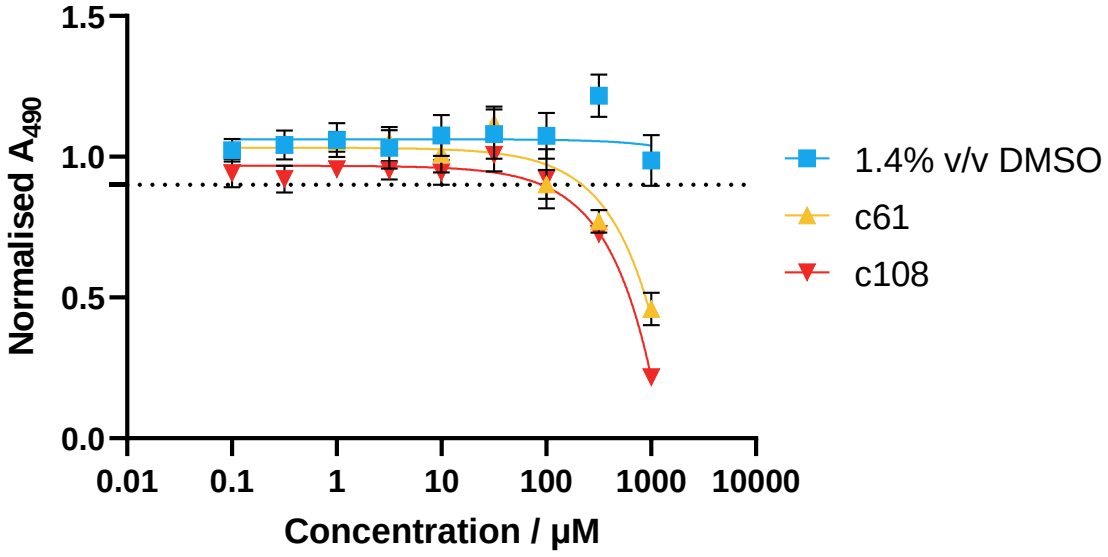


Figure 33 – MTS toxicity assay for α/β interface disrupting thiopyridones in Huh7 cells. The assay measured fluorescence relating to reduction of MTS dye, where lower fluorescence relative to untreated cells represents lower metabolic activity and, by extension, cell viability. CC10 values were extracted by interpolating mean x-values from the [Inhibitor] vs. Response non-linear fit applied using GraphPad Prism (v9.5.0). Upper toxicity limits were not found within the tested thiopyridone concentration range (0-1 mM). CC10: c61 = 220 µM; c108 = 91 µM.

The next step was to evaluate whether the thiopyridones that display *in vitro* activity against GluII in enzymology assays are capable of inhibiting the enzyme in live cells. To do so, I employed the HPLC-based FOS assay developed by the Butters group which monitors inhibition of ER GluI and GluII by determining the nature of cytosolic glycans released from the ER by PNGase F as part of the ERAD pathway (see Section 1.2.1).[52] Identification of triglucosylated glycans indicates inhibition of GluI, whereas di- or monoglucosylated shows inhibition of GluII. The FOS assay begins with treatment of Huh7 cells with thiopyridones using the CC10 values predicted in the previous MTS assay. The cytosolic glycans are then isolated following the protocol in Section 4.3.7 and analysed by HPLC. For this FOS assay, I compared the effect of thiopyridone treatment with that of *N*B-DNJ, a known GluII inhibitor that is able to penetrate to the ER to exert its inhibitory effect on both GluI and GluII. The data are presented in Figure 34, where α/β interface disrupting thiopyridones, c61 and c108, are seemingly unable

to inhibit GluII in live cells given the lack of glucosylated FOS present in these samples. In this assay, peaks in the NP-HPLC chromatogram are assigned to particular FOS based on their elution time, with full characterisation determined previously using glycosidase digests and MALDI-TOF MS.[52] For both c61 and c108, the NP-HPLC chromatogram closely resembles that of the negative control in which cells were treated with DMSO and are in stark contrast with the positive control, *N*B-DNJ. The *in vitro* inhibition of GluI and/or GluII by these thiopyridones was established in Section 4.4.6, suggesting that poor penetration of cell and/or ER membranes is likely responsible for the lack of GluII inhibition in Huh7 cells.

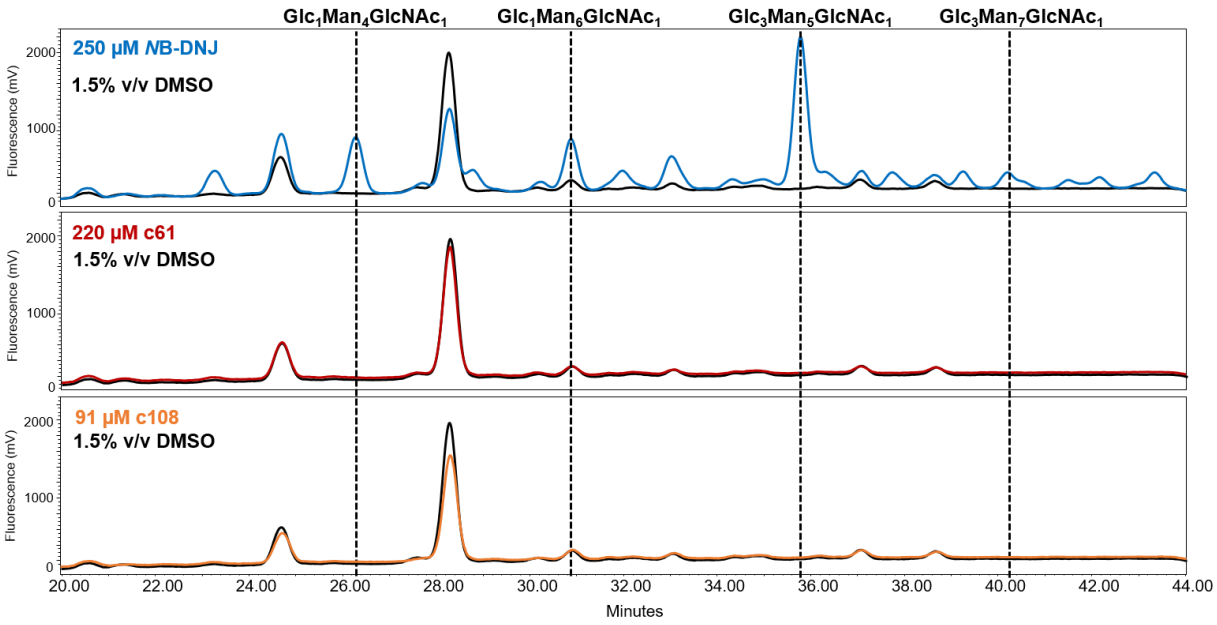


Figure 34 – Analysis of total cellular FOS by NP-HPLC following treatment of Huh7 cells with thiopyridones using *N*B-DNJ as a positive control. These data show an increase in the abundance of glucosylated FOS in the *N*B-DNJ control, while no obvious change is seen when cells are treated with c61 or c108 at their CC10 concentrations. This suggests that α/β interface disrupting thiopyridones are unable to inhibit GluII in live cells.

4.5 Conclusions

Previous work by the Zitzmann group had identified thiopyridones as potential GluII α/β interface disrupting small molecules, making use of a TR-FRET assay developed by Dr Mario

Hensen. Data from these assays appeared to show dose-dependent separation of the GluII subunits following treatment with thiopyridones. However, some uncertainty surrounded the results due to inconsistencies with the control assay suggesting that the α/β interface disruption may be overstated because the molecules interfere with the assay itself as well as with the GluII heterodimer. In this thesis, I present convincing data that confirm the legitimacy of the α/β interface disruption by thiopyridones shown by TR-FRET. Both SEC-MALS and MP allow for direct observation of GluII subunit separation, with the latter confirming the dose-dependent α/β interface disruption by FRET active thiopyridones. These techniques also verify the inactivity of thiopyridones in which the thiol group is blocked or replaced with oxygen, suggesting a significant role for this moiety in the protein-ligand interaction. The MP experiments whereby GluII is treated with increasing concentrations of thiopyridones allowed for calculation of XC50 values, which are in good agreement with FRET and provide confidence in the utility of the two techniques with respect to analysing the effect of the compounds on the GluII heterodimer. The SEC-MALS experiment, supported by SDS-PAGE and MS, not only corroborates the results of the TR-FRET and MP assays, but also demonstrates that GIIα aggregates under these experimental conditions. Protein aggregation was also observed during native MS experiments and limited the timeframe in which the technique can reliably measure GluII. The unbound GIIα is a likely culprit for this aggregation, providing further evidence for the known instability of GIIα in the absence of GIIβ.[54]

Thiopyridones were initially identified in a virtual screen of an interfacial pocket on GluII shown by X-ray crystallography to bind D-mannose. The *Mm*GluII$_{Tryp}$ crystal structure highlights the proximity of this pocket to C822, a residue bearing a free thiol moiety that was hypothesised to be a potential site for covalent binding to thiopyridones via a disulfide bond. Indeed, TR-FRET shows sensitivity of thiopyridone activity to alteration of their own thiol group suggests its involvement in GluII binding which could support the disulfide bond hypothesis. A

key aim of my research was to further investigate the GluII-thiopyridone binding mode to evaluate the legitimacy of this disulfide binding mode.

Many of the presented data provide loose evidence for the C822-thiopyridone disulfide bond. The interaction is indeed plausible, as MP experiments with a GIIα-LDLRa complex as well as N_3-TAMRA SDS-PAGE of GluII demonstrate that thiopyridones target GIIα rather than GIIβ. Native MS experiments confirmed that c61 engages in a specific interaction with GIIα with 1:1 stoichiometry and the same interaction is not observed for the inactive pyridone analogue, c28. Furthermore, the significant decrease in potency of c61 against the C822A GluII mutant shown by MP suggests a key involvement of this residue in the GIIα-thiopyridone interaction, confirming experimentally the computational prediction that the interfacial D-mannose binding pocket is indeed the most likely site of thiopyridone binding. The sensitivity of thiopyridone binding to the presence of a reducing agent (DTT) was demonstrated by MP and may support the disulfide bond hypothesis, as these conditions would favour cleavage of this bond to produce two free thiol groups. The most convincing evidence in favour of the formation of a protein-ligand disulfide bond comes from PRM-MS/MS experiments with GIIα following tryptic digestion, which reveal the presence of a C822 peptide covalently bound to c61. The peaks corresponding to this modified peptide are not present in the untreated GIIα sample, suggesting that the interaction is indeed real and not an artefact of the experiment. The interaction was also identified by MS following treatment of synthetic peptides containing a free cysteine with active thiopyridones, confirming that the formation of the proposed disulfide bond is indeed possible. Tangentially, MS data presented in this thesis also revealed the glycosylation of recombinant GluII, with GIIα found to bear an N- and core 1 O-linked glycan, while GIIβ carries a single core 1 O-linked glycan, though the O-glycans are unlikely to be functionally relevant to the endogenous, ER-resident enzyme.

Despite the ostensible possibility of a GIIα-thiopyridone disulfide bond, the data presented here more persuasively suggest that this is not the key mechanism of action for thiopyridones.

Firstly, while the proteomics experiments show c61 binding to a C822 peptide from GIIα in the form of MS2 daughter ions, their ion intensity and by extension their abundance is much lower than that of the unmodified C822 peptide. This would suggest that while C822 can form a disulfide bond with c61, it usually does not and thus the covalent interaction does not explain the profound effect thiopyridones have on the GluII α/β interface. The total inactivity of c28 with respect to α/β interface disruption was proposed to be a result of blocking the protein-ligand disulfide bond, however c61 still shows activity against the C822A mutant, albeit with lower potency. If the disulfide bond was truly the main interaction, then I would expect c61 to be as inactive against C822A as c28 is against GluII since both cases involve removal of one of the thiol groups required for this type of covalent binding. Data from an MP timecourse experiment with c61-treated GluII also suggests a reversible and therefore probably non-covalent mode of interaction between the protein and ligand, as the maximum α/β interface disruption by c61 proved to be concentration dependent. When using a final c61 concentration roughly equal to the MP and TR-FRET predicted XC50 value, I observed maximum disruption at around 50%. While this is the expected result given the definition of the XC50 parameter, one would expect a covalently binding α/β interface disruptor to eventually separate the GluII subunits entirely. Since an observable equilibrium is achieved under these experimental conditions, where c61 is present at roughly 20-fold molar excess relative to GluII, there must be an appreciable off rate for the compound suggesting that a covalent interaction is improbable.

The balance of evidence presented herein leans towards a non-covalent mode of action, but it remains unclear why interference with the thiol group on thiopyridones has such a profound impact on their capacity for α/β interface disruption. The close structural similarity between c61 and c28 and the fact that the compounds are isoelectronic would suggest that their non-covalent interactions with GluII should be largely the same. It is therefore puzzling that their activity towards GluII subunit separation should be so different. It is possible that the

increased hydrophilicity of c28 resulting from its more electronegative oxygen atom compared to sulfur in c61 (ChemDraw® - LogP: c28 = -0.08; c61 = 0.88) may affect the accessibility of the compound to the putative interfacial binding pocket, but this does not fully explain the difference as the pocket is known to accommodate D-mannose, which is considerably more hydrophilic than even c28 (ChemDraw® - LogP: D-mannose = -2.38). Obtaining structural data for GluII-thiopyridone complexes would clarify the binding mode and potentially explain the dramatic differences in activity between c61 and c28. Unfortunately, my efforts to crystallise this protein-ligand complex by co-crystallisation and soaking with thiopyridones, as well as using cross-linked GluII, were unsuccessful (see Appendix 9.10). This task is likely always to be problematic due to the inherent flexibility of GIIβ and lattice disrupting action of α/β interface disrupting thiopyridones complicates GluII crystal growth.

The true mode of thiopyridone binding to GluII remains unclear, but perhaps more important for the future of these small molecules as clinical leads is their ability to inhibit the catalytic activity of GIIα. Previous publications provide confidence that interface disruption should reduce GIIα activity, but the extent of the reduction following thiopyridone treatment had yet to be investigated. In *in vitro* enzymology assays using gp120 as a substrate glycoprotein, it was revealed that while GluII may be inhibited by c61, GluI is certainly inhibited bringing the selectivity of thiopyridones into question. The presence of 2-AA labelled glucosylated glycans in HPLC chromatograms following treatment of gp120 with GluI, GluII and c61 may indicate inhibition of both glucosidases, however incomplete mannosidase inhibition by Kif during recombinant expression of gp120 yielded significant amounts of $Man_8GlcNAc_2$ and $Man_7GlcNAc_2$ glycans on the substrate glycoprotein. As HPLC is not capable of effectively distinguishing between different hexose monosaccharides, it is difficult to ascertain the true identity of the observed glycans and therefore impossible to know the extent to which c61 inhibits GluI or GluII based on the HPLC data alone.

To assess the activity of c61 in living cells, I first performed an MTS assay which confirmed that α/β interface disrupting thiopyridones show acceptable cytotoxicity profiles in Huh7 cells, with CC10 values significantly higher than their *in vitro* XC50, suggesting that α/β interface disruption may occur at tolerable thiopyridone concentrations. However, the results of a FOS assay using the CC10 concentrations for the active thiopyridones imply that no inhibition of GluI or GluII occurs in live cells. Given their *in vitro* activity, it is possible that the observed lack of glucosidase inhibition for thiopyridones in Huh7 cells is a result of the activity of MANEA, a gene that is expressed in this cell line.[51,147] Since the proportion of glucosylated glycoproteins escaping ERQC/ERAD and reaching the Golgi is likely to be low, the observed inactivity of thiopyridones in the FOS assay is probably a result of their inability to cross the cell and/or ER membrane. Regardless, it would be prudent to make use of a different cell line that does not express MANEA, such as Calu-3, for future attempts at this FOS assay.[148] It may useful to explore different cell lines for this assay, as more clinically relevant primary cell lines may be more appropriate for analysing ER glucosidase activity than the significantly modified cancer cell lines that are typically used in these cellular assays for the sake of convenience. For thiopyridones to be seriously considered as clinical lead compounds, it is necessary to first ascertain their pharmacokinetic properties.

4.6 Future Work

4.6.1 α-Glucosidase II-Thiopyridone Binding Mode

The combination of evidence gathered in this thesis suggests that a disulfide bond between C822 and thiopyridones is unlikely to be the key mode of binding to GluII, despite MS evidence that the interaction is possible. For thiopyridones to be considered as clinical lead compounds, it is probable that some medicinal chemistry optimisation of protein-ligand interactions will be necessary. It is therefore important to better understand the true mode of binding of these molecules to GluII.

Data from MP experiments where various cysteine mutants of GluII were treated with active c61 revealed that the C822A mutant showed increased tolerance to the presence of the ligand with respect to α/β interface disruption, relative to the other cysteine mutants and the parent protein. These MP data indicate a role for C822 in the GluII-c61 interaction, even if proteomics data suggests that a disulfide bond does not significantly contribute. Native MS experiments showed a specific, 1:1 interaction between GluII and c61. Given more time, I would repeat the native MS experiments described in Sections 4.3.4 and 4.4.5 using the GluII cysteine mutants to confirm the observations from MP and to corroborate the importance of C822 for the GluII-c61 interaction. If the specific 1:1 protein-ligand interaction observed in native MS with h*Mm*GluII is lost for the C822A mutant, then we can be sure that this cysteine residue is pharmacologically relevant in this system. This approach could be applied to other residues in the interfacial D-mannose binding site in which c61 was predicted to bind *in silico* and confirmed experimentally using MP. By systematically removing residues in this pocket by SDM, it may be possible to use MP and native MS to determine the influence of their removal on α/β interface disruption by c61 and indirectly build a map of important GluII-c61 interactions.

In the absence of GluII-thiopyridone crystallography data, cryogenic electron microscopy (CryoEM) provides an alternate route towards gathering structural data for this protein ligand system.[149] Recent advances in this technique have enabled observation of proteins in complex with small molecules with resolution up to 2.2 Å, making it a valuable tool for in structure-based drug design.[150] One benefit of CryoEM over X-ray crystallography is that it does not require total homogeneity of the protein sample and may therefore be more appropriate for analysing proteins with intrinsic flexibility, such as GIIβ. This may provide a means of obtaining structural data for full-length GluII, which is discussed further in Appendix 9.12. In the context of α/β interface disruption by thiopyridones, it may be possible to obtain structural data of the GIIα-c61 complex using CryoEM as the technique is capable of imaging and sorting particles from a heterogeneous mixture on the electron micrograph. With this in mind, GluII subunit

separation following treatment of GluII with c61 and the resulting sample heterogeneity would not necessarily preclude analysis by CryoEM as it may be possible to select and analyse only the particles that correspond to lone GIIα. The GIIα particles that may be bound to c61 could be separated from GIIβ and indeed from any aggregates of GIIα that were observed under the experimental conditions for SEC-MALS and native MS.

4.6.2 Quantifying ER Glucosidase Inhibition by Thiopyridones

The *in vitro* enzymology assay with GluII using gp120 as a substrate showed that in the presence of c61, glucosylated glycans from gp120 are observed by HPLC after overnight incubation with GluI and GluII. These data indicate that GluII and/or GluI lost capacity for glucose trimming following c61 treatment. The observed inhibition of GluI by c61 is perhaps unsurprising given that the fragment was identified as a ligand for a D-mannose binding pocket at the GluII interface. It is therefore possible that c61 may bind any enzyme capable of binding mannose, so it is therefore necessary to determine the activity of c61 against other mannose lectins to assess the true selectivity of this chemotype for ER glucosidases.

The next steps in establishing α/β interface disrupting thiopyridones, in particular c61, as legitimate clinical lead compounds is to confirm and quantify their inhibition of GluII *in vitro* using the assay described in Section 4.4.6, but performing sequential glucose trimming of the substrate with GluI and GluII, rather than employing both enzymes simultaneously. Incubating gp120 with untreated and c61-treated GluI, followed by untreated and c61-treated GluII would confirm whether the thiopyridone completely or partially inhibits GluI at high concentrations and using the GluI-treated gp120 as the substrate in subsequent GluII treatment would tell us whether α/β interface disruption results in loss of GluII activity. Another approach to determining the nature of glucosidase activity in the presence of thiopyridones would be to collect HPLC peak fractions and analyse the glycans by MS and/or by use of glycosidase digests. Fully characterising the glycans would reveal the nature of GluI/GluII inhibition by c61.

Using the current data, the true extent of GluII inhibition by c61 is masked by the clear inhibition of GluI.

Quantification of GluI/GluII inhibition could be determined by repeating this assay with varying concentrations of c61 and analysing the released gp120 glycans by HPLC. By calculating the area under each peak relative to a $Man_9GlcNAc_2$ standard of known concentration, I can compare the relative amounts of $Man_9GlcNAc_2$ and $Glc_{1-3}Man_9GlcNAc_2$ obtained from the gp120 sample. These comparisons will allow for an estimation of IC50 based on the observed HPLC profile of glycans from gp120 without c61, a more clinically relevant metric than the XC50 described in FRET and MP experiments.

4.6.3 Thiopyridones as Host-Targeted Antivirals

The FOS assay outlined in Sections 4.3.9 and 4.4.7 showed that α/β interface disrupting thiopyridones are not capable of inhibiting GluII in living Huh7 cells. Since these molecules can inhibit GluII *in vitro*, it is likely that the lack of activity in live cells is a result of poor membrane permeability preventing them from reaching the ER.

To assess the membrane permeability of these compounds, I would suggest a parallel artificial membrane permeability assay (PAMPA) of the α/β interface disrupting thiopyridones.[151] Here, the compounds of interest will be placed in a donor well, before insertion of the donor well into an acceptor well that carries a lipid-infused artificial membrane. By analysing the content of the acceptor well after a period of incubation, typically by LC-MS/MS or UV-Vis absorbance, the degree to which the molecules cross the artificial membrane can be quantified. A full screen of active thiopyridones in this manner could inform the choice of compound to carry forward into cellular assays. It is possible that c61, the most potent *in vitro* α/β interface disruptor, may not be the most promising clinical candidate depending on the relative membrane permeability of the thiopyridones. For compounds that do not efficiently cross

membranes, it may be necessary to deliver them to the ER using polyunsaturated ER-targeting liposomes as previous work by the Zitzmann group has shown.[99]

According to the presented data, c61 inhibits GluI and does not efficiently cross cell membranes. Medicinal chemistry modification of thiopyridones is necessary to enhance their selectivity for GluII and membrane permeability to further their progress as valid clinical lead compounds. This process typically involves the systematic alteration of the functional groups on the molecule and assessing the impact these modifications have on protein-ligand interaction and on other biologically relevant factors such as aqueous solubility. SAR studies, such as those performed by Evotec (Section 4.1) focusing on the sulfur atom on thiopyridones, can reveal the key interacting moieties of the ligand. Medicinal chemistry modifications would continue until a thiopyridone analogue with optimal GluII selectivity, efficacy, cytotoxicity, solubility and steric/electronic profile is obtained. These steps require some understanding of the protein-ligand binding mode, highlighting the importance of the work proposed in Section 4.6.1. In the absence of experimental structural data, it is necessary to employ *in silico* strategies toward lead optimisation, in this instance making use of the computational model of the GluII-c61 complex identified in the initial VS of the interfacial D-mannose binding pocket.[152]

5. Synthesis of Novel α-Glucosidase II Active Site Inhibitors

5.1 Brief Introduction and Aims

The most obvious route towards novel inhibitors of GluII is to capitalise on the success of DNJ-derived iminosugars by further functionalising the existing scaffold. The DNJ-derived iminosugar that is currently most advanced in clinical trials in the context of HTA therapies and is generally well tolerated in humans is UV-4B.[85,86] This iminosugar has a low micromolar potency against both GluI and GluII that translates into micromolar IC50 values against various enveloped viruses *in vitro* and *in vivo*.[79,83] A recent report on phase 1 clinical trials with UV-4B revealed no dose-limiting toxicity in humans up to the maximum tested dose (1000 mg) and no dose-dependent increase in the occurrence of adverse effects (AEs).[84] However, gastrointestinal AEs were observed in dogs following a thrice daily 150 mg UV-4/kg/dose treatment; such side effects are consistent with the established AEs for patients treated with the approved DNJ-derived iminosugars miglustat and miglitol.[153,154] Furthermore, 10% of active human participants in the phase 1a trial reported mild nervous system AEs such as dizziness and issues with balance, though the incidence of these side effects was not dose-related. The trial suggests that potentially antiviral doses of UV-4B (146 to 293 mg/day human equivalent dose), based on the observed pharmacologically active dosing regimen in mouse models, could be safely administered to humans. That said, the typically limiting sample size in the phase 1a trial and the lack of an established antiviral dose of UV-4B in humans mean that we cannot yet confirm the safety and efficacy of the drug in human models. It is possible that at pharmacologically active doses, the severe gastrointestinal AEs observed in dogs thought to arise from off-target inhibition of intestinal disaccharidases could cause

similar problems in human participants. Despite the promising results from phase 1a clinical trials for UV-4B, identification of alternative iminosugars with improved selectivity, and possibly potency, relative to UV-4B would decrease the risk of dose-limiting AEs in human patients infected by enveloped viruses.

To improve selectivity of drugs for the ER glucosidases whose inhibition provides the desired HTA effect, it is important to target unique features of the enzyme. In the case of GluII, targeting the active site of GIIα with iminosugars comes with the risk of inhibiting other GH31 family enzymes that share analogous active site architecture. This off-target activity can lead to severe AEs such as the well-known gastrointestinal distress observed in humans following treatment with approved iminosugars miglustat and miglitol.[153,154] A key feature of GluII besides its heterodimeric quaternary structure discussed in Section 4, is the presence of the so-called "exclusion loop", a sequence of residues named for their ability to prevent the competitive inhibition of GluII by acarbose. Residues on the exclusion loop extend to the boundaries of the GIIα active site, with F307* showing interactions with the N-alkyl chain of UV-4B in crystal structures as shown in Figure 8. Similar interactions are formed with the F571 residue that contributes to the +2 subsite of the GIIα catalytic pocket, demonstrating the possibility of enhancing the selectivity of iminosugars for GluII by exploiting a larger proportion of the catalytic pocket (-1, +1, +2 and +3 subsites) and forming interactions with the exclusion loop unique to GluII among other GH31 family enzymes.[54] Indeed, many of the most potent GluII-inhibiting iminosugars incorporate an N-linked alkyl or aryl group that is capable of

* For clarity, residue numbering is based on the *Mm*GluII$_{Tryp}$ sequence seen in the crystal structure published by Caputo *et al.* (2016)[54], rather than the h*Mm*GluII sequence used experimentally in Section 4.

forming interactions with F286 and F550 on GIIα, with examples including NAP-DNJ and IHVR-19029 both of which show low micromolar potency in DENV infected cells.[95,99]

In collaboration with the XChem facility at DLS in 2016, a fragment-based lead discovery (FBLD) screen (unpublished) was carried out based on the *Mm*GluII$_{Tryp}$ crystal structure. The intention was to discover fragments which bound GluII allosterically to serve as the starting point for novel inhibitors. The screen involved 995 compounds, yielding twelve potential hits in various locations on the protein surface. Of these, two were shown to bind near the active site, fragment 027 (Figure 35) and 990 (Figure 36). The crystal structure of the *Mm*GluII$_{Tryp}$-027 showed greater ambiguity with respect to fragment interactions relative to the *Mm*GluII$_{Tryp}$-990, so 990 was chosen for further investigation. The 990 complex shows close proximity (~4 Å) of both the benzene ring and the 1,3,4-thiadiazole moieties on 990 to the side chain of exclusion loop residue F307, indicating potential for π-π stacking interactions. Furthermore, interactions between 990 and residues span the extended active site of GIIα (W402, W504, F652, F653, Figure 36) and could therefore help to improve the potency and selectivity of DNJ iminosugars for inhibition of GluII.

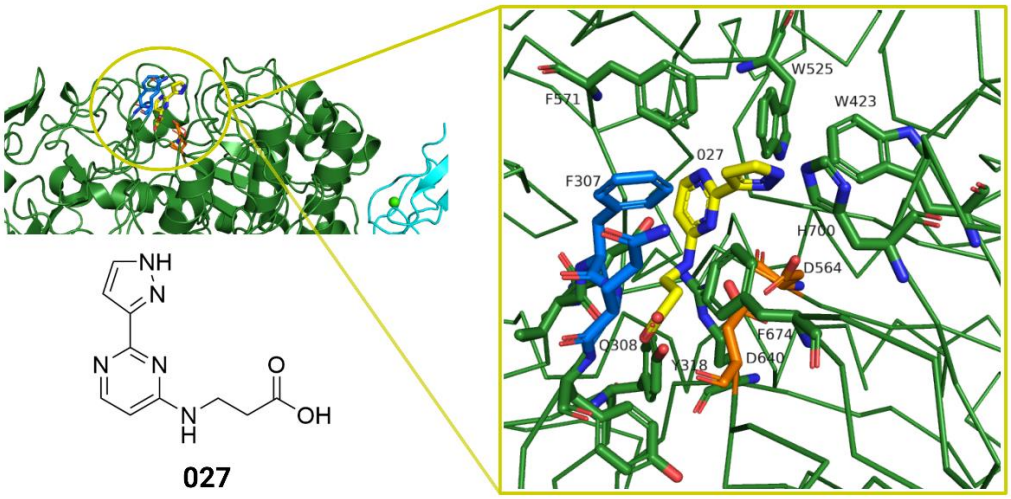


Figure 35 – Fragment 027, identified in an XChem FBLD screen that binds near the GluII active site. The 027-*Mm*GluII$_{Tryp}$ crystal structure shows close contacts between the fragment and the extended active site of GIIα. Interactions between 027 and the exclusion loop indicate the opportunity for its incorporation into novel selective iminosugar GluII inhibitors. Fragment 027: (3-((2-(1H-pyrazol-3-yl)pyrimidin-4-yl)amino)propanoic acid).

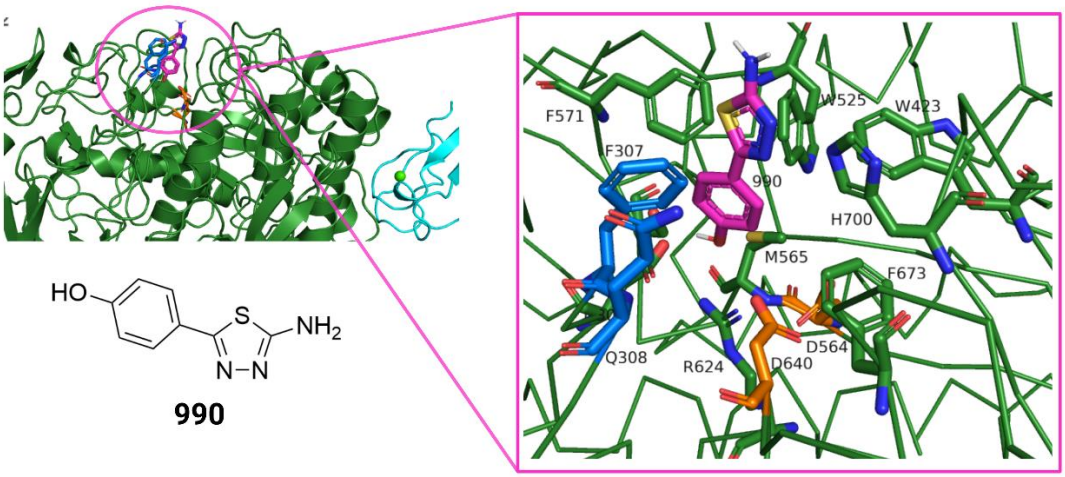


Figure 36 – Fragment 990, identified in an XChem FBLD screen that binds near the GluII active site. The 990-*Mm*GluII$_{Tryp}$ crystal structure shows unambiguous close contacts between the fragment and the extended active site of GIIα. Interactions between 990 and the exclusion loop indicate the opportunity for its incorporation into novel selective iminosugar GluII inhibitors. Fragment 990: 4-(5-amino-1,3,4-thiadiazol-2-yl)phenol.

I propose the target compounds shown in Figure 37, whereby 990 has been conjugated to DNJ via the endocyclic nitrogen atom, as possible novel DNJ-derived antivirals. In this chapter, I aim to synthesise these novel compounds and investigate the SAR between analogues of the target compound by varying linker lengths, benzene substitution patterns and composition of the 5-membered heterocyclic moiety.

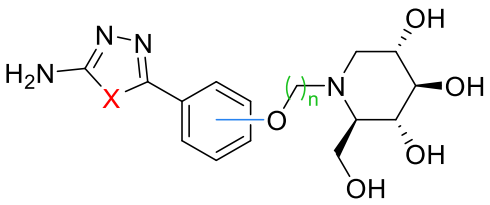


Figure 37 - Generic chemical structure of proposed novel DNJ-based iminosugars to be synthesised. The components to be varied for structure-activity relationship studies are colour coded: Red - X = S or O; Blue - Benzene substitution pattern = *meta* or *para*; Green - Length of alkyl linker, n = 3, 4, 5 or 6.

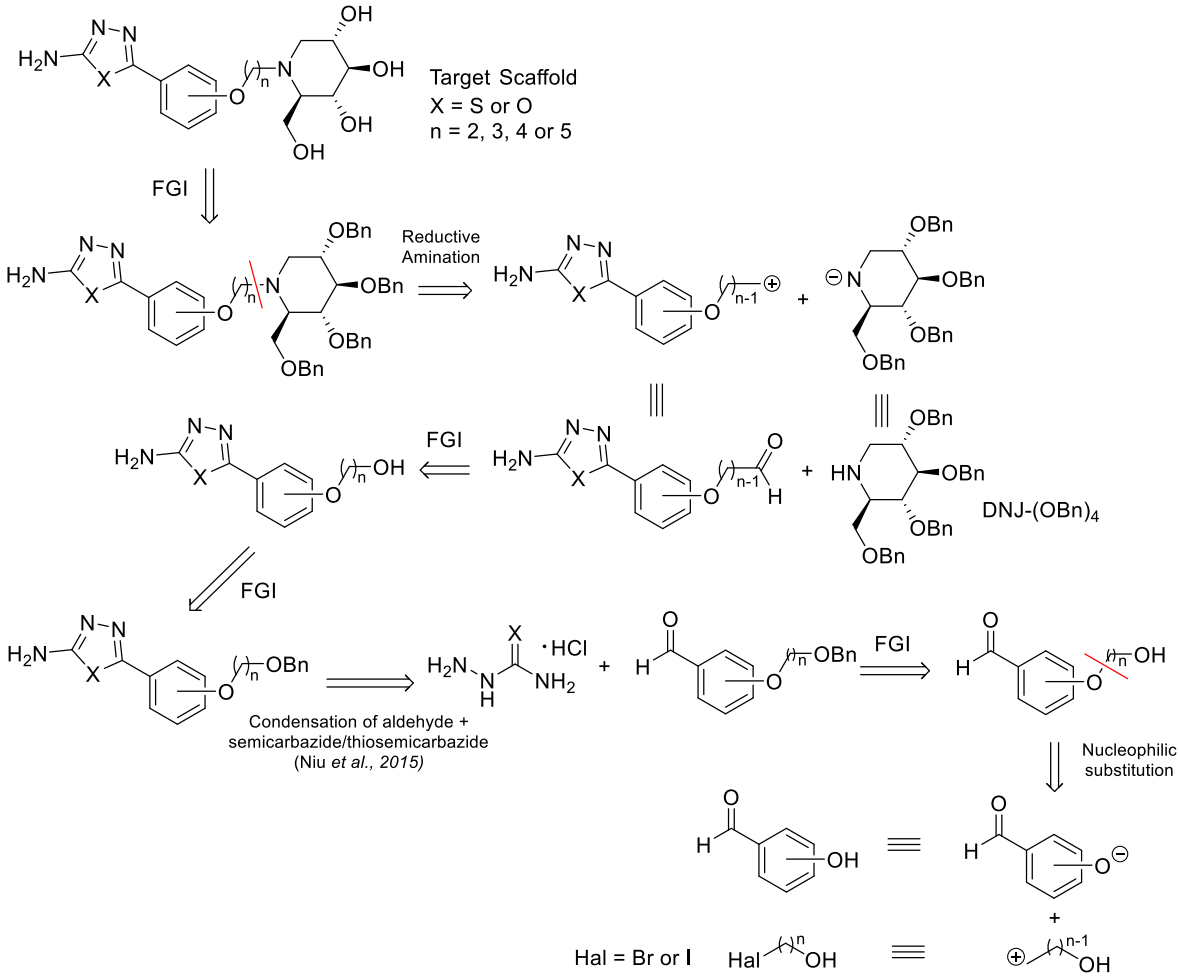
5.2 Proposed Synthesis

Before beginning the synthesis of my target compounds (Figure 37), it was first necessary to propose a plausible synthetic scheme. To do so, I performed a retrosynthetic analysis (RSA) on the target compounds in order to identify readily available starting materials and established protocols to achieve each synthetic step. When planning organic syntheses, RSA aims to reduce molecular complexity through the imaginary cleavage of bonds to yield smaller fragments (AKA synthons) that can inform the reactions necessary to construct the target compound. This provides a means by which one can formulate an efficient, convenient and flexible synthesis that takes into account the reaction selectivity of any intermediate compounds. By assessing the chemo-, regio- and stereoselectivity of intermediates, it is possible to consider steps that may require functional group interconversions (FGIs) and/or the addition of protecting groups to ensure accurate synthesis of the target compounds.[155]

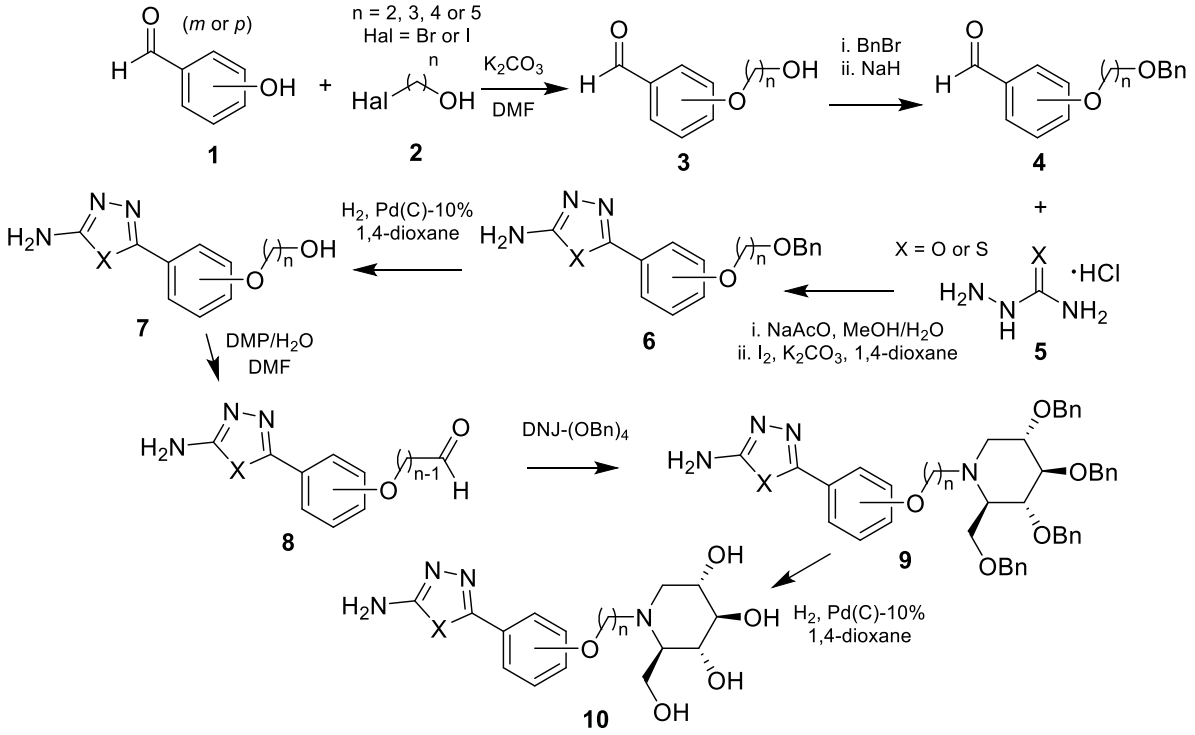
The RSA for the target DNJ-derived iminosugars is shown in Scheme 1. I identified that the simplest way to produce the starting materials would be to conjugate DNJ to an analogue of 990 with an attached linker, using the endocyclic nitrogen atom of DNJ as a nucleophile. Several syntheses of DNJ have been established and are discussed in Section 5.3.1.[156,157] The presence of multiple nucleophiles on DNJ necessitates the use of hydroxyl protecting groups to avoid low chemoselectivity which would drastically reduce the reaction efficiency. With benzyl-protected DNJ (DNJ-$(OBn)_4$) established as the nucleophile for the conjugation to 990, the 990 synthon needs to become a suitable electrophile. Given the lack of other electrophiles in the structure, the use of a reactive aldehyde could improve the rate of reaction with a nucleophile that is relatively sterically crowded and provides a direct route towards the uncomplicated alkyl linker in the target compound.

The next step of the RSA involved the deconvolution of the synthetic intermediate consisting of 990 and the alkyl linker with a terminal aldehyde. The reactivity of the aldehyde group may

contribute to lack of stability of this intermediate with respect to nucleophilic attack. It is therefore sensible to generate the aldehyde immediately before its reaction with DNJ-$(OBn)_4$ to enhance reaction efficiency and purity of the product. I reasoned that an FGI step whereby the aldehyde is obtained from a primary alcohol would reduce the risk of unintended reactivity of the electrophile. Next, I utilised the protocol for synthesis of 1,3,4-oxadiazoles and 1,3,4-thiadiazoles presented by Niu *et al.* in 2015.[158] This method is reported as an efficient and easily scalable route towards the desired functional group via the condensation of aldehydes with semicarbazide/thiosemicarbazide. The ostensible compatibility of this reaction with many types of aldehyde theoretically makes it a particularly convenient way of producing the 990 scaffold through use of a benzaldehyde starting material. Using benzaldehyde with a primary alcohol alkyl linker attached allows for a straightforward RSA C-O bond cleavage to arrive at hydroxybenzaldehyde as a nucleophilic synthon, however it would be sensible to add a protecting group to the primary alcohol to remove a nucleophile that may display undesirable reactivity. Using hydroxybenzaldehyde as a nucleophile requires the benzyl-protected primary alcohol linker to be the electrophile to allow for facile nucleophilic substitution, so I decided on linear haloalcohols as the readily available starting materials. In the forward synthesis, benzylation of the alkylated hydroxybenzaldehyde should be straightforward. The presented RSA provided a framework from which to begin the synthesis. The proposed synthesis is shown in Scheme 2.



Scheme 1 – A retrosynthetic analysis (RSA) for designing a synthetic scheme towards the target compounds shown in Figure X. Synthetic steps are either FGIs or bond cleavages, with the latter indicated as red lines bisecting the bonds to be broken. In the event of a bond cleavage, positively and negatively charged synthons are produced that represent electrophilic and nucleophilic starting materials, respectively. The synthesis of DNJ-$(OBn)_4$ and the condensation of an aldehyde with semicarbazide/thiosemicarbazide follow established protocols, so are not fully derived in this RSA.



Scheme 2 – Proposed generic synthesis of novel iminosugars combining DNJ with fragment 990. Variable starting materials include bromo- or iodo-alcohols with alkyl chains of different lengths, as well as semicarbazide or thiosemicarbazide depending on the desired heteroatom in the final product. Reagents are listed and the reactions are discussed further in Section 5.3.

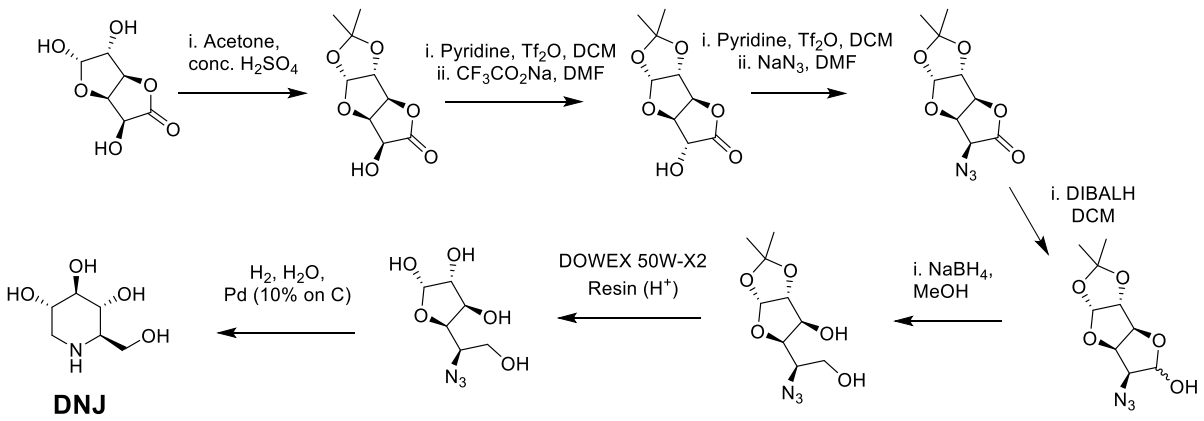
5.3 Results and Discussion

Detailed descriptions of the following syntheses can be found in Section 5.4, along with selected physical and chemical data.

5.3.1 Synthesis of 1-Deoxynojirimycin (DNJ)

The synthesis of DNJ was a necessary step for production of the target compounds, so I decided to employ a synthetic strategy published by Best *et al*. and previously used with success by Dr JL Kiappes, a former Zitzmann group member.[156] Scheme 3 shows the practical workflow used here. Unfortunately, the outlined synthesis proved to be complicated in my hands. In particular, the stereoinversion of the hydroxyl group in the second step of the synthesis consistently produced a clean, but unidentified product in moderate yields (< 40%). Poor solubility of the compound in the solvents used for silica column chromatography (ethyl

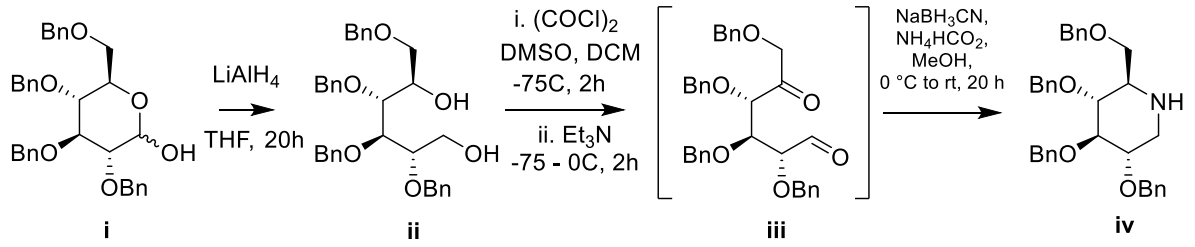
acetate (EtOAc) and petroleum ether (PE)) led to considerable loss of material. Physical data (^{1}H and ^{13}C NMR, MS – data not shown) did not match the expected results based on the literature values, confidently confirming that the product was not the desired intermediate. I hypothesised that insufficiently anhydrous conditions could have caused ring opening of the lactone ring, leading to a mixture of products in which the desired intermediate was not predominant. However, repeating the reaction with activated molecular sieves (3 Å, 8-12 mesh, Sigma Aldrich) to remove any contaminant water did not alter the results. My aim was to synthesise the proposed targets rather than investigate the mechanistic quirks of published reactions, so the DNJ synthesis in Scheme 3 was abandoned.



Scheme 3 - Synthetic route towards DNJ proposed by Best et al (2010).[156] In my hands, the stereoinversion step (second reaction) consistently produced a clean, unidentified product that did not match the literature data. This protocol was abandoned due to difficulty producing the desired intermediates.

In place of the Best *et al*. DNJ synthesis, I identified a shorter and seemingly more efficient strategy published by Wennekes *et al*. in 2007 which is outlined in Scheme 4 and described in detail in Section 5.4.1.[157] A key benefit of this alternative synthesis is that it directly produces benzyl-protected DNJ on a gram scale through use of a tetrabenzyl-D-glucopyranose starting material. Furthermore, the use of crude intermediates in subsequent reactions simplified the practical workflow and removed the risk of yield-limiting silica column chromatography that had reduced the efficiency of the previously attempted synthesis. Briefly, the Wennekes DNJ synthesis involves ring opening of 2,3,4,6-tetra-O-benzyl-D-glucopyranose via

$LiAlH_4$-mediated reduction. The resulting terminal alcohol groups were converted into aldehydes using a Swern oxidation, then the compound underwent a double reductive amination using ammonium formate (NH_4HCO_2) as a nitrogen source and sodium cyanoborohydride ($NaCNBH_3$) as a terminal hydride donor. Monitoring the reactions by thin-layer chromatography (TLC) (EtOAc:PE) I observed that the initial $LiAlH_4$-mediated reduction proceeded flawlessly with total consumption of the starting material. No unexpected side products were observed by TLC and the crude product was subjected to Swern oxidation without further purification. The oxidised intermediate was also not purified prior to the final reaction in which produced benzyl-protected DNJ as the major product by TLC (1:1 EtOAc:PE - R_f = 0.25, literature R_f = 0.25), though at least six other species were observable. Sequential silica columns (20% - 75% EtOAc in PE) were necessary to isolate the desired product and while this led to considerable inefficiency in the first attempt (pure yield < 10% over three steps), the large scale and relative practical ease allowed for significant improvements in subsequent syntheses (maximum yield = 65% over three steps). Analysis of the sample by MS and ^1H NMR showed clean production of the desired intermediate. With several grams of benzyl-protected DNJ in hand, I proceeded to the synthesis of the iminosugar targets.



Scheme 4 - Synthetic route towards DNJ proposed by Wennekes et al (2007). This strategy provided a facile, three-step synthesis to produce benzyl-protected DNJ (iv) with fair yield (65% over three steps).

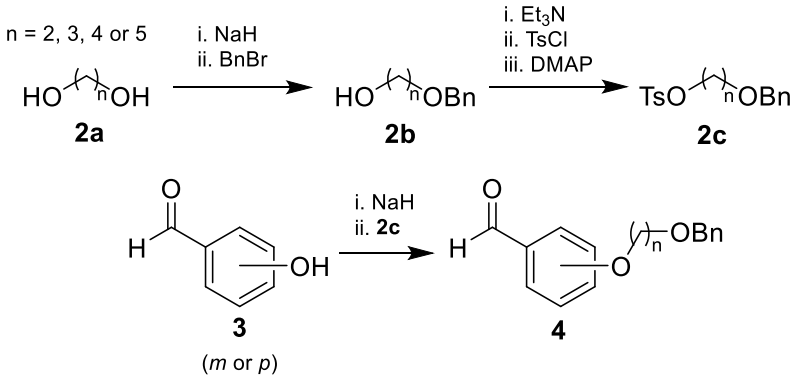
5.3.2 Alkylation of Hydroxybenzaldehyde

My initial attempts to conjugate the variable length alkyl linker involved the reaction of 3/4-hydroxybenzaldehyde with different haloalcohols and the subsequent benzyl protection of the resulting primary alcohol. Due to the availability of starting materials, I focused on the *meta* substituted benzene ring for my first attempts at the synthesis using 3-, 4- and 5-carbon linkers. The proposed reaction showed poor reproducibility with only the first attempt giving appreciable yields (> 40%), while subsequent attempts gave consistently poor yields (< 15%) and invariably demonstrated limited consumption of the 3-hydroxybenzaldehyde starting material by TLC. Of the tested linkers, 3C seemed to be the most amenable to these reaction conditions as the starting material had been mostly consumed after reacting for around 20 hours. On the other hand, the 4C and 5C linkers showed very little reaction progress in the same time frame and TLC showed that for 4C, and to a lesser extent 5C, many species were present among which the 3-hydroxybenzaldehyde starting material predominated. The relative inefficiency of the 4C and 5C reactions compared to 3C led me to hypothesise that the longer haloalcohols may be prone to intramolecular cyclisation under the given reaction conditions. This would explain the meagre consumption of 3-hydroxybenzaldehyde, though the hypothesis was difficult to prove as the reflux temperature used in the reaction (~150 °C) exceeded the boiling points (bp) of the suggested side products (4C – tetrahydrofuran (THF), bp = 66 °C; 5C – tetrahydropyran, bp = 88 °C). Limited access to appropriate glassware made collection of volatile side products difficult, so the hypothesis could not be properly examined. The presence of multiple species by TLC, particularly for the 4C linker, imply the presence of undesirable side reactions. It is possible that the haloalcohols are polymerising, forming ether chains of various lengths. This hypothesis explains the poor consumption of starting material and the presence of other species on the TLC plates. Separation of the species observed by TLC using silica column chromatography provided some intermediate 3 (Scheme 2) which was used in the subsequent benzylation step to give intermediate 4. This step was consistently unsuccessful after several attempts, with many undesirable side

products visible by TLC for all tested linker lengths. These species proved difficult to separate by silica column chromatography and having obtained prohibitively low yields (< 10%) of intermediate 4, I decided to abandon this synthetic route.

To make best use of the available starting materials, I decided on an alternative approach to hydroxybenzaldehyde alkylation using diols of various lengths in place of haloalcohols. The newly proposed synthesis is shown in Scheme 5. The use of diols necessitates additional FGI steps as the absence of a sufficiently electrophilic group on the diol precludes nucleophilic attack by the hydroxyl group on hydroxybenzaldehyde. Furthermore, the presence of two equally nucleophilic hydroxyl groups on the diol may increase the chance of undesirable nucleophilic attack on the benzaldehyde starting material. Consequently, the sequential benzylation and tosylation of the diols produced intermediate 2c (Scheme 5) which was used to alkylate 3, allowing for direct incorporation of the benzyl-protected linker. The benzylation reaction with 1,3-propanediol to produce 2b initially gave good yield (49%), with TLC identifying only two significant products. The more hydrophobic and therefore furthest running spot on TLC (9:1 PE:EtOAc) was assumed to be the doubly benzylated diol, while the lower running spot was confirmed to be 2b by ^{1}H NMR. Presuming the major side product to be the over-benzylated diol, I reduced the relative proportion of benzyl bromide (BnBr) in the reaction, using 5-fold excess of the diol. This improved the yield considerably (maximum yield = 87%) and the two products were easily separated using a silica column (0 – 60% EtOAc in PE). The FGI from 2b to 2c whereby the primary alcohol is replaced by a tosyl moiety is necessary for the introduction of a suitable leaving group to allow for nucleophilic attack by the hydroxyl group on 3. Owing to the reactivity of the tosylated alcohol, the crude material from this reaction was used in the subsequent alkylation of 3 without purification. This led to modest but appreciable yields for the alkylation step to produce intermediate 4 with a maximum yield of 30% using a 4C linker, following a silica column (0 – 60% EtOAc in PE). To streamline my synthetic procedure, I chose to focus only on this 4C linker compound with *meta*-substituted

benzene, reasoning that initially synthesising just one target compound to completion would better identify any problematic steps. This would save time in the event that an alternative synthetic strategy becomes necessary.



Scheme 5 – Alternative synthetic strategy for the alkylation of hydroxybenzaldehyde (3). This method was adopted following unsuccessful attempts using haloalcohols in place of the diol shown here. After successful synthesis of intermediate 4, I continued to follow the synthetic route outlined in Scheme 2.

5.3.3 Synthesis of 2-Amino-1,3,4-oxadiazoles and 2-Amino-1,3,4-thiadiazoles

The functionalisation of the benzaldehyde moiety in intermediate 4 with 2-amino-1,3,4-oxadiazoles/thiadiazoles would complete the synthesis of fragment 990 and variations thereon based on the target compounds depicted in Figure 37. A convenient strategy to introduce this functionality was published by Niu *et al*. in 2015, which involves reacting acidified semicarbazide/thiosemicarbazide (5, Scheme 2) with the aldehyde on intermediate 4 to yield intermediate 6 via via I_2-mediated oxidative cyclisation.[158] The protocol involves a simple two-step reaction without need of purification after the first step. This is a desirable approach towards the desired intermediate, with alternative strategies often involving toxic reagents such as mercury oxide, cyanogen bromide and triphenyl phosphine.[159–161] Niu *et al*. performed their presented synthesis on a variety of aldehydes, including several variously substituted aryl-aldehydes, yielding 2-amino-1,3,4-oxadiazoles at

high yield (85-99%) and 2-amino-1,3,4-thiadiazoles at lower, but still considerable yields (39-86%).

I attempted the Niu *et al*. protocol to produce oxadiazole and thiadiazole versions of intermediate 6 with a 4C linker (Section 5.4.3). The reaction proceeded as expected and I was able to cleanly produce both desired intermediates, however the presence of side products visible by TLC demanded purification by silica column chromatography (0-80% EtOAc in PE). The purification step proved inefficient and I obtained limited isolated yields of 52% for oxadiazole 6 and 30% for thiadiazole 6. By analysing the collected fractions (~10 ml) by MS (Agilent 6120 bench-top, single quadrupole), it became clear that the silica column provided poor resolution for separation of these intermediates, with compound 6 detected across fractions over an elution volume of 370 ml. The easily scalable synthesis by Niu *et al*. allowed me to cleanly produce sufficient intermediate 6 to carry forward to the next synthetic step. Should the remainder of my proposed synthesis prove successful, I will optimise this purification to provide a greener, more efficient route towards the target compounds.

5.3.4 Deprotection of the Benzylated Alkyl Linker

To remove the benzyl protecting group to form the primary alcohol linker (Scheme 2, Intermediate 6 - 4C, X=O), I decided to employ direct hydrogenation of the moiety using molecular hydrogen (H_2) and a palladium on charcoal catalyst (Pd(C) 10% wt. %), due to the relatively mild conditions used. The reaction proposed in Scheme 2 proved ineffective, with no significant consumption of the benzyl protected starting material after more than 12 hours, determined by TLC and MS. Several approaches were used to improve the efficacy of the reaction: Sequential sparging with N_2 then H_2 for > 45 minutes, a large flask and vigorous stirring to maximise the liquid surface area and encourage H_2 dissolution, activation of the Pd(C) catalyst (stirring in 1:1 H_2O:MeOH then drying under vacuum at high temperature), testing additional solvents (MeOH, EtOH, EtOAc and combinations thereof), alternative

hydrogen source (ammonium formate) with Pd(C), heating to reflux in EtOH. Eventually, I obtained intermediate 7 (Scheme 2) using similar conditions to those presented in Scheme 2, but with $Pd(OH)_2(C)$ (20 wt. %) as the catalyst and with the addition of a minimal volume of concentrated hydrochloric acid HCl. The acidification of the reaction mixture may have aided in hydrogenation by acidification of the oxygen atom on the benzyl group thus improving its leaving group ability, however the benzylated starting material was still not fully consumed after over 56 hours of reaction, with additional H_2, $Pd(OH)_2(C)$ and HCl added during this time. The product was separated from the starting material using a silica column (0 – 100% EtOAc in PE) and its identity confirmed by MS and ^{1}H NMR, with a yield of 68%. The MS spectrum shows the presence of a monomer ($[M+H]^+$ m/z 250.1) and dimer of intermediate 7 ($[2M+H]^+$ m/z 499.2) and ^{1}H NMR spectrum confirmed the presence of Intermediate 7. The obtained intermediate 7 was sufficient to attempt the subsequent reaction, but several iterations of this benzyl deprotection were necessary due to the low absolute yield (~70 mg per reaction). Due to the long reaction time, modest yield and use of dangerous reagents (H_2), it would be sensible to explore alternative hydrogenation methods, such as the use of concentrated aqueous acid or triethylsilane over Pd(C).[162,163] To assess the suitability of these alternative methods, it would be necessary to first explore the acid/base and/or redox sensitivity of the key oxadiazole/thiadiazole moiety to avoid unwanted degradation of the 990 moiety.

5.3.5 Dess-Martin Periodinane (DMP) Oxidation of the Primary Alcohol

Using the limited intermediate 7 obtained from the previous reaction, I attempted the FGI to introduce a terminal aldehyde for eventual conjugation to $DNJ\text{-}(OBn)_4$ via DMP oxidation. Theoretically, the well-established DMP oxidation provides a convenient conversion of primary alcohols to aldehydes under mild reaction conditions with high chemoselectivity. The rate of DMP oxidation can be enhanced by introduction of water before combination with the substrate due to the formation of a more reactive hydroxyl-containing intermediate, as proposed by Meyer and Schreiber in 1994.[164] This hydroxyl-containing intermediate will

decompose faster than the parent DMP molecule and therefore the accompanying alcohol oxidation should also be accelerated. I employed the protocol presented by Meyer and Schreiber, with the exception that I used dimethylformamide (DMF) in place of dichloromethane (DCM), due to the insolubility of intermediate 7 in the latter. Initially performing the reaction on a small scale (0.18 mmol Intermediate 7), I observed evidence of successful oxidation by MS and ^{1}H NMR of the crude reaction mixture. For the crude mixture, the MS spectrum showed the expected peak for intermediate 8 ([M+H] 248.1 m/z, [M+DMSO+H] 326.4 m/z) and an aldehyde peak was visible by ^{1}H NMR (1H, s, δ 9.72 ppm).

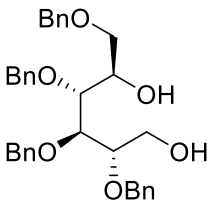
Having observed some success with the DMP oxidation, I repeated the reaction on a larger scale (0.56 mmol Intermediate 7). Unfortunately, the work-up for this reaction proved troublesome due to the precipitation of a profoundly insoluble solid product. Among the exhaustive list of aqueous and organic solvents tested, the residue was found to be only sparingly soluble in DMF and DMSO and not enough to confirm the identity of the residue by ^{1}H NMR and MS. All my efforts to solubilise the residue for analysis and possible inclusion in downstream synthetic steps eventually proved unsuccessful. It is possible that the change in solvent from the literature protocol (DCM to DMF) had some influence over the reaction. Indeed, as a somewhat nucleophilic solvent, DMF may not be compatible with the DMP oxidation, as amide resonance may allow the carbonyl on DMF to attack DMP and displace an acetate group. A similar interaction could be happening with the amine group on the heterocyclic moiety of 990 and it has been reported that DMP may not tolerate the presence of nucleophilic groups in the alcohol substrate.[165] For this reason, it would be prudent to repeat the reaction using an alternative solvent and it may be necessary to add a protecting group, such as tert-butyloxycarbonyl (Boc), to the amine on intermediate 6 or 7 (Scheme 2). Further evidence that DMF may be an incompatible solvent for this reaction is its known degradation when exposed to water to produce dimethylamine (DMA) and formic acid (FA); an oversight given that water is used here with the intention of accelerating the oxidation.[166]

Disregarding the insoluble residue from the reaction mixture, I continued with the work-up procedure stated in Section 5.4.5. I isolated the organic fraction from the liquid-liquid extraction and removed the solvent *in vacuo*, yielding a minute amount of solid residue that I analysed by ^{1}H NMR and MS. The observations previously described for the crude reaction mixture were consistent with for the dissolved residue, however the obtained mass was < 0.1 mg. It remains unclear precisely why the reaction proceeded in this manner, though it likely stems from the incompatibility of DMF as a solvent for this reaction as described above. The insolubility of the most abundant product in any of the tested solvents makes it difficult to confidently determine the outcome of this DMP oxidation. It is possible that the insoluble product is a result of some degradation of DMP in the presence of DMF, perhaps leading to production of 2-iodoxybenzoic acid (IBX), a synthetic precursor to DMP known to show limited solubility in most organic solvents.[167] Time constraints prevented me from further investigating this complication and my progress through the synthesis proposed in Schemes 2 and 5 was halted at this stage.

5.4 Materials and Methods

NB: The following section contains detail of selected syntheses and selected physical data can be found in Appendix 9.11.

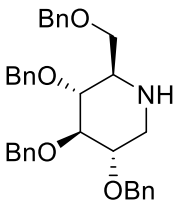
5.4.1 Synthesis of 1-Deoxynojirimycin (DNJ) – Wennekes *et al*.[157]



ii

2,3,4,6-Tetra-O-benzyl-D-glucitol (Scheme 4, Intermediate ii): To a dry 0.15 M solution of 2,3,4,6-tetra-O-benzyl-D-glucopyranose **1** (1 eq, 7.03 g, 13.0 mmol) in THF (85 mL), cooled to 0 °C, was added $LiAlH_4$ (3.5 eq, 1.73 g, 45.5 mmol) portionwise. The reaction mixture was stirred for 20 h, allowing it to warm to room temperature. The excess $LiAlH_4$ was quenched with MQ water at 0 °C. The mixture was diluted with EtOAc (150 mL) and washed with sat aq NH_4Cl (3 × 50 mL).

The organic phase was dried ($MgSO_4$) and concentrated under reduced pressure. The resulting product **2** was used crude in the next reaction.



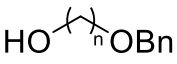
iv

2,3,4,6-Tetra-O-benzyl-1-deoxynojirimycin (Scheme 4, Intermediate iv): A solution of $(COCl)_2$ (4 eq, 6.6 g, 52 mmol) in DCM (52 mL, 1 M) was cooled to -78 °C. After dropwise addition of a 2 M solution of DMSO (5 eq, 4.62 mL, 65 mmol) in DCM (32.5 mL) over 10 min, the reaction mixture was stirred for 40 min while maintaining temperature below -70 °C. Next, a dry 0.5 M solution of crude 2 (1 eq., assumed 13 mmol) in DCM (26 mL) was added dropwise to the reaction mixture over 15 min, again maintaining the temperature of the reaction mixture below -70 °C. After stirring the reaction mixture for 2 hours below -65 °C, Et_3N (12 eq, 21.75 mL, 156 mmol) was added dropwise over 10 min, while keeping the reaction mixture below -65 °C. After addition of Et_3N, the reaction mixture was allowed to warm to -5 °C over 2 h.

The Swern reaction mixture was concentrated at a moderate temperature (approx. 30 °C) with simultaneous co-evaporation of toluene (3×). The residue was dissolved in MeOH (260 mL, 0.05 M relative to starting compound) and NH_4HCO_2 (20 eq, 16.4 g, 260 mmol) was added. The mixture was cooled to 0 °C and stirred until all NH_4HCO_2 had dissolved. Activated 3Å molecular sieves (10 g/mmol intermediate 2, 130 g) were added and reaction mixture was stirred for 40 min, after which $NaBH_3CN$ (4 eq, 3.27 g, 52 mmol) was added. The reaction mixture was kept at 0 °C for 1 h after which the cooling source was removed and the reaction was stirred for an additional 20 h. After removal of the molecular sieves over a glass microfibre filter (GF/D, Whatman), the filtrate was concentrated, dissolved in EtOAc (100 mL) and washed with sat aq NaHCO3 (100 mL). The aqueous phase was back-extracted with EtOAc (3 × 50 mL) and the combined organic layers were dried ($MgSO_4$) and concentrated. The resulting residue was purified by silica gel column chromatography (20 - 75% EtOAc in PE) to provide iv (4.42 g, 8.5 mmol) in 65% yield over three steps as a light yellow crystalline solid.

R_f = 0.25 (silica gel, 1:1 EtOAc:PE); ^{1}H NMR (600 MHz, DMSO-d_6): δ = 7.34-7.24 (m, 16H), 7.18-7.16 (m, 2H), 4.86 (d, J = 11.5 Hz, 1H), 4.73 (dd, J = 11.3, 5.0 Hz, 2H), 4.65-4.58 (m, 2H), 4.49-4.41 (m, 3H), 3.59 (dd, J = 9, 2.5 Hz, 1H), 3.47-3.40 (m, 3H), 3.22-3.18 (m, 2H) ppm; ^{13}C NMR (600 MHz, DMSO-d_6): δ = 139.5, 139.4, 139.1, 138.8, 128.7, 128.6, 128.2, 128.0, 127.9, 127.8, 127.7, 87.2, 80.6, 80.5, 75.0, 74.5, 72.8, 71.7, 70.5, 60.0, 48.1 ppm; LRMS (ESI, positive mode): 524.2 ($C_{34}H_{37}NO_4$ + H^+).

5.4.2 Alkylation of Hydroxybenzaldehyde

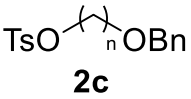


2b

Diol Benzylation (Scheme 5, Intermediate 2b): A solution of diol 2a (*n= 3 or 4,* 75 mmol, 5 equiv.) in DMF (20 mL) was cooled to 0 °C in an ice bath before portionwise addition of NaH (2 eq., 30 mmol, 0.72 g). The mixture was stirred for 30 min at 0 °C, then allowed to warm to room temperature and stirred for a further 30 min. The solution was cooled to -10 °C before addition of a solution of BnBr (1 eq., 15 mmol, 1.79 mL) in DMF (5 mL). The reaction mixture was stirred at room temperature for 18 h before cooling to 0 °C. The reaction was then quenched with ice-cold MQ water (150 mL) and Et_3N (1 equiv., 15 mmol, 2.10 mL). The product was extracted with ^{t}BuOMe (1 x 150 mL, 2 x 50 mL) and the organic layer washed with 20% brine (3 x 100 mL) then dried over $MgSO_4$ and concentrated *in vacuo*. The residue was diluted in minimal EtOAc and purified by silica flash column chromatography (0 – 60% EtOAc in PE) to yield a colourless oil.

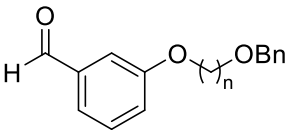
n = 3: R_f = 0.12 (silica gel, 3:1 PE:EtOAc), ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.37-7.26 (m, 5H), 4.53 (s, 2H), 3.80-3.77 (m, 2H), 3.66 (t, J = 6 Hz, 2H), 2.43 (s, 1H), 1.87 (quintet, J = 6 Hz, 2H) ppm; LRMS (ESI, positive mode): 189.0 ($C_{10}H_{14}O_2$ + Na^+).

n=4: R_f = 0.14 (silica gel, 4:1 Heptane:EtOAc); ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.37-7.26 (m, 5H), 4.52 (s, 2H), 3.62 (t, J = 6 Hz, 2H), 3.53 (t, J = 6 Hz, 2H), 2.55 (s, 1H), 1.74-1.64 (m, 4H) ppm; HRMS (ESI, positive): calcd for $C_{11}H_{16}O_2$ [M + H+]: 181.1223, found 181.1224.



2c

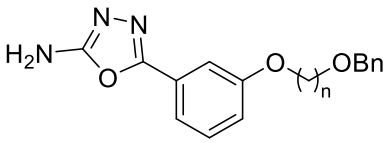
Tosylation of Benzyl-alcohol (Scheme 5, Intermediate 2c): To a solution of Benzyloxy-alcohol, 2b (1 eq., 7.5 mmol) in DCM (50 mL) was added Et_3N (1.3 eq., 9.8 mmol, 1.34 mL), 4-toluenesulfonyl chloride (TsCl) (1.1 eq., 8.1 mmol, 1.55 g) and 4-dimethylaminopyridine (DMAP) (0.1 eq., 0.75 mmol, 0.10 g), sequentially. The solution was stirred for 8 h at room temperature, then diluted in DCM (100 mL) and the organic layer washed with saturated $NaHCO_3$ (1 x 100 mL) and 20% brine (2 x 100 mL). The organic layer was dried over $MgSO_4$ and the solvent removed *in vacuo*. The product was used immediately in the subsequent reaction without purification.



Alkylation of 3-(hydroxyalkoxy)benzaldehyde (Scheme 5, Intermediate 4): A solution of tosylate intermediate, 2c (n = 4, 1 eq., 11 mmol, 3.68 g) in DMF (20 mL) was cooled to -10 °C before portionwise addition of NaH (2.2 eq., 24 mmol, 0.58 g). The solution was stirred at 0 °C for 1 h, or until H_2 evolution had ceased, before addition of a solution of 3-hydroxybenzaldehyde, 3 (1.1 eq., 12.1 mmol, 1.48 g) in DMF (5 mL). The reaction mixture was stirred under reflux at 150 °C for 18 h then cooled to room temperature, diluted in ^{t}BuOMe (150 mL) and washed with saturated $NaHCO_3$ (1 x 100 mL) and 20% brine (2 x 100 mL). The organic layer was dried over $MgSO_4$ and the solvent removed *in vacuo*. The residue was diluted in minimal EtOAc and purified by silica flash column chromatography (0 – 50% EtOAc in PE) to yield a colourless oil.

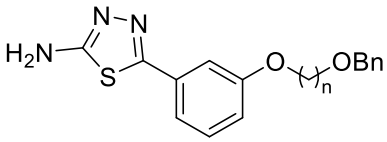
n = 4: R_f = 0.56 (silica gel, 3:2 PE:EtOAc); ^{1}H NMR (500 MHz, $CDCl_3$): δ = 9.97 (s, 1H), 7.46-7.42 (m, 2H), 7.37-7.25 (m, 7H), 7.17-7.15 (m, 1H), 4.53 (s, 2H), 4.04 (t, J = 6 Hz, 2H), 3.56 (t, J = 6 Hz, 2H), 1.93 (quintet, J = 7 Hz, 2H), 1.82 (quintet, J = 7 Hz, 2H) ppm; LRMS (ESI, positive mode): 189.0 ($C_{18}H_{20}O_3$ + Na^+).

5.4.3 Synthesis of 2-Amino-1,3,4-oxadiazoles and 2-Amino-1,3,4-thiadiazoles



2-Amino-1,3,4-oxadiazoles (Scheme 2, Intermediate 6 – X=O): To a stirred solution of semicarbazide hydrochloride (5 – X=O, 1 eq., 0.5 mmol, 55 mg) and sodium acetate trihydrate (1 eq., 0.5 mmol, 68 mg) in MQ water (1 mL) was added a solution of the aldehyde (4 – n=4, 1 eq., 0.5 mmol, 142 mg) in MeOH (1 mL). After being stirred at room temperature for 10 min, the solvent was evaporated under reduced pressure, and the resulting residue was redissolved in 1,4-dioxane (5 mL), followed by sequential addition of K_2CO_3 (3 eq., 1.5 mmol, 0.21 g) and iodine (1.2 eq., 0.6 mmol, 0.15 g). The reaction mixture was stirred at 80 °C until the conversion was complete (monitored by TLC in PE:EtOAc, 1−4.5 h). After being cooled to room temperature, it was treated with 5% $Na_2S_2O_3$ (20 mL) and extracted with DCM/MeOH (10:1, 10 mL × 4). The combined organic layer was dried over anhydrous $MgSO_4$ and concentrated. The given residue was purified through silica gel column chromatography using a mixture of EtOAc and petroleum ether as eluent to afford the corresponding 2-amino-1,3,4-oxadiazole (88 mg, 0.26 mmol) at 52% yield.

n = 4: R_f = 0.26 (silica gel, 1:3 PE:EtOAc); ^{1}H NMR (600 MHz, DMSO-d_6): δ = 7.42 (t, J = 8, 1H), 7.36-7.31 (m, 5H), 7.28-7.27 (m, 2H), 7.25 (s, 2H), 7.06-7.05 (m, 1H), 4.46 (s, 2H), 4.04 (t, J = 6 Hz, 2H), 3.50 (t, J = 6 Hz, 2H), 3.33 (s, 2H, H_2O), 1.83-1.78 (m, 2H), 1.73-1.68 (m, 2H) ppm; ^{13}C NMR (600 MHz, DMSO-d_6): δ = 159.4, 139.1, 131.0, 128.7, 127.9, 127.8, 126.0, 117.7, 117.4, 110.8, 72.3, 69.8, 67.9, 26.2, 26.1 ppm; LRMS (ESI, positive mode): 340.2 ($C_{19}H_{21}N_3O_3$ + H^+).

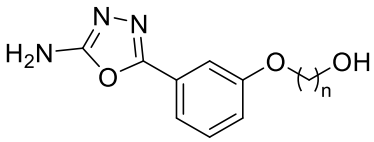


2-Amino-1,3,4-thiadiazoles (Scheme 2, Intermediate 6 – X=S): To a stirred solution of thiosemicarbazide (5 – X=S, 1 eq., 0.5 mmol, 46 mg) and acetic acid (1 eq., 0.5 mmol, 0.11 mL) in MQ water (1 mL) was added a solution of the aldehyde (4 – n=4, 1 eq., 0.5 mmol,

142 mg) in MeOH (1 mL). After being stirred at room temperature for 30 min, the solvent was evaporated under reduced pressure, and the resulting residue was redissolved in 1,4-dioxane (5 mL), followed by sequential addition of K_2CO_3 (3 eq., 1.5 mmol, 0.21 g) and iodine (1.2 eq., 0.6 mmol, 0.15 g). The reaction mixture was heated to reflux under nitrogen atmosphere until the conversion was complete (monitored by TLC, 1−4 h). After being cooled to room temperature, it was treated with 5% $Na_2S_2O_3$ (20 mL) and extracted with EtOAc (15 mL x 3). The combined organic layer was dried over anhydrous $MgSO_4$ and concentrated. The residue was purified through silica gel column chromatography (PE:EtOAc) to afford the corresponding 2-amino-1,3,4-thiadiazole (53 mg, 0.15 mmol) at 30% yield.

n = 4: R_f = 0.30 (silica gel, 1:3 PE:EtOAc); ^{1}H NMR (950 MHz, DMSO-d_6): δ = 7.42 (s, 2H), 7.36-7.31 (m, 5H), 7.29-7.28 (m, 1H), 7.28-7.25 (m, 2H), 6.98-6.97 (m, 1H), 4.46 (s, 1H), 4.03 (t, J = 6.5 Hz, 2H), 3.49 (t, J = 6.4 Hz, 2H), 3.36 (s, H_2O, 7H), 1.79 (q, J = 6.5 Hz, 2H), 1.70 (q, J = 6.5 Hz, 2H) ppm; ^{13}C NMR (950 MHz, DMSO-d_6): δ = 169.1, 159.4, 156.7, 139.1, 132.7, 130.8, 128.7, 127.9, 127.8, 119.4, 116.6, 111.9, 72.3, 69.7, 67.9, 26.3, 26.1 ppm; HRMS (ESI, positive): calcd for $C_{19}H_{21}N_3O_2S$ [M + H^+]: 356.1427, found 356.1480.

5.4.4 Deprotection of the Benzylated Alkyl Linker

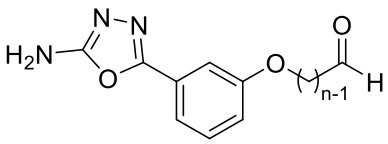


Hydrogenation of benzyl-protected primary alcohol (Scheme 2, Intermediate 7): To a solution of 2-amino-1,3,4-oxadiazole/2-amino-1,3,4-thiadiazole, 6 (n=4, X=O – 1 eq., 0.41 mmol, 0.14 g) in 1,4-dioxane (10 mL) was added an equal mass of $Pd(OH)_2(C)$ - 20 wt. % (0.14 g). The mixture was acidified using 12 M HCl (20 uL) and stirred vigorously under vacuum for 30 minutes. The stirred solution was sparged for 30 minutes with molecular nitrogen (N_2) using a balloon fitted with a needle and a separate vent needle, before again applying vacuum for a further 30 minutes. The N_2-vacuum cycle was repeated twice, before sparging the solution with molecular hydrogen (H_2) in a balloon for 1 hour, inflating the balloon

when necessary. The vent needle was removed and the solution left stirring vigorously for >24 hours. The mixture was filtered through glass microfiber filter (Grade GF/D, Whatman) to remove the Pd catalyst and the solvent removed *in vacuo*. The residue was dissolved in EtOAc and washed with MQ water (3 x 30 mL). The organic layer was extracted and dried over $MgSO_4$ and concentrated in vacuo. The residue was purified using silica column chromatography (0 – 100% EtOAc in PE), yielding Intermediate 7 (Scheme 2 - n=4, X=O, 70 mg, 0.28 mmol) at 68% yield.

n = 4, X = O: Rf = 0.09 (silica gel, EtOAc); ^{1}H NMR (600 MHz, DMSO-d_6): δ = 7.42 (t, J = 9 Hz, 1H), 7.36-7.34 (m, 1H), 7.28-7.26 (m, 3H), 7.07 (m, 1H), 4.48 (s, 1H), 4.02 (t, J = 6.5 Hz, 2H), 3.46 (t, J = 6.5 Hz, 2H), 3.38 (s, 5H, H_2O), 1.76 (quintet, J = 6.8 Hz, 2H), 1.57 (quintet, J = 6.5, 2H) ppm; HRMS (ESI, positive): calcd for $C_{12}H_{15}N_3O_3$ $[M + H]^+$: 250.1186, found 250.1220.

5.4.5 Dess-Martin Periodinane (DMP) Oxidation of the Primary Alcohol



7.2 µL of H_2O (0.40 mmol, 1.1 eq.) was solvated in 10 mL of DMF and mixed thoroughly by pipetting. The wet DMF was added dropwise to a vigorously stirring solution of Intermediate 7 (n=4. X=O, 1 eq., 0.36 mmol, 90 mg) and DMP (229 mg, 0.54 mmol, 1.5 eq.) in 3 mL of dry DMF. The clear solution grew cloudy toward the end of wet DMF addition, which required 30 min. The mixture was diluted with tBuOMe, then concentrated into a few mL of solvent by rotary evaporator. The residue was taken up in 30 mL of ether and then washed with 15 mL of 1:1 10% $Na_2S_2O_3$:saturated aqueous $NaHCO_3$, followed by 10 mL of H_2O and 10 mL of brine. The aqueous washings were back-extracted with 20 mL of Et_2O, and this organic layer was washed with H_2O and brine. The combined organic layers were dried with $MgSO_4$ and concentrated. The reaction yielded crude Intermediate 8 at negligible yield (< 1 mg) along with a profoundly insoluble solid for which no physical data was obtained.

n = 4, X = O: Rf = 0.23 (silica gel, EtOAc); HRMS (ESI, positive): calcd for $C_{12}H_{13}N_3O_3$ [M + H^+]: 248.1030, found 248.1060. *NB: No NMR data was collected for this compound due to prohibitively low yields of soluble material.*

5.5 Conclusions

Fragment 990, discovered to bind to GIIα near the active site and interact with exclusion loop residues, could be combined with DNJ in order to produce potent broad-spectrum HTAs that are more selective for GluII than existing iminosugar compounds. Crystal structures of *Mm*GluII$_{Tryp}$ in complex with both 990 and UV-4 show demonstrate that the two ligands could be conjugated using an alkyl linker in a way that allows both to adopt the observed poses to maximise the protein-drug interaction.

Here, I outline a synthetic strategy for the production of 990-DNJ iminosugars. Employing the DNJ synthesis proposed by Wennekes *et al.* in 2007, I was able to cleanly produce DNJ-(OBn)$_4$ in just three steps.[157] In my hands, this approach was more time and cost effective than the more convoluted Best *et al*. synthesis which proved ineffective and involved several more synthetic steps.[156] Furthermore, the Wennekes *et al*. synthesis produced DNJ-(OBn)$_4$ and therefore removed the need for an additional benzyl protection step prior to conjugation with the 990 fragment. The yield of DNJ-(OBn)$_4$ using the Wennekes *et al*. synthesis could be improved upon to increase the efficiency of the reaction. That said, the potentially low stereoselectivity of the final double reductive amination step could be responsible for the sub-optimal yield, in which case it may be difficult to improve without additional synthetic steps to encourage stereospecificity. In any case, it would be necessary to fully characterise any side products to elucidate the mechanism of unwanted side reactions such that they may be avoided in the future.

I present a convenient and flexible, albeit not optimised, synthetic route towards 990 with an attached alkyl linker. The benefits of the proposed strategy include the facile incorporation of variable linker lengths, benzene substitution patterns and nature of endocyclic heteroatoms (O or S). Sequential benzylation and tosylation of diols, both relatively efficient steps, allowed for facile gram scale production of alkylated hydroxybenzaldehyde, albeit with modest percentage yields. The subsequent introduction of the 2-amino-1,3,4-oxadiazole/2-amino-1,3,4-thiadiazole moieties using the method proposed by Niu *et al.* proved to be perhaps the most reliable of the synthetic steps I attempted in this thesis, cleanly affording the 990 analogues with attached alkyl linkers.[158] Again, the yields of this step were considerably lower than expected based on the source literature and this is likely due to poor resolution of the silica column with the EtOAc:PE solvent system used. Exploration of alternative eluent systems could improve the efficiency of this purification step. The removal of the benzyl protecting group to give an exposed primary alcohol (Scheme 2, Intermediate 7) gave reasonable percentage yield (68%), but low absolute yield (~70 mg) which limited the progress of downstream syntheses. This issue could be resolved by optimisation of the previous synthetic steps to afford more of the starting material (Scheme 2, Intermediate 6) for this reaction.

Optimisation of the synthesis proposed in Scheme 2 was an aim secondary to the primary one of establishing a valid synthetic route towards the DNJ-990 iminosugar targets, so the limited yields from the successful reactions are not a pressing concern. A larger problem was the unsuccessful oxidation of the primary alcohol in intermediate 7 to provide the aldehyde in intermediate 8, which precluded the conjugation of the 990-linker fragment to DNJ-$(OBn)_4$. It seems likely that the use of DMF, a decision made out of necessity due to limited solubility of intermediate 7, may have complicated the reaction and led to production of an unidentified and profoundly insoluble side product. The degradation of DMP in the presence of DMF may have produced IBX, a synthetic precursor to DMP with poor solubility in organic solvents, the

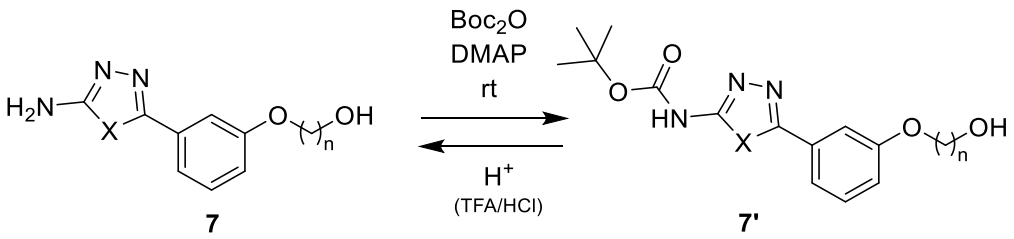
insolubility prevented its characterisation by NMR or MS and the precise nature of the residue remains unknown.[167] Other reasons why DMF was a poor choice of solvent include its own degradation in the presence of water, possibly leading to undesirable side reactions involving DMA/FA, and its nucleophilicity that may interfere with DMP oxidation by inactivating DMP itself.[165,166] The complications in the DMP oxidation likely result from the use of DMF and/or the presence of a nucleophilic amine group in Intermediate 7. Both factors are easily explored by identification of a more suitable solvent and possible Boc-protection of the amine prior to the oxidation step, providing a clear path forward with respect to continuation of Scheme 2. The work presented here provides an excellent platform from which to complete the proposed synthesis to yield 990-DNJ iminosugars and investigate their capacity for inhibition of GluII.

5.6 Future Work

5.6.1 Synthesis of 990-DNJ Iminosugars

The key aim of this organic synthesis was to produce the 990-DNJ target compounds (Figure 37) and this should continue to be the focus for future attempts. In the short term, it is necessary to repeat the oxidation of Intermediate 7 to give Intermediate 8 (Scheme 2) to allow for the conjugation of DNJ-$(OBn)_4$ to 990 via the variable alkyl linker. It appears that my choice of solvent (DMF) was a poor one and likely led to the production of the profoundly insoluble side product that complicated this reaction. My first priority would be to identify an alternative solvent that effectively dissolves Intermediate 7 and is compatible with DMP oxidation as this is theoretically the mildest and most convenient strategy that will likely preserve the rest of the molecule. Pyridine is a possible alternative to DMF and can increase the yields of carbonyl products when used as an additive.[168] Identification of alternative solvents is necessary, but may not address the problem that the nucleophilic amine on the heterocycle of the 990 moiety may interfere with DMP oxidation. Scheme 6 shows the Boc-protection of the free amine that may improve the compatibility of this substrate with DMP oxidation and improve its solubility

in solvents suitable for this reaction. The use of a nucleophilic catalyst such as DMAP may speed up this amine protection. Removal of a Boc protecting group typically involves a simple acid-catalysed carbamate hydrolysis, though it would be sensible to first determine the stability of the intermediates in acidic conditions to avoid unwanted modification of other moieties.[169]



Scheme 6 – Protection and deprotection of amino groups using Boc-anhydride (Boc_2O) and acid-catalysed carbamate hydrolysis, respectively.

Following successful production of Intermediate 8 (Scheme 2), the next step is its conjugation to DNJ-$(OBn)_4$. I anticipate difficulty in finding a solvent system that is capable of dissolving the two components in the same solvent due to the relatively high hydrophobicity of DNJ-$(OBn)_4$ relative to Intermediate 8, particularly for the analogues bearing a 2-amino-1,3,4-oxadiazole moiety. Use of acid or base catalysis for the condensation reaction that conjugates the two subunits of the compound may speed up the process, but the acid sensitivity of the Boc protecting group may mean that use of basic conditions is more appropriate in order to avoid dimerization of Intermediate 8 via nucleophilic attack by the free amine on the aldehyde. The benzyl deprotection step used to obtain Intermediate 7 (Section 5.4.4) may prove doubly useful for the final step as the addition of concentrated HCl may provide simultaneous Boc and benzyl deprotection to yield the target compounds.

5.6.2 Evaluation of GluII Inhibition by 990-DNJ Iminosugars

The primary purpose of the described synthesis is to produce iminosugar antivirals that demonstrate improved selectivity for GluII relative to the current best-in-class DNJ iminosugar, UV-4B. To determine the selectivity of the target compounds, I propose screening the

compounds against an array of glucosidases, including other members of the GH31 family to which GIIα and intestinal glucosidases belong. Initially, this would involve use of synthetic substrates such as 4-MUG in *in vitro* enzymology assays as described previously by the Zitzmann group.[54] This could also determine the potency of *in vitro* GluII inhibition for comparison to UV-4B. Screening the target compounds in this way would show whether these novel 990-DNJ iminosugars are capable of inhibiting GluII without significant inhibition of other glucosidases that may lead to AEs *in vivo*. This approach also saves time as only those 990-iminougars that show selectivity for GluII would be carried forward for more thorough investigation.

Provided at least one of the novel 990-DNJ iminosugars shows good selectivity for GluII, I would then assess their cytotoxicity and ability to inhibit GluII in the ER of living cells. To do so, I would employ the MTS and FOS assays described in Sections 4.3.8 and 4.3.9, respectively. If tolerable cytotoxicity and delivery to the ER was established, I would then quantify the inhibitory concentrations (IC50) against GluII for comparison with known iminosugars. Antiviral assays would then be used to test the 990-DNJ iminosugars for their activity against a range of enveloped viruses as previously described by the Zitzmann group and others.[79,83,96]

5.6.3 An Opportunity for Covalent Active Site Inhibitors of α-Glucosidase II

The off-target activity of iminosugars is known to cause AEs, particularly gastrointestinal distress, *in vivo*.[153,154] Many attempts have been made to improve the selectivity of DNJ-derived iminosugars, however there is a risk that the promiscuity of the DNJ functional group among glucosidases may confer an unavoidable lack of selectivity to these compounds. Development of potent and selective GluII active site inhibitors that do not carry an iminosugar moiety may remove the off-target activity associated with known iminosugar antivirals.

Fragments that bind near the GluII active site, such as 990, have been discovered through crystallography-based fragment screens (unpublished work by the Zitzmann group). It may be possible to further functionalise these fragments to improve their selectivity for the GluII active site and establish a starting point for iminosugar-free ligands of GIIα. With suitably selective fragments in hand, I propose their conjugation to electrophilic warheads to exploit the presence of C502, a residue that reaches the edge of the GIIα catalytic pocket and carries a free thiol group in the *Mm*GluII$_{Tryp}$ crystal structure (Figure 38). As summarised in Section 1.4.3, I have written an extensive literature review exploring the rise in the use of covalent inhibitors in drug discovery which revealed that through sensible design, it is possible to develop targeted covalent inhibitors (TCIs) that inhibit their target with enhanced potency and duration of action relative to non-covalent inhibitors.[170] TCIs utilise selective, non-covalent protein-ligand interactions to correctly orient the ligand for covalent binding to a nucleophilic amino acid side chain via a relatively unreactive electrophilic warhead moiety. In the context of cysteine residues, several of these warheads have been described in the literature, with the perhaps the most well-known being acrylamides which have been incorporated into several clinically approved drugs.[171]

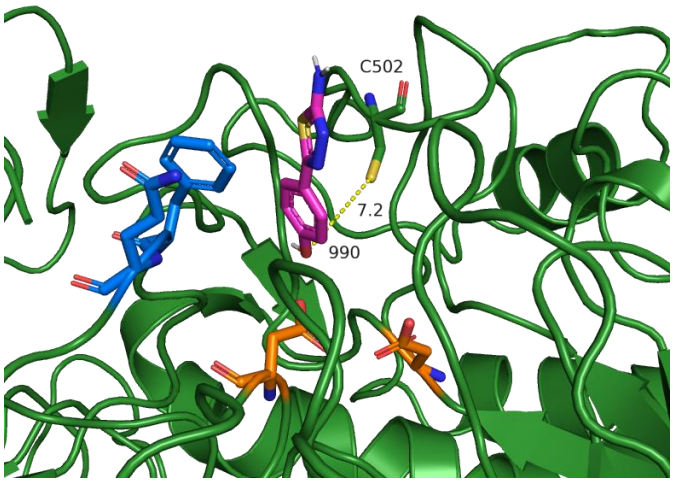


Figure 38 – Crystal structure of *Mm*GluII in complex with 990. C502 exists as a free thiol, roughly 7 Å from the phenolic oxygen atom on 990. Functionalisation of 990 at the phenolic hydroxyl group to include a cysteine-reactive electrophilic warhead may yield covalently binding active site inhibitors of GluII without need of a promiscuous iminosugar moiety.

I speculatively suggest that the design and synthesis of covalently binding small molecules containing fragments known to bind in the catalytic pocket of GIIα is a possible strategy towards the development of iminosugar-free, selective GluII inhibitors. The proposal is an ambitious one, however, as the presence of covalently binding moieties without suitably target-selective scaffolds may bind to nucleophilic residues indiscriminately, leading to drug related idiosyncratic toxicity. However, literature precedent for the long-term achievability of such a goal is bountiful and may warrant further investigation.

6. Investigating the Role of the Mannose 6-phosphate Receptor Homology (MRH) Domain on α-Glucosidase II Enzymatic Activity

6.1 Brief Introduction & Aims

Despite the absence of convincing structural data of GluII-glycan complex that includes the MRH domain, its importance for the efficient catalytic activity of GluII is clear. A deeper understanding of the structure of this complex could provide new insights into targetable interactions with the aim of slowing down GluII activity in the ER to combat infection by enveloped viruses. It is broadly assumed that the MRH domain of GIIβ binds the same glycan as is bound in the active site on GIIα, though there is currently no structural data to support this assumption.[66,67] This conceptual model, shown in Figure 39, neatly explains the observed enhancement of GluII activity by the MRH domain via orientation of the A-arm of the $Glc_{1\text{-}2}Man_9$ substrate glycan into the active site on GIIα, but depending on the glycoprotein substrate and the quality of its fold at the time of interaction with GluII, this interaction may become significantly sterically crowded. Furthermore, many of the glycoproteins processed in the ER bear several glycans and this model neglects the possibility that GIIβ interacts with some secondary glycan rather than the one bound by GIIα, as proposed by Deprez *et al*. in 2005.[172] One could postulate that this type of interaction would increase the rate of glucose hydrolysis by bringing the nascent glycoprotein into closer proximity with GIIα, meaning that the flexible and elongated GIIβ is effectively 'fishing' for substrate glycans in the protein-rich pond that is the ER, using the MRH domain as its hook. In this theoretical model, the catalytic

efficiency of GIIα is increased by the MRH domain of GIIβ as the secondary glycan interaction can increase the effective concentration of the substrate near the active site. In the absence of the GIIβ, and therefore the MRH domain, GIIα activity may be limited by the requirement of chance encounters with the substrate glycoproteins. This speculative suggestion highlights the value of investigating the precise structure of a GluII-glycan complex to assess the plausibility of the current conceptual model in real space and explore alternative hypotheses. As well as aiming to gather experimental structural data using cryogenic electron microscopy (CryoEM – Appendix 9.12), I decided to explore the interaction of the GluII with substrate glycans *in silico*.

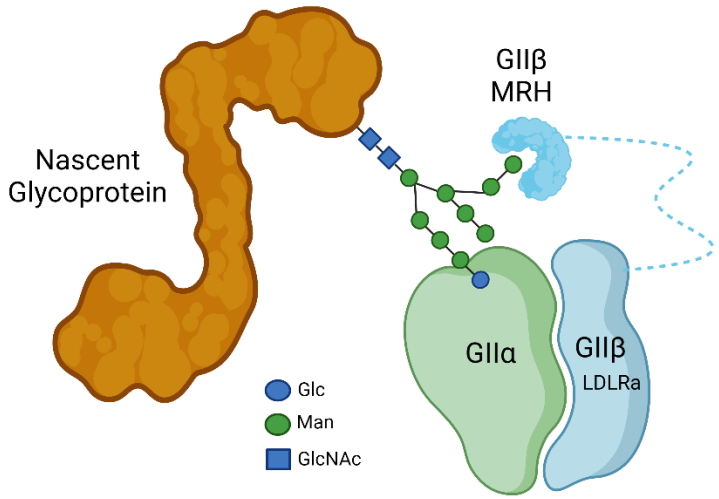


Figure 39 – A schematic diagram of GluII in complex with a substrate glycoprotein. This represents the current conceptual model in which GIIα and the MRH domain of GIIβ bind the same glycan simultaneously via the A-arm and B- or C-arm of the glycan, respectively.

6.2 Materials & Methods

Molecular dynamics simulations were performed using the charmm36 force field and Gromacs software. The multicomponent system was established using the CHARMM-GUI web-sever and followed the standard protocol for energy minimisation and equilibration.[173,174] The MD

system contained a single molecule of both *Mm*GluII$_{Tryp}$ and N-glycosylated phaseolin, 82,288 water molecules and 56 Na^+ ions.

6.3 Results & Discussion

To date, no experimental structural data has been obtained for full length GluII. In lieu of experimental structures, it is worth exploring the dynamics of GluII *in silico* to determine the likely role of the MRH domain on GIIβ. As the *Mm*GluII$_{Tryp}$ crystal structure contains around 90% of the full length GIIα, this was used as a starting point for *in silico* analysis. However, the same crystal structure shows only the LDLRa domain for GIIβ, representing only 17% of the subunit. Figure 40 shows the AlphaFold prediction of the full length murine GIIβ structure superimposed onto the *Mm*GluII$_{Tryp}$ crystal structure and reveals a particularly long single α-helix (SAH) spanning 107 amino acids from T119-L225.[69,70] Typically 40-200 residues long, SAHs occur naturally in roughly 4% of proteins, though few have been experimentally characterised, and often form in peptide sequences rich in acidic and basic residues (Glu - E, Arg - R, Lys - K).[175] Indeed, these residues comprise roughly 45% of the putative SAH seen in the AlphaFold GIIβ structure (48/107 residues), so the prediction may be legitimate. SAH domains display high kinetic stability and diverse functionality in different proteins and their properties likely depend on the relative K:R content, since the helix is formed as a result of E-R or E-K salt bridges. It is unclear whether the predicted SAH on GIIβ has any direct influence over GluII activity, though the expected rigidity of this domain tentatively suggests some mechanical function by acting as a constant force spring, allowing variable separation between the domains connected by it.[176] The AlphaFold GIIβ structure highlights interesting structural features, although when superimposed onto the experimental structure of *Mm*GluII$_{Tryp}$ it produces a molecule occupying an expansive volume which would make any molecular dynamics (MD) simulation with this molecule very computationally expensive and may not be suitable for this application.

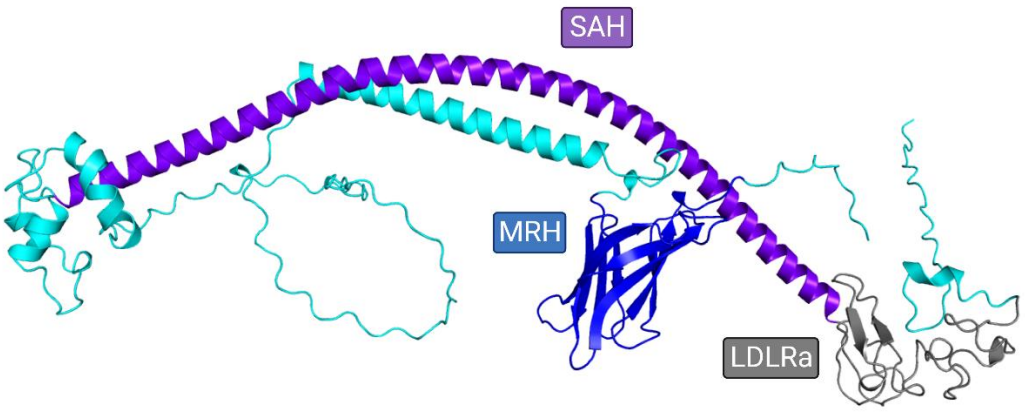


Figure 40 – AlphaFold prediction of full length murine GIIβ (Uniprot: O08795*)* with key domains labelled.[69,70]

To simplify the MD simulations, I propose examination of a three-body model to study MD of *Mm*GluII$_{Tryp}$ in combination with a substrate glycoprotein and the MRH domain of GIIβ. The first step here was construction of an *in silico* model of *Mm*GluII$_{Tryp}$ in complex with a substrate glycoprotein. The monomer of β-type phaseolin was chosen as the substrate glycoprotein due to its relatively small size (47.6 kDa) and simple glycosylation profile, carrying only one N-glycan on its N252 residue (Uniprot ID: P02853; PDB: 1PHS).[177] Since the crystal structure of phaseolin does not show the GluII substrate glycan, I used the glycan reader and modeller tool from CHARMM-GUI to construct the $Glc_1Man_9GlcNAc_2$ onto residue N252 as seen in Figure 41.[174] Energy minimisation simulations were run on the resulting glycosylated phaseolin structure to obtain the most energetically favourable pose for the glycoprotein and attached glycans. With the minimised phaseolin structure in hand, the multicomponent assembler tool, again from CHARMM-GUI, was used to combine phaseolin with *Mm*GluII$_{Tryp}$ into a single system for MD simulations.[173] Here, the A-arm of the substrate glycan on phaseolin was first manually positioned close to the catalytic pocket of GIIα to promote the native catalytic interaction. With the MD parameters used here, the two proteins did not form any significant interaction and drifted away from one another over the course of the trajectory. Importantly, this revealed that the glycan on the surface of phaseolin was not able to access the active site on GIIα to form the native enzyme-substrate complex. The catalytic pocket of GIIα is deep and shows limited solvent accessibility from the surface, meaning that the MD simulation reflects

true difficulty for the substrate glycan to access the pocket. The mannose binding pocket on the MRH domain is more open in structure, tentatively supporting the notion that the MRH-mannose interaction may form first and conformational change in GIIβ could encourage the glycan into the catalytic pocket of GIIα by increasing the effective substrate concentration at the active site.[67]

A key limitation of the MD simulation presented here is that no constraints were applied to force the A-arm of the substrate glycan into the GIIα active site. The interaction of interest between GluII and phaseolin via the glycan was not observed, reducing the utility of the model. The simulation did display the dynamics of the N-glycan on phaseolin in solution, which showed the anticipated flexibility. To extract useful information from this trajectory, I used the gromacs simulation package to identify the poses of the glycosylated phaseolin in which the distance between the terminal glucose of the A-arm and terminal mannose on the B- or C-arms of the glycan are maximised. Here, poses with Glc-Man distances over 2.9 nm were considered to provide a practical number of structures to manually examine (9 A-arm-B-arm poses; 5 A-arm-C-arm poses) as shown in Figure 41.

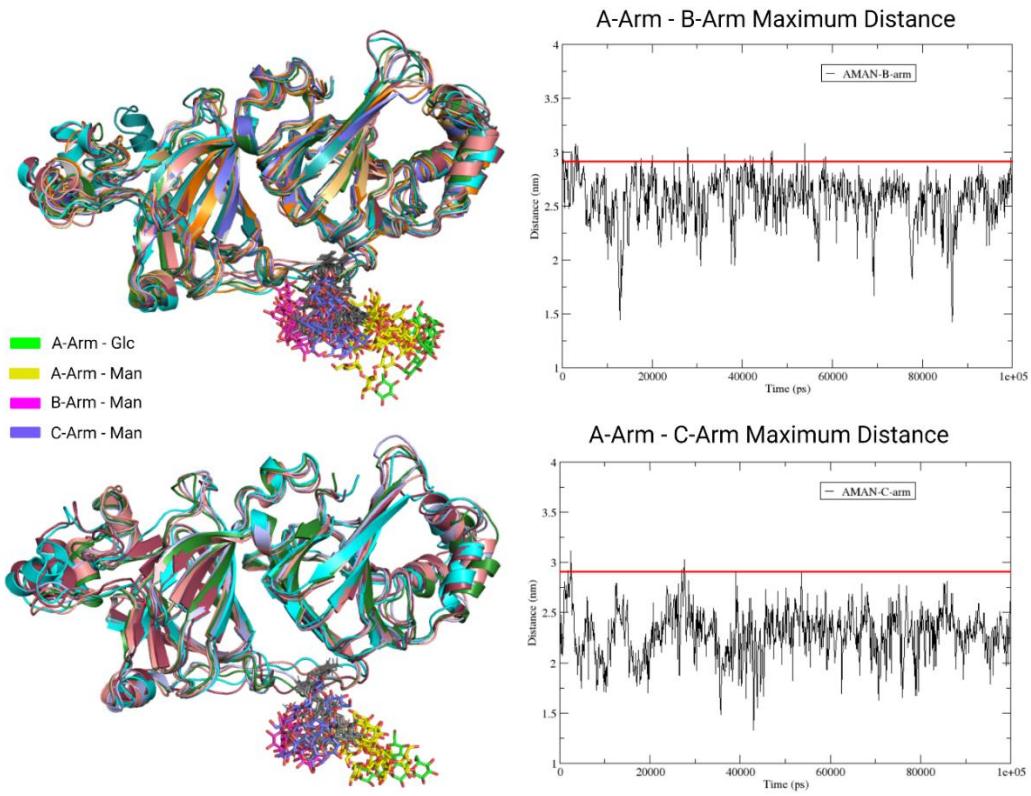


Figure 41 – Various overlaid poses of phaseolin bearing a $Glc_1Man_9GlcNAc_2$ glycan, shown in stick representation, at the N252 residue extracted from an MD simulation of phaseolin with $MmGluII_{Tryp}$. Distances (>2.9 nm) between the centre of mass of the terminal mannose residue in either the B- or C-arm from the terminal glucose of the A-arm were calculated and poses extracted corresponding to these distances.

Reasoning that the N-glycan conformations with the largest distance between the terminal residues of the A-arm and B-/C-arms would allow maximum space for the conceptual three-body system in which the catalytic GIIα and the MRH domain of GIIβ are both bound simultaneously to the N-glycan, I generated rigid superpositions using the LSQKAB tool on the CCP4 Cloud server to assess the plausibility of the proposed interaction.[178,179] To produce these images, I used the crystal structure of $MmGluII_{Tryp}$ with a disaccharide mimic derived from D-glucal bound in the active site of GIIα (PDB: 5HJO). I manually evaluated the superpositions to find any significant steric clashes between the N-glycan and either of the proteins in the system. Using this approach, I identified four poses in which the N-glycan shows little or no steric clash with GluII (3 A-arm-B-arm poses; 1 A-arm-C-arm pose). Any minimal clashes were deemed acceptable as minor repositioning of the flexible glycan

structure would relieve the steric strain in the true system in solution. The poses are shown in Figure 42 and generally show reasonable exposure of terminal mannose residues to the solvent, indicating that the three-body interaction in which GIIα and the MRH domain of GIIβ simultaneously bind the same substrate glycan may be possible. Of the complexes shown in Figure 42, the one derived from the A-arm-C-arm distance measurement is particularly well oriented for this three-body interaction based on the superposition of the full length GIIβ predicted by AlphaFold. In this pose, the terminal mannose of the B-arm on the N-glycan is positioned roughly 18 Å away from the mannose binding residues of the MRH domain (Q428, R470, E489, Y495), as seen in Figure 43, suggesting that minimal rearrangement of the MRH domain would facilitate its binding to the substrate glycan.[67] To my knowledge this is the first computational model of full length GluII bound to a client glycoprotein with feasible interactions between the N-glycan on the substrate and both the active site of GIIα and the MRH domain of GIIβ.

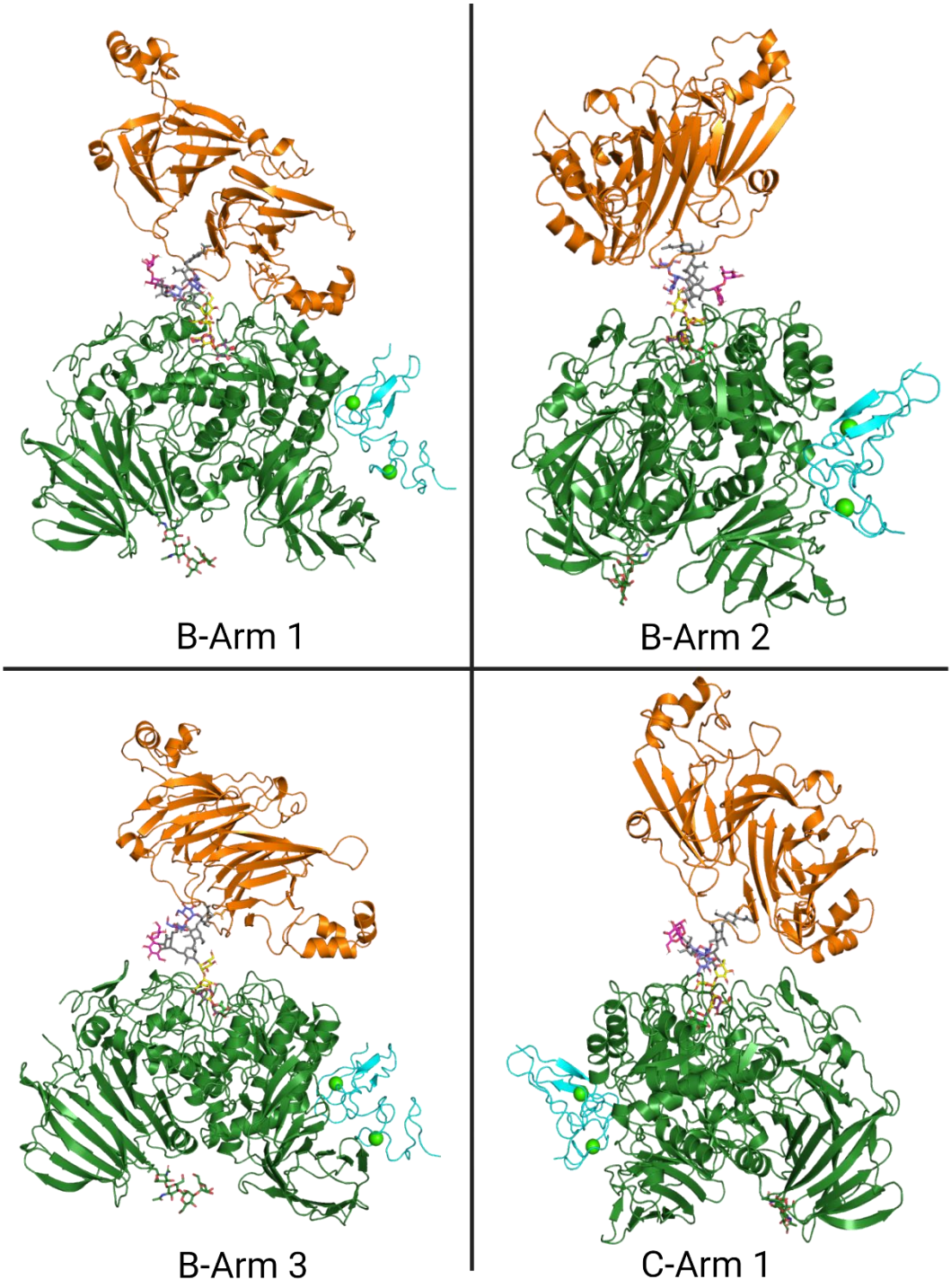


Figure 42 – Complexes of glycosylated phaseolin (orange) with $MmGluII_{Tryp}$ (GIIα – dark green; GIIβ – cyan). The terminal Glc-α(1,3)-Man residues of the A-arm of the glycan on phaseolin (yellow sticks) have been superimposed onto the α(1,3)-disaccharide substrate bound in the active site of GIIα (PDB: 5HJO). The B- and C-arms are shown as magenta and light blue sticks, respectively. In each of the poses shown, the B- and/or C-arm of the glycan are positioned such that interaction with the MRH domain on GIIβ (not shown) may be possible. Core residues of the glycan are shown as grey sticks.

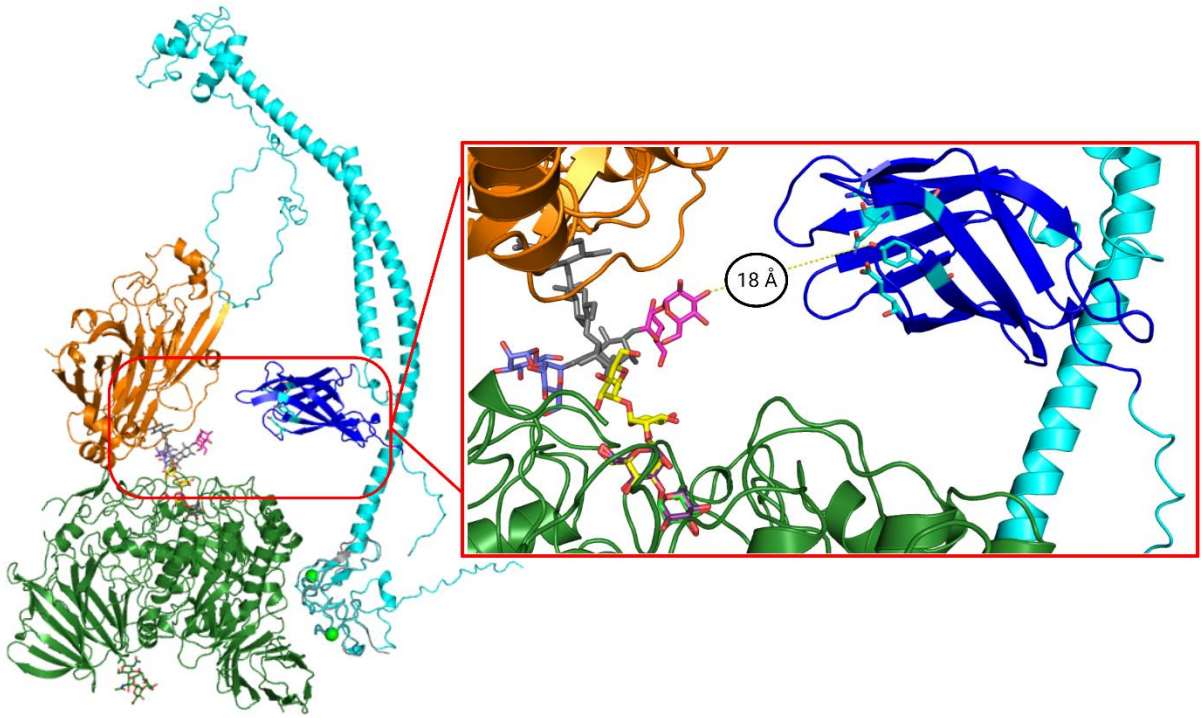


Figure 43 – A superposition of the AlphaFold prediction of full length GIIβ onto pose C-Arm 1 from Figure 42. Full length GIIβ is shown in cyan, with the MRH domain in dark blue. Mannose binding residues in the MRH domain are shown as cyan sticks and are positioned roughly 18 Å from the terminal mannose of the B-arm of the glycan (magenta sticks). The zoomed image shows the superposition of the glycan A-arm (yellow sticks = Man; light green sticks = Glc) onto the α(1,3)-disaccharide substrate (purple sticks) bound in the active site of GIIα (PDB: 5HJO). Remaining representations are as follows: Phaseolin – orange; GIIα – dark green; LDLRa from *Mm*GluII$_{Tryp}$ – grey; core N-glycan residues – grey sticks; C-arm glycan residues – light blue sticks.

6.4 Conclusions

Despite seeing no significant interaction between glycosylated phaseolin and *Mm*GluII$_{Tryp}$ in MD simulations, the presented *in silico* analysis of GluII in complex with a substrate glycoprotein provide good understanding of the dynamics of the glycan in solution. From this, I was able to extract poses in which the terminal residues of the A-arm and B- or C-arms were most distant from one another, therefore allowing the best chance at a three-body interaction where GIIα and the MRH domain both interact with the same substrate glycan. By superimposing the terminal A-arm residues of the glycan from these poses, I identified a number of structures in which there is minimal steric clash between *Mm*GluII$_{Tryp}$ and phaseolin.

Of these structures, one demonstrates suitable orientation of both phaseolin and the B-arm of its N-glycan to allow for the concomitant interaction of both GIIα and the MRH domain. The presented static models are limited by the lack of MD displayed, however they tentatively support the assertion that GluII can engage in multi-domain interactions with substrate glycans to improve the catalytic activity of GIIα, thus providing a useful starting point for future MD simulations. The mechanistic function of the MRH domain remains elusive, as the presented models do not prove or disprove the hypothesis that the MRH domain improves catalytic activity by first binding a mannose residue on a distant substrate glycoprotein before bringing it into closer proximity with GIIα. It is also plausible that formation of the three-body GIIα:substrate:MRH interaction occurs at the GluII active site without initial substrate recruitment by GIIβ as the static models suggest that this interaction is sterically feasible. Further investigation into the mechanism of action of full length GluII is required to assess both hypotheses and the preliminary models presented here offer a valuable starting point for further *in silico* analysis.

Initial efforts to experimentally solve the structure of full length GluII using cryoEM are presented in Appendix 9.12. The attempts at cryoEM used glutaraldehyde cross-linked (GXL) GluII with the rationale that this would rigidify the flexible GIIβ such that some level of homogeneity was achieved. The cryoEM data collected showed low resolution and lack of consistently identified conformations of GXL GluII prevented the grouping of similar 2D images for the construction of a 3D projection. The experiment was attempted only once, so repetitions would be necessary to improve the data or categorically rule out GXL GluII as an appropriate choice of protein for cryoEM.

6.5 Future Work

The static models of GluII bound to a client glycoprotein presented here provide a good starting point for further *in silico* analysis. To improve the utility of these models and progress

Investigating the Role of the MRH Domain on α-Glucosidase II Enzymatic Activity

towards the original aim of acquiring MD simulations of the enzyme-substrate complex, I would repeat the initial MD experiment using the sterically optimal static model. Applying constraints to the A-arm of the N-glycan on phaseolin such that it begins docked in the active site of GIIα should encourage retention of the complex such that the MD simulation provide insight into the dynamics of the complex rather than of the individual proteins.

Building upon the MD simulation of the enzyme-substrate complex, the next step would be to introduce the MRH domain of GIIβ into the system. The introduction of a third binding partner will increase the computational complexity and will likely require a more dedicated, long-term analysis to optimise the parameters. To simplify the model, constraints can be applied such that the mannose binding pocket of the MRH domain is occupied by the terminal mannose from the B- or C-arm of the N-glycan on phaseolin, according to the published structure of this interaction.[67] Analysis of the energetics of this complex may inform us about the true role of the MRH domain and may permit a more detailed and biologically relevant analysis than is possible with the static models presented here. The ultimate goal is to build a computational model of full length GluII in complex with a substrate glycoprotein. This may be achieved by using the three-body model of phaseolin, *Mm*GluII$_{Tryp}$ and the MRH domain as structural starting points around which full length GIIβ can be fitted. Here, the AlphaFold prediction of GIIβ could be used in MD simulations where the predicted orientations of the LDLRa and MRH domains are known based on the previous iterations of the three-body model. Ideally, the *in silico* study outlined here would help us to understand the structure and dynamics of the full length enzyme-substrate complex of GluII which may improve the current knowledge of the enzymes mechanism of action, particularly with respect to the involvement of the MRH domain. Furthermore, the computational model could inform us on novel interactions that may be targeted by small molecules to modulate the function of GluII.

7. Closing Remarks

Disruption of the life cycle of enveloped viruses by inhibiting key ERQC proteins, such as GluII, is a validated approach towards the production of broad spectrum HTAs. Traditionally, attempts to achieve GluII inhibition have focused on iminosugar competitive inhibitors, many of which are proven to be antiviral *in vitro* and *in vivo* against a range of viruses. A notable drawback of iminosugar antivirals is their apparent low selectivity, as they often display off-target activity against other glucosidases to cause significant AEs, particularly gastrointestinal distress. To overcome this, I have investigated alternatives to the traditional methods in order to identify selective and potent GluII inhibitors for clinical consideration.

My main focus has been on thiopyridones, a class of compounds identified in a virtual screen of an interfacial D-mannose binding site on GluII that had been shown previously to disrupt the α/β interface. This interface disruption is, in theory, an avenue towards GluII inhibition that does not rely on a promiscuously binding functional group such as DNJ, thereby allowing for more selective inhibition of the enzyme that targets its heterodimeric structure, a feature unique to GluII among other GH31 enzymes. Using SEC-MALLS and MP, I confirmed the dose-dependent α/β interface disruption by thiopyridones and provided experimental evidence supporting the computationally predicted binding site. These experiments corroborated previous assertions that α/β interface disrupting activity by this chemotype requires a free thiol group. This observation led to the hypothesis that thiopyridones may bind GluII in covalent disulfide interactions with the free cysteine residue (C822) proximal to the putative binding site. However, MS and MP experiments provide convincing evidence that the GluII-thiopyridone interaction is non-covalent, though the true mode of binding remains unclear. Furthermore, *in vitro* enzymology assays seem to reveal significant inhibition of GluI by thiopyridones, meaning that these compounds may not be as selective for GluII as we had hoped. Nevertheless, the profound impact that thiopyridones have on the GluII α/β interface,

often at low micromolar concentrations, should not be ignored and this chemical scaffold represents an excellent starting point from which to develop potent inhibitors of GluII with the desired selectivity for their target.

Inhibition of GluII via disruption of the α/β interface is a novel concept in preliminary stages of development. In the meantime, it is worthwhile to improve upon the validated approach by producing new DNJ-derived iminosugar antivirals. To do this, I have designed a new molecular scaffold that conjugates DNJ to fragment 990, a 4-(5-amino-1,3,4-thiadiazol-2-yl)phenol compound that was identified binding near the GluII active site, making significant interactions with the exclusion loop, another unique feature of GluII relative to other GH31 enzymes. Targeting this peptide to extend the interactions between the ligand and the GluII active site may improve the selectivity of these compounds for GluII. Here, I outline a valid synthetic strategy to produce the 990 moiety attached to an alkyl linker of variable length for eventual conjugation to DNJ. This final conjugation is yet to be attempted, but lessons learned from my attempts provide a solid platform from which to attempt the final synthetic step. It would be necessary to screen the target compounds against a range of glucosidases to determine their selectivity for GluII, but contributing to the pool of potentially antiviral iminosugars provides the opportunity to identify the next best-in-class DNJ-based HTA.

The importance of the MRH domain of GIIβ is well established, however the intrinsic flexibility of this subunit has confounded attempts to solve its tertiary structure. In lieu of convincing structural data, I set out to examine the protein *in silico* to elucidate the possible catalytic conformations of the full length enzyme. Preliminary attempts to simulate the molecular dynamics of GluII in complex with a substrate glycoprotein proved challenging due to the lack of a significant PPI between the enzyme and its substrate, whose interaction is dominated by GluII-glycan interactions. By analysing the dynamics of the substrate glycan, I built a static model of GluII in complex with the substrate glycan. Superimposition of the AlphaFold prediction of full length GIIβ allowed for production of an *in silico* model of full length GluII and

revealed the possibility for both GIIα and the MRH domain of GIIβ to interact with the same substrate glycan simultaneously. To my knowledge, this is the first example of such a model and establishes a framework from which to develop more sophisticated simulations to further explore the dynamics of the catalytic activity of GluII.

8. References

1. Howard, C. R. & Fletcher, N. F. Emerging virus diseases: can we ever expect the unexpected? *Emerg. Microbes Infect.* **1**, 1–9 (2012).
2. Centres for Disease Control and Prevention. Zika Virus: Statistics and Maps. https://www.cdc.gov/zika/reporting/index.html (2019).
3. Chippaux, J.-P. Outbreaks of Ebola virus disease in Africa: the beginnings of a tragic saga. *J. Venom. Anim. Toxins Incl. Trop. Dis.* **20**, 44 (2014).
4. Centres for Disease Control and Prevention. History of Ebola Virus Disease (EVD) Outbreaks. https://www.cdc.gov/vhf/ebola/history/chronology.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fvhf%2Febola%2Foutbreaks%2Fhistory%2Fchronology.html (2022).
5. Rutz, C. *et al.* COVID-19 lockdown allows researchers to quantify the effects of human activity on wildlife. *Nat. Ecol. Evol.* **4**, 1156–1159 (2020).
6. Johns Hopkins University (JHU). COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). https://coronavirus.jhu.edu/map.html (2022).
7. European Centre for Disease Prevention and Control. COVID-19 situation update worldwide, updated 7 April 2022. https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases (2022).
8. Gao, Y. *et al.* Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science (80-.).* **368**, 779–782 (2020).
9. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).
10. Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280.e8 (2020).
11. Yuan, M. *et al.* A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science* **368**, 630–633 (2020).
12. Touret, F. *et al.* In vitro screening of a FDA approved chemical library reveals potential inhibitors of SARS-CoV-2 replication. *Nat. Sci. Reports* **10**, (2020).
13. Ou, X. *et al.* Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat. Commun.* **11**, (2020).
14. Polack, F. P. *et al.* Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N. Engl. J. Med.* **383**, 2603–2615 (2020).
15. Baden, L. R. *et al.* Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N. Engl. J. Med.* **384**, 403–416 (2021).
16. Voysey, M. *et al.* Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet* **397**, 99–111 (2021).

17. Gouglas, D. *et al.* Estimating the cost of vaccine development against epidemic infectious diseases: a cost minimisation study. *Lancet Glob. Heal.* **6**, e1386–e1396 (2018).

18. Yung, C.-F. *et al.* Dengue serotype-specific differences in clinical manifestation, laboratory parameters and risk of severe disease in adults, singapore. *Am. J. Trop. Med. Hyg.* **92**, 999–1005 (2015).

19. Cuevas, J. M., Geller, R., Garijo, R., López-Aldeguer, J. & Sanjuán, R. Extremely High Mutation Rate of HIV-1 In Vivo. *PLOS Biol.* **13**, e1002251 (2015).

20. Andrews, N. *et al.* Covid-19 Vaccine Effectiveness against the Omicron (B.1.1.529) Variant. *N. Engl. J. Med.* 1–15 (2022) doi:10.1056/nejmoa2119451.

21. Lou, Z., Sun, Y. & Rao, Z. Current progress in antiviral strategies. *Trends Pharmacol. Sci.* **35**, 86–102 (2014).

22. Su, F. & Ioannou, G. N. The impact of direct-acting antiviral therapy for hepatitis C on hepatocellular carcinoma risk. *Curr. Hepatol. reports* **17**, 377–384 (2018).

23. Gonzales Zamora, J. A. Adverse Effects of Direct Acting Antivirals in HIV/HCV Coinfected Patients: A 4-Year Experience in Miami, Florida. *Diseases* **6**, 51 (2018).

24. Wu, P.-S. *et al.* An Investigation of the Side Effects, Patient Feedback, and Physiological Changes Associated with Direct-Acting Antiviral Therapy for Hepatitis C. *Int. J. Environ. Res. Public Health* **16**, 4981 (2019).

25. Ahmed, A. & Felmlee, D. J. Mechanisms of Hepatitis C Viral Resistance to Direct Acting Antivirals. *Viruses* **7**, 6716–6729 (2015).

26. Sanjuán, R. & Domingo-Calap, P. Mechanisms of viral mutation. *Cell. Mol. Life Sci.* **73**, 4433–4448 (2016).

27. Ji, X. & Li, Z. Medicinal chemistry strategies toward host targeting antiviral agents. *Med. Res. Rev.* **40**, 1519–1557 (2020).

28. Griffiths, E. C., Pedersen, A. B., Fenton, A. & Petchey, O. L. The nature and consequences of coinfection in humans. *J. Infect.* **63**, 200–206 (2011).

29. World Health Organization. WHO model list of essential medicines - 22nd list, 2021. *Tech. Doc.* 2021 (2021).

30. Leyssen, P., Balzarini, J., De Clercq, E. & Neyts, J. The predominant mechanism by which ribavirin exerts its antiviral activity in vitro against flaviviruses and paramyxoviruses is mediated by inhibition of IMP dehydrogenase. *J. Virol.* **79**, 1943–1947 (2005).

31. Bauer, R. A. Covalent inhibitors in drug discovery: From accidental discoveries to avoided liabilities and designed therapies. *Drug Discov. Today* **20**, 1061–1073 (2015).

32. Rey, F. A. & Lok, S.-M. Common Features of Enveloped Viruses and Implications for Immunogen Design for Next-Generation Vaccines. *Cell* **172**, 1319–1334 (2018).

33. Brun, J. *et al.* Assessing Antigen Structural Integrity through Glycosylation Analysis of the SARS-CoV-2 Viral Spike. *ACS Cent. Sci.* **7**, 586–593 (2021).

34. Dwek, R. A., Butters, T. D., Platt, F. M. & Zitzmann, N. Targeting glycosylation as a therapeutic approach. *Nat. Rev. Drug Discov.* **1**, 65–75 (2002).

35. Hammond, C., Braakman, I. & Helenius, A. Role of N-linked oligosaccharide

recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. *Proc. Natl. Acad. Sci.* **91**, 913–917 (1994).

36. Chen, W., Helenius, J., Braakman, I. & Helenius, A. Cotranslational folding and calnexin binding during glycoprotein synthesis. *Proc. Natl. Acad. Sci.* **92**, 6229–6233 (1995).

37. Tatu, U., Hammond, C. & Helenius, A. Folding and oligomerization of influenza hemagglutinin in the ER and the intermediate compartment. *EMBO J.* **14**, 1340–1348 (1995).

38. Hebert, D. N., Foellmer, B. & Helenius, A. Calnexin and calreticulin promote folding, delay oligomerization and suppress degradation of influenza hemagglutinin in microsomes. *EMBO J.* **15**, 2961–2968 (1996).

39. Kelleher, D. J. & Gilmore, R. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **16**, 47R-62R (2006).

40. Ruggiano, A., Foresti, O. & Carvalho, P. Quality control: ER-associated degradation: protein quality control and beyond. *J. Cell Biol.* **204**, 869–879 (2014).

41. Xu, C. & Ng, D. T. W. Glycosylation-directed quality control of protein folding. *Nat. Rev. Mol. Cell Biol.* **16**, 742–752 (2015).

42. Galli, C., Bernasconi, R., Soldà, T., Calanca, V. & Molinari, M. Malectin Participates in a Backup Glycoprotein Quality Control Pathway in the Mammalian ER. *PLoS One* **6**, e16304 (2011).

43. Chen, Y. *et al.* Role of malectin in Glc(2)Man(9)GlcNAc(2)-dependent quality control of α1-antitrypsin. *Mol. Biol. Cell* **22**, 3559–3570 (2011).

44. Qin, S.-Y. *et al.* Malectin forms a complex with ribophorin I for enhanced association with misfolded glycoproteins. *J. Biol. Chem.* **287**, 38080–38089 (2012).

45. Sellaththurai, S. *et al.* Molecular and functional insights into a novel teleost malectin from big-belly seahorse Hippocampus abdominalis. *Fish Shellfish Immunol.* **99**, 483–494 (2020).

46. Hebert, D. N., Foellmer, B. & Helenius, A. Glucose trimming and reglucosylation determine glycoprotein association with calnexin in the endoplasmic reticulum. *Cell* **81**, 425–433 (1995).

47. Hammond, C. & Helenius, A. Quality control in the secretory pathway: retention of a misfolded viral membrane glycoprotein involves cycling between the ER, intermediate compartment, and Golgi apparatus. *J. Cell Biol.* **126**, 41–52 (1994).

48. Ninagawa, S. *et al.* EDEM2 initiates mammalian glycoprotein ERAD by catalyzing the first mannose trimming step. *J. Cell Biol.* **206**, 347–356 (2014).

49. Lamriben, L., Graham, J. B., Adams, B. M. & Hebert, D. N. N-Glycan-based ER Molecular Chaperone and Protein Quality Control System: The Calnexin Binding Cycle. *Traffic* **17**, 308–326 (2016).

50. George, G. *et al.* Purified EDEM3 or EDEM1 alone produces determinant oligosaccharide structures from M8B in mammalian glycoprotein ERAD. *Elife* **10**, (2021).

51. Hamilton, S. R. *et al.* Intact {alpha}-1,2-endomannosidase is a typical type II membrane protein. *Glycobiology* **15**, 615–624 (2005).

52. Alonzi, D. S., Neville, D. C. A., Lachmann, R. H., Dwek, R. A. & Butters, T. D.

Glucosylated free oligosaccharides are biomarkers of endoplasmic- reticulum alpha-glucosidase inhibition. *Biochem. J.* **409**, 571–580 (2008).

53. Alonzi, D. S., Scott, K. A., Dwek, R. A. & Zitzmann, N. Iminosugar antivirals: The therapeutic sweet spot. *Biochem. Soc. Trans.* **45**, 571–582 (2017).

54. Caputo, A. T. *et al.* Structures of mammalian ER α-glucosidase II capture the binding modes of broad-spectrum iminosugar antivirals. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E4630–E4638 (2016).

55. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).

56. van de Laarschot, L. F. M. *et al.* Novel GANAB variants associated with polycystic liver disease. *Orphanet J. Rare Dis.* **15**, 302 (2020).

57. Gao, H. *et al.* PRKCSH/80K-H, the protein mutated in polycystic liver disease, protects polycystin-2/TRPP2 against HERP-mediated degradation. *Hum. Mol. Genet.* **19**, 16–24 (2010).

58. De Masi, R. & Orlando, S. GANAB as a Novel Biomarker in Multiple Sclerosis: Correlation with Neuroinflammation and IFI35. *Pharmaceuticals* vol. 14 (2021).

59. Shin, G. C. *et al.* PRKCSH contributes to tumorigenesis by selective boosting of IRE1 signaling pathway. *Nat. Commun.* **10**, (2019).

60. Khaodee, W., Udomsom, S., Kunnaja, P. & Cressey, R. Knockout of glucosidase II beta subunit inhibits growth and metastatic potential of lung cancer cells by inhibiting receptor tyrosine kinase activities. *Sci. Rep.* **9**, 10394 (2019).

61. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

62. Kaushal, G. P., Pastuszak, I., Hatanaka, K. & Elbein, A. D. Purification to homogeneity and properties of glucosidase II from mung bean seedlings and suspension-cultured soybean cells. *J. Biol. Chem.* **265**, 16271–16279 (1990).

63. Stigliano, I. D., Caramelo, J. J., Labriola, C. A., Parodi, A. J. & D'Alessio, C. Glucosidase II beta subunit modulates N-glycan trimming in fission yeasts and mammals. *Mol. Biol. Cell* **20**, 3974–3984 (2009).

64. Arendt, C. W. & Ostergaard, H. L. Two distinct domains of the beta-subunit of glucosidase II interact with the catalytic alpha-subunit. *Glycobiology* **10**, 487–492 (2000).

65. Stigliano, I. D., Alculumbre, S. G., Labriola, C. A., Parodi, A. J. & D'Alessio, C. Glucosidase II and N-glycan mannose content regulate the half-lives of monoglucosylated species in vivo. *Mol. Biol. Cell* **22**, 1810–1823 (2011).

66. Olson, L. J. *et al.* 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. *J. Biol. Chem.* **288**, 16460–16475 (2013).

67. Olson, L. J. *et al.* Crystal Structure and Functional Analyses of the Lectin Domain of Glucosidase II: Insights into Oligomannose Recognition. *Biochemistry* **54**, 4097–4111 (2015).

68. Totani, K., Ihara, Y., Matsuo, I. & Ito, Y. Effects of Macromolecular Crowding on Glycoprotein Processing Enzymes. *J. Am. Chem. Soc.* **130**, 2101–2107 (2008).

69. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature*

596, 583–589 (2021).

70. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).

71. Ward, B. K. *et al.* The endoplasmic reticulum-associated protein, OS-9, behaves as a lectin in targeting the immature calcium-sensing receptor. *J. Cell. Physiol.* **233**, 38–56 (2018).

72. D'Alessio, C. & Dahms, N. M. Glucosidase II and MRH-domain containing proteins in the secretory pathway. *Curr. Protein Pept. Sci.* **16**, 31–48 (2015).

73. US National Library of Medicine. A Study to Evaluate the Safety, Antiviral Effect, and Pharmacokinetics of Celgosivir in Combination With Peginterferon Alfa-2b and Ribavirin in Treatment-Naïve Patients With Chronic Hepatitis C. *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/NCT00332176?term=celgosivir&draw=2&rank=5 (2008).

74. US National Library of Medicine. A Randomized, Double-Blind Active-Controlled, Dose-Ranging Study of the Safety and Efficacy of Chronically Administered MDL 28,574A in the Treatment of HIV-Infected Patients. *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/NCT00002151?term=celgosivir&draw=2&rank=8 (2005).

75. US National Library of Medicine. A Study of the Safety and Efficacy of Chronically Administered MDL 28,574A in the Treatment of HIV-Infected Patients. *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/NCT00002150?term=celgosivir&draw=2&rank=9 (2005).

76. Warfield, K. L. *et al.* The Iminosugar UV-4 is a Broad Inhibitor of Influenza A and B Viruses ex Vivo and in Mice. *Viruses* **8**, 71 (2016).

77. US National Library of Medicine. Safety and Pharmacokinetics of UV-4B Solution Administered Orally as Multiple Ascending Doses to Healthy Subjects. *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/NCT02696291?term=UV-4&draw=2&rank=2 (2020).

78. US National Library of Medicine. Study to Determine the Safety, Tolerability and Pharmacokinetics of UV-4B Solution Administered Orally in Healthy Subjects (UV). *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/NCT02061358?term=UV-4&draw=2&rank=3 (2016).

79. Warfield, K. L. *et al.* Inhibition of endoplasmic reticulum glucosidases is required for in vitro and in vivo dengue antiviral activity by the iminosugar UV-4. *Antiviral Res.* **129**, 93–98 (2016).

80. Stavale, E. J., Vu, H., Sampath, A., Ramstedt, U. & Warfield, K. L. In vivo therapeutic protection against influenza A (H1N1) oseltamivir-sensitive and resistant viruses by the iminosugar UV-4. *PLoS One* **10**, e0121662 (2015).

81. Chang, J. *et al.* Small molecule inhibitors of ER α-glucosidases are active against multiple hemorrhagic fever viruses. *Antiviral Res.* **98**, 432–440 (2013).

82. Ma, J. *et al.* Enhancing the antiviral potency of ER α-glucosidase inhibitor IHVR-19029 against hemorrhagic fever viruses in vitro and in vivo. *Antiviral Res.* **150**, 112–122 (2018).

83. Clarke, E. C., Nofchissey, R. A., Ye, C. & Bradfute, S. B. The iminosugars celgosivir,

castanospermine and UV-4 inhibit SARS-CoV-2 replication. *Glycobiology* **31**, 378–384 (2021).

84. Callahan, M. *et al.* Randomized single oral dose phase 1 study of safety, tolerability, and pharmacokinetics of Iminosugar UV-4 Hydrochloride (UV-4B) in healthy subjects. *PLoS Negl. Trop. Dis.* **16**, e0010636 (2022).

85. Shearer, J. *et al.* Investigational New Drug Enabling Nonclinical Safety Assessment of the Iminosugar UV-4, a Broad-Spectrum Host-Targeted Antiviral Agent. *Int. J. Toxicol.* **41**, 182–200 (2022).

86. Shearer, J. *et al.* Investigational New Drug Enabling Nonclinical Safety Pharmacology Assessment of the Iminosugar UV-4, a Broad-Spectrum Host-Targeted Antiviral Agent. *Int. J. Toxicol.* **41**, 201–211 (2022).

87. Durantel, D. Celgosivir, an alpha-glucosidase I inhibitor for the potential treatment of HCV infection. *Curr. Opin. Investig. Drugs* **10**, 860–870 (2009).

88. Low, J. G. *et al.* Efficacy and safety of celgosivir in patients with dengue fever (CELADEN): a phase 1b, randomised, double-blind, placebo-controlled, proof-of-concept trial. *Lancet. Infect. Dis.* **14**, 706–715 (2014).

89. European Medicines Agency. Galafold. https://www.ema.europa.eu/en/medicines/human/EPAR/galafold (2016).

90. U.S. Food & Drug Administration. Drug Approval Package: Glyset (Miglitol) Tablets. https://www.accessdata.fda.gov/drugsatfda_docs/nda/99/020682S001_Glyset.cfm#:~:text=Approval Date%3A 8%2F16%2F1999 (2001).

91. European Medicines Agency. Zavesca. https://www.ema.europa.eu/en/medicines/human/EPAR/zavesca (2009).

92. Tierney, M. *et al.* The tolerability and pharmacokinetics of N-butyl-deoxynojirimycin in patients with advanced HIV disease (ACTG 100). The AIDS Clinical Trials Group (ACTG) of the National Institute of Allergy and Infectious Diseases. *J. Acquir. immune Defic. Syndr. Hum. retrovirology Off. Publ. Int. Retrovirology Assoc.* **10**, 549–553 (1995).

93. Ichikawa, S., Sakiyama, H., Suzuki, G., Hidari, K. I. & Hirabayashi, Y. Expression cloning of a cDNA for human ceramide glucosyltransferase that catalyzes the first glycosylation step of glycosphingolipid synthesis. *Proc. Natl. Acad. Sci.* **93**, 4638–4643 (1996).

94. Treiber, A., Morand, O. & Clozel, M. The pharmacokinetics and tissue distribution of the glucosylceramide synthase inhibitor miglustat in the rat. *Xenobiotica.* **37**, 298–314 (2007).

95. Ma, J. *et al.* Ester Prodrugs of IHVR-19029 with Enhanced Oral Exposure and Prevention of Gastrointestinal Glucosidase Interaction. *ACS Med. Chem. Lett.* **8**, 157–162 (2017).

96. Kiappes, J. L. *et al.* ToP-DNJ, a Selective Inhibitor of Endoplasmic Reticulum α-Glucosidase II Exhibiting Antiflaviviral Activity. *ACS Chem. Biol.* **13**, 60–65 (2018).

97. Balsitis, S. J. *et al.* Tropism of dengue virus in mice and humans defined by viral nonstructural protein 3-specific immunostaining. *Am. J. Trop. Med. Hyg.* **80**, 416–424 (2009).

98. Wang, X. & Quinn, P. J. The location and function of vitamin E in membranes (review).

Mol. Membr. Biol. **17**, 143–156 (2000).

99. Miller, J. L. *et al.* Liposome-mediated delivery of iminosugars enhances efficacy against dengue virus in vivo. *Antimicrob. Agents Chemother.* **56**, 6379–6386 (2012).

100. Karade, S. S. *et al.* N-Substituted Valiolamine Derivatives as Potent Inhibitors of Endoplasmic Reticulum α-Glucosidases I and II with Antiviral Activity. *J. Med. Chem.* **64**, 18010–18024 (2021).

101. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).

102. Duarte, J. M., Srebniak, A., Schärer, M. A. & Capitani, G. Protein interface classification by evolutionary analysis. *BMC Bioinformatics* **13**, 334 (2012).

103. Jurrus, E. *et al.* Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112–128 (2018).

104. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38**, W529-33 (2010).

105. Celniker, G. *et al.* ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. *Isr. J. Chem.* **53**, 199–206 (2013).

106. Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).

107. Arendt, C. W. & Ostergaard, H. L. Identification of the CD45-associated 116-kDa and 80-kDa proteins as the alpha- and beta-subunits of alpha-glucosidase II. *J. Biol. Chem.* **272**, 13117–13125 (1997).

108. Sekar, R. B. & Periasamy, A. Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *J. Cell Biol.* **160**, 629–633 (2003).

109. Banning, C. *et al.* A Flow Cytometry-Based FRET Assay to Identify and Analyse Protein-Protein Interactions in Living Cells. *PLoS One* **5**, e9344 (2010).

110. Bajar, B. T., Wang, E. S., Zhang, S., Lin, M. Z. & Chu, J. A Guide to Fluorescent Protein FRET Pairs. *Sensors (Basel).* **16**, (2016).

111. Xiong, J. *et al.* Development of a Time-Resolved Fluorescence Resonance Energy Transfer Ultrahigh-Throughput Screening Assay for Targeting the NSD3 and MYC Interaction. *Assay Drug Dev. Technol.* **16**, 96–106 (2018).

112. Schaap, M. C. A., Guimaraes, A. M. R., Wilderspin, A. F. & Wells, G. Protocol for a Steady-State FRET Assay in Cancer Chemoprevention. in *Strano S. (eds) Cancer Chemoprevention. Methods in Molecular Biology, vol 1379* 165–179 (Humana Press, New York, NY, 2016).

113. Berrow, N. S. *et al.* A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* **35**, e45 (2007).

114. Gasteiger, E. *et al.* Protein Identification and Analysis Tools on the ExPASy Server BT - The Proteomics Protocols Handbook. in (ed. Walker, J. M.) 571–607 (Humana Press, 2005). doi:10.1385/1-59259-890-0:571.

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

116. Ortega Arroyo, J. *et al.* Label-Free, All-Optical Detection, Imaging, and Tracking of a Single Protein. *Nano Lett.* **14**, 2065–2070 (2014).

117. Asor, R. & Kukura, P. Characterising biomolecular interactions and dynamics with mass photometry. *Curr. Opin. Chem. Biol.* **68**, 102132 (2022).

118. Catoire, L. J., Warnet, X. L. & Warschawski, D. E. Micelles, Bicelles, Amphipols, Nanodiscs, Liposomes, or Intact Cells: The Hitchhiker's Guide to the Study of Membrane Proteins by NMR BT - Membrane Proteins Production for Structural Analysis. in (ed. Mus-Veteau, I.) 315–345 (Springer New York, 2014). doi:10.1007/978-1-4939-0662-8_12.

119. Olerinyova, A. *et al.* Mass Photometry of Membrane Proteins. *Chem* **7**, 224–236 (2021).

120. Foley, E. D. B., Kushwah, M. S., Young, G. & Kukura, P. Mass photometry enables label-free tracking and mass measurement of single proteins on lipid bilayers. *Nat. Methods* **18**, 1247–1252 (2021).

121. Li, Y., Struwe, W. B. & Kukura, P. Single molecule mass photometry of nucleic acids. *Nucleic Acids Res.* **48**, e97–e97 (2020).

122. Fineberg, A., Surrey, T. & Kukura, P. Quantifying the Monomer–Dimer Equilibrium of Tubulin with Mass Photometry. *J. Mol. Biol.* **432**, 6168–6172 (2020).

123. Soltermann, F. *et al.* Quantifying Protein–Protein Interactions by Molecular Counting with Mass Photometry. *Angew. Chemie Int. Ed.* **59**, 10774–10779 (2020).

124. Wu, D. & Piszczek, G. Measuring the affinity of protein-protein interactions on a single-molecule level by mass photometry. *Anal. Biochem.* **592**, 113575 (2020).

125. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **1**, 2876–2890 (2006).

126. Andrews, S. S. & Tretton, J. Physical Principles of Circular Dichroism. *J. Chem. Educ.* **97**, 4370–4376 (2020).

127. Gao, K., Oerlemans, R. & Groves, M. R. Theory and applications of differential scanning fluorimetry in early-stage drug discovery. *Biophys. Rev.* **12**, 85–104 (2020).

128. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. **2**, 2212–2221 (2007).

129. Marty, M. T. *et al.* Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Anal. Chem.* **87**, 4370–4376 (2015).

130. Chen, H. *et al.* A Small Covalent Allosteric Inhibitor of Human Cytomegalovirus DNA Polymerase Subunit Interactions. *ACS Infect. Dis.* **3**, 112–118 (2017).

131. Yu, Y.-Q., Gilar, M., Lee, P. J., Bouvier, E. S. P. & Gebler, J. C. Enzyme-Friendly, Mass Spectrometry-Compatible Surfactant for In-Solution Enzymatic Digestion of Proteins. *Anal. Chem.* **75**, 6023–6028 (2003).

132. Suder, P., Bierczynska, A., König, S. & Silberring, J. Acid-labile surfactant assists in-solution digestion of proteins resistant to enzymatic attack. *Rapid communications in mass spectrometry : RCM* vol. 18 822–824 (2004).

133. Lauber, M. A. *et al.* Rapid Preparation of Released N-Glycans for HILIC Analysis Using a Labeling Reagent that Facilitates Sensitive Fluorescence and ESI-MS Detection. *Anal. Chem.* **87**, 5401–5409 (2015).

134. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).

135. Diament, B. J. & Noble, W. S. Faster SEQUEST searching for peptide identification from tandem mass spectra. *J. Proteome Res.* **10**, 3871–3879 (2011).

136. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319 (2016).

137. Generalized Proteoform Meta-analysis Database. gpmDB. https://gpmdb.thegpm.org/ (2023).

138. Pino, L. K. *et al.* The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. *Mass Spectrom. Rev.* **39**, 229–244 (2020).

139. Gupta, R. & Brunak, S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac. Symp. Biocomput.* 310–322 (2002).

140. Steentoft, C. *et al.* Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* **32**, 1478–1488 (2013).

141. Kim, M.-S. & Pandey, A. Electron transfer dissociation mass spectrometry in proteomics. *Proteomics* **12**, 530–542 (2012).

142. Zhou, M. *et al.* Higher-order structural characterisation of native proteins and complexes by top-down mass spectrometry. *Chem. Sci.* **11**, 12918–12936 (2020).

143. Warfield, K. L. *et al.* Targeting Endoplasmic Reticulum α-Glucosidase I with a Single-Dose Iminosugar Treatment Protects against Lethal Influenza and Dengue Virus Infections. *J. Med. Chem.* **63**, 4205–4214 (2020).

144. Pollock, S. *et al.* Uptake and trafficking of liposomes to the endoplasmic reticulum. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **24**, 1866–1878 (2010).

145. Liu, Y., Jia, H.-R., Han, X. & Wu, F.-G. Endoplasmic reticulum-targeting nanomedicines for cancer therapy. *Smart Mater. Med.* **2**, 334–349 (2021).

146. Berridge, M. V, Herst, P. M. & Tan, A. S. B. T.-B. A. R. Tetrazolium dyes as tools in cell biology: New insights into their cellular reduction. in vol. 11 127–152 (Elsevier, 2005).

147. Klijn, C. *et al.* A comprehensive transcriptional portrait of human cancer cell lines. *Nat. Biotechnol.* **33**, 306–312 (2015).

148. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).

149. Renaud, J.-P. *et al.* Cryo-EM in drug discovery: achievements, limitations and prospects. *Nat. Rev. Drug Discov.* **17**, 471–492 (2018).

150. Bartesaghi, A. *et al.* 2.2 Å resolution cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor. *Science* **348**, 1147–1151 (2015).

151. He, S., Zhiti, A., Barba-Bon, A., Hennig, A. & Nau, W. M. Real-Time Parallel Artificial Membrane Permeability Assay Based on Supramolecular Fluorescent Artificial Receptors . *Frontiers in Chemistry* vol. 8 (2020).

152. de Souza Neto, L. R. *et al.* In silico Strategies to Support Fragment-to-Lead Optimization in Drug Discovery . *Frontiers in Chemistry* vol. 8 (2020).

153. US National Library of Medicine. Miglustat on MedlinePlus. *MedlinePlus*

https://medlineplus.gov/druginfo/meds/a604015.html (2018).

154. US National Library of Medicine. Miglitol on MedlinePlus. *MedlinePlus* https://medlineplus.gov/druginfo/meds/a601079.html (2018).

155. Corey, E. J. The Logic of Chemical Synthesis: Multistep Synthesis of Complex Carbogenic Molecules (Nobel Lecture). *Angew. Chemie Int. Ed. English* **30**, 455–465 (1991).

156. Best, D. *et al.* Looking glass inhibitors: scalable syntheses of DNJ, DMDP, and (3R)-3-hydroxy-l-bulgecinine from d-glucuronolactone and of l-DNJ, l-DMDP, and (3S)-3-hydroxy-d-bulgecinine from l-glucuronolactone. DMDP inhibits β-glucosidases and β-galactosidases whereas l. *Tetrahedron: Asymmetry* **21**, 311–319 (2010).

157. Wennekes, T. *et al.* Development of Adamantan-1-yl-methoxy-Functionalized 1-Deoxynojirimycin Derivatives as Selective Inhibitors of Glucosylceramide Metabolism in Man. *J. Org. Chem.* **72**, 1088–1097 (2007).

158. Niu, P. *et al.* Synthesis of 2-amino-1,3,4-oxadiazoles and 2-amino-1,3,4-thiadiazoles via sequential condensation and I2-mediated oxidative C–O/C–S bond formation. *J. Org. Chem.* **80**, 1018–1024 (2015).

159. Omar, A.-M. M. E. & Aboulwafa, O. M. Synthesis and anticonvulsant properties of a novel series of 2-substituted amino-5-aryl-1,3,4-oxadiazole derivatives. *J. Heterocycl. Chem.* **21**, 1415–1418 (1984).

160. Dumčiūtė, J. *et al.* Synthesis and ring transformations of 1-amino-1,2,3,9a-tetrahydroimidazo[1,2-a]indol-2(9H)-ones. *Tetrahedron* **62**, 3309–3319 (2006).

161. PATEL, N. B. & PATEL, J. C. Synthesis and Antimicrobial Activity of 3-(1,3,4-Oxadiazol-2-yl)quinazolin-4(3H)-ones. *Scientia Pharmaceutica* vol. 78 171–194 (2010).

162. Grrene, T. W. & Wuts, P. G. M. Protection for the Hydroxyl Group, Including 1,2- and 1,3-Diols. in *Greene's Protective Groups in Organic Synthesis* 16–366 (2006). doi:https://doi.org/10.1002/9780470053485.ch2.

163. Mandal, P. K. & McMurray, J. S. Pd−C-Induced Catalytic Transfer Hydrogenation with Triethylsilane. *J. Org. Chem.* **72**, 6599–6601 (2007).

164. Meyer, S. D. & Schreiber, S. L. Acceleration of the Dess-Martin Oxidation by Water. *J. Org. Chem.* **59**, 7549–7552 (1994).

165. Linderman, R. J. & Graves, D. M. Oxidation of fluoroalkyl-substituted carbinols by the Dess-Martin reagent. *J. Org. Chem.* **54**, 661–668 (1989).

166. Magtaan, J. K., Devocelle, M. & Kelleher, F. Regeneration of aged DMF for use in solid-phase peptide synthesis. *J. Pept. Sci.* **25**, e3139 (2019).

167. Frigerio, M., Santagostino, M. & Sputore, S. A User-Friendly Entry to 2-Iodoxybenzoic Acid (IBX). *J. Org. Chem.* **64**, 4537–4538 (1999).

168. Yusubov, M. S. *et al.* 2-Iodoxybenzoic Acid Tosylates: the Alternative to Dess–Martin Periodinane Oxidizing Reagents. *Adv. Synth. Catal.* **359**, 3207–3216 (2017).

169. Greene, T. W. & Wuts, P. G. M. Protection for the Amino Group. in *Protective Groups in Organic Synthesis* 494–653 (1999). doi:https://doi.org/10.1002/0471220574.ch7.

170. Lonsdale, R. & Ward, R. A. Structure-based design of targeted covalent inhibitors. *Chem. Soc. Rev.* **47**, 3816–3830 (2018).

171. Huang, F., Han, X., Xiao, X. & Zhou, J. Covalent Warheads Targeting Cysteine Residue: The Promising Approach in Drug Development. *Molecules* **27**, 7728 (2022).

172. Deprez, P., Gautschi, M. & Helenius, A. More than one glycan is needed for ER glucosidase II to allow entry of glycoproteins into the calnexin/calreticulin cycle. *Mol. Cell* **19**, 183–195 (2005).

173. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).

174. Park, S.-J. *et al.* CHARMM-GUI Glycan Modeler for modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology* **29**, 320–331 (2019).

175. Wolny, M. *et al.* Characterization of long and stable de novo single alpha-helix domains provides novel insight into their stability. *Sci. Rep.* **7**, 44341 (2017).

176. Wolny, M. *et al.* Stable Single α-Helices Are Constant Force Springs in Proteins. *J. Biol. Chem.* **289**, 27825–27835 (2014).

177. Lawrence, M. C., Izard, T., Beuchat, M., Blagrove, R. J. & Colman, P. M. Structure of Phaseolin at 2·2 Å Resolution: Implications for a Common Vicilin/Legumin Structure and the Genetic Engineering of Seed Storage Proteins. *J. Mol. Biol.* **238**, 748–776 (1994).

178. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D* **67**, 235–242 (2011).

179. Krissinel, E., Uski, V., Lebedev, A., Winn, M. & Ballard, C. Distributed computing for macromolecular crystallography. *Acta Crystallogr. Sect. D* **74**, 143–151 (2018).

180. Arkin, M. R. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat. Rev. Drug Discov.* **3**, 301–317 (2004).

181. Roth, G. J., Stanford, N. & Majerus, P. W. Acetylation of prostaglandin synthase by aspirin. *Proc. Natl. Acad. Sci. U. S. A.* **72**, 3073–3076 (1975).

182. Yocum, R. R., Rasmussen, J. R. & Strominger, J. L. The mechanism of action of penicillin. Penicillin acylates the active site of Bacillus stearothermophilus D-alanine carboxypeptidase. *J. Biol. Chem.* **255**, 3977–3986 (1980).

183. Singh, J., Petter, R. C., Baillie, T. A. & Whitty, A. The resurgence of covalent drugs. *Nat. Rev. Drug Discov.* **10**, 307–317 (2011).

184. Jozwiak-Bebenista, M. & Nowak, J. Z. Paracetamol: Mechanism of action, applications and safety concern. *Acta Pol. Pharm. - Drug Res.* **71**, 11–23 (2014).

185. Ran, X. & Gestwicki, J. E. Inhibitors of protein–protein interactions (PPIs): an analysis of scaffold choices and buried surface area. *Curr. Opin. Chem. Biol.* **44**, 75–86 (2018).

186. Vasudevan, A. *et al.* Chapter One - Covalent binders in drug discovery. in *Progress in Medicinal Chemistry - Volume 58* (2019).

187. Khan, A., Munir, M., Aiman, S., Wadood, A. & Khan, A. The in silico identification of small molecules for protein-protein interaction inhibition in AKAP-Lbc–RhoA signaling complex. *Comput. Biol. Chem.* **67**, 84–91 (2017).

188. Awoonor-Williams, E., Walsh, A. G. & Rowley, C. N. Modeling covalent-modifier drugs. *Biochim. Biophys. Acta - Proteins Proteomics* **1865**, 1664–1675 (2017).

189. Willems, H., De Cesco, S. & Svensson, F. Computational Chemistry on a Budget:

Supporting Drug Discovery with Limited Resources. *J. Med. Chem.* **63**, (2020).

190. Ballone, A., Centorrino, F. & Ottmann, C. 14-3-3: A Case Study in PPI Modulation. *Molecules* **23**, 1386 (2018).

191. Cheng, S.-S., Yang, G.-J., Wang, W., Leung, C.-H. & Ma, D.-L. The design and development of covalent protein-protein interaction inhibitors for cancer treatment. *J. Hematol. Oncol.* **13**, 26 (2020).

192. Dugan, A. *et al.* Discovery of Enzymatic Targets of Transcriptional Activators via in Vivo Covalent Chemical Capture. *J. Am. Chem. Soc.* **138**, 12629–12635 (2016).

193. Degorce, F. *et al.* HTRF: A technology tailored for drug discovery - a review of theoretical aspects and recent applications. *Curr. Chem. Genomics* **3**, 22–32 (2009).

194. Salem, M., Mauguen, Y. & Prangé, T. Revisiting glutaraldehyde cross-linking: the case of the Arg-Lys intermolecular doublet. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **66**, 225–228 (2010).

195. Gorrec, F. The MORPHEUS II protein crystallization screen. *Acta Crystallogr. Sect. F* **71**, 831–837 (2015).

196. Tang, X. & Bruce, J. E. Chemical Cross-Linking for Protein–Protein Interaction Studies BT - Mass Spectrometry of Proteins and Peptides: Methods and Protocols. in (eds. Lipton, M. S. & Paša-Tolic, L.) 283–293 (Humana Press, 2009). doi:10.1007/978-1-59745-493-3_17.

197. Arora, B., Tandon, R., Attri, P. & Bhatia, R. Chemical Crosslinking: Role in Protein and Peptide Science. *Curr. Protein Pept. Sci.* **18**, 946–955 (2017).

9. Appendices

9.1 Literature Review Summary: Covalent and Protein-Protein Interface-targeting Drugs

Historically, both the therapeutic targeting of PPIs and the use of covalent warheads in drug discovery have met with significant resistance, despite sporadic examples of their use in 20th century literature. In the case of PPI modulation, typical interface architecture being large and flat meant that traditional ideas with regards to protein-ligand interactions did not seem to apply. This, in combination with the rarity of natural products that could act as medicinal chemistry starting points and the complications in differentiating true biological PPIs from artefactual crystallographic interfaces, led to the description of PPIs as “undruggable” targets.[180] Covalent protein-ligand interactions were perhaps less difficult to achieve, with early examples such as Aspirin and penicillin inadvertently utilising covalent modification to exert therapeutic activity, but the risk of idiosyncratic drug-related toxicity as a result of the unpredictable off-target reactivity of non-selective covalent warheads was perceived to be too great.[181–183] In this regard, paracetamol became the poster child for caution towards covalent inhibitors as a result of the appreciable hepatotoxicity of its metabolites at high doses.[184] Moreover, the ubiquity of endogenous nucleophiles, such as glutathione, in the body results in the scavenging of reactive electrophiles such that undesirably large or frequent doses are required for any observable therapeutic activity.[170]

Despite the underlying resistance towards these avenues of drug discovery, both have gained considerable traction since the turn of the century, with several examples of covalent inhibitors and PPI modulators approved for clinical use or in late stage clinical trials.[185,186] The

resurgence of covalent inhibition and PPI modulation can be, in part, attributed to the significant advances in computational biochemistry such that precise structure-activity relationships can be identified and rational, structure-based drug design can be achieved with relative ease, provided 3-D structural data is available.[187–189] The combination of advanced computational biochemistry and sophisticated experimental techniques have been used to great effect in capitalising on the potential benefits of covalent/PPI modulators while minimising the associated risks and challenges. While many PPIs deserve the "undruggable" moniker due to their large buried interfacial areas, the revelation that some PPIs are not as flat as once assumed, but rather studded with complementary pockets, known as hot spots, that contribute a disproportionate amount of the binding energy allowed for attempts at targeting these hot spots with small molecules, rather than attempting to target the relatively large area that constitutes the entire PPI. Consequently, researchers have been able to make use of the highly specific architecture of PPIs to produce particularly selective modulators of biological processes that would be otherwise intractable. PPI modulators of various structure and function have been developed and show considerable promise in the development of novel therapeutics for a variety of diseases and conditions, including cancers, pathogenic infection and neurological disorders.[185] Furthermore, a more detailed understanding of proteins with a high level of connectivity, so-called 'hub proteins', opens the door for the use of PPI modulators of just one target protein in order to affect many different cellular pathways.[190] In addition to the potential clinical applications, it is also possible that the use of PPI modulators as chemical probes into signalling pathways controlled by the hub protein of interest could advance our understanding of the hub protein itself and of the implications of its modulation on a more holistic level, potentially identifying novel, therapeutically relevant PPIs in the process.

The incorporation of covalently binding small molecules into the catalogue of possible therapeutics was revolutionised by the inception of targeted covalent inhibitors (TCIs) where

the initial, non-covalent protein-ligand interaction is optimised, providing selectivity for the target nucleophilic residue and orienting the electrophilic warhead on the ligand such that the covalent bond forms readily, even when the warhead has intrinsically low reactivity. The contribution of computational biochemistry cannot be overstated here, as it allows for the rational design of the ligand in order to fine tune the non-covalent interactions. This allows for potentially irreversible, and therefore near permanent, inhibition of the target protein in a selective and specific manner. This can confer significant advantages over temporary, non-covalent inhibition, which include the ability to use lower and less frequent doses, which improves both the risk of idiosyncratic drug-related toxicity and the convenience of the dosing schedule for the patient. Moreover, the use of carefully chosen electrophilic warheads with low intrinsic reactivity further decreases the risk of off-target toxicity as well as the proportion of the compound that is scavenged by endogenous nucleophiles, improving the pharmacokinetics/pharmacodynamics relationship.[170] Despite these benefits, the development of covalent inhibitors is not without its challenges, and much of the original scepticism towards it is well-founded and should be carefully considered, with particular heed paid to the risk of unpredictable off-target activity. However, as with PPI modulators, the incidences of *ab initio* covalent inhibition in the literature has dramatically increased in the last decade and there are many examples of FDA-approved covalent drugs treating conditions such as cancer and infectious disease, as well as cardiovascular, gastrointestinal and CNS complications.[186]

The combination of these two strategies represents an exciting future prospect for drug discovery, with many examples of covalently binding PPI modulators emerging in the literature, with particular focus on their use in oncology and as transcriptional regulators.[191,192] In theory, this combination could provide a highly desirable drug profile, with the high selectivity and specificity associated with successful PPI modulators synergistically integrating with the potential for enhanced potency, duration of effect and low dosages provided by

covalently binding small molecules. While the enhanced duration of effect of covalent inhibitors may have detrimental effects in some cases, particularly when the target protein is essential for cellular proliferation, it can theoretically be exploited to great effect in situations where permanent inhibition of a particular process is desirable, as is often the case in infectious disease. Such a situation is exemplified by HTAs, which seek to eliminate or nullify viral infections by inhibiting host processes necessary for viral replication.[27] Covalently acting HTAs would be beneficial as the extended inhibition of the target could greatly increase the likelihood of viral elimination at relatively small doses, and targeting relevant PPIs could provide selectivity for the desired target, thereby limiting off-target activity against other structurally similar host proteins, or those belonging to the same family. The extensive toolkit of modern *in* silico analytical techniques could focus *ab initio* design of molecules bearing both these features to produce highly potent and highly selective drugs against both well-known and completely novel protein targets. To date, there are many examples of PPI modulators, covalent inhibitors and a combination thereof in various stages of clinical development and the boom in the literature for both classes of therapeutic agents since the turn of the century will likely be followed by a wave of regulatory approvals in this exciting field.

9.2 Fluorescence resonance energy transfer (FRET) assay

Table I – Experimental settings for TR-FRET assays conducted by Dr Mario Hensen. *HTRF – homogenous time resolved fluorescence; Exc. – excitation; Ems – Emission.*

Variable	Setting
Donor Fluorophore (mAb)	Terbium, Tb (Anti-His_6)
Concentration of Donor Fluorophore	0.25 nM
Acceptor Fluorophore (mAb)	XL665 (Anti-FLAG)[193]
Concentration of Acceptor Fluorophore	2.5 nM
Detection type	HTRF, ratio (A655 nm/A620 nm) * 10,000
Top mirror	LANCE/DELFIA Dual Enh (D400/D630)
Exc. Filter	UV2 (TRF) 320
Exc. Wavelength	320 nm
Ems. Filter	APC 665
Ems. Wavelength	665 nm
2nd Ems. Filter	Europium 615
2nd Ems. wavelength	615 nm
Measurement height	3 mm
Cycle	2000
Delay	150 µsec
No. of flashes	200
No. of flashes for 2nd detector	20
No. of sequence windows	1
Total time of windows	800 µsec
Exc. light strength	100%

9.3 Molecular Biology: DNA Constructs & Sequences

9.3.1 Humanised murine *GANAB* and *PRKCSH*

The following are constructs in the pHLsec vector and were provided by Dr Mario Hensen.

Red-Organge – pHLsec secretion signal. Red peptide is removed during protein processing in ER/Golgi, so is not present in the purified protein.

Yellow – Glycine spacer introduced during cloning to maintain reading frame.

Green – His_6/FLAG-tag

Cyan – SDM for humanisation

Secretion Signal-MmGANAB-F724G-His_6-2xstop:

ATGGGGATCCTTCCCAGCCCTGGGATGCCTGCGCTGCTCTCCCTCGTGAGCCTTCTCTCCGTGCTGCTGATGGGTTGCGT
AGCTGAAACCGGTGGCCTTGCTGTGGATAGAAGCAACTTTAAGACCTGTGATGAGAGTTCCTTTTGCAAACGGCAGCGAA
GCATTCGGCCAGGCCTCTCTCCTTACCGTGCCTTGCTGGACACTCTGCAGCTTGGTCCTGATGCTCTTACAGTCCATCTGA
TCCATGAAGTCACCAAGGTGCTGCTTGTGCTGGAACTCCAGGGCCTTCAGAAGAACATGACTCGGATCAGGATCGATGAG
CTAGAGCCCCGGCGGCCTCGATACCGAGTGCCAGATGTTTTAGTGGCTGACCCCCCCACAGCTAGGCTTTCAGTCTCTGG
CCGTGATGACAACAGTGTGGAGCTAACAGTGGCTGAGGGACCCTACAAAATCATTTTGACAGCACAGCCATTCCGCCTTG
ACCTGCTAGAAGATCGAAGCCTCCTGCTCAGTGTCAATGCCCGAGGACTTATGGCCTTTGAGCACCAGAGGGCCCCCAGG
GTCCCTTTCTCGGATAAAGTTAGTCTCGCGCTCGGTAGCGTGTGGGATAAGATCAAGAACCTTTTCTCTAGGCAAGAATCA
AAAGACCCAGCTGAAGGCAATGGAGCCCAGCCTGAAGCAACACCTGGGGATGGTGACAAGCCAGAGGAGACCCAGGAAA
AGGCTGAGAAGGATGAGCCAGGAGCCTGGGAGGAGACATTCAAAACACATTCTGACAGCAAGCCTTATGGCCCCACGTCT
GTAGGTTTGGACTTTTCTCTGCCAGGAATGGAACATGTGTATGGGATCCCTGAGCATGCTGACAGCCTGAGACTGAAGGT
CACTGAGGGCGGTGAGCCGTACCGCCTGTACAATTTGGATGTGTTCCAGTATGAGCTGAACAACCCAATGGCTCTATATG
GGTCTGTGCCTGTGCTCCTGGCACACAGCTTTCATCGAGACCTGGGCATCTTCTGGCTTAATGCTGCTGAGACTTGGGTT
GATATATCCTCCAACACGGCTGGGAAGACCCTGTTTGGGAAGATGCTTGATTACCTGCAGGGCTCTGGGGAGACTCCACA
GACAGACATTCGTTGGATGTCAGAGAGTGGCATTATTGATGTTTTCCTAATGCTTGGCCCTTCGGTTTTTGATGTCTTTAGG
CAGTATGCTAGTCTCACAGGGACCCAGGCATTGCCCCCACTCTTCTCCCTCGGCTATCACCAGAGTCGCTGGAACTACCG
GGATGAGGCTGATGTTTTGGAAGTGGATCAGGGTTTTGATGATCACAACATGCCTTGTGATGTCATTTGGTTGGACATTGA
ACATGCTGATGGCAAGCGGTACTTCACTTGGGACCCCACCCGATTTCCTCAGCCCCTCAATATGCTTGAGCACTTGGCTTC
CAAGAGGCGGAAGCTGGTGGCCATTGTGGACCCCCACATCAAGGTAGACTCTGGCTACCGAGTTCACGAAGAATTGCGAA
ACCACGGGCTGTATGTTAAAACTCGGGATGGCTCTGATTACGAGGGCTGGTGCTGGCCAGGCTCAGCTAGTTACCCTGAC
TTCACTAATCCAAGGATGAGGGCCTGGTGGTCTAACATGTTCAGCTTTGACAATTATGAGGGTTCAGCTCCTAATCTTTATG
TTTGGAATGACATGAATGAACCGTCTGTGTTCAATGGTCCTGAGGTCACCATGTTGAAGGATGCTGTGCATTATGGAGGCT
GGGAGCACCGGGACATCCATAACATCTATGGCTTATATGTGCACATGGCGACTGCTGATGGGCTAATACAGCGCTCTGGG
GGCATAGAGCGTCCCTTTGTCCTGAGTAGGGCTTTCTTCTCAGGCTCCCAGCGCTTTGGAGCTGTGTGGACAGGGGACAA
CACTGCCGAATGGGATCATTTGAAGATCTCTATCCCTATGTGTCTCAGCCTGGCACTGGTGGGGCTTTCCTTCTGTGGAGC
GGATGTGGGTGGCTTCTTCAAGAACCCAGAGCCAGAGCTGCTTGTGCGCTGGTACCAAATGGGTGCCTACCAGCCGTTCT
TTCGGGCTCATGCCCACTTGGACACTGGGCGGCGAGAGCCCTGGCTGTTAGCGTCTCAATACCAAGATGCAATCCGAGAT
GCCTTGGGCCAGCGATATTCTTTGCTGCCCTTCTGGTATACCCTCTTCTATCAAGCTCACAAGGAAGGGTTTCCTGTCATG
AGGCCCCTCTGGGTACAGTATCCTGAGGATATGTCTACCTTCAGTATAGAGGATCAGTTCATGCTTGGTGATGCACTCCTT
ATTCACCCTGTATCGGATGCTGGGGCCCACGGAGTGCAGGTCTATTTGCCTGGCCAAGAAGAAGTGTGGTATGACATTCA
GAGCTATCAGAAGCATCATGGGCCCCAGACCTTGTATCTGCCAGTAACTTTGAGCAGTATCCCCGTGTTCCAGCGTGGCG
GAACCATTGTGCCTCGATGGATGCGTGTGAGGCGCTCTTCAGACTGTATGAAGGACGATCCTATCACTCTCTTTGTTGCTC
TCAGTCCCCAGGGTACTGCCCAAGGAGAACTCTTTCTAGATGATGGACACACCTTTAACTATCAGACTCGCCATGAGTTCC
TGTTGCGGCGGTTCTCTTTCTCTGGCAGCACACTGGTCTCTAGTTCAGCAGACCCCAAAGGCCACCTTGAGACACCTATTT
GGATTGAGCGAGTAGTCATCATGGGGGCTGGAAAGCCAGCAGCTGTGGTGCTCCAGACCAAAGGATCCCCTGAAAGTCG
CCTGTCCTTCCAGCATGACCCTGAGACCTCAGTGCTGATATTGCGTAAACCTGGTGTCAGCGTGGCATCCGACTGGAGTA
TTCATCTTCGAGGTACCAAGCACCACCATCACCATCACTAATGA

Translation:

MGILPSPGMPALLSLVSLLSVLLMGCVAETGGLAVDRSNFKTCDESSFCKRQRSIRPGLSPYRALLDTLQLGPDALTVHLIHEVTK
VLLVLELQGLQKNMTRIRIDELEPRRPRYRVPDVLVADPPTARLSVSGRDDNSVELTVAEGPYKIILTAQPFRLDLLEDRSLLLSVN
ARGLMAFEHQRAPRVPFSDKVSLALGSVWDKIKNLFSRQESKDPAEGNGAQPEATPGDGDKPEETQEKAEKDEPGAWEETFK
THSDSKPYGPTSVGLDFSLPGMEHVYGIPEHADSLRLKVTEGGEPYRLYNLDVFQYELNNPMALYGSVPVLLAHSFHRDLGIFW
LNAAETWVDISSNTAGKTLFGKMLDYLQGSGETPQTDIRWMSESGIIDVFLMLGPSVFDVFRQYASLTGTQALPPLFSLGYHQS
RWNYRDEADVLEVDQGFDDHNMPCDVIWLDIEHADGKRYFTWDPTRFPQPLNMLEHLASKRRKLVAIVDPHIKVDSGYRVHEE
LRNHGLYVKTRDGSDYEGWCWPGSASYPDFTNPRMRAWWSNMFSFDNYEGSAPNLYVWNDMNEPSVFNGPEVTMLKDAVH
YGGWEHRDIHNIYGLYVHMATADGLIQRSGGIERPFVLSRAFFSGSQRFGAVWTGDNTAEWDHLKISIPMCLSLALVGLSFCGA
DVGGFFKNPEPELLVRWYQMGAYQPFFRAHAHLDTGRREPWLLASQYQDAIRDALGQRYSLLPFWYTLFYQAHKEGFPVMRP
LWVQYPEDMSTFSIEDQFMLGDALLIHPVSDAGAHGVQVYLPGQEEVWYDIQSYQKHHGPQTLYLPVTLSSIPVFQRGGTIVPR
WMRVRRSSDCMKDDPITLFVALSPQGTAQGELFLDDGHTFNYQTRHEFLLRRFSFSGSTLVSSSADPKGHLETPIWIERVVIMG
AGKPAAVVLQTKGSPESRLSFQHDPETSVLILRKPGVSVASDWSIHLRGTKHHHHHH**

Secretion Signal-FLAGtag-MmPRKCSH-L88P-S90N-2xstop:

ATGGGGATCCTTCCCAGCCCTGGGATGCCTGCGCTGCTCTCCCTCGTGAGCCTTCTCTCCGTGCTGCTGATGGGTTGCGT
AGCTGAAACCGGTGACTACAAAGACGATGACGACAAGGGCGCTGTAGAAGTTAAGAGACCCCGGGGCGTTTCCCTCAGC
AACCATCACTTCTATGAAGAATCTAAACCTTTCACCTGTTTGGACGGCACAGCCACCATCCCATTCGATCAGGTGAACGAC
GACTACTGCGACTGTAAGGACGGTTCAGATGAGCCTGGCACAGCTGCTTGTCCCAATGGCAGCTTTCACTGCACCAACAC
TGGGTACAAGCCCTTGTACATCCCCTCCAACCGGGTCAATGATGGGGTATGTGACTGCTGTGATGGCACAGACGAGTACA
ACAGCGGCACGGTCTGCGAGAACACCTGCAGAGAGAAGGGTCGCAAGGAAAAAGAGTCCCTGCAGCAGCTGGCGGAAGT
CACCCGTGAAGGGTTCCGCCTGAAGAAGATTCTCATTGAGGAGTGGAAGACAGCCCGGGAAGAAAAGCAGAGTAAGCTTC
TTGAGCTTCAGGCTGGAAAGAAGTCTCTGGAAGACCAGGTAGAAACACTGCGGGCAGCGAAAGAAGAAGCAGAGAGGCC
AGAGAAGGAGGCCAAGGACCAGCACCGGAAGCTGTGGGAAGAGCAGCAAGCTGCTGCCAAGGCCCGGCGGGAACAGGA
GCGGGCAGCCAGTGCCTTCCAGGAACTTGACGACAACATGGATGGGATGGTCTCGCTGGCTGAGTTACAGACTCACCCG
GAGCTGGACACAGATGGAGATGGAGCGCTGTCTGAGGAGGAGGCCCAGGCCCTTCTCAGTGGAGACACACAGACTGACA
CCACCTCCTTCTATGACCGTGTCTGGGCTGCCATCAGGGACAAGTACCGCTCTGAGGTCCCGCCCACTGACATACCTGTT
CCGGAGGAGACTGAGCCCAAAGAGGAAAAGCCACCAGTGTTGCCACCCACAGAGGAGGAGGAAGAGGAGGAGGAGGAG
CCAGAAGAAGAGGAGGAGGAAGAGGAAGAGGAGGAGGAGGCTCCGCCCCCACTGCAGCCCCCACAGCCTCCCAGCCCC
ACAGAGGATGAGAAGATGCCGCCCTATGATGAGGAGACCCAGGCCATCATCGATGCTGCACAGGAGGCCCGGAGTAAGT
TTGAGGAAGTCGAACGGTCCTTGAAAGAGATGGAAGAGTCCATCAGGAGTTTGGAACAAGAGATCTCCTTTGATTTCGGTC
CCTCTGGAGAGTTTGCATATCTCTACAGCCAATGCTACGAACTCACCACCAATGAGTACGTCTACCGGCTTTGCCCCTTCA
AACTGGTCTCCCAGAAACCCAAACATGGGGGCTCCCCGACCAGCCTGGGCACATGGGGCTCCTGGGCTGGCCCTGATCA
TGACAAGTTCAGTGCGATGAAGTACGAGCAGGGCACGGGCTGTTGGCAGGGCCCCAACCGATCCACCACAGTGCGCCTG
CTGTGTGGCAAAGAGACTGTGGTGACCAGCACCACGGAGCCCAGTCGCTGTGAGTACCTCATGGAGCTGATGACACCAG
CAGCCTGCCCAGAGCCGCCACCAGAAGCACCCAGTGATGGGGACTGATAA

Translation:

MGILPSPGMPALLSLVSLLSVLLMGCVAETGDYKDDDDKGAVEVKRPRGVSLSNHHFYEESKPFTCLDGTATIPFDQVNDDYCD
CKDGSDEPGTAACPNGSFHCTNTGYKPLYIPSNRVNDGVCDCCDGTDEYNSGTVCENTCREKGRKEKESLQQLAEVTREGFR
LKKILIEEWKTAREEKQSKLLELQAGKKSLEDQVETLRAAKEEAERPEKEAKDQHRKLWEEQQAAAKARREQERAASAFQELDD
NMDGMVSLAELQTHPELDTDGDGALSEEEAQALLSGDTQTDTTSFYDRVWAAIRDKYRSEVPPTDIPVPEETEPKEEKPPVLPP
TEEEEEEEEEPEEEEEEEEEEEAPPPLQPPQPPSPTEDEKMPPYDEETQAIIDAAQEARSKFEEVERSLKEMEESIRSLEQEISF
DFGPSGEFAYLYSQCYELTTNEYVYRLCPFKLVSQKPKHGGSPTSLGTWGSWAGPDHDKFSAMKYEQGTGCWQGPNRSTTV
RLLCGKETVVTSTTEPSRCEYLMELMTPAACPEPPPEAPSDGD**

9.3.2 Cysteine mutants of Humanised *GANAB*

The following insert sequences, designed by myself based upon the humanised *GANAB* sequence (9.2.1) and synthesised by Twist Bioscience using the pOPING expression vector.

Red-Organge – pOPING secretion signal. Red peptide is removed during protein processing in ER/Golgi, so is not present in the purified protein.

Yellow – Glycine spacer introduced during cloning to maintain reading frame.

Green – His_6-tag

Cyan – Cys SDM

hGANAB – C19/25A:

ATGGGGATCCTTCCCAGCCCTGGGATGCCTGCGCTGCTCTCCCTCGTGAGCCTTCTCTCCGTGCTGCTGATGGGTTGCGT
AGCTGAAACCGGTGGCCTTGCTGTGGATAGAAGCAACTTTAAGACCGCTGATGAGAGTTCCTTTGCCCAAACGGCAGCGAA
GCATTCGGCCAGGCCTCTCTCCTTACCGTGCCTTGCTGGACACTCTGCAGCTTGGTCCTGATGCTCTTACAGTCCATCTGA
TCCATGAAGTCACCAAGGTGCTGCTTGTGCTGGAACTCCAGGGCCTTCAGAAGAACATGACTCGGATCAGGATCGATGAG
CTAGAGCCCCGGCGGCCTCGATACCGAGTGCCAGATGTTTTAGTGGCTGACCCCCCCACAGCTAGGCTTTCAGTCTCTGG
CCGTGATGACAACAGTGTGGAGCTAACAGTGGCTGAGGGACCCTACAAAATCATTTTGACAGCACAGCCATTCCGCCTTG
ACCTGCTAGAAGATCGAAGCCTCCTGCTCAGTGTCAATGCCCGAGGACTTATGGCCTTTGAGCACCAGAGGGCCCCCAGG
GTCCCTTTCTCGGATAAAGTTAGTCTCGCGCTCGGTAGCGTGTGGGATAAGATCAAGAACCTTTTCTCTAGGCAAGAATCA
AAAGACCCAGCTGAAGGCAATGGAGCCCAGCCTGAAGCAACACCTGGGGATGGTGACAAGCCAGAGGAGACCCAGGAAA
AGGCTGAGAAGGATGAGCCAGGAGCCTGGGAGGAGACATTCAAAACACATTCTGACAGCAAGCCTTATGGCCCCACGTCT
GTAGGTTTGGACTTTTCTCTGCCAGGAATGGAACATGTGTATGGGATCCCTGAGCATGCTGACAGCCTGAGACTGAAGGT
CACTGAGGGCGGTGAGCCGTACCGCCTGTACAATTTGGATGTGTTCCAGTATGAGCTGAACAACCCAATGGCTCTATATG
GGTCTGTGCCTGTGCTCCTGGCACACAGCTTTCATCGAGACCTGGGCATCTTCTGGCTTAATGCTGCTGAGACTTGGGTT
GATATATCCTCCAACACGGCTGGGAAGACCCTGTTTGGGAAGATGCTTGATTACCTGCAGGGCTCTGGGGAGACTCCACA
GACAGACATTCGTTGGATGTCAGAGAGTGGCATTATTGATGTTTTCCTAATGCTTGGCCCTTCGGTTTTTGATGTCTTTAGG
CAGTATGCTAGTCTCACAGGGACCCAGGCATTGCCCCCACTCTTCTCCCTCGGCTATCACCAGAGTCGCTGGAACTACCG
GGATGAGGCTGATGTTTTGGAAGTGGATCAGGGTTTTGATGATCACAACATGCCTTGTGATGTCATTTGGTTGGACATTGA
ACATGCTGATGGCAAGCGGTACTTCACTTGGGACCCCACCCGATTTCCTCAGCCCCTCAATATGCTTGAGCACTTGGCTTC
CAAGAGGCGGAAGCTGGTGGCCATTGTGGACCCCCACATCAAGGTAGACTCTGGCTACCGAGTTCACGAAGAATTGCGAA
ACCACGGGCTGTATGTTAAAACTCGGGATGGCTCTGATTACGAGGGCTGGTGCTGGCCAGGCTCAGCTAGTTACCCTGAC
TTCACTAATCCAAGGATGAGGGCCTGGTGGTCTAACATGTTCAGCTTTGACAATTATGAGGGTTCAGCTCCTAATCTTTATG
TTTGGAATGACATGAATGAACCGTCTGTGTTCAATGGTCCTGAGGTCACCATGTTGAAGGATGCTGTGCATTATGGAGGCT
GGGAGCACCGGGACATCCATAACATCTATGGCTTATATGTGCACATGGCGACTGCTGATGGGCTAATACAGCGCTCTGGG
GGCATAGAGCGTCCCTTTGTCCTGAGTAGGGCTTTCTTCTCAGGCTCCCAGCGCTTTGGAGCTGTGTGGACAGGGGACAA
CACTGCCGAATGGGATCATTTGAAGATCTCTATCCCTATGTGTCTCAGCCTGGCACTGGTGGGGCTTTCCTTCTGTGGAGC
GGATGTGGGTGGCTTCTTCAAGAACCCAGAGCCAGAGCTGCTTGTGCGCTGGTACCAAATGGGTGCCTACCAGCCGTTCT
TTCGGGCTCATGCCCACTTGGACACTGGGCGGCGAGAGCCCTGGCTGTTAGCGTCTCAATACCAAGATGCAATCCGAGAT
GCCTTGGGCCAGCGATATTCTTTGCTGCCCTTCTGGTATACCCTCTTCTATCAAGCTCACAAGGAAGGGTTTCCTGTCATG
AGGCCCCTCTGGGTACAGTATCCTGAGGATATGTCTACCTTCAGTATAGAGGATCAGTTCATGCTTGGTGATGCACTCCTT
ATTCACCCTGTATCGGATGCTGGGGCCCACGGAGTGCAGGTCTATTTGCCTGGCCAAGAAGAAGTGTGGTATGACATTCA
GAGCTATCAGAAGCATCATGGGCCCCAGACCTTGTATCTGCCAGTAACTTTGAGCAGTATCCCCGTGTTCCAGCGTGGCG
GAACCATTGTGCCTCGATGGATGCGTGTGAGGCGCTCTTCAGACTGTATGAAGGACGATCCTATCACTCTCTTTGTTGCTC
TCAGTCCCCAGGGTACTGCCCAAGGAGAACTCTTTCTAGATGATGGACACACCTTTAACTATCAGACTCGCCATGAGTTCC
TGTTGCGGCGGTTCTCTTTCTCTGGCAGCACACTGGTCTCTAGTTCAGCAGACCCCAAAGGCCACCTTGAGACACCTATTT
GGATTGAGCGAGTAGTCATCATGGGGGCTGGAAAGCCAGCAGCTGTGGTGCTCCAGACCAAAGGATCCCCTGAAAGTCG
CCTGTCCTTCCAGCATGACCCTGAGACCTCAGTGCTGATATTGCGTAAACCTGGTGTCAGCGTGGCATCCGACTGGAGTA
TTCATCTTCGAGGTACCAAGCACCACCATCACCATCACTAATGA

Translation:

MGILPSPGMPALLSLVSLLSVLLMGCVAETGGLAVDRSNFKTADESSFAKRQRSIRPGLSPYRALLDTLQLGPDALTVHLIHEVTK
VLLVLELQGLQKNMTRIRIDELEPRRPRYRVPDVLVADPPTARLSVSGRDDNSVELTVAEGPYKIILTAQPFRLDLLEDRSLLLSVN
ARGLMAFEHQRAPRVPFSDKVSLALGSVWDKIKNLFSRQESKDPAEGNGAQPEATPGDGDKPEETQEKAEKDEPGAWEETFK
THSDSKPYGPTSVGLDFSLPGMEHVYGIPEHADSLRLKVTEGGEPYRLYNLDVFQYELNNPMALYGSVPVLLAHSFHRDLGIFW
LNAAETWVDISSNTAGKTLFGKMLDYLQGSGETPQTDIRWMSESGIIDVFLMLGPSVFDVFRQYASLTGTQALPPLFSLGYHQS
RWNYRDEADVLEVDQGFDDHNMPCDVIWLDIEHADGKRYFTWDPTRFPQPLNMLEHLASKRRKLVAIVDPHIKVDSGYRVHEE
LRNHGLYVKTRDGSDYEGWCWPGSASYPDFTNPRMRAWWSNMFSFDNYEGSAPNLYVWNDMNEPSVFNGPEVTMLKDAVH
YGGWEHRDIHNIYGLYVHMATADGLIQRSGGIERPFVLSRAFFSGSQRFGAVWTGDNTAEWDHLKISIPMCLSLALVGLSFCGA
DVGGFFKNPEPELLVRWYQMGAYQPFFRAHAHLDTGRREPWLLASQYQDAIRDALGQRYSLLPFWYTLFYQAHKEGFPVMRP
LWVQYPEDMSTFSIEDQFMLGDALLIHPVSDAGAHGVQVYLPGQEEVWYDIQSYQKHHGPQTLYLPVTLSSIPVFQRGGTIVPR
WMRVRRSSDCMKDDPITLFVALSPQGTAQGELFLDDGHTFNYQTRHEFLLRRFSFSGSTLVSSSADPKGHLETPIWIERVVIMG
AGKPAAVVLQTKGSPESRLSFQHDPETSVLILRKPGVSVASDWSIHLRGTKHHHHHH**

hGANAB – C423A:

ATGGGGATCCTTCCCAGCCCTGGGATGCCTGCGCTGCTCTCCCTCGTGAGCCTTCTCTCCGTGCTGCTGATGGGTTGCGT
AGCTGAAACCGGTGGCCTTGCTGTGGATAGAAGCAACTTTAAGACCTGTGATGAGAGTTCCTTTTGCAAACGGCAGCGAA
GCATTCGGCCAGGCCTCTCTCCTTACCGTGCCTTGCTGGACACTCTGCAGCTTGGTCCTGATGCTCTTACAGTCCATCTGA
TCCATGAAGTCACCAAGGTGCTGCTTGTGCTGGAACTCCAGGGCCTTCAGAAGAACATGACTCGGATCAGGATCGATGAG
CTAGAGCCCCGGCGGCCTCGATACCGAGTGCCAGATGTTTTAGTGGCTGACCCCCCCACAGCTAGGCTTTCAGTCTCTGG
CCGTGATGACAACAGTGTGGAGCTAACAGTGGCTGAGGGACCCTACAAAATCATTTTGACAGCACAGCCATTCCGCCTTG
ACCTGCTAGAAGATCGAAGCCTCCTGCTCAGTGTCAATGCCCGAGGACTTATGGCCTTTGAGCACCAGAGGGCCCCCAGG
GTCCCTTTCTCGGATAAAGTTAGTCTCGCGCTCGGTAGCGTGTGGGATAAGATCAAGAACCTTTTCTCTAGGCAAGAATCA
AAAGACCCAGCTGAAGGCAATGGAGCCCAGCCTGAAGCAACACCTGGGGATGGTGACAAGCCAGAGGAGACCCAGGAAA
AGGCTGAGAAGGATGAGCCAGGAGCCTGGGAGGAGACATTCAAAACACATTCTGACAGCAAGCCTTATGGCCCCACGTCT
GTAGGTTTGGACTTTTCTCTGCCAGGAATGGAACATGTGTATGGGATCCCTGAGCATGCTGACAGCCTGAGACTGAAGGT
CACTGAGGGCGGTGAGCCGTACCGCCTGTACAATTTGGATGTGTTCCAGTATGAGCTGAACAACCCAATGGCTCTATATG
GGTCTGTGCCTGTGCTCCTGGCACACAGCTTTCATCGAGACCTGGGCATCTTCTGGCTTAATGCTGCTGAGACTTGGGTT
GATATATCCTCCAACACGGCTGGGAAGACCCTGTTTGGGAAGATGCTTGATTACCTGCAGGGCTCTGGGGAGACTCCACA
GACAGACATTCGTTGGATGTCAGAGAGTGGCATTATTGATGTTTTCCTAATGCTTGGCCCTTCGGTTTTTGATGTCTTTAGG
CAGTATGCTAGTCTCACAGGGACCCAGGCATTGCCCCCACTCTTCTCCCTCGGCTATCACCAGAGTCGCTGGAACTACCG
GGATGAGGCTGATGTTTTGGAAGTGGATCAGGGTTTTGATGATCACAACATGCCTGCTGATGTCATTTGGTTGGACATTGA
ACATGCTGATGGCAAGCGGTACTTCACTTGGGACCCCACCCGATTTCCTCAGCCCCTCAATATGCTTGAGCACTTGGCTTC
CAAGAGGCGGAAGCTGGTGGCCATTGTGGACCCCCACATCAAGGTAGACTCTGGCTACCGAGTTCACGAAGAATTGCGAA
ACCACGGGCTGTATGTTAAAACTCGGGATGGCTCTGATTACGAGGGCTGGTGCTGGCCAGGCTCAGCTAGTTACCCTGAC
TTCACTAATCCAAGGATGAGGGCCTGGTGGTCTAACATGTTCAGCTTTGACAATTATGAGGGTTCAGCTCCTAATCTTTATG
TTTGGAATGACATGAATGAACCGTCTGTGTTCAATGGTCCTGAGGTCACCATGTTGAAGGATGCTGTGCATTATGGAGGCT
GGGAGCACCGGGACATCCATAACATCTATGGCTTATATGTGCACATGGCGACTGCTGATGGGCTAATACAGCGCTCTGGG
GGCATAGAGCGTCCCTTTGTCCTGAGTAGGGCTTTCTTCTCAGGCTCCCAGCGCTTTGGAGCTGTGTGGACAGGGGACAA
CACTGCCGAATGGGATCATTTGAAGATCTCTATCCCTATGTGTCTCAGCCTGGCACTGGTGGGGCTTTCCTTCTGTGGAGC
GGATGTGGGTGGCTTCTTCAAGAACCCAGAGCCAGAGCTGCTTGTGCGCTGGTACCAAATGGGTGCCTACCAGCCGTTCT
TTCGGGCTCATGCCCACTTGGACACTGGGCGGCGAGAGCCCTGGCTGTTAGCGTCTCAATACCAAGATGCAATCCGAGAT
GCCTTGGGCCAGCGATATTCTTTGCTGCCCTTCTGGTATACCCTCTTCTATCAAGCTCACAAGGAAGGGTTTCCTGTCATG
AGGCCCCTCTGGGTACAGTATCCTGAGGATATGTCTACCTTCAGTATAGAGGATCAGTTCATGCTTGGTGATGCACTCCTT
ATTCACCCTGTATCGGATGCTGGGGCCCACGGAGTGCAGGTCTATTTGCCTGGCCAAGAAGAAGTGTGGTATGACATTCA
GAGCTATCAGAAGCATCATGGGCCCCAGACCTTGTATCTGCCAGTAACTTTGAGCAGTATCCCCGTGTTCCAGCGTGGCG
GAACCATTGTGCCTCGATGGATGCGTGTGAGGCGCTCTTCAGACTGTATGAAGGACGATCCTATCACTCTCTTTGTTGCTC
TCAGTCCCCAGGGTACTGCCCAAGGAGAACTCTTTCTAGATGATGGACACACCTTTAACTATCAGACTCGCCATGAGTTCC
TGTTGCGGCGGTTCTCTTTCTCTGGCAGCACACTGGTCTCTAGTTCAGCAGACCCCAAAGGCCACCTTGAGACACCTATTT
GGATTGAGCGAGTAGTCATCATGGGGGCTGGAAAGCCAGCAGCTGTGGTGCTCCAGACCAAAGGATCCCCTGAAAGTCG
CCTGTCCTTCCAGCATGACCCTGAGACCTCAGTGCTGATATTGCGTAAACCTGGTGTCAGCGTGGCATCCGACTGGAGTA
TTCATCTTCGAGGTACCAAGCACCACCATCACCATCACTAATGA

Translation:

MGILPSPGMPALLSLVSLLSVLLMGCVAETGGLAVDRSNFKTCDESSFCKRQRSIRPGLSPYRALLDTLQLGPDALTVHLIHEVTK
VLLVLELQGLQKNMTRIRIDELEPRRPRYRVPDVLVADPPTARLSVSGRDDNSVELTVAEGPYKIILTAQPFRLDLLEDRSLLLSVN
ARGLMAFEHQRAPRVPFSDKVSLALGSVWDKIKNLFSRQESKDPAEGNGAQPEATPGDGDKPEETQEKAEKDEPGAWEETFK
THSDSKPYGPTSVGLDFSLPGMEHVYGIPEHADSLRLKVTEGGEPYRLYNLDVFQYELNNPMALYGSVPVLLAHSFHRDLGIFW
LNAAETWVDISSNTAGKTLFGKMLDYLQGSGETPQTDIRWMSESGIIDVFLMLGPSVFDVFRQYASLTGTQALPPLFSLGYHQS
RWNYRDEADVLEVDQGFDDHNMPADVIWLDIEHADGKRYFTWDPTRFPQPLNMLEHLASKRRKLVAIVDPHIKVDSGYRVHEE

LRNHGLYVKTRDGSDYEGWCWPGSASYPDFTNPRMRAWWSNMFSFDNYEGSAPNLYVWNDMNEPSVFNGPEVTMLKDAVH
YGGWEHRDIHNIYGLYVHMATADGLIQRSGGIERPFVLSRAFFSGSQRFGAVWTGDNTAEWDHLKISIPMCLSLALVGLSFCGA
DVGGFFKNPEPELLVRWYQMGAYQPFFRAHAHLDTGRREPWLLASQYQDAIRDALGQRYSLLPFWYTLFYQAHKEGFPVMRP
LWVQYPEDMSTFSIEDQFMLGDALLIHPVSDAGAHGVQVYLPGQEEVWYDIQSYQKHHGPQTLYLPVTLSSIPVFQRGGTIVPR
WMRVRRSSDCMKDDPITLFVALSPQGTAQGELFLDDGHTFNYQTRHEFLLRRFSFSGSTLVSSSADPKGHLETPIWIERVVIMG
AGKPAAVVLQTKGSPESRLSFQHDPETSVLILRKPGVSVASDWSIHLRGTKHHHHHH**

hGANAB – C502A:

ATGGGTTGCGTAGCTGAAACCGGTGGCCTTGCTGTGGATAGAAGCAACTTTAAGACCTGTGATGAGAGTTCCTTTTGCAAA
CGGCAGCGAAGCATTCGGCCAGGCCTCTCTCCTTACCGTGCCTTGCTGGACACTCTGCAGCTTGGTCCTGATGCTCTTAC
AGTCCATCTGATCCATGAAGTCACCAAGGTGCTGCTTGTGCTGGAACTCCAGGGCCTTCAGAAGAACATGACTCGGATCA
GGATCGATGAGCTAGAGCCCCGGCGGCCTCGATACCGAGTGCCAGATGTTTTAGTGGCTGACCCCCCCACAGCTAGGCT
TTCAGTCTCTGGCCGTGATGACAACAGTGTGGAGCTAACAGTGGCTGAGGGACCCTACAAAATCATTTTGACAGCACAGC
CATTCCGCCTTGACCTGCTAGAAGATCGAAGCCTCCTGCTCAGTGTCAATGCCCGAGGACTTATGGCCTTTGAGCACCAG
AGGGCCCCCAGGGTCCCTTTCTCGGATAAAGTTAGTCTCGCGCTCGGTAGCGTGTGGGATAAGATCAAGAACCTTTTCTCT
AGGCAAGAATCAAAAGACCCAGCTGAAGGCAATGGAGCCCAGCCTGAAGCAACACCTGGGGATGGTGACAAGCCAGAGG
AGACCCAGGAAAAGGCTGAGAAGGATGAGCCAGGAGCCTGGGAGGAGACATTCAAAACACATTCTGACAGCAAGCCTTAT
GGCCCCACGTCTGTAGGTTTGGACTTTTCTCTGCCAGGAATGGAACATGTGTATGGGATCCCTGAGCATGCTGACAGCCT
GAGACTGAAGGTCACTGAGGGCGGTGAGCCGTACCGCCTGTACAATTTGGATGTGTTCCAGTATGAGCTGAACAACCCAA
TGGCTCTATATGGGTCTGTGCCTGTGCTCCTGGCACACAGCTTTCATCGAGACCTGGGCATCTTCTGGCTTAATGCTGCTG
AGACTTGGGTTGATATATCCTCCAACACGGCTGGGAAGACCCTGTTTGGGAAGATGCTTGATTACCTGCAGGGCTCTGGG
GAGACTCCACAGACAGACATTCGTTGGATGTCAGAGAGTGGCATTATTGATGTTTTCCTAATGCTTGGCCCTTCGGTTTTTG
ATGTCTTTAGGCAGTATGCTAGTCTCACAGGGACCCAGGCATTGCCCCCACTCTTCTCCCTCGGCTATCACCAGAGTCGCT
GGAACTACCGGGATGAGGCTGATGTTTTGGAAGTGGATCAGGGTTTTGATGATCACAACATGCCTTGTGATGTCATTTGGT
TGGACATTGAACATGCTGATGGCAAGCGGTACTTCACTTGGGACCCCACCCGATTTCCTCAGCCCCTCAATATGCTTGAGC
ACTTGGCTTCCAAGAGGCGGAAGCTGGTGGCCATTGTGGACCCCCACATCAAGGTAGACTCTGGCTACCGAGTTCACGAA
GAATTGCGAAACCACGGGCTGTATGTTAAAACTCGGGATGGCTCTGATTACGAGGGCTGGGCCTGGCCAGGCTCAGCTAG
TTACCCTGACTTCACTAATCCAAGGATGAGGGCCTGGTGGTCTAACATGTTCAGCTTTGACAATTATGAGGGTTCAGCTCC
TAATCTTTATGTTTGGAATGACATGAATGAACCGTCTGTGTTCAATGGTCCTGAGGTCACCATGTTGAAGGATGCTGTGCAT
TATGGAGGCTGGGAGCACCGGGACATCCATAACATCTATGGCTTATATGTGCACATGGCGACTGCTGATGGGCTAATACA
GCGCTCTGGGGGCATAGAGCGTCCCTTTGTCCTGAGTAGGGCTTTCTTCTCAGGCTCCCAGCGCTTTGGAGCTGTGTGGA
CAGGGGACAACACTGCCGAATGGGATCATTTGAAGATCTCTATCCCTATGTGTCTCAGCCTGGCACTGGTGGGGCTTTCCT
TCTGTGGAGCGGATGTGGGTGGCTTCTTCAAGAACCCAGAGCCAGAGCTGCTTGTGCGCTGGTACCAAATGGGTGCCTAC
CAGCCGTTCTTTCGGGCTCATGCCCACTTGGACACTGGGCGGCGAGAGCCCTGGCTGTTAGCGTCTCAATACCAAGATGC
AATCCGAGATGCCTTGGGCCAGCGATATTCTTTGCTGCCCTTCTGGTATACCCTCTTCTATCAAGCTCACAAGGAAGGGTT
TCCTGTCATGAGGCCCCTCTGGGTACAGTATCCTGAGGATATGTCTACCTTCAGTATAGAGGATCAGTTCATGCTTGGTGA
TGCACTCCTTATTCACCCTGTATCGGATGCTGGGGCCCACGGAGTGCAGGTCTATTTGCCTGGCCAAGAAGAAGTGTGGT
ATGACATTCAGAGCTATCAGAAGCATCATGGGCCCCAGACCTTGTATCTGCCAGTAACTTTGAGCAGTATCCCCGTGTTCC
AGCGTGGCGGAACCATTGTGCCTCGATGGATGCGTGTGAGGCGCTCTTCAGACTGTATGAAGGACGATCCTATCACTCTC
TTTGTTGCTCTCAGTCCCCAGGGTACTGCCCAAGGAGAACTCTTTCTAGATGATGGACACACCTTTAACTATCAGACTCGC
CATGAGTTCCTGTTGCGGCGGTTCTCTTTCTCTGGCAGCACACTGGTCTCTAGTTCAGCAGACCCCAAAGGCCACCTTGAG
ACACCTATTTGGATTGAGCGAGTAGTCATCATGGGGGCTGGAAAGCCAGCAGCTGTGGTGCTCCAGACCAAAGGATCCCC
TGAAAGTCGCCTGTCCTTCCAGCATGACCCTGAGACCTCAGTGCTGATATTGCGTAAACCTGGTGTCAGCGTGGCATCCG
ACTGGAGTATTCATCTTCGAGGTACCAAGCACCACCATCACCATCACTAATGA

Translation:

MGILPSPGMPALLSLVSLLSVLLMGCVAETGGLAVDRSNFKTCDESSFCKRQRSIRPGLSPYRALLDTLQLGPDALTVHLIHEVTK
VLLVLELQGLQKNMTRIRIDELEPRRPRYRVPDVLVADPPTARLSVSGRDDNSVELTVAEGPYKIILTAQPFRLDLLEDRSLLLSVN
ARGLMAFEHQRAPRVPFSDKVSLALGSVWDKIKNLFSRQESKDPAEGNGAQPEATPGDGDKPEETQEKAEKDEPGAWEETFK
THSDSKPYGPTSVGLDFSLPGMEHVYGIPEHADSLRLKVTEGGEPYRLYNLDVFQYELNNPMALYGSVPVLLAHSFHRDLGIFW
LNAAETWVDISSNTAGKTLFGKMLDYLQGSGETPQTDIRWMSESGIIDVFLMLGPSVFDVFRQYASLTGTQALPPLFSLGYHQS
RWNYRDEADVLEVDQGFDDHNMPCDVIWLDIEHADGKRYFTWDPTRFPQPLNMLEHLASKRRKLVAIVDPHIKVDSGYRVHEE
LRNHGLYVKTRDGSDYEGWAWPGSASYPDFTNPRMRAWWSNMFSFDNYEGSAPNLYVWNDMNEPSVFNGPEVTMLKDAVH
YGGWEHRDIHNIYGLYVHMATADGLIQRSGGIERPFVLSRAFFSGSQRFGAVWTGDNTAEWDHLKISIPMCLSLALVGLSFCGA
DVGGFFKNPEPELLVRWYQMGAYQPFFRAHAHLDTGRREPWLLASQYQDAIRDALGQRYSLLPFWYTLFYQAHKEGFPVMRP
LWVQYPEDMSTFSIEDQFMLGDALLIHPVSDAGAHGVQVYLPGQEEVWYDIQSYQKHHGPQTLYLPVTLSSIPVFQRGGTIVPR
WMRVRRSSDCMKDDPITLFVALSPQGTAQGELFLDDGHTFNYQTRHEFLLRRFSFSGSTLVSSSADPKGHLETPIWIERVVIMG
AGKPAAVVLQTKGSPESRLSFQHDPETSVLILRKPGVSVASDWSIHLRGTKHHHHHH**

hGANAB – C822A:

ATGGGGATCCTTCCCAGCCCTGGGATGCCTGCGCTGCTCTCCCTCGTGAGCCTTCTCTCCGTGCTGCTGATGGGTTGCGT
AGCTGAAACCGGTACCGGCCTTGCTGTGGATAGAAGCAACTTTAAGACCTGTGATGAGAGTTCCTTTTGCAAACGGCAGC
GAAGCATTCGGCCAGGCCTCTCTCCTTACCGTGCCTTGCTGGACACTCTGCAGCTTGGTCCTGATGCTCTTACAGTCCATC
TGATCCATGAAGTCACCAAGGTGCTGCTTGTGCTGGAACTCCAGGGCCTTCAGAAGAACATGACTCGGATCAGGATCGAT
GAGCTAGAGCCCCGGCGGCCTCGATACCGAGTGCCAGATGTTTTAGTGGCTGACCCCCCCACAGCTAGGCTTTCAGTCTC
TGGCCGTGATGACAACAGTGTGGAGCTAACAGTGGCTGAGGGACCCTACAAAATCATTTTGACAGCACAGCCATTCCGCC
TTGACCTGCTAGAAGATCGAAGCCTCCTGCTCAGTGTCAATGCCCGAGGACTTATGGCCTTTGAGCACCAGAGGGCCCCC
AGGGTCCCTTTCTCGGATAAAGTTAGTCTCGCGCTCGGTAGCGTGTGGGATAAGATCAAGAACCTTTTCTCTAGGCAAGAA
TCAAAAGACCCAGCTGAAGGCAATGGAGCCCAGCCTGAAGCAACACCTGGGGATGGTGACAAGCCAGAGGAGACCCAGG
AAAAGGCTGAGAAGGATGAGCCAGGAGCCTGGGAGGAGACATTCAAAACACATTCTGACAGCAAGCCTTATGGCCCCACG
TCTGTAGGTTTGGACTTTTCTCTGCCAGGAATGGAACATGTGTATGGGATCCCTGAGCATGCTGACAGCCTGAGACTGAAG
GTCACTGAGGGCGGTGAGCCGTACCGCCTGTACAATTTGGATGTGTTCCAGTATGAGCTGAACAACCCAATGGCTCTATAT
GGGTCTGTGCCTGTGCTCCTGGCACACAGCTTTCATCGAGACCTGGGCATCTTCTGGCTTAATGCTGCTGAGACTTGGGT
TGATATATCCTCCAACACGGCTGGGAAGACCCTGTTTGGGAAGATGCTTGATTACCTGCAGGGCTCTGGGGAGACTCCAC
AGACAGACATTCGTTGGATGTCAGAGAGTGGCATTATTGATGTTTTCCTAATGCTTGGCCCTTCGGTTTTTGATGTCTTTAG
GCAGTATGCTAGTCTCACAGGGACCCAGGCATTGCCCCCACTCTTCTCCCTCGGCTATCACCAGAGTCGCTGGAACTACC
GGGATGAGGCTGATGTTTTGGAAGTGGATCAGGGTTTTGATGATCACAACATGCCTTGTGATGTCATTTGGTTGGACATTG
AACATGCTGATGGCAAGCGGTACTTCACTTGGGACCCCACCCGATTTCCTCAGCCCCTCAATATGCTTGAGCACTTGGCTT
CCAAGAGGCGGAAGCTGGTGGCCATTGTGGACCCCCACATCAAGGTAGACTCTGGCTACCGAGTTCACGAAGAATTGCGA
AACCACGGGCTGTATGTTAAAACTCGGGATGGCTCTGATTACGAGGGCTGGTGCTGGCCAGGCTCAGCTAGTTACCCTGA
CTTCACTAATCCAAGGATGAGGGCCTGGTGGTCTAACATGTTCAGCTTTGACAATTATGAGGGTTCAGCTCCTAATCTTTAT
GTTTGGAATGACATGAATGAACCGTCTGTGTTCAATGGTCCTGAGGTCACCATGTTGAAGGATGCTGTGCATTATGGAGGC
TGGGAGCACCGGGACATCCATAACATCTATGGCTTATATGTGCACATGGCGACTGCTGATGGGCTAATACAGCGCTCTGG
GGGCATAGAGCGTCCCTTTGTCCTGAGTAGGGCTTTCTTCTCAGGCTCCCAGCGCTTTGGAGCTGTGTGGACAGGGGACA
ACACTGCCGAATGGGATCATTTGAAGATCTCTATCCCTATGTGTCTCAGCCTGGCACTGGTGGGGCTTTCCTTCTGTGGAG
CGGATGTGGGTGGCTTCTTCAAGAACCCAGAGCCAGAGCTGCTTGTGCGCTGGTACCAAATGGGTGCCTACCAGCCGTTC
TTTCGGGCTCATGCCCACTTGGACACTGGGCGGCGAGAGCCCTGGCTGTTAGCGTCTCAATACCAAGATGCAATCCGAGA
TGCCTTGGGCCAGCGATATTCTTTGCTGCCCTTCTGGTATACCCTCTTCTATCAAGCTCACAAGGAAGGGTTTCCTGTCAT
GAGGCCCCTCTGGGTACAGTATCCTGAGGATATGTCTACCTTCAGTATAGAGGATCAGTTCATGCTTGGTGATGCACTCCT
TATTCACCCTGTATCGGATGCTGGGGCCCACGGAGTGCAGGTCTATTTGCCTGGCCAAGAAGAAGTGTGGTATGACATTC
AGAGCTATCAGAAGCATCATGGGCCCCAGACCTTGTATCTGCCAGTAACTTTGAGCAGTATCCCCGTGTTCCAGCGTGGC
GGAACCATTGTGCCTCGATGGATGCGTGTGAGGCGCTCTTCAGACGCTATGAAGGACGATCCTATCACTCTCTTTGTTGCT
CTCAGTCCCCAGGGTACTGCCCAAGGAGAACTCTTTCTAGATGATGGACACACCTTTAACTATCAGACTCGCCATGAGTTC
CTGTTGCGGCGGTTCTCTTTCTCTGGCAGCACACTGGTCTCTAGTTCAGCAGACCCCAAAGGCCACCTTGAGACACCTATT
TGGATTGAGCGAGTAGTCATCATGGGGGCTGGAAAGCCAGCAGCTGTGGTGCTCCAGACCAAAGGATCCCCTGAAAGTC
GCCTGTCCTTCCAGCATGACCCTGAGACCTCAGTGCTGATATTGCGTAAACCTGGTGTCAGCGTGGCATCCGACTGGAGT
ATTCATCTTCGAGGTACCAAGCATCACCATCACCATCACTAATGA

Translation:

MGILPSPGMPALLSLVSLLSVLLMGCVAETGTGLAVDRSNFKTCDESSFCKRQRSIRPGLSPYRALLDTLQLGPDALTVHLIHEVT
KVLLVLELQGLQKNMTRIRIDELEPRRPRYRVPDVLVADPPTARLSVSGRDDNSVELTVAEGPYKIILTAQPFRLDLLEDRSLLLSV
NARGLMAFEHQRAPRVPFSDKVSLALGSVWDKIKNLFSRQESKDPAEGNGAQPEATPGDGDKPEETQEKAEKDEPGAWEETF
KTHSDSKPYGPTSVGLDFSLPGMEHVYGIPEHADSLRLKVTEGGEPYRLYNLDVFQYELNNPMALYGSVPVLLAHSFHRDLGIF
WLNAAETWVDISSNTAGKTLFGKMLDYLQGSGETPQTDIRWMSESGIIDVFLMLGPSVFDVFRQYASLTGTQALPPLFSLGYHQ
SRWNYRDEADVLEVDQGFDDHNMPCDVIWLDIEHADGKRYFTWDPTRFPQPLNMLEHLASKRRKLVAIVDPHIKVDSGYRVHE
ELRNHGLYVKTRDGSDYEGWCWPGSASYPDFTNPRMRAWWSNMFSFDNYEGSAPNLYVWNDMNEPSVFNGPEVTMLKDAV
HYGGWEHRDIHNIYGLYVHMATADGLIQRSGGIERPFVLSRAFFSGSQRFGAVWTGDNTAEWDHLKISIPMCLSLALVGLSFCG
ADVGGFFKNPEPELLVRWYQMGAYQPFFRAHAHLDTGRREPWLLASQYQDAIRDALGQRYSLLPFWYTLFYQAHKEGFPVMR
PLWVQYPEDMSTFSIEDQFMLGDALLIHPVSDAGAHGVQVYLPGQEEVWYDIQSYQKHHGPQTLYLPVTLSSIPVFQRGGTIVP
RWMRVRRSSDAMKDDPITLFVALSPQGTAQGELFLDDGHTFNYQTRHEFLLRRFSFSGSTLVSSSADPKGHLETPIWIERVVIM
GAGKPAAVVLQTKGSPESRLSFQHDPETSVLILRKPGVSVASDWSIHLRGTKHHHHHH**

9.4 Materials & Methods – Molecular Biology

9.4.1 Molecular Cloning & DNA Production

All GluII constructs used in this study are based on the ER alpha-glucosidase II heterodimer from *Mus musculus* (*Mm*GIIα UniProt: Q8BHN3, GANAB_mouse, *Mm*GIIβ UniProt: O08795, PRKCSH_mouse).

For the purposes of investigating interface disruptors, *Mm*GluII was 'humanised' (h*Mm*GluII) within the GIIα:GIIβ PPI based on the *Mm*GluII PDB structure (PDB-ID: 5F0E) and the UniProt entries for GANAB_human (*Hs*GIIα, Q14697) and PRKCSH_human (*Hs*GIIβ, P14314). In order to more closely mimic the PPI of the human protein, for which there is currently no structural data on the PDB, selected interface residues in *Mm*GluII were replaced by their equivalent in the human protein. The mutations are as follows: αF724G, βL88P and βS90N. This allowed for observation of a more clinically relevant interface with respect to drug discovery for use in humans, while maintaining the reliability of structural comparisons to existing *Mm*GluII structures (see Section 1.4.2).

Constructs of humanised GIIα and GIIβ in pHLsec were kindly provided by Dr Mario Hensen, a former member of the Zitzmann group. This was used for protein expression in mammalian cells (Section 3) and served as a starting point for further GluII mutagenesis. In this project, I used exclusively h*Mm*GluII for all experiments, unless stated explicitly otherwise. Residue numbering for h*Mm*GluII differs from that of *Mm*GluII$_{Tryp}$, explaining any discrepancy for equivalent residues.

9.4.2 Subcloning h*Mm*GluII into Cloning Vector

In an attempt to ease the site directed mutagenesis of the GluII gene, the inserts from the hGIIα/hGIIβ in pHLsec constructs were first subcloned into the cloning vector Litmus28i, as

the shorter length may improve the cloning efficiency. Amplification of the target genes was done by polymerase chain reaction (PCR) using the primer pairs listed in Table II. These primers were designed using the New England Biolabs (NEB) NEBtools™ NEBuilder® assembly tool and incorporate an extension sequence for insertion into the destination vector. Solutions were mixed according to the 'Q5 High-Fidelity DNA Polymerase' (M0491) protocol provided by NEB and the reaction mix incubated in a Proflex PCR system (Applied Biosystems) with thermocycling adjusted as stated in the NEB Q5 protocol.

Table II - DNA primers for subcloning of GluII genes into Litmus28i.

Construct	Primer Direction	Sequence (5′-3′)	Tm / °C	Ta / °C
*hMm*GANAB	Forward	AGGGCAGATCTTCGAATGCATCTTGCTGTGGATAGAAGC	59.9	58.6
	Reverse	AGGCCTTGACTAGAGGGTACCTCGAAGATGAATACTCCAG	57.6	

9.4.3 Gibson Assembly

Amplified genes mentioned in Section 9.4.1 were used for Gibson assembly (GA) into linearised destination vectors following the manufacturers' protocol (NEB - NEBuilder® HiFi DNA Assembly Master Mix, Takara Bio - In-Fusion® HD Cloning Kit). Prior to GA, vectors were linearised via restriction digest according to the manufacturer's protocol and the linearised vector purified by gel extraction from an agarose gel. The GA product was transformed into *E. coli* DH5a cells (NEB) following a standard heat shock protocol. Positive colonies were selected by colony PCR, the DNA isolated (QIAprep® Spin Miniprep Kit, standard protocol) and the identity of the DNA verified by sequencing.

9.4.4 DNA Agarose Gels

Agarose gels for analysis and separation of DNA constructs were made of 0.8% w/v agarose in 1X TAE buffer. After microwave-assisted dissolution of the agarose, nucleic acid gel stain was added (GelRed® – Biotium, SYBR™ Safe - Invitrogen™) and gels were poured into moulds. Once cured, samples mixed with 6x Purple loading dye (NEB) were loaded and the gels run at 150 V for 50-60 min in 1X TAE buffer and DNA bands were visualised by UV excitation. For gel extraction to purify separated DNA bands, I followed the protocol from the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel).

9.4.5 Site Directed Mutagenesis

Site directed mutagenesis (SDM) was primarily carried out in Litmus28i. Later attempts to perform SDM of GluII construct directly from expression vectors are outlined in Section 9.4.7. Primers for nucleotide modification were designed using the NEBaseChanger™ tool (NEB). The two-step mutagenesis protocol involves the PCR amplification of the target gene and cloning vector using primers that incorporate the desired mutation (Table III), followed by the circularisation of the linear construct using the KLD reaction mix from NEB. An aliquot of the PCR product (1 µl) was mixed with 10X KLD enzyme mix (1 µl) and 2X KLD reaction buffer (5 µl) and MilliQ water was added to a total of 10 µl. The mixture was incubated for 5 min at RT, followed by 30-60 min at 37˚C and the final product was used for standard heat shock transformation into *E. coli* DH5a cells for vector amplification. Positive colonies were selected by colony PCR, the DNA isolated (QIAprep® Spin Maxiprep Kit, standard protocol) and the identity of the DNA verified by sequencing.

Table III - DNA Primers for site directed mutagenesis of GluII constructs

Construct	Primer Direction	Sequence (5′-3′)	Tm / °C	Ta / °C
*hMm*GANAB C19/25A	Forward	TTCCTTTGCTAAACGGCAGCGAAGCATT	59.9	58.6
	Reverse	CTCTCATCAGCGGTCTTAAAGTTGCTTCTATCC	57.6	
*hMm*GANAB C423A	Forward	CAACATGCCTGCTGATGTCATTTGGTTG	59.9	58.6
	Reverse	TGATCATCAAAACCCTGATC	57.6	
*hMm*GANAB C502A	Forward	CGAGGGCTGGGCTTGGCCAGGCT	59.9	58.6
	Reverse	TAATCAGAGCCATCCCGAG	57.6	
*hMm*GANAB C822A	Forward	CTCTTCAGACGCTATGAAGGACGATCCTATCACTCTCTTTG	59.9	58.6
	Reverse	CGCCTCACACGCATCCAT	57.6	

9.4.6 Subcloning *hMm*GluII into Expression Vector

Using the fully sequenced mutant *hMm*GluII-Litmus28i constructs, the target gene was amplified by PCR using the primers and destination vectors (pHLsec, pOPING). Gibson assembly was then performed..

9.4.7 Direct SDM in Expression Vector Constructs

In parallel with the subcloning strategy outlined above, I also attempted to perform SDM directly on *hMm*GluII-pOPING constructs using the Agilent Technologies QuikChange II XL Site-Directed Mutagenesis Kit and the associated QuikChange Primer Design tool. Forward and reverse primers covering the same sequence on opposite strands of the template DNA both contain the desired mutation flanked on either side by 10-15bp. After mixing these primers with the other components as outlined in the manufacturer's protocol, the DNA was incubated in a Proflex PCR system (Applied Biosystems) with thermocycling adjusted as stated in the Agilent QuikChange II XL protocol. Following this PCR amplification, the reaction

mixture was treated with the *DpnI* restriction enzyme and incubated at 37°C for 1 hour to digest the non-mutated parent construct. The *DpnI*-treated DNA was used to transform XL10-Gold Ultracompetent Cells following the QuikChange heat shock protocol. Positive colonies appear blue on LB–ampicillin agar plates containing 80 µg/ml X-gal and 20 mM IPTG and the DNA isolated (QIAprep® Spin Maxiprep Kit, standard protocol) and the identity of the DNA verified by sequencing.

9.5 Site Directed Mutagenesis on h*Mm*GluII

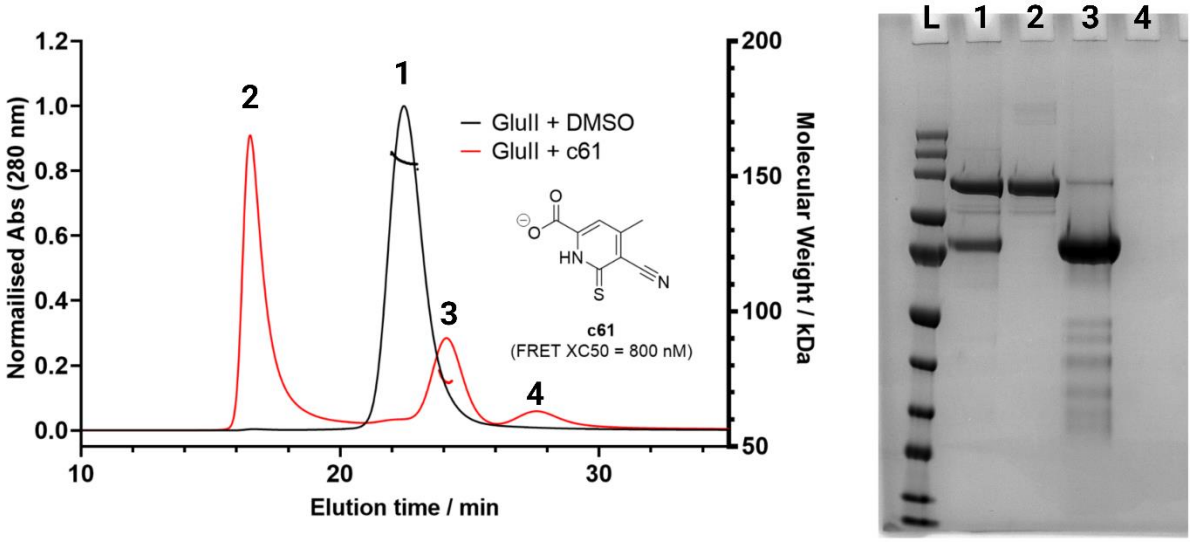
My cloning efforts to produce the mutant *GANAB* constructs proved more challenging and time-consuming than I had expected. Initially, I followed the established protocol used by Dr Alessandro T Caputo and Dr Mario Hensen to produce humanised GluII constructs for both GIIα (F724G) and GIIβ (βL88P, βS90N). This involves PCR extraction of the humanised *GANAB* (*hGANAB)* gene from existing pHLsec constructs used for protein expression and their subsequent insertion into a cloning vector (Litmus28i, 2823 bp) by GA. The rationale here is to perform SDM on the Litmus28i cloning construct as it is smaller than the pHLsec expression vector (4632 bp). Shorter vectors means less DNA synthesis during PCR amplification for SDM where the primers bind within the insert and synthesise the full length construct as linear DNA. In theory, using a shorter construct means a lower risk of errors during DNA synthesis and also saves time on the PCR runs themselves. The initial production of the four mutant *hGANAB*-Litmus28i proceeded without issue.

With the mutant *hGANAB*-Litmus28i cloning constructs in hand, I then attempted to insert them back into the pHLsec expression vector using GA. I encountered numerous problems during this step, foremost among them being consistently poor colony growth post-transformation into DH5α competent cells, with almost no detectable inclusion of the *hGANAB* insert by colony PCR for the vast majority of attempts. I used different strains of competent *E.coli* cells from different sources (DH5α – commercial (NEB) and ‘homebrew’; HST08 – commercial (Takara Bio)) and was able to quickly rule out sub-optimal cells as a source of complication because control transformations with different constructs invariably produced hundreds of healthy colonies. This left the DNA production itself as the reason for poor results. In the few *E.coli* colonies that showed insert incorporation by colony PCR, sequencing revealed that several point mutations were introduced as well as the intended ones, indicating an issue with the experimental steps associated with GA. I attempted GA using different commercial kits (NEB, Takara Bio), experimenting with different insert:vector

ratios and incubation temperatures as well as using at least two pairs of GA primers during PCR amplification of the insert, but was unable to obtain healthy colonies containing the desired construct. To troubleshoot further, I attempted thorough quality control experiments to investigate the pHLsec vector template and mutant *hGANAB* inserts. The initial GA of *hGANAB* gene into Litmus28i was successful and given absence of observable problems with the subsequent PCR amplification of the gene that used *hGANAB*-Litmus28i as a template, I concluded that the insert itself likely was not the root of the problem. It appeared that introduction of pHLsec brought complications and this was supported by the consistently inefficient linearisation of the vector with digestion enzymes (AgeI + KpnI). I noticed improvement with the transformation efficacy when I substituted the pHLsec vector for pOPING (Addgene), chosen on the basis that it carries the same secretion signal and C-terminal hexahistidine tag that made the two practically interchangeable. Following GA of *hGANAB*-pOPING, I observed markedly more colonies after transformation into competent cells, though colony PCR still showed very little incorporation of the gene. To check my starting materials for GA, I ran an agarose gel of the insert and linearised pOPING vector and saw no band for the insert. This prompted an investigation into the gel extraction following PCR amplification in which I discovered that during the gel extraction and purification protocol (Macherey-Nagel), I lost almost all of the PCR amplified insert such that it was not visible on subsequent agarose gels, despite the fact that the sample gave a modest concentration reading when measured on a nanodrop spectrophotometer (A260). To test this, I repeated the PCR amplification and omitted the gel purification step, opting instead for an in solution PCR clean up protocol (Macherey-Nagel). This significantly improved the results of the *hGANAB*-pOPING GA, but simultaneous *hGANAB*-pHLsec continued to prove unsuccessful. It is unclear why the gel extraction and purification protocol was so inefficient, but with the adapted workflow I observed several colonies with the desired insert following colony PCR and selected several for DNA sequencing.

Despite improvements in the GA efficiency using pOPING and a more crude PCR amplification mixture, I was able to isolate colonies containing the *hGANAB* insert with the target cysteine mutations for C19/25A, C423A and C822A, but C502A colonies invariably contained at least one other significant point mutation. However, even these promising sequences all demonstrated a common gap in the sequencing data around the F724G humanisation mutation meaning no fully correct DNA sequences were obtained. Assuming that this issue came from complicated DNA secondary structure confounding the sequencing experiments, rather than from an incorrect sequence, I performed a PCR insertion experiment to replace the F724G codon with an alternate glycine codon. The resulting construct showed the same sequencing gap of around 20 bp. In the meantime, I attempted an alternate method of SDM whereby the mutation is introduced directly into the expression vector using the Agilent QuikChange II XL kit and new primers designed using the online tool provided by Agilent. Unfortunately, this did not improve upon the very low cloning efficiency for *hGANAB*-pOPING and I was unable to produce the designed constructs.

9.6 Protein Identification Following SEC-MALS



SEC-MALS - Peak 2:
(Trypsin digest)

Protein ID: GIIα
71% Sequence Coverage

(DDA MS/MS)

SEC-MALS - Peak 3:
(aLP digest)

Protein ID: GIIβ
59% Sequence Coverage

(DDA MS/MS)

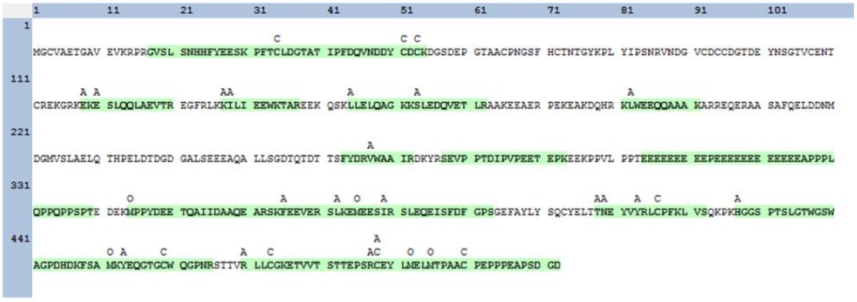


Figure I – Protein characterisation of SEC-MALS peaks from the experiment described in Section 4.4.1. Peak fractions from SEC (top left) were applied to SDS-PAGE (top right). Bands were excised and in-gel digestion performed prior to DDA MS/MS. Unique peptides corresponding to GIIα and GIIβ were identified, confirming α/β interface disruption and aggregation of GIIα.

9.7 Proteomics Experiments – Evotec

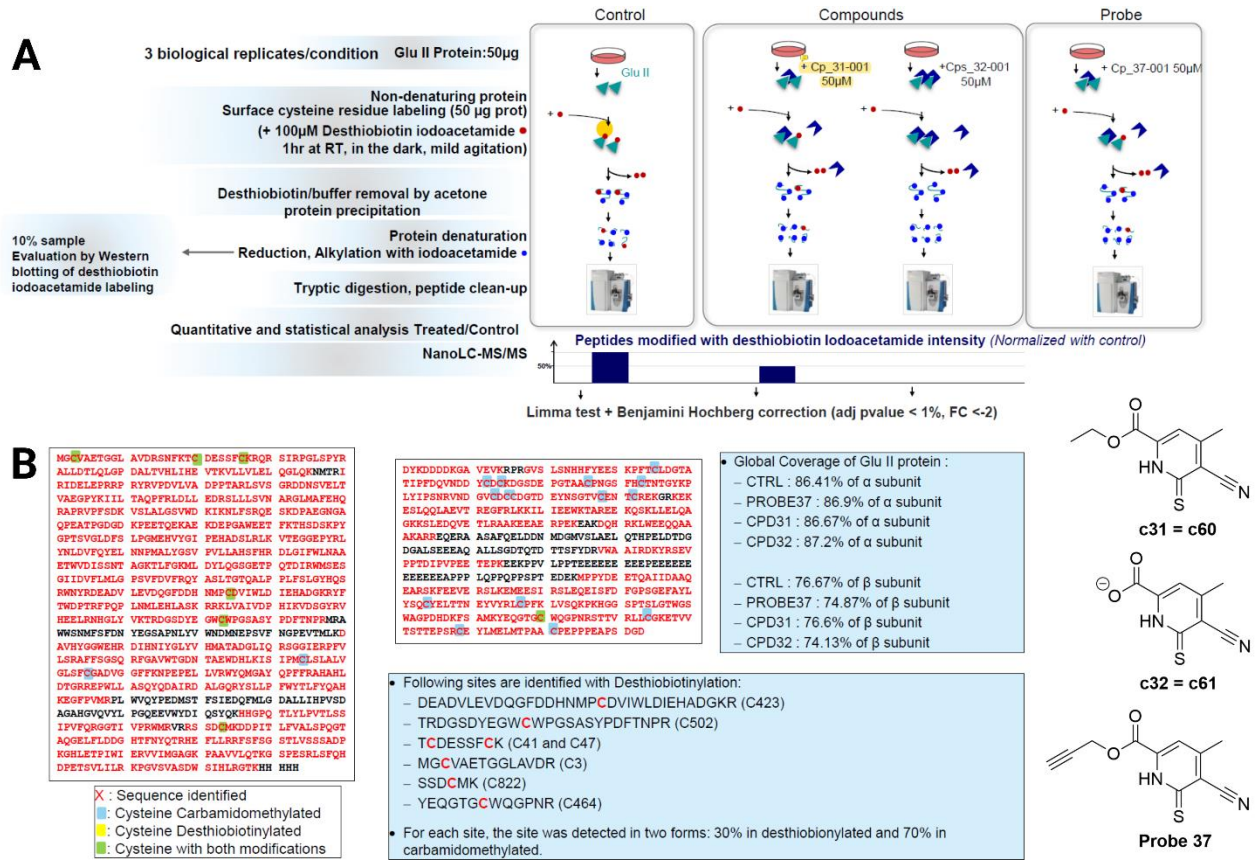


Figure II – DTB-IAA assay developed by Evotec and used to identify cysteine residues that may be modified by thiopyridones. These experiments are discussed in Section 4.4.2.1.

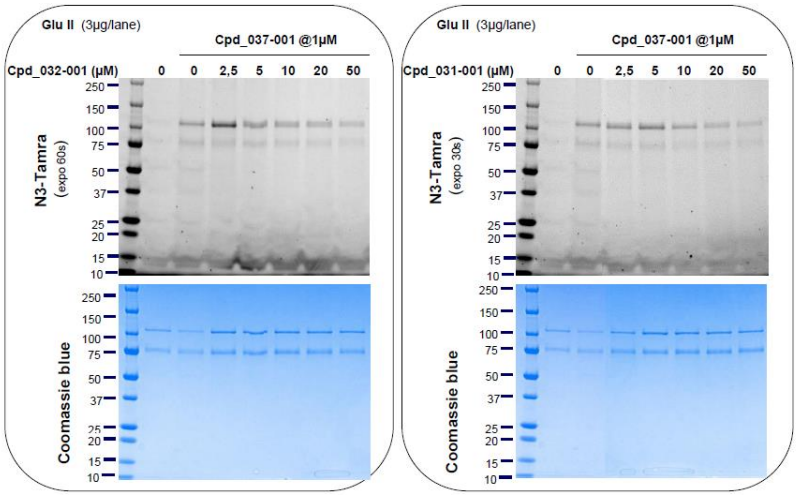


Figure III – SDS-PAGE N_3-TAMRA assay described in 4.4.3 shows dose-dependent displacement of probe 37 by active thiopyridones 31 (c60) and 32 (c61).

9.8 DDA MS/MS Sequence Coverage – GIIα

MS experiments were carried out by Dr Bevin Gangadharan (Zitzmann group) and are described in Section 4.4.2.2. Text in red indicates peptides identified in DDA MS/MS, while black text shows peptides that were not identified. The long peptide containing C822 is highlighted as bold text.

Expected c61 site – C822

Known disulfide bond

Other cysteines

DDA GIIα with c61 (Trypsin):

Coverage: 79.245% - (756 / 954)

mgcvaetggl avdrsnfktc dessfckrqr SIRPGLSPYR ALLDTLQLGP DALTVHLIHE VTKVLLVLEL
QGLQKnmtrI RIDELEPRrp rYRVPDVLVA DPPTARLSVS GRDDNSVELT VAEGPYKIIL TAQPFRLDLL
EDRSLLLSVN ARGLMAFEHQ RaprVPFSDK VSLALGSVWD KIKNLFSRQE SKDPAEGNGA QPEATPGDGD
KPEETQEKAE KDEPGAWEET FKTHSDSKPY GPTSVGLDFS LPGMEHVYGI PEHADSLRLK VTEGGEPYRL
YNLDVFQYEL NNPMALYGSV PVLLAHSFHR DLGIFWLNAA ETWVDISSNT AGKTLFGKML DYLQGSGETP
QTDIRWMSES GIIDVFLMLG PSVFDVFRQY ASLTGTQALP PLFSLGYHQS Rwnyrdeadv levdqgfddh
nmpcdviwld iehadgkRYF TWDPTRFPQP LNMLEHLASK RrKLVAIVDP HIKVDSGYRV HEELRNHGLY
VKtrdgsdye gwcwpgsasy pdftnprmrA WWSNMFSFDN YEGSAPNLYV WNDMNEPSVF NGPEVTMLKD
AVHYGGWEHR DIHNIYGLYV HMATADGLIQ RSGGIERPFV LSRAFFSGSQ RFGAVWTGDN TAEWDHLKis
ipmclslalv glsfcgadvg gffkNPEPEL LVRWYQMGAY QPFFRAHAHL DTGRREPWLL ASQYQDAIRD
ALGQRYSLLP FWYTLFYQAH KEGFPVMRpl wvqypedmst fsiedqfmlg dallihpvsd agahgvqvyl
pgqeevwydi qsyqkHHGPQ TLYLPVTLSS IPVFQRGGTI VPRWMRvr**RS SDCMKDDPIT LFVALSPQGT**
AQGELFLDDG HTFNYQTRHE FLLRRFSFSG STLVSSSADP KGHLETPIWI ERVVIMGAGK PAAVVLQTKG
SPESRLSFQH DPETSVLILR KPGVSVASDW SIHLRgtkhh hhhh

DDA untreated GIIα (Trypsin):

Coverage: 77.673% - (741 / 954)

mgcvaetggl avdrsnfktc dessfckrqr SIRPGLSPYR ALLDTLQLGP DALTVHLIHE VTKVLLVLEL
QGLQKnmtrI RIDELEPRrp rYRVPDVLVA DPPTARLSVS GRDDNSVELT VAEGPYKIIL TAQPFRLDLL
EDRSLLLSVN ARGLMAFEHQ RaprVPFSDK VSLALGSVWD KIKNLFSRQE SKDPAEGNGA QPEATPGDGD
KPEETQEKAE KDEPGAWEET FKTHSDSKPY GPTSVGLDFS LPGMEHVYGI PEHADSLRLK VTEGGEPYRL
YNLDVFQYEL NNPMALYGSV PVLLAHSFHR DLGIFWLNAA ETWVDISSNT AGKTLFGKML DYLQGSGETP
QTDIRWMSES GIIDVFLMLG PSVFDVFRQY ASLTGTQALP PLFSLGYHQS Rwnyrdeadv levdqgfddh

nmpcdviwld iehadgkRYF TWDPTRFPQP LNMLEHLASK RrKLVAIVDP HIKVDSGYRV HEELRNHGLY
VKTRdgsdye gwCwpgsasy pdftnprmra wwsnmfsfdn yegsapnlyv wndmnepsvf ngpevtmlkD
AVHYGGWEHR DIHNIYGLYV HMATADGLIQ RSGGIERPFV LSRAFFSGSQ RFGAVWTGDN TAEWDHLKis
ipmclslalv glsfcgadvg gffkNPEPEL LVRWYQMGAY QPFFRAHAHL DTGRREPWLL ASQYQDAIRD
ALGQRYSLLP FWYTLFYQAH KEGFPVMRpl wvqypedmst fsiedqfmlg dallihpvsd agahgvqvyl
pgqeevwydi qsyqkHHGPQ TLYLPVTLSS IPVFQRGGTI VPRWMRvr**RS SDCMKDDPIT LFVALSPQGT**
AQGELFLDDG HTFNYQTRHE FLLRRFSFSG STLVSSSADP KGHLETPIWI ERVVIMGAGK PAAVVLQTKG
SPESRLSFQH DPETSVLILR KPGVSVASDW SIHLRgtkhh hhhh

DDA GIIα with C61 (Chymotrypsin):

Coverage: 73.061% - (697 / 954)

mgCvaetggl AVDRSNFktC dessfCkrqr sirpglspyR ALLDTLqlgp daltvhlIHE VTKVLLvlel
qglqknmTRI RIDELEPRRP RYRVPDVLVA DPPTARLSVS GRDDNSVELT VAEGPYKIIL TAQPFRLDLL
EDRSLLLSVN ARGLMAFEHQ RAPRVPFsdk vslalgsvwd kiknlfsrqe skdpaegnga qpeatpgdgd
kpeetqekae kdepgaweet fKTHSDSKPY GPTSVGLDFS LPGMEHVYGI PEHADSLRLK VTEGGEPYrl
yNLDVFQYEL NNPMALYGSV PVLLAHSFHR DLGIFwlnaa etwVDISSNT AGKTLFgkml dylqgsgetp
qtdirwMSES GIIDVFLMLG PSVFDVFrqy ASLTGTQALP PLFSLGYHQS RWNYRDEADV Levdqgfddh
nmpCdviwlD IEHADGKRYF TWDPTRFPQP LNMLEHLASK RRKLVAIVDP HIKVDSGYRV HEELRNHGLY
VKTRDGSDYE GWCwpgsasy pdftnprmra wwSNMFSFDN YEGSAPNLYV WNDMNEPSVF NGPEVTMLKD
AVHYGGWEHR DIHNIYGLYV HMATADGLIQ RSGGIERPFV LSRAFFSGSQ RFGAVWTGDN TAEWDHLKIS
IPMClSLALV GLSFCgadvg gfFKNPEPEL LvrwYQMGAY QPFFRAHAHL dtgrrepwLL ASQYQDAIRD
ALGQRYSLLP FWYtlfYQAH KEGFPVMRPL WVQYPEDMST FSIEDQFMLG DALLIHPVSD AGAHGVQVYL
PGQEEVWYDI QSYQKHHGPQ TLYlpvtlss ipvfQRGGTI VPRWMrvrrs sdcmKDDPIT LFVALSPQGT
AQGELFLDDG HTFnyqtrhe fllrrfSFSG STLVSSSADP KGHLETPIWI ERVVIMGAGK PAAVVLQTKG
SPESRLSFQH DPETSVLILR KPGVSVASDW SIHLrgtkhh hhhh

DDA untreated GIIα (Chymotrypsin):

Coverage: 88.05% - (840 / 954)

mgCvaetggl AVDRSNFktC dessfCkrqr sirpglspyR ALLDTLQLGP DALTVHLIHE VTKVLLVLEL
QGLqknmTRI RIDELEPRRP RYRVPDVLVA DPPTARLSVS GRDDNSVELT VAEGPYKIIL TAQPFRLDLL
EDRSLLLSVN ARGLMAFEHQ RAPRVPFSDK VSLALGSVWD KIKNLFSRQE SKDPAEGNGA QPEATPGDGD
KPEETQEKAE KDEPGAWEET FKTHSDSKPY GPTSVGLDFS LPGMEHVYGI PEHADSLRLK VTEGGEPYrl
yNLDVFQYEL NNPMALYGSV PVLLAHSFHR DLGIFWLNAA ETWVDISSNT AGKTLFGKML DYLQGSGETP
QTDIRWMSES GIIDVFLMLG PSVFDVFRQY ASLTGTQALP PLFSLGYHQS RWNYRDEADV Levdqgfddh
nmpCdviwLD IEHADGKRYF TWDPTRFPQP LNMLEHLASK RRKLVAIVDP HIKVDSGYRV HEELRNHGLY
VKTRDGSDYE GWCwpgsasy PDFTNPRMra wwSNMFSFDN YEGSAPNLYV WNDMNEPSVF NGPEVTMLKD
AVHYGGWEHR DIHNIYglyV HMATADGLIQ RSGGIERPFV LSRAFFSGSQ RFGAVWTGDN TAEWDHLKIS
IPMClSLALV GLSFCgadvg gffKNPEPEL LVRWYQMGAY QPFFRAHAHL dtgrrepwLL ASQYQDAIRD
ALGQRYSLLP FWYTLFYQAH KEGFPVMRPL WVQYPEDMST FSIEDQFMLG DALLIHPVSD AGAHGVQVYL
PGQEEVWYDI QSYQKHHGPQ TLYLPVTLSS IPVFQRGGTI VPRWMrvrrs sdcmKDDPIT LFVALSPQGT
AQGELFLDDG HTFNYQTRHE FllrrfSFSG STLVSSSADP KGHLETPIWI ERVVIMGAGK PAAVVLQTKG
SPESRLSFQH DPETSVLILR KPGVSVASDW SIHLrgtkhh hhhh

DDA GIIα with C61 (α-Lytic Protease):

Coverage: 88.47% - (844 / 954)

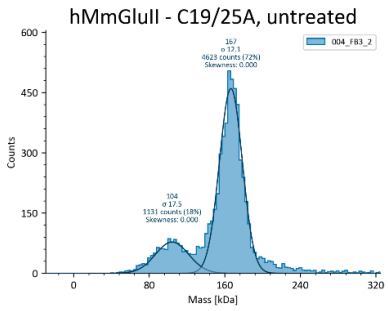
mgcvaetggl avDRSNFKTc dessfckRQR SIRPGLSPYR ALLDTLQLGP DALTVHLIHE VTkVLLVLEL
QGLQKNMTRI RIDELEPRRP RYRVPDVLVA DPPTARLSVS GRDDNSVELT VAEGPYKIIL TAQPFRLDLL
EDRSLLLSVN ARGLMAFEHQ RAPRVPFSDK VSLALGSVWD KIKNLFSRQE SKDPAEGNGA QPEATPGDGD
KPEETQEKAE KDEPGAWEET FKTHSDSKPY GPTSVGLDFS LPGMEHVYGI PEHADSLRLK VTEGGEPYRL
YNLDVFQYEL NNPMALYGSV PVLLAHSFHR DLGIFWLNAA ETWVDISSnt aGKTLFGKML DYLQGSGETP
QTDIRWMSES GIIDVFLMLG PSVFDVFRQY ASLTGTQALP PLFSLGyhqs RWNYRDEADV LEVDQGfddh
nmpcdvIWLD IEHADGKRYF TWDPTRFPQP LNMLEHLASK RRKLVAIVDP HIKVDSGYRV HEELRNHGLY
VKTrdgsdye gwcwpgSASY PDFTNPRMRA WWSNMFSFDN YEGSAPNLYV WNDMNEPSVF NGPEVTMLKD
AVHYGGWEHR DIHNIYGLYV HMATADGLIQ RSGGIERPFV LSRAFFSGSQ RFGAVWTGDN TAEWDHLKIS
ipmclslalv glsfcgADVG GFFKNPEPEL LVRWYQMGAY QPFFRAHAHL DTGRREPWLL ASQYQDAIRD
ALGQRYSLLP FWYTLFYQAH KEGFPVMRPL WVQYPEDMST FSIEDQFMLG DALLIHPVSD AGAHGVQVYL
PGQEEVWYDI QSyqkhhgpq tlylpvtlss ipvFQRGGTI VPRWMRVRRS sdcmkddpit lfvALSPQGT
AQGELFLDDG HTFNYQTRHE FLLRRFSFSG STlvSSSADP KGHLETPIWI ERVVIMGAGK PAAVVLQTKG
SPESRLSFQH DPETSVLILR KPGVSVASDW SIHLRGTkhh hhhh

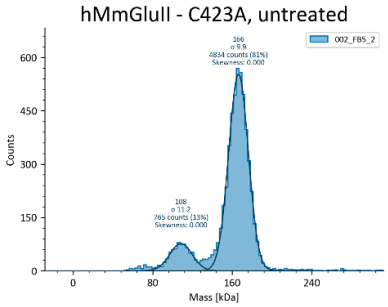
DDA untreated GIIα (α-Lytic Protease):

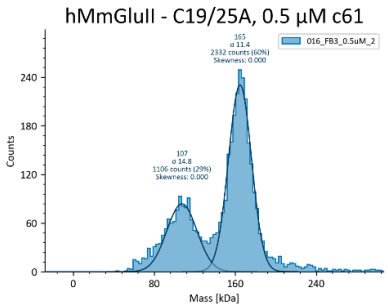
Coverage: 87.421% (834 / 954)

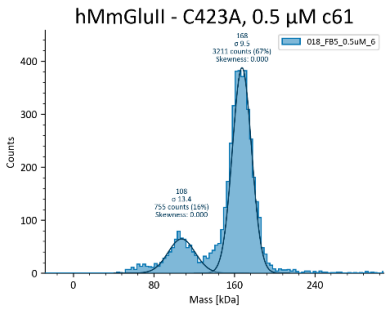
mgcvaetggl aVDRSNFKTc dessfckrqr sIRPGLSPYR ALLDTLQLGP DALTVHLIHE VTkVLLVLEL
QGLQKNMTRI RIDELEPRRP RYRVPDVLVA DPPTARLSVS GRDDNSVELT VAEGPYKIIL TAQPFRLDLL
EDRSLLLSVN ARGLMAFEHQ RAPRVPFSDK VSLALGSVWD KIKNLFSRQE SKDPAEGNGA QPEATPGDGD
KPEETQEKAE KDEPGAWEET FKTHSDSKPY GPTSVGLDFS LPGMEHVYGI PEHADSLRLK VTEGGEPYRL
YNLDVfqyel nnpmaLYGSV PVLLAHSFHR DLGIFWLNAA ETWVDISSnt aGKTLFGKML DYLQGSGETP
QTDIRWMSES GIIDVFLMLG PSVFDVFRQY ASLTGTQALP PLFSLGyhqs RWNYRDEADV LEVDQGfddh
nmpcdvIWLD IEHADGKRYF TWDPTRFPQP LNMLEHLASK RRKLVAIVDP HIKVDSGYRV HEELRNHGLY
VKTrdgsdye gwcwpgSASY PDFTNPRMRA WWSNMFSFDN YEGSAPNLYV WNDMNEPSVF NGPEVTMLKD
AVHYGGWEHR DIHNIYGLYV HMATADGLIQ RSGGIERPFV LSRAFFSGSQ RFGAVWTGDN TAEWDHLKIS
ipmclslalv glsFCGADVG GFFKNPEPEL LVRWYQMGAY QPFFRAHAHL DTGRREPWLL ASQYQDAIRD
ALGQRYSLLP FWYTLFYQAH KEGFPVMRPL WVQYPEDMST FSIEDQFMLG DALLIHPVSD AGAHGVQVYL
PGQEEVWYDI QSyqkhhgpq tlylpvtlss ipvFQRGGTI VPRWMRVRRS sdcmkddpit lfvALSPQGT
AQGELFLDDG HTFNYQTRHE FLLRRFSFSG STlvSSSADP KGHLETPIWI ERVVIMGAGK PAAVVLQTKG
SPESRLSFQH DPETSVLILR KPGVSVASDW SIHLRGTkhh hhhh

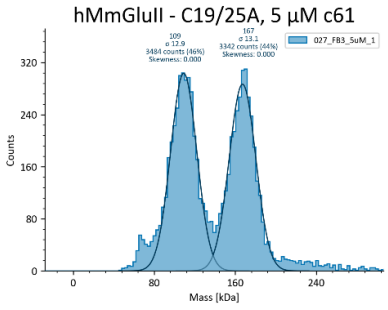
9.9 Representative Histograms from MP Experiments with Cysteine Mutants of h*Mm*GluII + c61

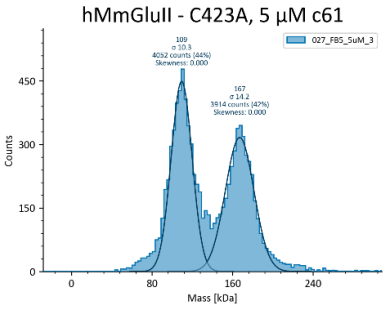


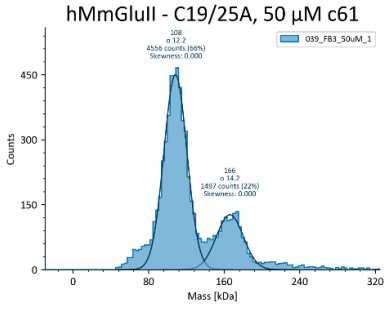


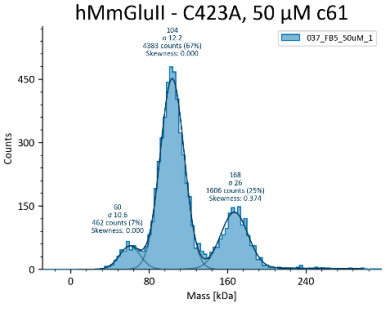


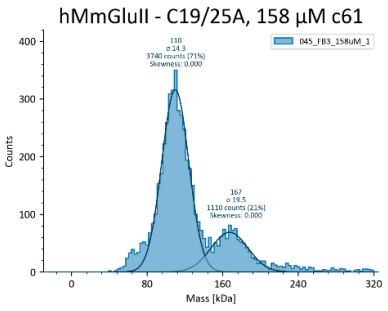


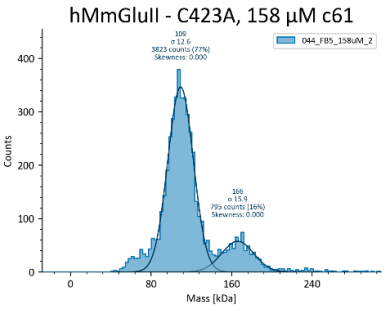


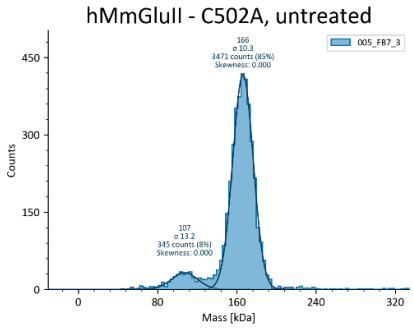


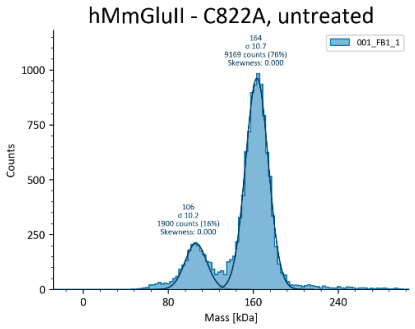


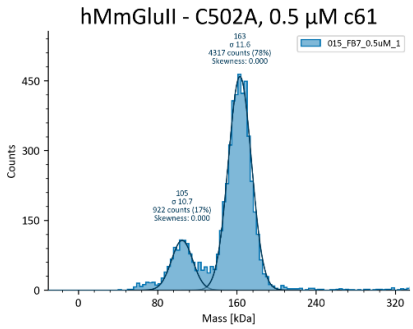


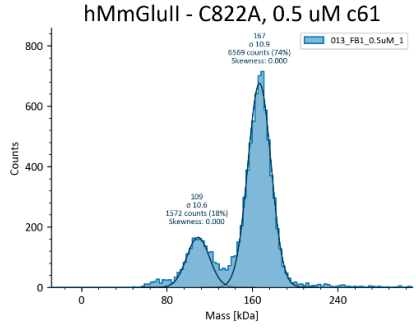


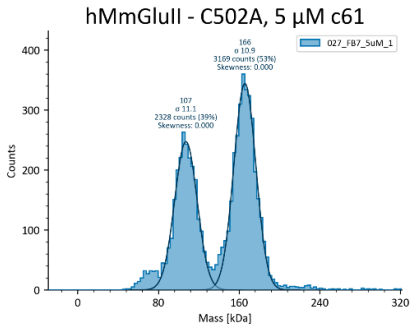


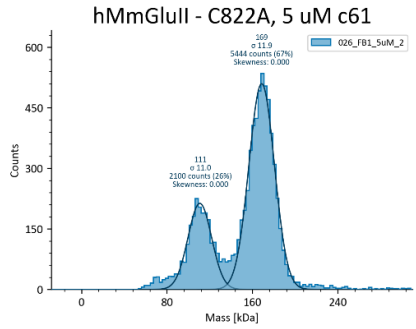

hMmGluII - C502A, untreated
005_FB7_3
166
σ 10.3
3471 counts (85%)
Skewness: 0.000
107
σ 13.2
345 counts (8%)
Skewness: 0.000
Counts
Mass [kDa]

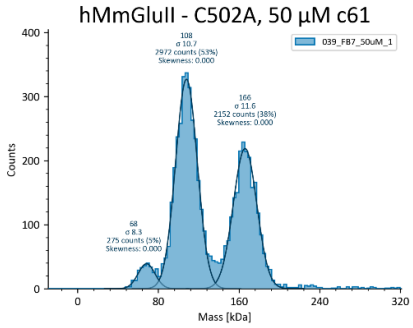

hMmGluII - C822A, untreated
001_FB1_1
164
σ 10.7
9169 counts (76%)
Skewness: 0.000
106
σ 10.2
1900 counts (16%)
Skewness: 0.000
Counts
Mass [kDa]

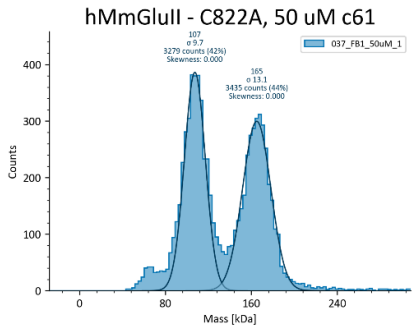

hMmGluII - C502A, 0.5 µM c61
015_FB7_0.5uM_1
163
σ 11.6
4317 counts (78%)
Skewness: 0.000
105
σ 10.7
922 counts (17%)
Skewness: 0.000
Counts
Mass [kDa]

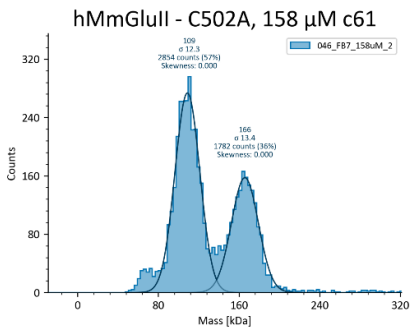

hMmGluII - C822A, 0.5 uM c61
013_FB1_0.5uM_1
167
σ 10.9
6569 counts (74%)
Skewness: 0.000
109
σ 10.6
1572 counts (18%)
Skewness: 0.000
Counts
Mass [kDa]

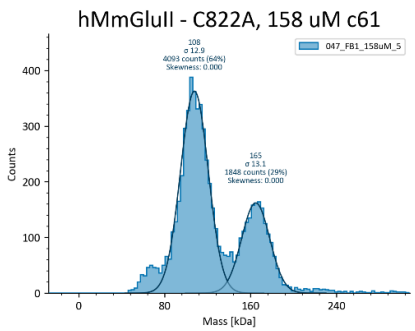

hMmGluII - C502A, 5 µM c61
027_FB7_5uM_1
166
σ 10.9
3169 counts (53%)
Skewness: 0.000
107
σ 11.1
2328 counts (39%)
Skewness: 0.000
Counts
Mass [kDa]


hMmGluII - C822A, 5 uM c61
026_FB1_5uM_2
169
σ 11.9
5444 counts (67%)
Skewness: 0.000
111
σ 11.0
2100 counts (26%)
Skewness: 0.000
Counts
Mass [kDa]


hMmGluII - C502A, 50 µM c61
039_FB7_50uM_1
108
σ 10.7
7972 counts (53%)
Skewness: 0.000
166
σ 11.6
2152 counts (38%)
Skewness: 0.000
68
σ 8.3
275 counts (5%)
Skewness: 0.000
Counts
Mass [kDa]


hMmGluII - C822A, 50 uM c61
037_FB1_50uM_1
107
σ 9.7
3279 counts (42%)
Skewness: 0.000
165
σ 13.1
3435 counts (44%)
Skewness: 0.000
Counts
Mass [kDa]


hMmGluII - C502A, 158 µM c61
046_FB7_158uM_2
109
σ 12.3
2834 counts (57%)
Skewness: 0.000
166
σ 13.4
1782 counts (36%)
Skewness: 0.000
Counts
Mass [kDa]


hMmGluII - C822A, 158 uM c61
047_FB1_158uM_5
108
σ 12.9
4093 counts (64%)
Skewness: 0.000
165
σ 13.1
1848 counts (29%)
Skewness: 0.000
Counts
Mass [kDa]

9.10 X-Ray Crystallography

9.10.1 Materials & Methods

Preparation of glutaraldehyde cross-linked GluII for crystallography screens: Glutaraldehyde cross-linked GluII was incubated with c61 at 200-fold molar excess for 1 h on ice before application to SEC using a S200 Superdex 10/300 increase column with 20mM HEPES pH 7.5, 120mM NaCl as the eluent. Peak fractions were combined and concentrated to roughly 5.5 mg/ml then centrifuged at 13,300 rpm for 40 min at room temperature. The cross-linked GluII was set up for crystallisation in two commercial screens, Morpheus® I and Morpheus® II from Molecular Dimensions, using MRC 2-drop 96-well SwissSci crystallisation plates. Sitting drops were set up using a Mosquito system (TTP Labtech) with a protein-to-crystallisation condition drop ratio of 1:1. All crystallisation plates were stored in a Rigaku Gallery HT (Rigaku) plate imager maintained at 18°C. The visible and UV images of the drops, automatically acquired by the imager at regular intervals, were examined for the presence of protein crystals.

A single crystal appeared after roughly 6 weeks in one of the conditions in Morpheus® II screen (0.1M Buffer System 6, 90mM LiNaK 32.5% Precipitant Mix 6), but failed to diffract when checked on a synchrotron beamline at Diamond Light Source (DLS).

Crystallography experiments with native MmGluII$_{Tryp}$: Crystallography conditions previously established by Dr Mario Hensen were used here and are summarised in Table IV. Crystal seeding was performed whereby small *Mm*GluII$_{Tryp}$ crystals, found in varying conditions of a Morpheus® HT-96 screen (Molecular Dimensions) were crushed using a glass rod, taken up in 1 µl crystallisation solution (Table IV) and pooled in a 500 µl Eppendorf tube before vortexing for 4 min. Dilutions at 1:10 and 1:100 ratios were prepared using the same crystallisation solution as before and samples vortexed again. 96-well MRC 2-well crystallisation plates were set up at a 2:1:1 ratio (protein:crystallisation solution:seeding solution) and stored in a plate imager at 18°C as described above.

Table IV – Crystallisation conditions for MmGluII$_{Tryp}$. All solutions were purchased from Molecular Dimensions.

Crystal Form	Crystallisation Solution
Orthorhombic	32% ethylene glycol/PEG 8000 mix, 0.1M ethylene glycol mix, 0.1 buffer system 2, pH 7.25
Trigonal	30% glycerol/PEG 4000 mix, 0.12 M alcohol mix, 0.1 M buffer system 1 pH 6.5

MmGluIItryps crystals were soaked with thiopyridones c28 (inactive) or c61 active,) solubilised in DMSO at 100 mM and subsequently diluted to 10 mM with soaking solution (16% PEG 8000, 20% PEG 400, 0.05 M Morpheus® carboxylic acids mix and 0.1 M Morpheus® buffer system 2 pH 7.2). Depending on the crystal form used, the compound containing soaking solution was diluted with the corresponding crystallisation solution (Table X). 0.2-0.5 µl of the thiopyridone solutions were transferred into a 96-well MRC 2-well crystallisation plate and up to five *Mm*GluII$_{Tryp}$ crystals placed into the solution and incubated at RT for >1 h before cryo-cooling. Few crystals survived the soaking process and those that did failed to diffract when checked on a synchrotron beamline at DLS.

9.10.2 Results & Discussion

Acquiring structural data of full length GluII using X-ray crystallography has proven difficult for a number of reasons. The inherent flexibility of full length GIIβ is the primary cause of the difficulty, hence the success with obtaining crystal structures of trypsin-digested GluII ($GluII_{Tryp}$) which removes the majority of GIIβ from the heterodimer.[54] A crystal structure of full length GluII could elucidate the interaction of the MRH domain of GIIβ with the substrate glycans, which is discussed further in Appendix 9.12. In the context of thiopyridones, crystal structures would confirm beyond doubt the true protein-ligand binding mode. However, the interface disruption introduces further problems for crystallography because it is likely to destroy any GluII crystals that form by disturbing the crystal lattice. During my DPhil, I continued what previous Zitzmann group members had started to try to obtain GluII-thiopyridone crystals.

The first X-ray crystallography experiment I attempted utilised the inactive pyridone, c28, which is structurally analogous to the active compound, c61, save for the replacement of a sulfur atom with oxygen (Figure 14). I reasoned that their structural similarity should mean that they would bind to GluII at the same location, forming similar non-covalent interactions, but crucially c28 would not disrupt the α/β interface as shown in previous data presented here. This means that GluII-c28 crystal structures could provide detail about the binding mode of active thiopyridones without the lattice disruption associated with subunit separation. To explore this, I attempted $GluII_{Tryp}$-c28 co-crystallisation as well as soaking of $GluII_{Tryp}$ crystals in a solution of c28. In both cases, I used crystallography conditions that have previously yielded X-ray-diffracting $GluII_{Tryp}$ crystals. Unfortunately, neither experiment gave $GluII_{Tryp}$-c28 crystals and this approach was abandoned. In parallel, I also attempted a $GluII_{Tryp}$ crystal soak using a solution of c61 whereby the crystals were exposed to c61 very briefly (< 1 s) and immediately frozen using liquid nitrogen in an attempt to trap the complex before the α/β interface disruption could occur. This experiment was also unsuccessful.

Obtaining crystal structures of GluII-thiopyridone complexes may be possible if the separation of GluII subunits by the compounds is avoided. Glutaraldehyde cross-linking (GXL) is an established method of non-specifically cross-linking proteins using positively charged amine groups on lysine side chains in attempt to improve the robustness of protein crystals.[194] In the context of GluII, GXL may be beneficial for two reasons. Firstly, it may rigidify full length GIIβ such that full length GluII could crystallise, facilitating X-ray crystallography without need for trypsin digest. Secondly, cross-linking theoretically prevents GluII subunit separation following treatment with thiopyridones so may allow for trapping of the GluII-thiopyridone complex in a manner compatible with crystal growth. GXL of GluII was performed according to the protocol in Section 9.12.1 and the cross-linked GluII heterodimer was separated from higher order oligomers by SEC. Figure IV shows the SEC profile for the glutaraldehyde-treated GluII sample and the SDS-PAGE gel used to identify the cross-linked GluII heterodimer. Both techniques reveal the presence of higher molecular weight species resulting from non-specific cross-linking of higher order oligomers. To minimise these high molecular weight species, a low concentration of GluII was used during glutaraldehyde treatment to ensure that the bulk of cross-linking reactions were intramolecular with respect to the GluII heterodimer.

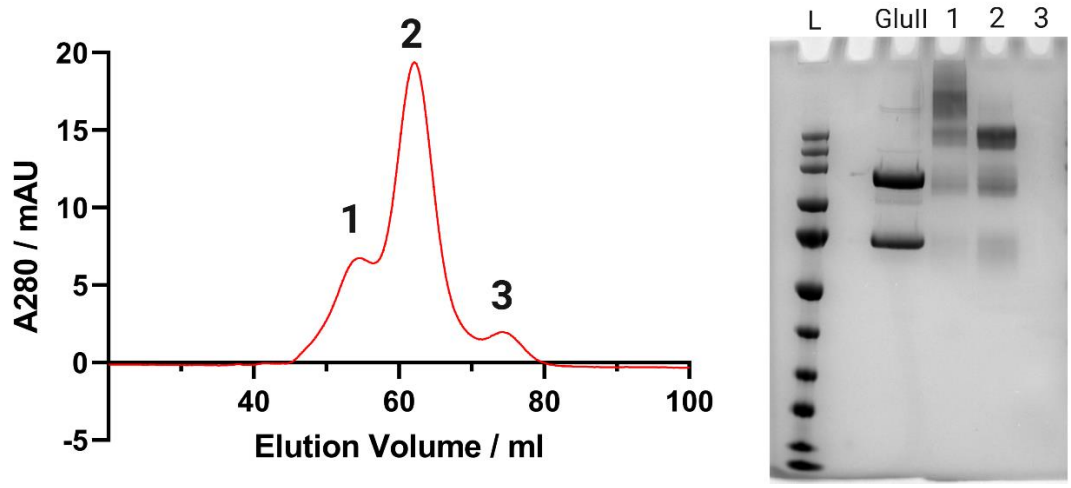


Figure IV – Evaluation of the GXL GluII by SEC (left) and SDS-PAGE. Both techniques show the presence of high molecular weight species resulting from non-specific cross-linking of many equivalents of GluII or its subunits.

With the cross-linked GluII heterodimer in hand, I performed a crystallography screen of 96 conditions (Morpheus® II), co-crystallising with c61.[195] This screen yielded a single UV-absorbing, rod-shaped crystal that unfortunately did not display sufficient X-ray diffraction for crystallographic study (Figure V). The low success rate for this crystallographic screen may suggest that glutaraldehyde cross-linking does not resolve any of the known issues associated with growing crystals of full length GluII. Indeed, the low specificity of glutaraldehyde cross-linking and conformational flexibility of GIIβ likely mean that the cross-linked heterodimer is trapped in a variety of different poses and therefore does not show the homogeneity required for crystal lattice formation. It is possible that X-ray crystallography is unsuitable for the study of full length GluII, though this experiment was only attempted once and it would be prudent to repeat the crystallography screen using variations on the conditions that produced the only observable crystal. An attempt to simultaneously acquire crystal structures of full length GluII *and* a GluII-thiopyridone complex may have been overly ambitious since neither one has been successfully achieved to date.

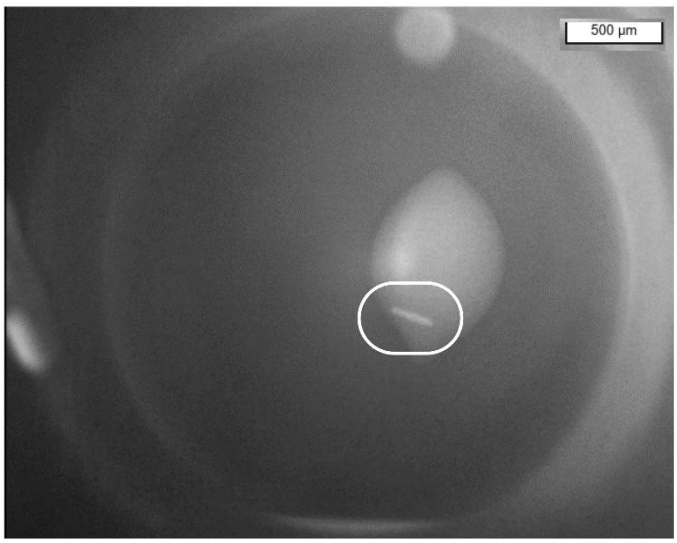


Figure V – Image of a crystal observed in a Morpheus 2 screen with GXL GluII and c61. The crystal growth conditions are listed in Section 9.10.1. The crystal did not display sufficient X-ray diffraction for data acquisition. A thorough screen of similar crystal growth conditions may yield more crystals to be tested with X-ray crystallography.

9.10.3 Future Work

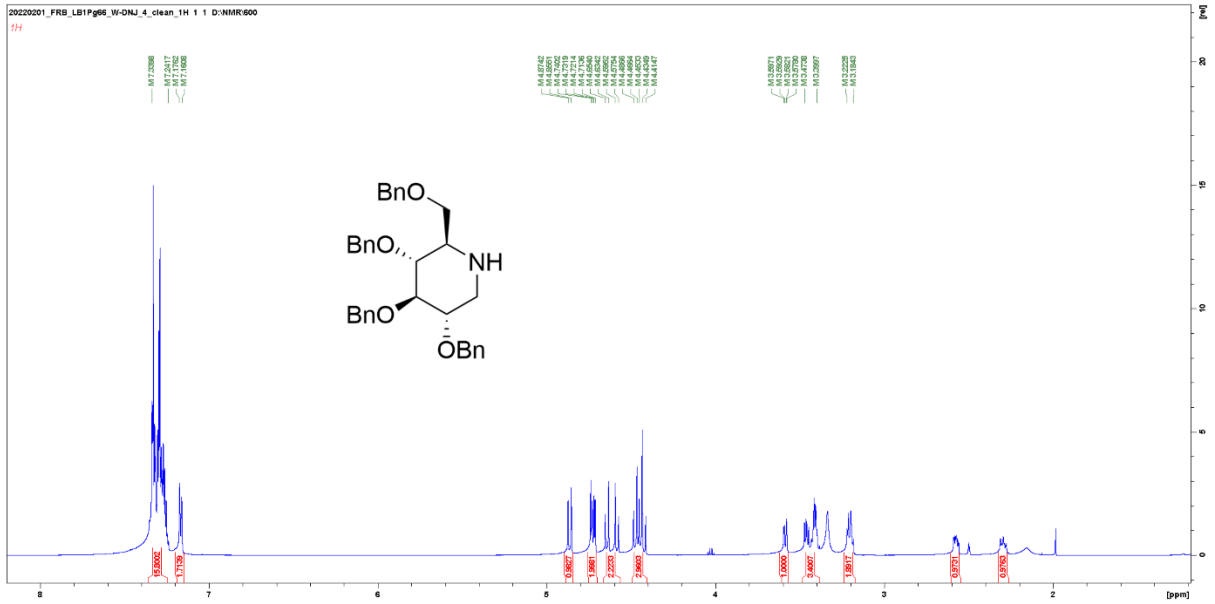
My attempts to crystallise GluII with c61 or c28 were ultimately unsuccessful but far from exhaustive and would be worth pursuing further. GXL GluII was treated with c61 and in one set of crystallography conditions, a UV-absorbing, rod-shaped crystal grew but did not sufficiently diffract X-rays to produce any structural data. It would be worth probing these conditions further in attempt to grow more cross-linked GluII crystals in complex with c61. That said, it is possible that the inherent flexibility of GIIβ means that even GXL GluII would adopt different conformations that reduce its suitability for X-ray crystallography. Attempts have been made by the Zitzmann group to use glutaraldehyde to cross-link *Mm*GluII$_{Tryp}$ with which previous crystallography experiments have been successful, but these attempts proved unsuccessful. The sparsity of lysine residues, particularly on GIIβ, in *Mm*GluII$_{Tryp}$ may explain the unsuccessful glutaraldehyde cross-linking for this construct. I propose an investigation into alternative cross-linking methodologies for *Mm*GluII$_{Tryp}$ as trypsin digestion this remains our only successful strategy towards obtaining protein crystals of GluII to date. Many different

types of protein cross-linkers are described in the literature and one or more of these may provide a GluII construct that is resistant to α/β interface disruption when treated with c61.[196,197] An alternative, speculative strategy for crystallisation of c61 in complex with GluII would be to simply combine the two binding partners and attempt to grow crystals in spite of the α/β interface disruption. It is highly unlikely that this would provide any crystals of the GluII heterodimer, but there is a chance that c61-bound GIIα may crystallise in a way that allows observation of the protein-ligand interaction in the absence of GIIβ, which is anyway not the target of c61 based on the data presented herein.

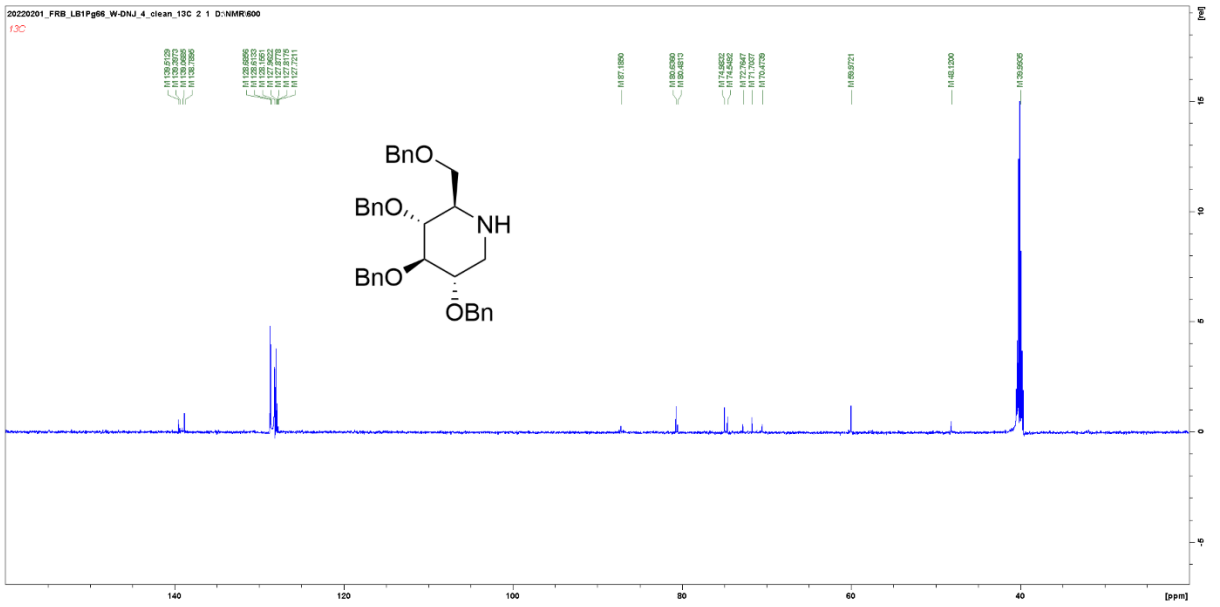
9.11 Selected Physical Data for Synthetic Intermediates

9.11.1 Synthesis of 1-Deoxynojirimycin (DNJ) – Wennekes *et al*.[157]

DNJ-(OBn)$_4$ - ^{1}H NMR, 600 MHz, d$_6$-DMSO:

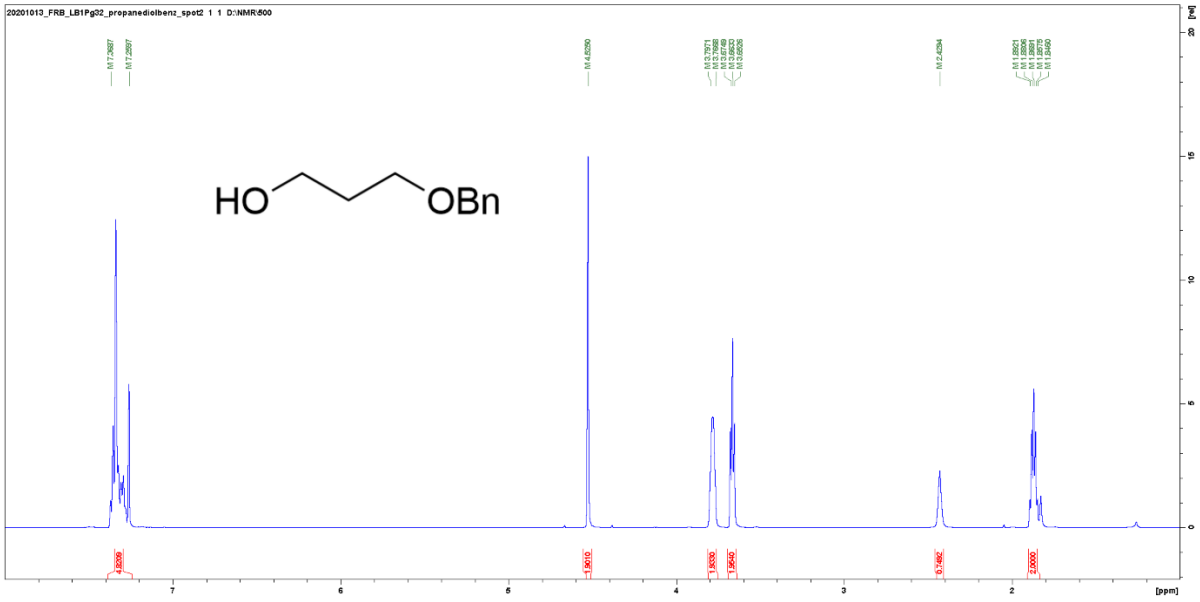


DNJ-(OBn)$_4$ - ^{13}C NMR, 600 MHz, d$_6$-DMSO:

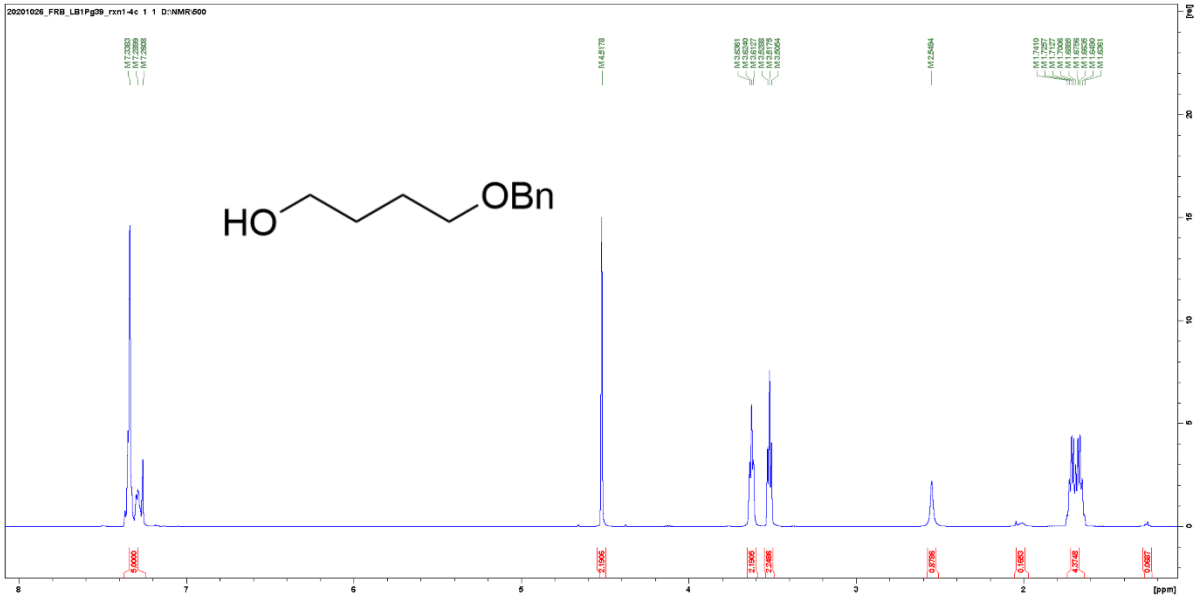


9.11.2 Alkylation of Hydroxybenzaldehyde

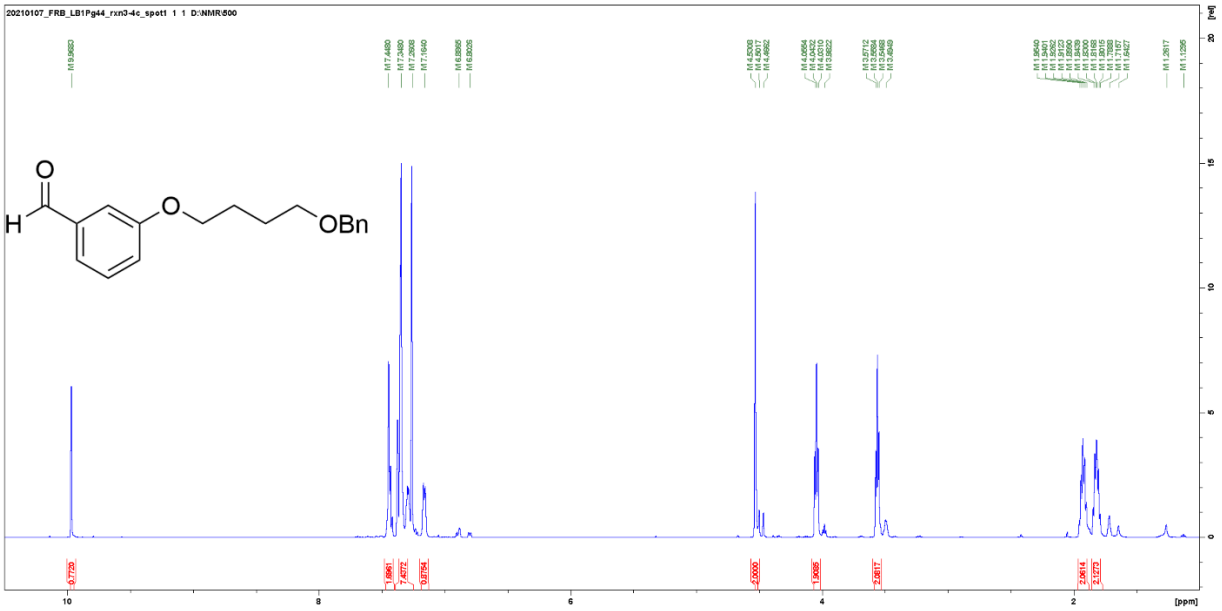
n = 3 – ^{1}H NMR, 500 MHz, $CDCl_3$:



n = 4 – ^{1}H NMR, 500 MHz, $CDCl_3$:

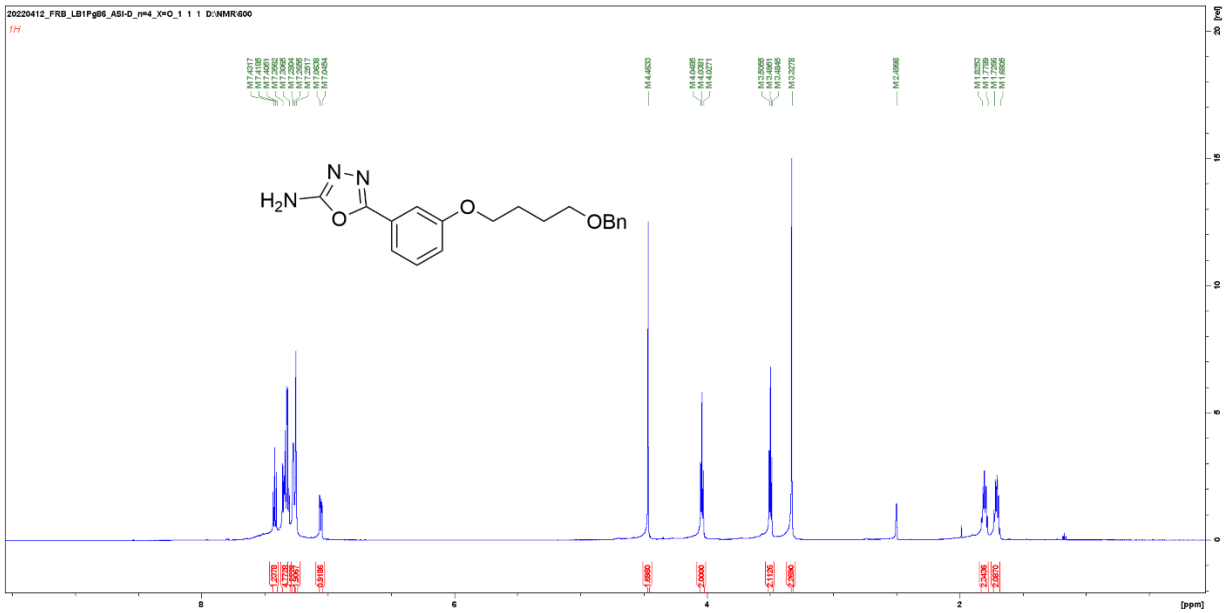


n = 4 - ^{1}H NMR, 500 MHz, $CDCl_3$:

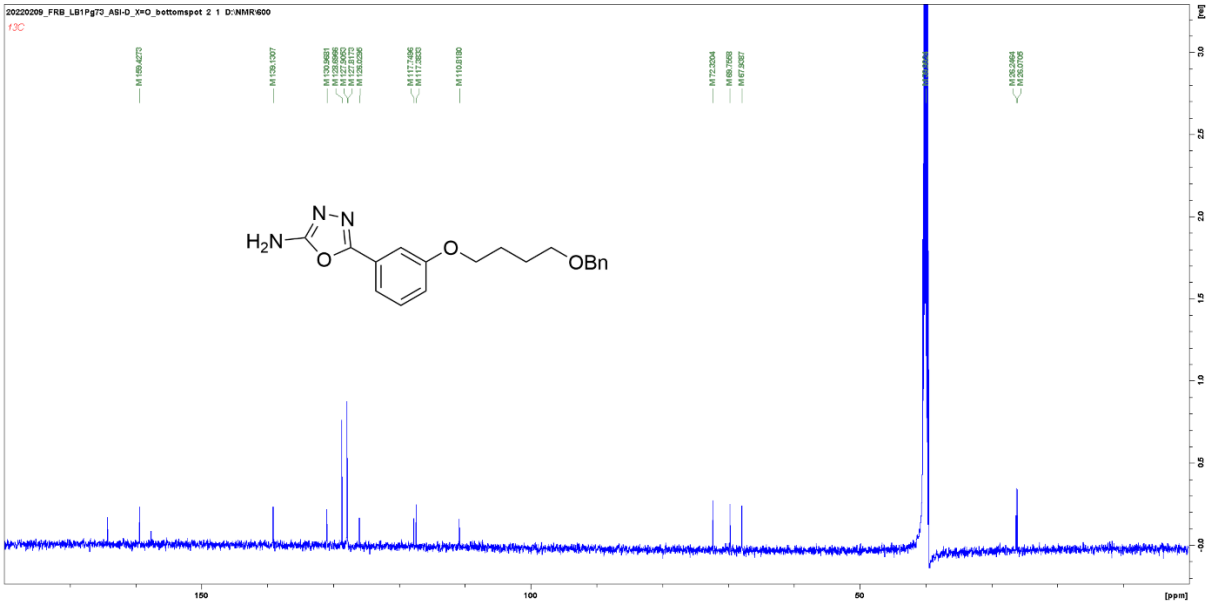


9.11.3 Synthesis of 2-Amino-1,3,4-oxadiazoles and 2-Amino-1,3,4-thiadiazoles

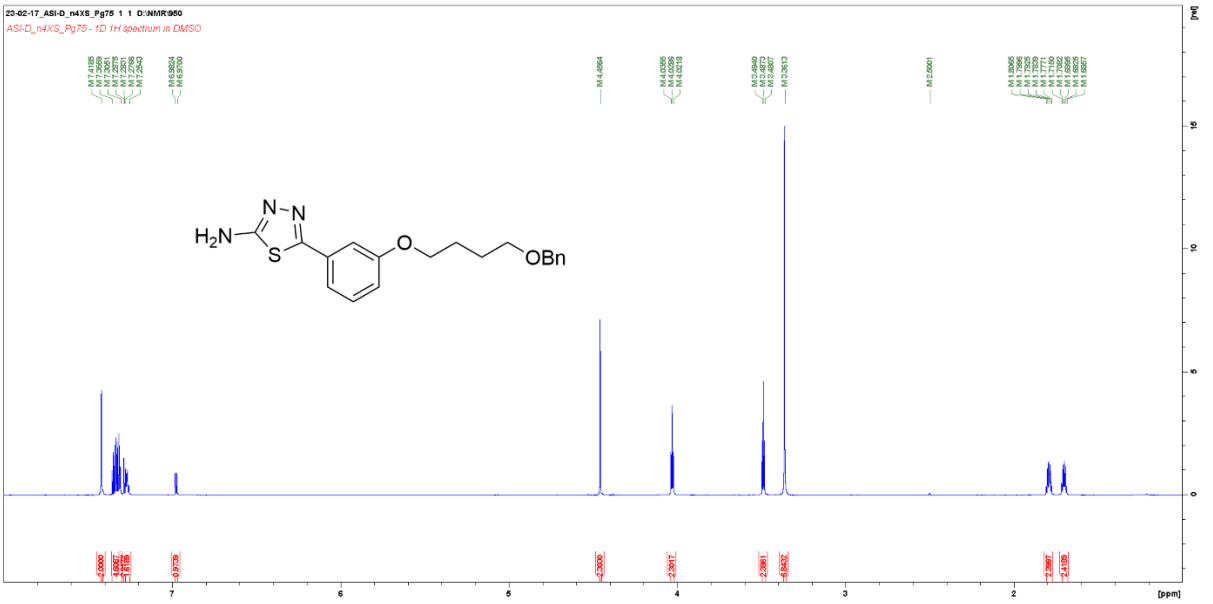
n = 4, X = O - ^{1}H NMR, 600 MHz, d_6-DMSO:



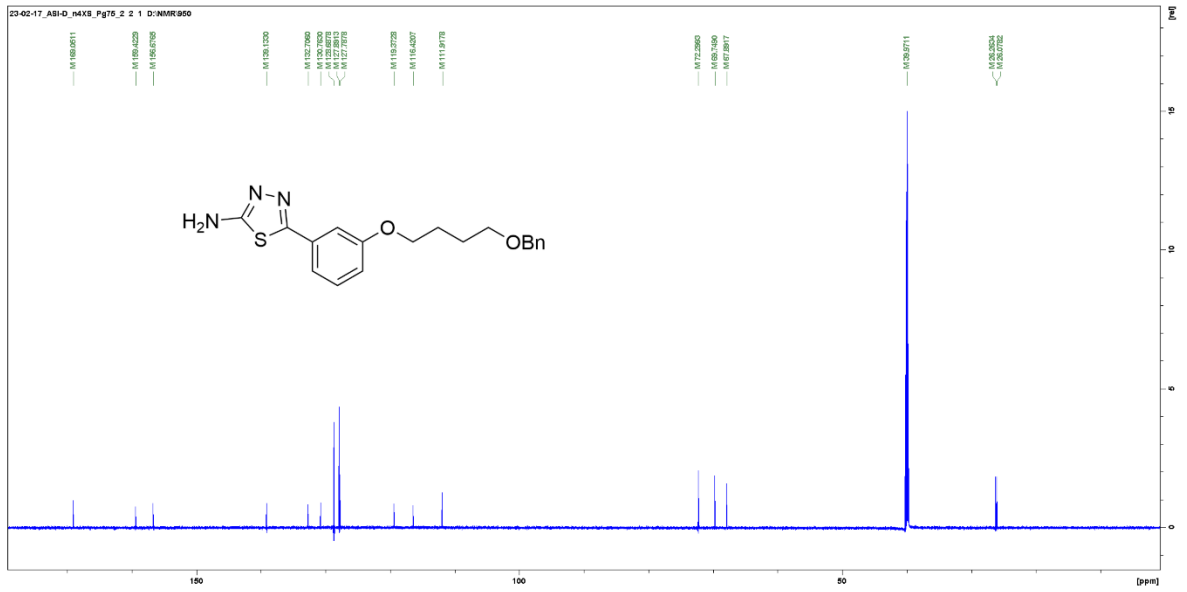
n = 4, X = O - ^{13}C NMR, 600 MHz, d_6-DMSO:



n = 4, X = S - ^{1}H NMR, 950 MHz, d_6-DMSO:

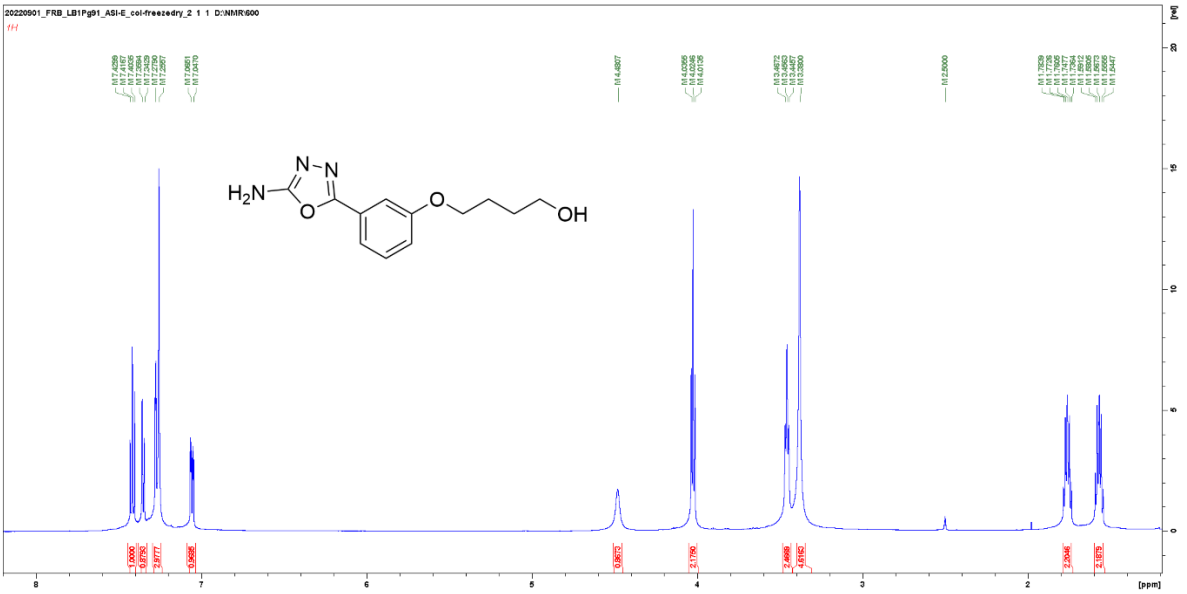


n = 4, X = S - ^{13}C NMR, 950 MHz, d_6-DMSO:



9.11.4 Deprotection of the Benzylated Alkyl Linker

n = 4, X = O - ^{1}H NMR, 600 MHz, d_6-DMSO:



9.12 Cryogenic Electron Microscopy (CryoEM) on Glutaraldehyde Cross-Linked (GXL) GluII

9.12.1 Materials & Methods

Glutaraldehyde cross-linking of GluII: A solution of GluII was mixed with 0.1% v/v glutaraldehyde in PBS and further diluted in PBS to final concentration of 0.2 mg/ml GluII and 0.005% v/v glutaraldehyde. The mixture was incubated at 4°C overnight before quenching of the excess glutaraldehyde with 10 mM Tris buffer (pH 8.5). The protein was concentrated to roughly 3 mg/ml in Vivaspin® 20 spin concentrators (50,000 MWCO, PES, Sartorius) and the buffer exchanged into 20mM HEPES pH 7.5, 120mM NaCl in preparation for SEC.

Cryo-EM grid preparation of glutaraldehyde cross-linked (GXL) GluII: Quantfoil 1.2/1.3 300mesh copper grids (Agar Scientific) were used for sample preparation. Grids were glow discharged before sample application. Sample application was carried out on a vitrobot (Thermo Fisher) operated at 4°C and 100% humidity. The grids were blotted for 2.5-5.0s and vitrified in liquid ethane. All cryo-grid preparations used 4.3 µL of protein sample in the concentration range 0.6-0.8mg/ml.

Grid Screening: Screening of the grids were carried on an Arctica microscope (Thermo Fisher Scientific) available at the COSMIC Cryo-EM facility at the Department of Biochemistry, University of Oxford. The microscope was operated at a voltage of 200 kV during the screening. Selected images from one of the grids are shown in Section 9.12.2. Further data collection and processing is required to determine the possibility of solving the structure of GluII heterodimer through cryo-EM.

9.12.2 Results & Discussion

The intrinsic flexibility of GIIβ has confounded efforts to obtain crystal structures of the full length GluII heterodimer. CryoEM is an alternative approach to collect structural data that displays a higher tolerance to conformational heterogeneity in a protein sample and may be an opportunity to examine the tertiary and quaternary structure of GluII with resolution comparable to X-ray crystallography.[149,150] In Appendix 9.10, I outline my attempts to crystallise full length GXL GluII. These experiments did not yield crystals and it is possible that the flexible GIIβ is locked into a number of different conformations following glutaraldehyde treatment and the heterogeneity this introduces may be incompatible with X-ray crystallography. However, cryoEM may be capable of distinguishing between these different conformations during the image processing stage, where 2D projections of the sample molecules imaged on the electron micrograph are grouped together based on their observed shape and structure. Once grouped, the images are averaged to enhance the signal-to-noise ratio and *in silico* computation can determine the angular relationship between the averaged images to construct a 3D projection of the particle of interest. The 3D projections can then be classified to identify different conformations of the same molecule. In the context of GXL GluII, it may be possible to solve the structure of different conformations of the protein using cryoEM, provided that the sample contains a sufficient number of particles of each one.

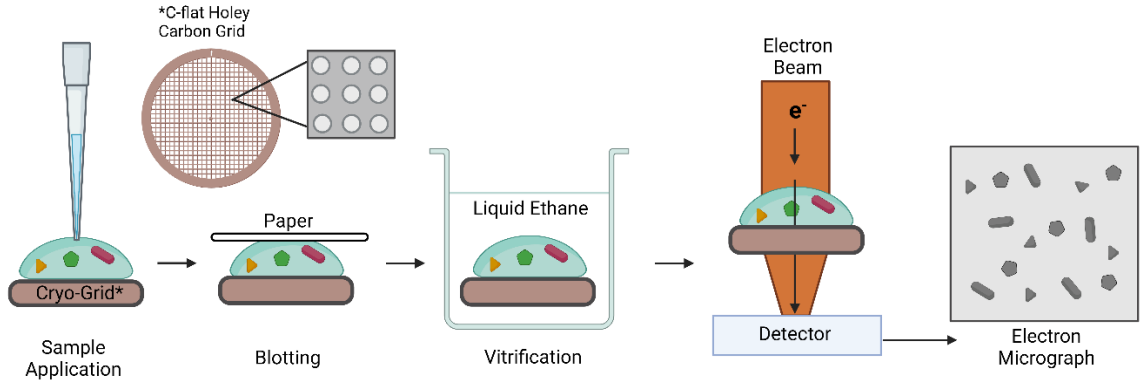


Figure VI – A schematic diagram of cryoEM workflow using a heterogeneous protein sample. Sample preparation was carried out using a vitrobot instrument (Thermo Fisher). Once the electron micrograph has been acquired, image processing software can select particles of the same shape and perform 2D classification followed by 3D reconstruction to produce models of the tertiary and quaternary structure of the proteins of interest.

To analyse GXL GluII, the sample was first applied to SEC to separate the GIIα:GIIβ heterodimer from the larger non-specific GXL complexes and grid preparation performed as described in Appendix 9.12.1. Images of the prepared grids from these preliminary experiments are shown in Figure VII and indicate that the particle distribution on the grids is uniform, however in some cases the thickness of the layer of vitrified ice that surrounds the grid caused problems. When the layer of ice is too thin, the protein sample does not completely cover the hole in the grid and the blank space indicates particle distribution is not uniform. If the ice layer is too thick, the resolution on the electron micrograph will be poorer. A balance between particle distribution and good resolution can be achieved by altering the blotting time and/or the volume of protein sample loaded. So far, electron micrographs are yet to be analysed and optimisation of the grid preparation protocol is necessary.

A limitation of this approach is that the observed GXL GluII conformations may not be biologically relevant. The glutaraldehyde connects free amine groups on lysine side chains non-specifically with a 4C linker, introducing constraints on the connections made based on distance and location. While this may inform us of the full length structure of GluII, in particular GIIβ, the conformations observed here may not provide any mechanistic insight with respect to the role of the MRH domain on GIIβ during catalysis.

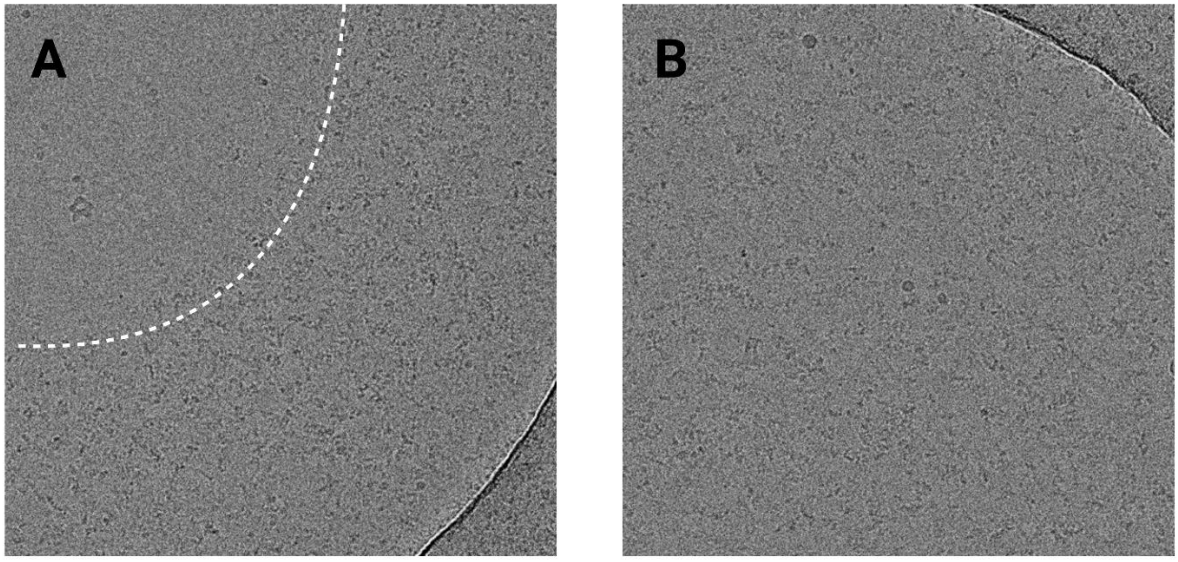


Figure VII – Electron micrographs obtained from cryoEM of GXL GluII. (A) An example of a hole in the grid where the layer of vitrified ice is too thin, leaving a gap in the centre of the hole indicated by a dashed line. (B) Uniform distribution of particles across the hole in the grid.

9.12.3 Future Work

In combination with repeating/continuing cryoEM experiments outlined above, I propose performing similar experiments using native GluII as opposed to GXL GluII. Given the capacity of cryoEM to parse particles from heterogeneous protein samples based on shared conformation, it may be possible to probe whether or not GIIβ is continuously flexible or exists in discrete, mechanistically relevant conformations characterised by local energy minima of the tertiary structure. The observation of several discrete conformations of GIIβ in full length GluII could greatly improve the current understanding of not only the complete GluII α/β interface, but also the mechanism by which the MRH domain is involved in the glucosidase function of the enzyme. Successful conformational analysis of full length GIIβ by cryoEM would require a sufficient number of each pose in the GluII sample to group them effectively, so in the case of continuous flexibility this technique may not be suitable.

If analysis of untreated full length GluII by cryoEM as described above proves ineffective, it may be necessary to first trap the protein into catalytically relevant conformations. This could be achieved in several ways. Firstly, treatment of GluII with a substrate glycan mimic that binds in the catalytic pocket but bears a non-cleavable terminal saccharide residue. An example of this type of substrate analogue would be a $Glc_1Man_9GlcNAc_2$ glycan with the terminal residues of the A-arm replaced by a non-cleavable D-glucal transglucosylation product to mimic the initial enzyme-substrate complex. Alternatively, using a 5F-Glc fluoride containing glycan of the same structure which would undergo the initial nucleophilic attack by the catalytic D543 residue on GIIα without subsequent hydrolysis, thereby trapping the enzyme-substrate complex as a covalent intermediate. Both approaches were used with success by Caputo *et al.* in the context of mono- and disaccharide mimetics. Application of these approaches with larger glycans is chemically feasible but obtaining these bespoke modified GluII substrate glycans may be expensive. A potentially cheaper approach to trapping the native enzyme-substrate complex would be to use SDM to produce inactive mutants of GluII, allowing the use of an unmodified GluII substrate glycan. Provided these experiments returned structural data of full length GluII in complex with a native-like substrate, it may be possible to determine the role of the MRH domain during catalysis and the plausibility of a nascent glycoprotein within the same system.

9.13 Circular Dichroism –gp120

By treating HEK293-F cells with the glucosidase inhibitor, NB-DNJ, and mannosidase inhibitor, Kif, during the expression of gp120 I was able to produce a glycoprotein substrate for GluI bearing predominantly Glc_3Man_9 N-glycans. Treatment of this substrate with GluI generated the GluII substrate glycans (Glc_2Man_9). Theoretically, this treatment inhibits the calnexin cycle for the transfected HEK293-F cells and may interfere with the quality of the fold for the recombinant protein. To test this, I analysed gp120 with and without glycosidase inhibitor treatment during expression by CD (Figure VIII). The results show that the two recombinant gp120 proteins have very similar CD profiles, indicating that the glycosidase inhibitors have not the significantly affected the protein secondary structure. It is possible that glycosidase treatment interferes with the variable loop regions of gp120, though CD is not suitable for detecting these differences. Given that the ER glucosidases cleave glucose residues independently of protein structure, the quality of the gp120 fold was not explored further in the context of the enzymology assay outlined in Section 4.4.6.

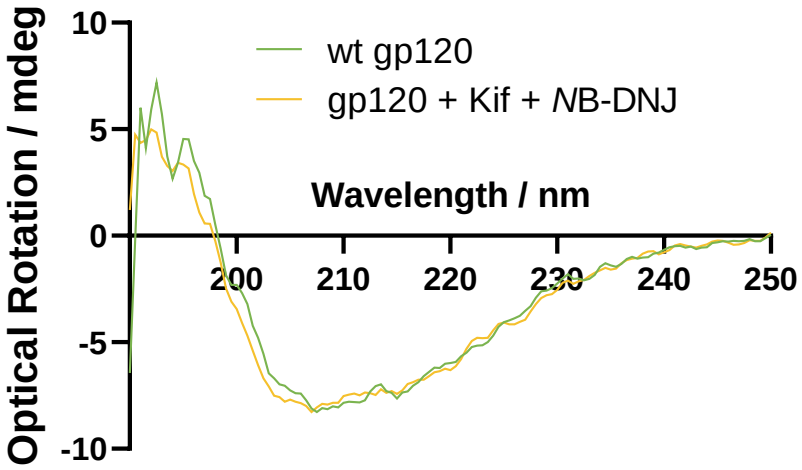


Figure VIII - CD spectra of gp120 expressed in HEK293-F cells with and without glycosidase inhibitors. Recombinant gp120 was expressed according to the protocol outlined in Sections 4.3.7.