

A Fragment-Based Drug Discovery Approach Targeting Human UDP-Galactose-4- Epimerase

William Browne

Imperial College London (ICL)
Department of Chemistry
PhD Thesis
2023

Supervisors:
Dr. Benjamin Schumann, ICL
Dr. Jacob Bush, GSK



Declaration of Originality

I, William Browne, confirm that the work discussed here are my own. Information, data, and results included from published material or collaborators have been properly credited and referenced. The project was undertaken at the Francis Crick Institute and funded by Imperial College London and GSK.

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution No Derivatives 4.0 International Licence (CC BY-ND).

Under this licence, you may copy and redistribute the material in any medium or format for both commercial and non-commercial purposes. This on the condition that; you credit the author and do not distribute modified versions of the work.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

O-GalNAc glycosylation is a ubiquitous post translational modification and plays a crucial role in a variety of inter- and intracellular processes. However, our understanding of the precise mechanisms by which O-GalNAc glycans act is still limited, in part due to a lack of appropriate chemical tools. Mucins are a family of heavily O-GalNAc glycosylated glycoproteins that have pro-tumorigenic properties and are thus a well validated therapeutic target. Essential to the function and expression of mucins are their characteristic dense array of O-GalNAc glycans. A small molecule inhibitor of O-GalNAc glycosylation will provide a useful chemical tool for the study of glycans and may offer a novel method of therapeutic intervention against mucin expressing cancers.

The enzyme UDP-Galactose-4-epimerase (GALE) is required for the biosynthesis of the essential O-GalNAc glycosylation precursor UDP-GalNAc. A small molecule inhibitor of GALE would thus inhibit the O-GalNAc biosynthetic pathway and subsequent glycosylation. Therefore, this manuscript describes the screening and optimisation of a GALE inhibitor using a fragment-based drug discovery approach. Non-covalent fragment screening was performed with cutting edge crystallographic screening. An initial hit fragment was elaborated using structure guided design to a nanomolar binder and low micromolar inhibitor. In parallel, covalent fragment screening has been used to validate GALE as tractable by covalent molecules. With this validation, a ligand first approach was used to develop a covalent inhibitor of GALE from the identified non-covalent fragment binders.

In summary, two small molecule lead inhibitors of GALE with orthogonal mechanisms of inhibition have been developed towards GALE. Providing a robust starting point from which to develop tool compounds suitable for application in cells.

Acknowledgments

I must start by thanking the many people who without their support, patience, and willingness to listen to me vent, I would not have managed to produce this manuscript. It certainly takes a village to produce a PhD. First to my supervisor, Dr. Ben Schumann, who has defined my scientific career not only by giving me the opportunity and extraordinary privilege of joining his group, but by setting such an exemplary standard for what an academic supervisor and mentor should be. Your dedication and commitment to those who work with you is something I hope to emulate.

To all those in the Schumann lab (who are now far too many to list) I would not have achieved a fraction of what is discussed here without you. Special mention of course to Caden, who worked so tirelessly on this project with me. Thank you to Bea and Sophie for managing to sit either side of me as I was writing, yoyoing between calm and stressed. The entire lab has made the time during my PhD not just academically rewarding but a moment in my life that I will always reflect on with a smile. I consider myself so fortunate to have been able to run so many successful and failed experiments with you all for company.

I am also extremely grateful to those in the structural biology, high throughput, and chemical biology STP platforms at the Crick, who consistently go above and beyond. Particularly Chloe Roustan who helped an uncertain chemist find their feet in a biology lab, and Dr. Stephane Mouilleron and Dr. Andy Purkiss who have contributed so much time and energy to the project. I am also extremely grateful to my industrial supervisor Dr. Jacob Bush, and unofficial supervisor Dr. Jonathan Pettinger, for all their mentoring, guidance, and commitment to making things work.

More personally, I must thank the friends I have from time in Southampton. They have been a rock that I can cling to when things are difficult. As well as a source of much needed distraction, “Shreddies”, and great memories. I am so fortunate to have the support of such wonderful friends. I must also express how grateful I am to my two housemates, Michael and Vicky, who have been both a pleasure to live and always been around to lean on. You both

somehow made student living in London a wholly positive experience, in addition to keeping me relaxed when writing.

Last and foremost I must thank my amazing family who have kept me on my feet throughout not just the last four years but all the ones before too. Especially to my Mum and Dad, I have to say how grateful I am for all your belief and hard work. A PhD takes a lot of patience and resilience; I have no doubt that I developed these through years spent learning from you both.

Table of Contents

Declaration of Originality	1
Copyright Declaration.....	2
Abstract	3
Acknowledgments	4
Table of Contents.....	6
List of Figures	9
List of Tables.....	12
Acronyms and Abbreviations	13
Chapter 1 – Thesis Introduction.....	15
1.1 Glycosylation	15
1.1.1 Glycans and Glycosylation Overview	15
1.1.2 O -GalNAc Glycans.....	17
1.2 Mucins	20
1.2.1 Mucin Structure and Biosynthesis.....	20
1.2.2 Mucin Function in Healthy/Normal Cells	22
1.2.3. Mucin Function in Cancer	25
1.2.4 Mucins as a Therapeutic Target.....	28
1.3 UDP-Galactose-4-Epimerase (GALE)	30
1.3.1 Overview	30
1.3.2 GALE as a Target for O-GalNAc Glycosylation Inhibition	31
1.3.3 GALE “Druggability”	34
1.4 Inhibitor Development	36
1.4.1 Phenotypic and Target-Based Screening	36
1.4.2 Fragment-Based Drug Discovery	38
1.5 Project Aims.....	40
1.5.1 Aims.....	40
Chapter 2 Protein Expression, Characterisation and Crystallography	41
2.1 Chapter Introduction	41
2.1.1 Protein Expression, Purification and Characterisation	41
2.1.2 Crystallography in Drug Discovery	44
2.2 Results and Discussion.....	46
2.2.1 Activity Characterisation	46
2.2.2 GALE Crystallography – Initial Screen	48
2.2.3 GALE ITC Characterisation	54
2.2.4 GALE Ammonium Sulfate Precipitation	56
2.2.5 GALE Crystallography – Seeding.....	59
2.3 Chapter 2 Summary	62

2.4 Methods	63
2.4.1 Cloning and Protein Expression	63
2.4.2 Epimerisation Assay	63
2.4.3 Protein ITC Characterisation	64
2.4.4 Protein Crystallography	65
2.5 Chapter 2 Crystal Structure “Table 1” Validation	67
3. Chapter 3 Non-Covalent Fragment Screening and Elaboration	70
3.1 Chapter Introduction	70
3.1.1 Chapter Overview.....	70
3.1.2 Fragment Screening Methods	70
3.2 Results and Discussion.....	73
3.2.1 DSI-Poised Library Thermal Shift Fragment Screening	73
3.2.2 Crystallographic Screening and Hit Validation.....	78
3.2.3 Elaboration of the fragment X0213 toward a GALE lead compound	88
3.2.4 WBX4 Elaboration	97
3.3 Chapter 3 Summary	106
3.4 Methods	107
3.4.1 DSF Thermal Shift Assay Optimisation	107
3.4.2 DSF Thermal Shift Screening	107
3.4.3 XChem Fragment Screening	108
3.4.4 XChem Crystal Data Collection	108
3.4.5 ICM Virtual Screening.....	109
3.4.6 ITC Fragment Assays.....	110
3.4.7 Fragment Inhibition Assays	110
3.4.8 Expression and Purification of GALE from E.coli	111
3.4.9 Data Collection for X0213 Analogue Structures.....	111
3.5 Chapter 3 Crystal Structure “Table 1” Validation	113
Chapter 4 Covalent Fragment Screening and Characterisation.....	117
4.1 Chapter Introduction	117
4.1.1 Covalent Fragments.....	117
4.1.2 Covalent Inhibitors	118
4.1.3 Activity-Based Protein Profiling (ABPP)	120
4.1.4 GALE as a Covalent Fragment Target.....	121
4.2 Results and Discussion.....	123
4.2.1 Aryl Sulfonyl Fluoride Library Screening.....	123
4.2.2 GSK Covalent Fragment Hit Validation	129
4.2.3 GSK Covalent Fragment Modification Site ID Experiments	133
4.2.4 XChem Covalent Fragment Analogues	136
4.2.5 XChem Covalent Analogue Design and Docking.....	138
4.2.6 XChem Covalent Analogue Intact-LCMS Screening and SAR	140
4.2.7 XChem Covalent Analogue Docking Evaluation.....	143
4.2.8 Effect of WB-C1 on GALE Activity	145
4.3 Chapter 4 Summary.....	146
4.4 Methods	148
4.4.1 Intact LCMS	148
4.4.2 Covalent Fragment Screening	149
4.4.3 Covalent Fragment Validation	149
4.4.4 Covalent Fragment Kinetics Assays	149

4.4.5 Covalent Fragment Activity Assays	150
4.4.6 Tryptic Digest LCMS/MS	150
4.4.7 Covalent Fragment Docking	151
Chapter 5 Conclusions and Future Work	152
5.1 Conclusions and Summary	152
5.1.1 Aims.....	152
5.1.2 Protein Characterisation and Crystallography	152
5.1.3 Non-Covalent Fragments.....	152
5.1.4	154
5.2 Future Work	155
5.2.1 Non-Covalent Fragments.....	155
5.2.2 Covalent Fragments.....	155
References.....	158

List of Figures

Figure 1: N- and O-Linked Glycosylation.	16
Figure 2: O-GalNAc glycan biosynthetic pathway.....	17
Figure 3: Common core and truncated/cancer associated O-glycans.	19
Figure 4: Structure and function of transmembrane and secreted mucins.	21
Figure 5: Representation of the differences in glycan structure between healthy (left) and cancer associated mucins (right).	26
Figure 6: A) GALE epimerisation mechanism. B) Leloir pathway.	31
Figure 7: GALE KO K562 cell phenotype.	33
Figure 8: GALE UDP-Sugar binding site structure.	35
Figure 9: Structures and IC ₅₀ values for the inhibitors of GALE identified by the Bertozzi and Ferguson labs.	36
Figure 10: FBDD, rule of 3 and elaboration strategies.....	39
Figure 11: GALE fusion construct and purification.....	42
Figure 12: Illustration depicting reversible ligand binding with K _D	44
Figure 13: GALE Michaelis-Menten activity validation.	47
Figure 14: Initial crystallography condition screening results.	49
Figure 15: Crystal morphology summary from initial crystallography condition screening.	50
Figure 16: Structures obtained for crystals A3, H6 and D10.	53
Figure 17: Insect cell expressed GALE ITC characterisation.	55
Figure 18: (NH ₄) ₂ SO ₄ precipitated GALE ITC assay thermogram (left) and Integral plot (right) with UDP-Glc at 250 μM	57
Figure 19: Crystal structures obtained for crystals C11 and F8.....	58

Figure 20: Hit rate summary for crystallography plates set up using crystal seeding.	60
Figure 21: Representative images showing crystal morphology from crystal seeding experiments. ...	61
Figure 22: Crystal structures obtained for crystals E11 and D9.	61
Figure 23: Thermal shift assay condition optimisation.	74
Figure 24: Thermal shift fragment screening workflow.	75
Figure 25: Thermal shift fragment screening results.	76
Figure 26: Fragment screening hit results, ranking, and structures.	77
Figure 27: Illustration of the XChem screening process.	79
Figure 28: XChem screening fragment structures and X0121 and X0213 binding sites.	80
Figure 29: XChem hit X0213-protein interaction analysis.	82
Figure 30: XChem hit X0121-protein interaction analysis.	83
Figure 31: XChem hit X0373-GALE crystal structures.	84
Figure 32: X0213 biophysical characterisation	86
Figure 33: X0121 and X0213 inhibition assay results.	87
Figure 34: X0213 analogue series one structures and characterisation workflow.	90
Figure 35: Analogue series one docking, thermal shift, and inhibition assays results.	91
Figure 36: WBX4 ITC assay characterisation.	93
Figure 37: ITC assay results for <i>E. coli</i> expressed GALE with UDP-Glc showing both thermogram and integral plots.	94
Figure 38: Crystal and docking structures for fragment WBX4.	96
Figure 39: WBX2 chemical and crystal structure	97
Figure 40: Structures of WBX4 and its analogues separated by modification.	98

Figure 41: Analogues series two characterisation.	101
Figure 42: WBX9 ITC assay results with thermogram and integral plots.	102
Figure 43: Analogue series 2 crystal structures and binding interaction analysis.	105
Figure 44: Covalent fragment kinetics.	118
Figure 45: FDA approved covalent drugs.	120
Figure 46: Activity based protein profiling.	121
Figure 47: Location of covalent fragment target Tyr211 in GALE.	122
Figure 48: A) Intact protein LCMS assay.	123
Figure 49: Covalent fragment screening results.	125
Figure 50: Change in GALE binding site structure upon substrate binding.	126
Figure 51: Covalent fragment hit docking results and structures.	128
Figure 52: Covalent fragment hit validation and kinetics assay results.	131
Figure 53: GSK Covalent fragment hit inhibition assays.	133
Figure 54: GSK covalent fragment protein modification site ID experiment results.	136
Figure 55: Non-covalent to covalent fragment candidates and structures.	137
Figure 56: Covalent compounds WB-C1 – WB-C8 structures and virtual docking results.	139
Figure 57: Covalent fragment intact-LCMS results.	142
Figure 58: Covalent fragment docking evaluation.	144
Figure 59: Covalent fragment WB-C1 effect on GALE activity.	146
Figure 60: Non-covalent inhibitor development summary.	154
Figure 61: Future elaboration strategy for fragment WB-C1.	156

List of Tables

Table 1: GALE kinetics parameters.....	47
Table 2: Diffracting crystal crystallisation conditions from initial condition screen (1EK5 based).....	50
Table 3: Structure validation and information for the crystals analysed from wells A3, H6 and D10. ("Table 1" values).	51
Table 4: ITC results for different GALE preparations.	55
Table 5: Structure validation and information for the ammonium sulfate precipitated GALE crystals analysed from wells F8 and C11.....	58
Table 6: Structure validation summary for the GALE crystals grown using seed stock from wells D9 and E11.	62
Table 7: HPLC assay elution conditions.	64
Table 8: Full statistical validation for crystals A3, H6 and D10.	67
Table 9: Full statistical validation for crystals F8 and C11.	68
Table 10: Full statistical validation for crystals D9 and E11.....	69
Table 11: ITC assay results for X0213 binding to recombinant GALE.....	85
Table 12: X0213 and analogues ΔT_m values calculated from DMSO controls.....	92
Table 13: WBX4 and analogues ΔT_m values calculated from DMSO controls.....	102
Table 14: Full "Table 1" validation for the XChem hit crystal structures.	113
Table 15: Full "Table 1" validation for the structures of WBX4 and WBX2.	114
Table 16: Full "Table 1" validation for the structures of WBX17 and WBX15.	115
Table 17: Full "Table 1" validation for the structures of WBX12 and WBX9.	116
Table 18: GSK covalent fragment site ID experiment fragment modified peptides.	135
Table 19: Solvent B% change with time	148

Acronyms and Abbreviations

ABPP	Activity-Based Protein Profiling
ADC	Antibody-Drug Conjugate
AEBSF	4-(2-Aminoethyl) benzenesulfonyl fluoride
AGX1	Alanine-Glyoxylate Aminotransferase 1
ATP	Adenosine Triphosphate
BAC	Bacterial Artificial Chromosome
CAR-T	Chimeric Antigen Receptor T-cell
CHO	Chinese Hamster Ovary
CNS	Central Nervous System
COX	Cyclooxygenase
DLS	Diamond Light Source
DMSO	Dimethyl sulfoxide
DNA	Deoxyribosenucleic acid
DSF	Differential Scanning Fluorimetry
DSI	Diamond-Structural genomics consortium-iNEXT
DTT	Dithiothreitol
ECEPP	Empirical Conformational Energy Program for Peptides
EDTA	Ethylenediamine <i>N,N,N',N'</i> -Tetraacetic Acid
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
EMT	Epithelial-Mesenchymal Transition
ERK	Extracellular signal-Regulated Kinases
ESA	Early Stage Assessment
ESI	Electrospray Ionisation
FBDD	Fragment-Based Drug Discovery
FBS	Fetal Bovine Serum
FDA	Food and Drug Administration
FF	Force Field
GALE	UDP-Galactose-4-Epimerase
GST	Gluthathione-S-Transferase
HAC	Heavy Atom Count
HCD	Higher-Energy C-trap Dissociation
HEK	Human Embryonic Kidney
HER2	Human Epidermal Growth Factor Receptor-2
HPLC	High Performance Liquid Chromatography
HRV	Human Rhinovirus
IAA	Iodoacetamide
ICAM	Intercellular Adhesion Molecule
ICM	Internal Coordinate Mechanics

IsoTOP	Isotopic Tandem Orthogonal Proteolysis
ITC	Isothermal Titration Calorimetry
KRAS	Kirsten Rat Sarcoma
LCMS	Liquid Chromatography Mass Spectrometry
MCS	Multiple Cloning Site
MMFF	Merck Molecular Force Field
MOI	Multiplicity Of Infection
MRC	Medical Research Council
MS	Mass Spectrometry
MWCO	Molecular Weight Cut Off
NAD	Nicotinamide Adenine Dinucleotide
NADH	1,4-Dihydronicotinamide Adenine Dinucleotide
NMR	Nuclear Magnetic Resonance
NSAID	Non-Steroidal Anti-Inflammatory Drug
NSCLC	Non-Small Cell Lung Cancer
PAMP	Pathogen-Associated Molecular Patterns
PanDDA	Pan-Dataset Density Analysis
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEG	Polyethylene Glycol
POI	Protein of Interest
PRR	Pattern Recognition Receptor
PTS	Proline Threonine Serine
RMSD	Relative Mean Standard Deviation
RTK	Receptor Tyrosine Kinase
SAR	Structure Activity Relationship
SBDD	Structure-Based Drug Design
SDS-PAGE	Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis
SEA	Sea urchin sperm protein, Enterokinase, and Agrin
SPR	Surface Plasmon Resonance
TCEP	Tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic Acid
TLR	Toll-Like Receptor
UAP1	UDP- <i>N</i> -acetylglucosamine pyrophosphorylase 1
UDP	Uridine Diphosphate
UTP	Uridine Triphosphate
VLS	Virtual Ligand Screening
XCE	XChem Explorer

Chapter 1 – Thesis Introduction

1.1 Glycosylation

1.1.1 Glycans and Glycosylation Overview

Glycosylation is a ubiquitous post translational modification that can broadly be defined as the addition of carbohydrates (monosaccharide or polysaccharide) to other biomolecules such as proteins and lipids via a glycosidic bond.¹ Glycosylation is essential to the activity and structure of other biomolecules and is involved in a vast array of cellular processes across all forms of life.² For proteins, glycosylation is introduced either in the nucleocytoplasm or - much more abundantly - in the secretory pathway from suitable precursors. The precursors are activated with anomeric phosphates, either as nucleotide-sugars or dolichol pyrophosphate-linked sugars. The biosyntheses of the activated precursors are complex, but most can either be made through salvage pathways or interconversion of nucleotide-sugars, with the *de novo* synthesis of most tracing back to glucose.²

Once in the secretory pathway, the extent to which a protein is glycosylated is largely controlled through factors such as 1) the protein amino acid sequence, 2) expression levels of glycosyltransferases (GTs, carbohydrate adding enzymes) and glycosidases (carbohydrate removing enzymes), 3) activated sugar nucleotide concentrations (the substrate of most GTs). 4) Spatial factors i.e. the access of glycosylation enzymes to the protein substrate. The complex interplay between these factors makes glycosylation a highly dynamic process.^{3,4}

Broadly speaking, protein glycosylation in mammalian cells can be split into two main types; N-Linked glycosylation and O-Linked glycosylation.⁵ N-linked glycosylation is defined by the formation of an *N*-glycosidic bond between an *N*-acetylglucosamine (GlcNAc) sugar and the side chain amide group of Asn (**Figure 1A**).⁶ O-linked glycans are defined by a glycosidic bond between a monosaccharide and the side chain oxygen of a Ser, Thr or Tyr residue in a protein backbone (**Figure 1B**).⁶ O-linked glycosylation in mammalian cells is dominated by O-GalNAc glycosylation to *N*-acetylgalactosamine (secretory pathway) and O-GlcNAc glycosylation

(nucleocytoplasm). After the addition of the initial monosaccharide in O-GalNAc glycosylation, elaboration with other sugars can then occur.

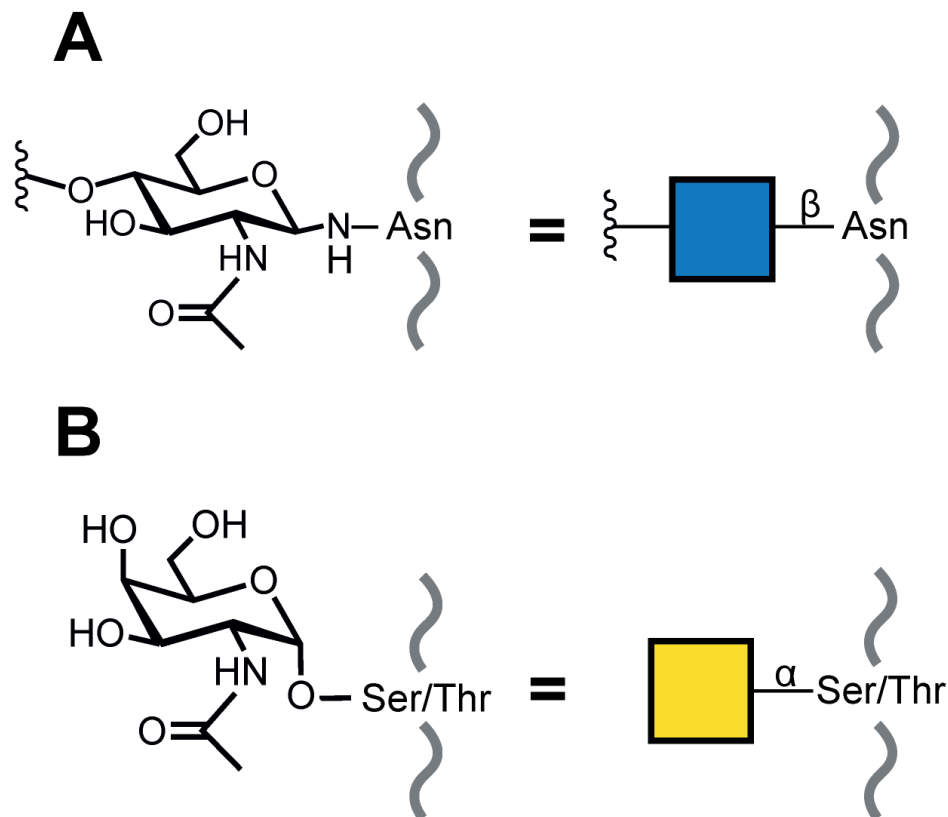


Figure 1: N- and O-Linked Glycosylation.

A) Structure of the glycan-peptide linkage of N-linked glycans with symbol used for GlcNAc (blue square). B) Structure of the glycan-peptide linkage of O-linked glycans with symbol used for GalNAc (yellow square).

Both types of glycosylation serve two broad functions: 1) structural roles, including protein stability, folding, and solubility. 2) Signalling and recognition roles, where glycans enable Cell-cell and cell-matrix interactions.^{7,8} The interactions between cell surface glycans and the receptors on other cells are key to cellular communication, particularly in the context of immune response and cell adhesion.⁹ It is therefore unsurprising that, as with most dynamic and essential physiological processes, dysregulation of glycosylation is associated with a range of diseases.^{5,10}

1.1.2 O-GalNAc Glycans

1.1.2.1 Initiation

O-GalNAc glycans are the major type of O-glycan found on the surface of mammalian cells. Essential to O-GalNAc glycosylation is the availability of the activated form of GalNAc, uridine diphosphate *N*-acetylgalactosamine (UDP-GalNAc).¹¹ UDP-GalNAc is synthesised in the cytosol by two possible routes: 1) a salvage pathway that reacts uridine triphosphate (UTP) with GalNAc-1-phosphate (GalNAc-1P), catalysed by UDP-*N*-acetylglucosamine pyrophosphorylase 1 (UAP1 or AGX1) (**Figure 2**).¹² 2) Direct epimerisation of the *gluco*-configured analogue, UDP-GlcNAc by the enzyme UDP-Galactose-4-epimerase (GALE) (**Figure 2**).¹²

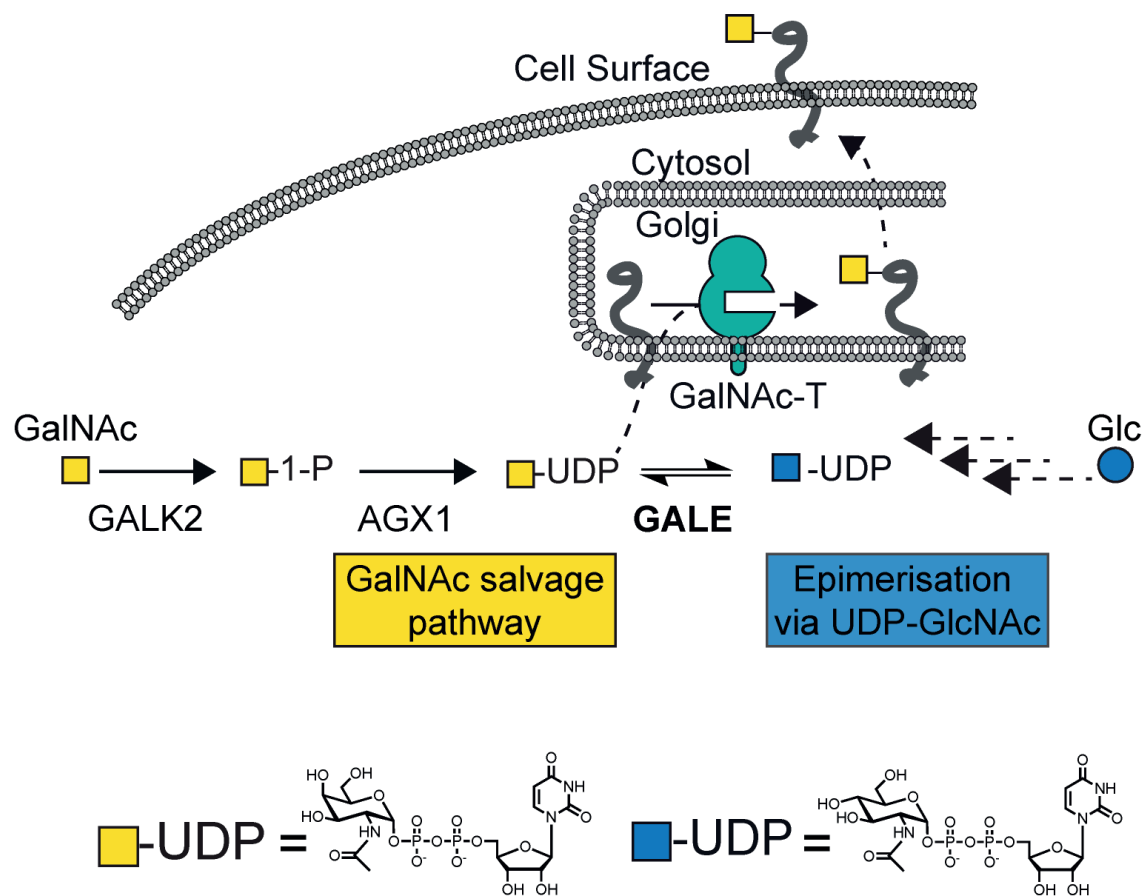


Figure 2: O-GalNAc glycan biosynthetic pathway.

Figure includes UDP-GalNAc biosynthesis from GalNAc and via epimerisation of UDP-GlcNAc. Peptides are subsequently glycosylated in the Golgi before translocation to the cell surface.

UDP-GalNAc is the substrate of a class of GTs called the polypeptide *N*-acetylgalactosaminyl transferases (GalNAc-Ts).¹³ The GalNAc-Ts are a family of at least twenty isoenzymes that

transfer GalNAc from UDP-GalNAc to a Ser, Thr or Tyr in a protein. Within the GalNAc-T family, there is considerable redundancy, with many having overlapping substrate specificities.¹³ Crucially, all the GalNAc-Ts are found in the Golgi.³ Since UDP-GalNAc is biosynthesised in the cytosol, UDP-GalNAc is transported into the Golgi by specialised antiporters such as SLC35D1.¹⁴

1.1.2.2 O-GalNAc Glycan Structures

As mentioned above, all O-GalNAc glycans begin with a glycosidic bond between a GalNAc sugar and a Ser or Thr in a protein. The addition of a single GalNAc to a Ser or Thr is defined as the Tn antigen (**Figure 3**).^{15,16} After this initial step, additional sugars are sequentially added by a range of GTs as the protein passes through the Golgi.³ Thus, the process of glycan elaboration and the structures produced are defined by GT activity.¹⁷ The results of this are that there are defined initial O-GalNAc glycan core structures, core 1-4 (**Figure 3**).¹⁸ Each core can be further elaborated with additions of GalNAc, GlcNAc, galactose (Gal), glucose (Glc), fucose (Fuc) and sialic acid (Sia).¹⁹ Core 1 and 2 glycans are the most common and are found across all human cell types. Unelaborated core 1 glycans are referred to as T antigens (**Figure 3**). Core 3 and Core 4 are predominantly expressed in gastrointestinal and bronchial cells.² The Tn and T antigens are often sialylated on tumour cells, forming the STn and ST antigens, respectively that are truncated compared to O-glycans on healthy cells.²⁰

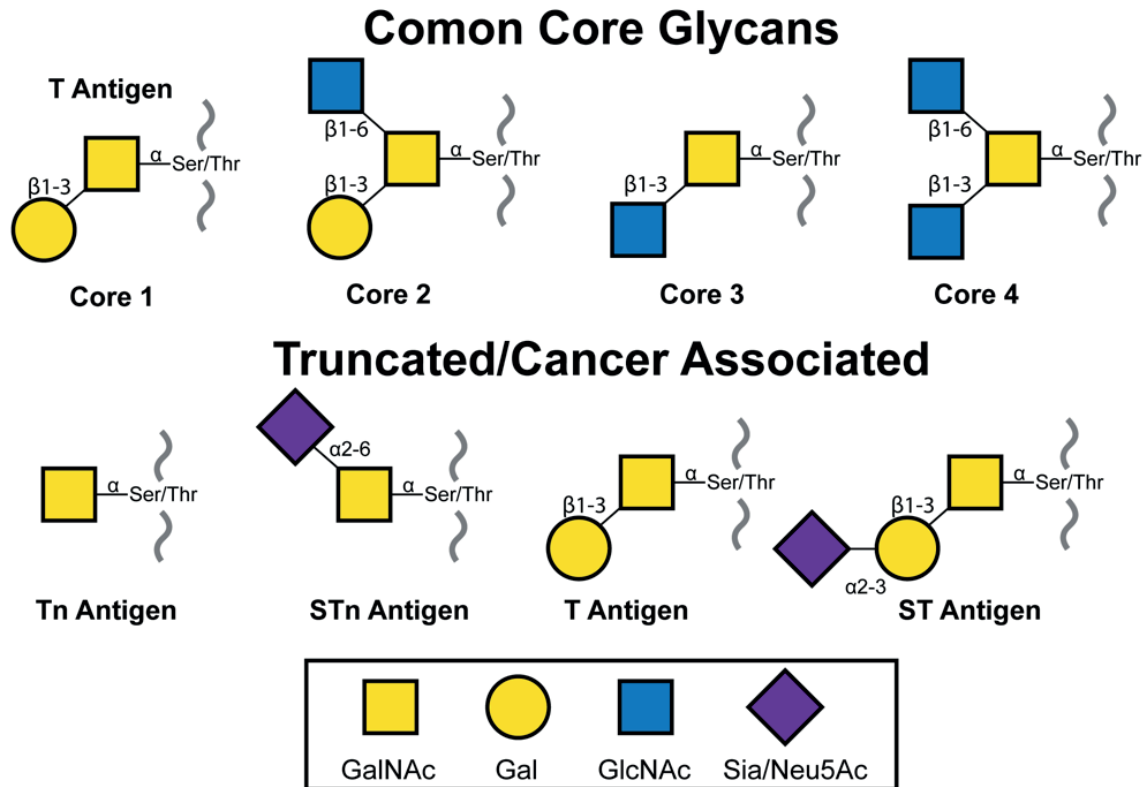


Figure 3: Common core and truncated/cancer associated O-glycans.

Core glycan structures (Top) and truncated/cancer associated glycan structures (bottom). Key shows (from left to right) GalNAc, Galactose, GlcNAc, and sialic acid (N-acetylneuramic acid).

1.1.2.3 O-GalNAc Glycans in Cancer

Truncated, tumour-associated O-glycans are significantly upregulated in a range of cancers and as such are of significant therapeutic interest.¹⁵ This was first noted by *G.F. Springer et al* who reported the increased expression of the Tn antigen in >90% of breast cancers.²¹ This was later expanded to link both the T and Tn antigens to cancer progression and differentiation.²² More recently a study by *T. B. Rømer et al* looked at the expression of the truncated O-glycans in >700 healthy and cancer tissue cores from breast, colon, lung, pancreas, skin, CNS and mesenchymal tissue.²³ The Tn antigen was overexpressed in the majority of breast, colorectal and pancreatic tumours. The STn antigen was predominantly upregulated in pancreatic and colorectal cancers. Notably, no Tn antigen was detected in any of the healthy tissue samples, while the STn was only observed in healthy gastrointestinal samples. This study highlights the dramatic changes in glycosylation between healthy and cancerous cells. Part of the reason for the increase in truncated glycans is that many cancers carry mutations that inhibit glycan elaboration. The enzyme core 1 β 3-galactosyltransferase (T-synthase) is responsible for

elaborating the Tn antigen to the T antigen.²⁴ T-synthase activity is thus linked to the amount of Tn antigen. The expression of T-synthase is controlled by a molecule chaperon called Cosmc.²⁵ Loss of function mutations in Cosmc have been observed in human cancers with high Tn and STn expression.²⁶

Unsurprisingly, the dramatic change in O-GalNAc glycosylation pattern between healthy and cancerous tissue is not superficial. The expression levels of the STn and Tn antigen have been seen to correlate with tumour progression and are thus associated with poor patient prognosis.^{27,28} Unfortunately, despite the clear connection between cancer progression and truncated glycans, a consensus on how they affect tumour development has not been reached.

Crucially, whilst many proteins are O-GalNAc glycosylated, so-called mucin-type glycoproteins (herein just mucins) are responsible for the bulk of the O-GalNAc glycans present on the surfaces of epithelial and many cancer cells. Therefore, in recent years mucins have gained attention as therapeutic targets in cancer.²⁹

1.2 Mucins

1.2.1 Mucin Structure and Biosynthesis

Mucins are a family of heavily O-GalNAc glycosylated glycoproteins (**Figure 4**). They have a unique “bottle brush” like structure consisting of a central protein backbone with a dense array of O-GalNAc glycan side chains.³⁰ The glycan structures that decorate them are mostly elaborated core 1-4. >50% of the molecular weight of a mucin is from the O-GalNAc glycans they carry.³¹ The essential structural feature that enables such extensive glycosylation are domains rich in Pro, Thr, Ser (PTS domains), often organised in tandem repeats.³² These PTS domains are highly variable with regards to length and sequence.

There are at least twenty-one mucins identified to date and these can be separated into two categories, secreted and membrane-bound. (**Figure 4**). Both of which have been heavily studied and reported on in several reviews.^{29,30,32–34} It is worth noting that whilst mucins have core similarities; in the case of some mucins, there are notable variations in structure,

expression profile, and function, not all of which can be discussed in this introduction. Additionally, there are also many so-called mucin-type glycoproteins that contain a mucin-like domain as part of a larger structure. As such, mucins are discussed below in as generalised a way as possible, with reference to examples of the most notable mucins.

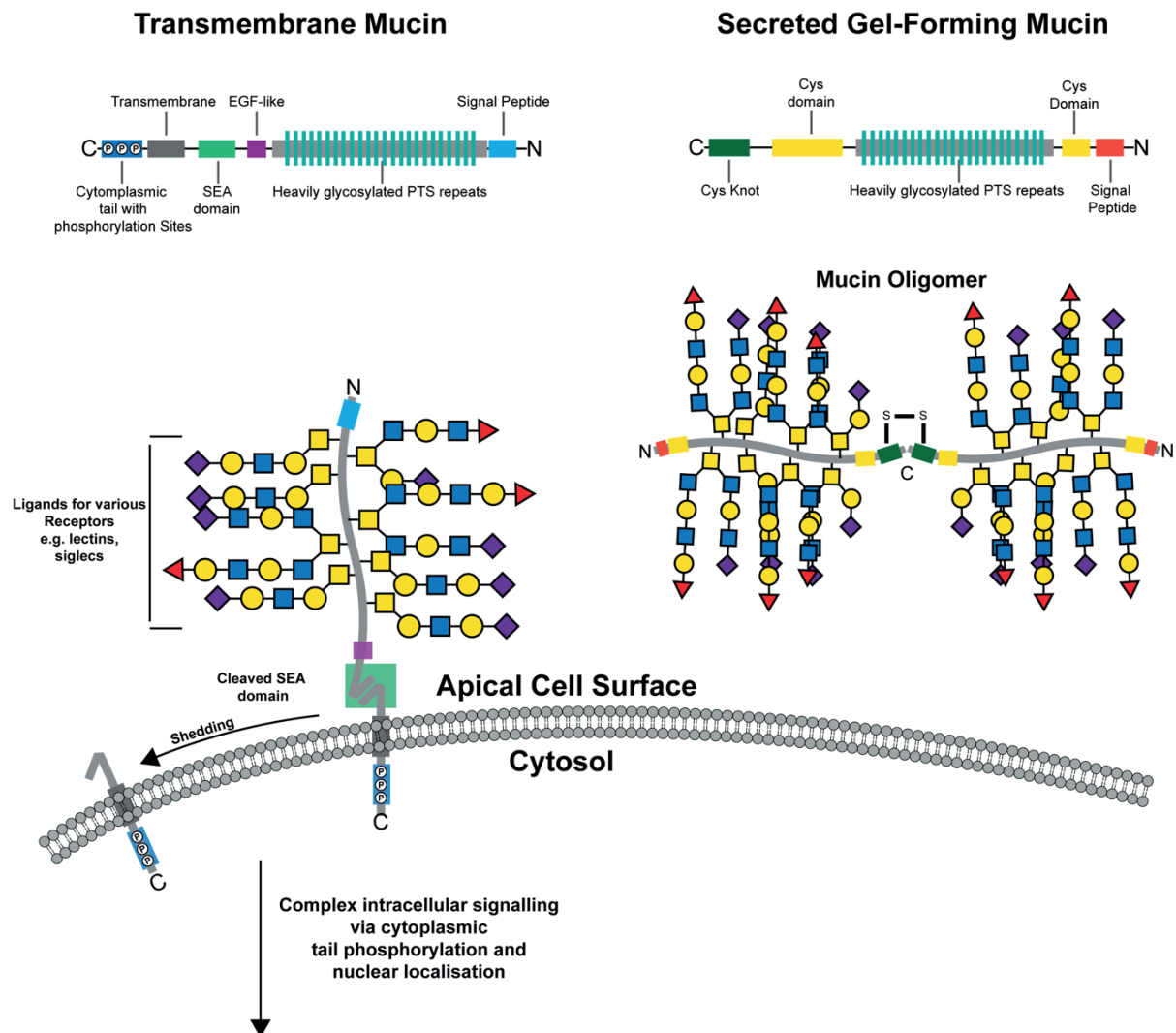


Figure 4: Structure and function of transmembrane and secreted mucins.

Diagram shows the simplified general structures of the transmembrane and secreted mucins (top). Diagram of the transmembrane and secreted mucins on the cell surface, including shedding of the extracellular domain, and signalling via the transmembrane cytoplasmic tail.

Transmembrane mucins consist of an extracellular N-terminal domain (and are thus type-1 membrane proteins) and a C-terminal cytoplasmic tail.³³ Generally, from N- to C- termini, the transmembrane mucins consist of PTS repeats, epidermal growth factor (EGF)-like, and sea

urchin sperm protein, enterokinase, and agrin (SEA) domains (**Figure 4**).^{30,35} However, the order of and presence of additional domains will vary between specific mucins (e.g. MUC4).³⁶ The SEA domains are common in many cell surface glycoproteins and have autoproteolytic properties.^{37,38} Mucins with SEA domains will undergo autoproteolysis to form heterodimers. Differences in the two monomers between mucins results in varied physiological roles.

The secreted mucins can be sub-divided further to gel-forming (e.g. MUC2, MUC5AC, MUC5B, MUC6) and non-gel-forming types (e.g. MUC7), with the former being the more common.^{39–43} Gel forming mucins are secreted primarily from salivary and stomach glands, as well as other intestinal and stomach cells.³⁰ In addition to the typical densely glycosylated PTS repeat regions, secreted gel forming mucins also have Cys rich regions that can be found from N- and C- termini (**Figure 4**). More specifically Cys rich regions called Cys knots (CKs) are often identified close to the C-terminal. These Cys rich regions enable oligomerisation and the build-up of visco-elastic gels.⁴⁴ Predictably, the non-gel forming secreted mucins lack these CK regions and so do not oligomerise.

Both secreted and transmembrane mucins are translated from single but distinct gene products. Transmembrane mucins undergo autoproteolytic cleavage in the endoplasmic reticulum (ER).^{37,45} This forms the mucin heterodimer via non-covalent interactions and the dimer is then trafficked to the Golgi. For secreted mucins, dimerisation also occurs via the CK regions.⁴⁴ Both types of mucins then pass through the Golgi where the PTS regions are glycosylated. Secreted mucins are then packaged into vesicles for secretion, whilst transmembrane mucins are transported to the cell surface.⁴⁶ Notably, in healthy epithelial cells transmembrane mucins are expressed only on the apical surface.³³

1.2.2 Mucin Function in Healthy/Normal Cells

1.2.2.1 Overview

In healthy cells, the primary function of mucins is in cell protection. Mucins are typically expressed by epithelial and immune cells, and their biological roles have been extensively reviewed by *J.P.M. Van Putten and K. Strijbis*.³³ Mucins form a major component of mucous and thus act to protect epithelial cells from damage, particularly in the event of inflammation

during immune responses. Notably, cell surface mucins serve additional functions over their secreted forms. The heavily O-glycosylated extracellular domain mainly acts as a protective barrier; however, this region can be released to leave only the short C-terminal unit. The C-terminal transmembrane region has been seen to act as a cell surface receptor, having roles in cell growth, signalling and oncogenesis. The diverse roles mucins and their domains are best illustrated by MUC1, the most extensively studied mucin.

1.2.2.2 Extracellular N-Domain and Glycans

The EC domain of mucins expressed in healthy cells has the primary function of cell protection, much like their secreted form. This protective role usually occurs in relation to infection.^{47,48}

For example, in a MUC1 deficient murine model infected with the bacteria *H. pylori* (a bacteria linked to pathologies such as stomach ulcers, gastritis and lymphoma) the induced expression of MUC1 resulted in significantly decreased *H. pylori* colonisation. This protective effect was not only due to steric interference of MUC1, but also through *H. pylori* binding MUC1 glycans via the blood group antigen A and sialic acid binding adhesins of *H. pylori*, and subsequent shedding of the extracellular domain as a decoy.⁴⁹ Additionally, MUC1 was also seen to reduce the severity of *H. pylori* infection due to its expression on immune cells that resulted in selective suppression of inflammatory cytokines via reduced inflammasome activation.⁵⁰

In another study, the flagellin on the surface of *P. aeruginosa* was identified as an adhesin for the MUC1 extracellular domain.⁵¹ This resulted in phosphorylation of Ser residues in the cytoplasmic tail and subsequent activation of the extracellular signal-regulated kinase (ERK) pathway (discussed further below).⁵²

The interaction of MUC1 was also observed with intracellular adhesion molecule-1 (ICAM-1) which resulted in increased migration of MUC1-expressing cells *in vitro*, suggesting a function for MUC1 in the regulation of cellular adhesion and motility.⁵³ However, this was observed in cancer cells and the role of MUC1 expression on the motility of immune cells is unclear.

It can therefore be said that in healthy cells, the densely glycosylated extracellular domains of mucins act largely as a defensive coating over epithelial cells, acting as decoys or barriers for invading pathogens. However, these domains are also linked to cellular adhesion, motility and inflammation through receptor and ligand binding, which results in internal signalling via the cytoplasmic tail of the C-terminal domain. Notably, our understanding of precisely what factors regulate the extracellular domain shedding is still poor. The mechanism by which binding of the extracellular domain affects the transmembrane N-terminal is also lacking.

1.2.2.3 C terminal domain

The C-terminal domain has been linked more directly to signalling, often as a result of shedding of the extracellular domain. In all transmembrane mucins, the cytoplasmic tail contains putative phosphorylation sites. The effects of phosphorylation have been most extensively shown with MUC1; several studies demonstrated that MUC1 C-terminal interactions with other proteins are under the control of these phosphorylation sites.^{54,55} The effects of phosphorylation on MUC1 and other mucins have been extensively reviewed by *K. C. Kim and E. P. Lillehoj, P.K. Singh and M. A. Hollingsworth*.^{56,57} In keeping with their protective roles, the mucin cytoplasmic tails often appear to be involved in signalling in response to infection, often to limit inflammation. This was demonstrated in a study where dendritic cell response was shown to be downregulated by MUC1 expression.⁵⁸

Moreover, toll-like receptors (TLRs) are a class of proteins that play a crucial role in the innate immune systems defence against invading pathogens and can be found on both epithelial and immune cells (e.g. macrophages). TLRs are a type of pattern recognition receptor (PRR) and are primarily responsible for recognising molecular patterns associated with pathogens. These patterns, known as pathogen-associated molecular patterns (PAMPs), will bind to TLRs through various signalling cascades and trigger an immune response (often through cytokine production).⁵⁹ Above it was mentioned that *P. aeruginosa* infection results in binding via the flagellin of the pathogen to the extracellular glycans chains of MUC1, which can subsequently result in the activation of signalling pathways via the cytoplasmic tail.⁶⁰ It was then shown that in response to infection, *MUC1*-null mice exhibited an increased immune response over *MUC1* expressing mice.⁶⁰ The authors demonstrated that *MUC1* knockdown enhanced flagellin-

induced interleukine-8 (IL-8) production in primary epithelial cells. Meanwhile, MUC1 expression in HEK293T cells inhibited the TLR5 dependent production of IL-8, thus reducing the innate immune response. This effect was removed upon deletion of the MUC1 cytoplasmic tail. The TLR activation signalling eventually converges in the NF- κ B pathway. The regulation of cytokine expression by MUC1 is thought to occur through suppression of the NF- κ B pathway.

It should be noted that whilst MUC1 has been well characterised for its role in downregulating inflammation, this is only one example of the role of mucins. In contrast to MUC1, MUC13 has been seen to have pro-inflammatory properties in response to infectious stimuli.⁶¹ This highlights that the full scope mucin function is as yet still poorly defined owing to the complexity and cross-talk of the signalling pathways involved.

What can be stated is that mucins in healthy cells are primarily associated with cell protection through control of the extracellular environment and providing a physical barrier to invading pathogens. However, a more sophisticated role for transmembrane mucins exists through their ability to modulate immune responses by engaging in complex signalling pathways via the extensive phosphorylation sites of their C-terminal domain. This is clearly displayed by the effect of mucins on modulating the immune response, through their expression on both epithelial cells and immune cells.

1.2.3. Mucin Function in Cancer

1.2.3.1 Overview

Due to the important role mucins have in cell protection and signalling, it is perhaps unsurprising that studies have also shown both secreted and transmembrane mucins are frequently exploited by a range of cancers (though predominantly epithelial cancers), being linked to drug resistance, angiogenesis, immune regulation, and metastasis.^{32,62,63} This is exemplified by the number of tumour associated mucin antigens found in serum that show a direct correlation with tumour size and poor prognosis. For example, in breast cancer (MUC1), ovarian carcinoma (MUC16) colorectal and pancreatic adenocarcinomas (MUC1).⁶⁴ Broadly speaking, the pro-tumorigenic properties of mucins can be divided into several categories: 1)

mechanical effects of the extracellular domain, 2) specific interactions between the glycans of the PTS repeats with other molecules and cells, 3) through signal transduction via the cytoplasmic tail.

The aberrant glycosylation and expression of mucins is characteristic in cancer and facilitates a number of oncogenic processes. The mucins of cancer cells have two notable differences to mucins expressed by healthy cells: 1) Epithelial cancer cells depolarise, and therefore start expressing mucins across the entire cell surface. 2) Healthy mucins are mostly decorated with core 2 glycans. In contrast, cancer associated MUC1 predominantly expresses truncated Tn, STn, T and ST O-glycans (**Figure 5**).⁶⁵ Many of these changes are most notably observed for mucins, such as MUC1, MUC4 and MUC16. The role these changes have in carcinogenesis is best illustrated with several examples.

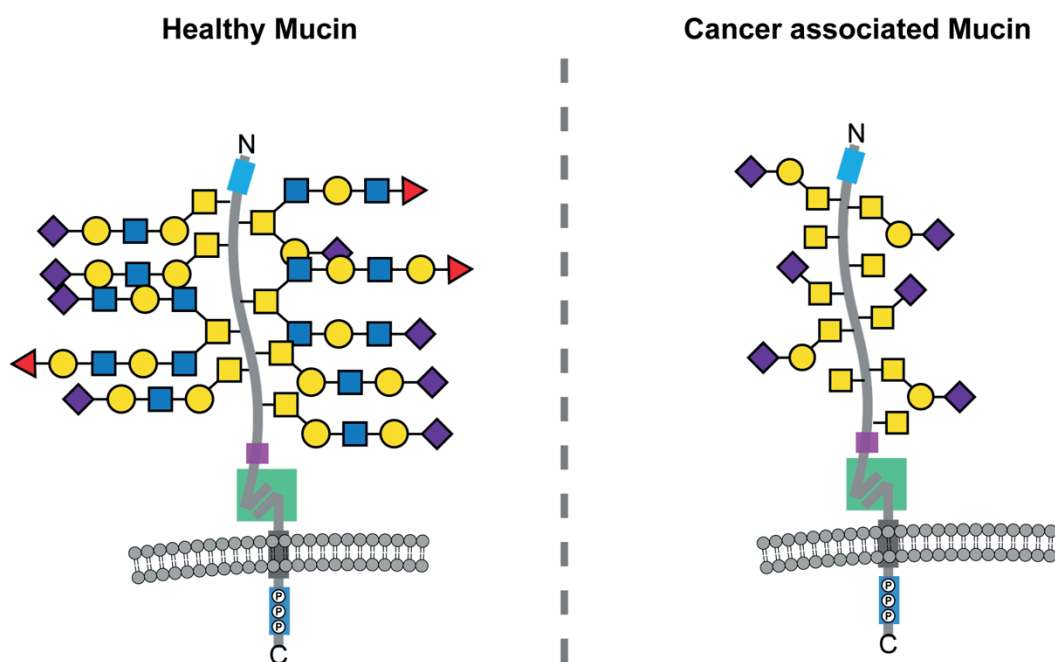


Figure 5: Representation of the differences in glycan structure between healthy (left) and cancer associated mucins (right).

1.2.3.2 Mucins, Metastasis, and Immune Evasion

Essential to the process of metastasis is the ability of tumour cells to selectively alter their adhesive properties. For metastasis to occur, a tumour cell must first detach from a primary tumour and migrate to the blood stream (intravasation), then penetrate the endothelial layer (extravasation).⁶⁶ Mucins have been linked to both processes.

Intravasation requires the permeabilisation of the rigid endo- and epithelial layers, which is maintained through junctions formed by E-cadherin.⁶⁷ The epithelial-mesenchymal transition (EMT) is the by process where epithelial cells depolarise and develop fibroblast-like properties and thus lose adhesion to the neighbouring cells. Characteristically, this results in loss of E-cadherin function.⁶⁸ In a mouse pancreatic cancer model, MUC1 expression was found to induce the EMT transition due to a loss of E-cadherin function.⁶⁹ This was thought to occur both through physical disruption by the extracellular glycosylated domain (now expressed on the entire surface due to depolarisation) and through the signalling of the cytoplasmic tail. Thus, resulting in increased cancer invasiveness.

For extravasation, circulating tumour cell survival is dependent on their ability to adhere in the low adhesion environment of endothelial cells.⁶⁶ E-selectin is a C-type lectin expressed by endothelial cells that binds to sialylated O-glycans and promotes the recruitment of immune cells to endothelial cells, where the leukocyte can then extravasate to the site of inflammation.^{70,71} E-Selectin is thus expressed by endothelial cells in response to inflammatory cytokines.⁷¹ The process of inflammation and extravasation are therefore integrally linked. Cancer cells will express ligands that bind to E-selectin enabling their attachment, followed by integrin activation and binding to ICAM molecules on endothelial cells. If the endothelial cells do not express the required selectin, then the cancer cell may release cytokines to stimulate E-selectin release.⁷²

The overexpression of mucins was seen to promote the formation of focal adhesion points through increased integrin clustering in breast cancer cells lines, thus enhancing attachment to endothelial cells.⁷³ The increase in integrin clustering was found to correlate with increased mucin length and expression. As has been discussed, mucins can suppress immune response

and inflammation through cytokine production and immune cell inactivation. However, somewhat paradoxically, cancer-associated MUC1 can lead to proinflammatory responses via dysregulation of intracellular signaling.⁷⁴ Thus, expression of mucins by cancer cells could be used to promote inflammation and thus E-selectin expression. This may have a synergistic effect with mucin expression effect on increase integrin clustering and focal adhesion formation to promote metastasis. Finally, mucins and other glycoproteins containing cancer associated glycans can directly bind selectins and sialic acid binding immunoglobulin like lectins (siglecs) on immune cells. This has been demonstrated to act as an immune checkpoint that results in cancer cell recognition as “self” and consequently immunosuppression.^{75,76}

In the context of cancer, mucins have been found to enhance cancer survival and progression through their ability to regulate cellular adhesion, inflammation, and immune recognition. Each of these properties are a result of the glycosylated extracellular N- terminal domain and transmembrane C-terminal. Despite extensive and thorough research, much of the precise mechanism by which mucins provide these (and other) benefits is still uncertain, owing to the range of mucins expressed and the complexity of the signalling pathways involved. Regardless, targeting mucins provides an exciting potential therapeutic approach for treating a variety of cancers.

1.2.4 Mucins as a Therapeutic Target

1.2.4.1 *Emerging Therapies*

Due to the alterations in glycan structures on cancer associated mucins, their therapeutic targeting to date has mainly consisted of antibody-mediated immune therapies including neutralising monoclonal antibodies, chimeric antigen receptor T (CAR-T) cells, vaccines, and antibody-drug conjugates. There is one notable exception to this trend. Due to the absence of a clear active site or enzymatic pocket on mucins, effective inhibition of the mucin domain has not been achieved with a small molecule inhibitor. However, there is one peptide inhibitor (GO-203) targeting the C-terminal domain of MUC1 in the clinic.⁷⁷ The C-terminal tail of MUC1 has been shown to dimerise in some cancers resulting in import to the nucleus. This dimerization occurs through a Cys-Gln-Cys (CQC) motif.⁷⁸ Inhibition of dimerization was achieved by mutation to Ala-Gln-Ala (AQA) and resulted in attenuated tumorigenicity and

growth in breast cancer cells.⁷⁸ GO-203 was shown to inhibit dimerization of this region and provide similar therapeutic benefit.

The most promising example of a mucin targeting ADC is BM7-PE. This ADC comprises an antibody conjugated to *Pseudomonas* exotoxin A and is currently in phase I/II clinical studies for colorectal cancer.⁷⁹ MUC1 is also considered a potential antigen for CAR-T cell therapies. Cancer associated mucins express dramatically increased levels of the Tn antigen compared to healthy tissue they make for ideal antigens. An excellent example of this is the Tn-MUC1 CAR currently in phase I clinical trials for solid tumours.⁸⁰ Finally, MUC1 is also an extremely attractive target for vaccines owing to its overexpression and aberrant glycosylation that both exposes more of the core peptide but also present tumour associated antigens (Tn, STn). The range of vaccines attempted are extensively reviewed by *T. Gao et al* and includes peptide-based subunit, viral vector, and dendritic cell vaccines. Similar approaches have been attempted with MUC1 tumour associated glycopeptides.⁸¹ Unfortunately, whilst many of these have entered phase I-III clinic trials, none have so far entered widespread clinic use. This is thought to be due to the immunosuppressing tumour microenvironment, which is in part promoted by mucins themselves. MUC1 peptide targeting vaccines may show more effect if the immune suppressing glycan epitopes were removed.

A final mucin targeting therapy was developed by the Bertozzi group that conjugates an enzyme that selectively digests mucins (mucinase) with a cancer targeting nanobody (HER2).⁸² This removes the bulky mucin extracellular domains from the cancer cell surface. This resulted in significantly increased breast and leukaemia death due to targeting from natural killer (NK) immune cells.

Mucins, particularly MUC1, are a well validated and highly relevant therapeutic target for a range of cancers. To date, targeting mucins has included vaccines, ADCs, and peptides. However, no mucin targeting therapy has yet achieved FDA approval for widespread use. A more thorough understanding of mucin function is clearly required as are possible alternative therapies. We hypothesise that an inhibitor of O-GalNAc glycosylation may provide a useful chemical tool to further characterise the role that glycans on mucins (and associated proteins) glycans have in promoting cancer progression and could provide a novel approach to inhibiting

mucin function. This final hypothesis has already been demonstrated to some degree by the GalNAc mimetic, benzyl- α -GalNAc.⁸³ This compound acts as a decoy substrate for the glycosyl transferases that would elaborate the Tn antigen. This was shown to not only reduce O-GalNAc glycosylation but also inhibit metastatic potential of pancreatic cancer cells expressing MUC1.⁸⁴ Unfortunately, this compound requires millimolar concentrations to inhibit glycan extension glycosylation, which somewhat limits its applicability. Crucially, this compound does not prevent the addition of GalNAc to proteins (the Tn antigen).

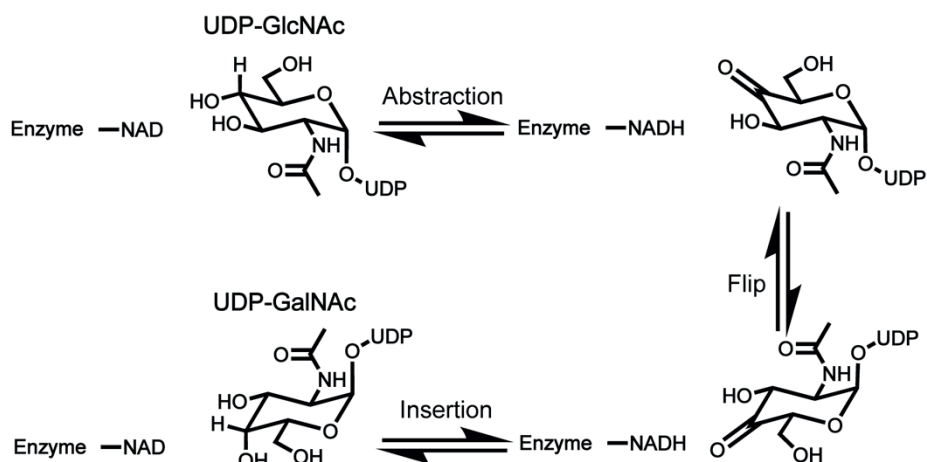
1.3 UDP-Galactose-4-Epimerase (GALE)

1.3.1 Overview

The enzyme UDP-Galactose-4-epimerase (GALE) is a cytosolic metabolic enzyme found in all kingdoms of life. GALE is primarily responsible for two reactions. 1) The conversion of UDP-Gal to UDP-Glc and vice versa, all GALE homologs perform this function. 2) The same conversion is also performed with UDP-GalNAc and UDP-GlcNAc in mammalian GALE.⁸⁵ Both these functions are unique to GALE and no isoform exists in the human genome. Both reactions are performed via a hydride abstraction from the C4 position and subsequent reinsertion, mediated by the co-factor nicotinamide adenine dinucleotide (NAD) (**Figure 6A**). The major role of GALE is in galactose metabolism via the Leloir pathway (**Figure 6B**), the process by which galactose is metabolised to glucose-1-phosphate.⁸⁶ However, more broadly, GALE is responsible for maintaining sugar nucleotide levels.⁸⁷ The only *de novo* biosynthetic pathway for UDP-GalNAc is from UDP-GlcNAc through epimerisation of GALE. Thus, GALE occupies a crucial bottleneck in UDP-GalNAc biosynthesis, and subsequently O-GalNAc glycosylation.

A

GALE Mechanism

**B**

Leloir Pathway

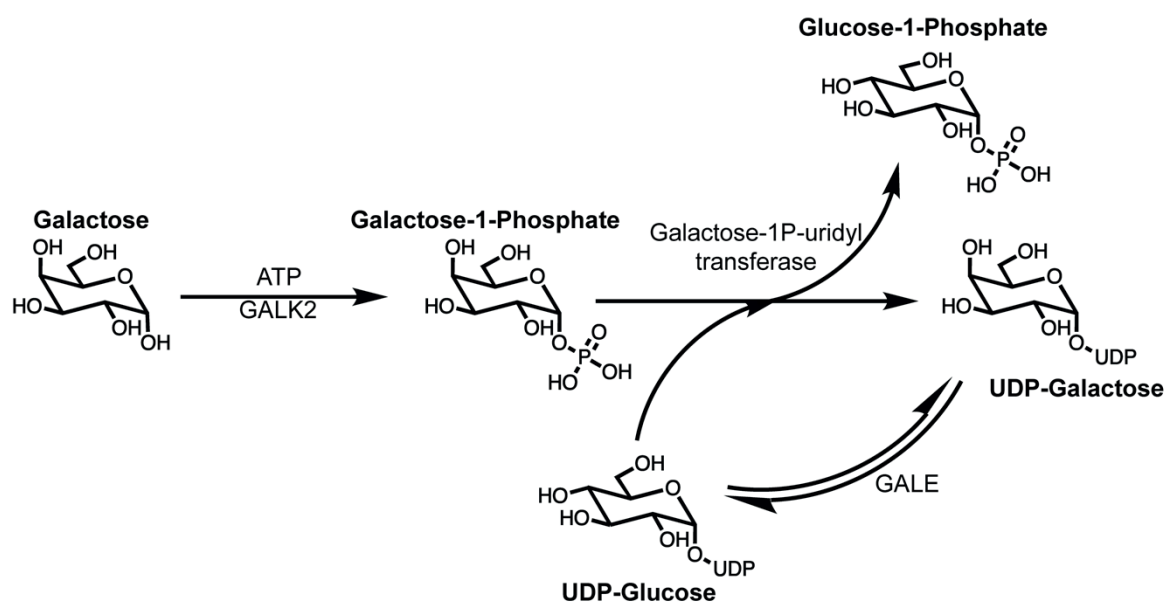


Figure 6: A) GALE epimerisation mechanism. B) Leloir pathway.

1.3.2 GALE as a Target for O-GalNAc Glycosylation Inhibition

The bottleneck occupied by GALE has been utilised by several groups to generate GALE KO cell lines, Chinese hamster ovary (CHO) and human embryonic kidney (HEK), that are devoid of O-GalNAc glycans owing to their inability to biosynthesise UDP-GalNAc.^{88,89} Notably, despite

salvage pathways which enable UDP-GalNAc production directly from endogenous GalNAc, serum levels of GalNAc were not sufficient to restore O-GalNAc glycosylation, illustrating that a GALE inhibitor cannot be circumvented by serum GalNAc. However, when the HEK GALE-KO cells were cultured in 10% FBS with supplied GalNAc, glycosylation was restored. This suggests that serum levels of Gal are sufficient for glycosylation, but not serum levels of GalNAc. Thus, a GALE inhibitor might result in sufficiently depleted levels of GalNAc but not Gal, which is essential given Gal is also required for N-glycans. Consequently, both cell lines displayed no change in their N-glycome. Finally, both cell lines appear to exhibit normal growth rates in culture with no GalNAc, highlighting the role O-GalNAc glycans play in multicellular processes.

GALE KO K562 (chronic myeloid leukaemia) suspension cells have been generated in the Schumann lab. Mucins and other heavily O-GalNAc glycosylated proteins have been extensively linked to enhancing cell survival in suspension.^{73,90} Loss of these glycoproteins results in cells dropping out of suspension.^{82,90} It is therefore unsurprising that GALE KO K562 cells drop out of suspension and aggregate. This effect can be reversed by supplying cells with free GalNAc (**Figure 7**). This phenotype would suggest that GALE inhibition with a small molecule could provide a novel method for targeting mucin expressing cancers (such as breast cancer) by preventing their characteristic expression of O-GalNAc glycans and glycoproteins. Whilst there may be the potential for a therapeutic window with such an inhibitor, owing to the increased expression and overreliance of cancer cells to mucins and O-GalNAc, toxicity issues would be of concern.

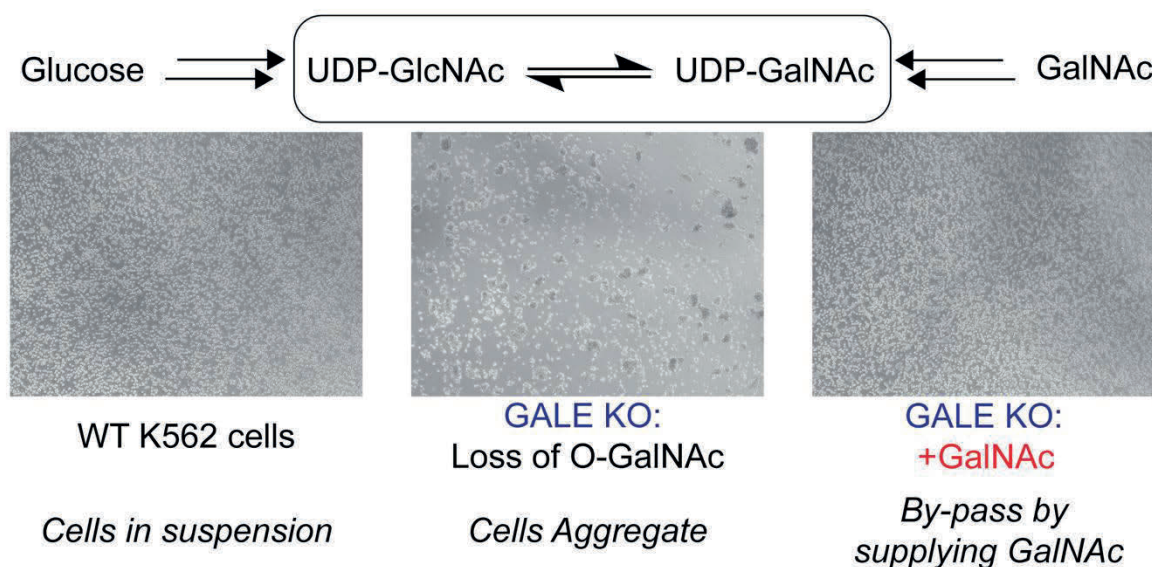


Figure 7: GALE KO K562 cell phenotype.

Images show GALE KO and WT K562 cell line either in the presence or absence of GalNAc. Images taken by Dr. Edgar Gonzalez-Rodriguez.

Type III galactosemia is a congenital condition that arises from GALE deficiencies due to certain point mutations.⁹¹ Type III galactosemia is typically categorised into either generalised or peripheral. Generalised is characterised by low (or zero) GALE activity in all tissue and results in renal, liver, and ocular damage.⁹² In contrast, in peripheral type III galactosemia, GALE activity is only lost in blood and circulating white blood cells and the side effects are far less severe, usually being identified only through newborn screening.

It is possible that adverse side effects similar to those of type III galactosemia could be observed from GALE inhibition by a small molecule. However, several factors bear consideration. First, the extreme side effects observed for generalised type III galactosemia are thought to arise from the build-up of toxic galactose metabolites due to the position of GALE in the Leloir pathway.⁸⁶ By far the major source of galactose in the human diet is from the lactose in milk, this is especially true for infants who have an exclusively milk diet. Thus, the treatment used for patients is a restricted galactose diet.⁹² Such a dietary regime could be implemented alongside a GALE targeting drug. Secondly, patients with type III galactosemia have had loss of GALE activity from the start of embryonic development and effects are likely far more extreme than would be expected in an adult. Personal discussions with clinicians specialised in rare diseases suggest that GALE galactosemia majorly affects early development rather than adulthood. Finally, generalised type III galactosemia patients have loss of GALE

activity from all tissue types. As is becoming routine for many small molecule therapies, a GALE drug could be targeted to specific tissues with modern targeted delivery methods such as antibody drug conjugates and nanoparticles.⁹³

1.3.3 GALE “Druggability”

GALE may be an attractive target for inhibiting O-GalNAc glycosylation as a chemical tool and potentially a therapeutic from a biological position. The next question is whether GALE itself is suitable to small molecule targeting, i.e. how “druggable” is GALE. Enzyme druggability is dependent on several factors.

Firstly, classical competitive inhibitors (inhibitors that bind to the same location as the reacting substrate) rely on an active site that has a defined pocket, one with amino acids that facilitate the formation of van der Waals interactions, hydrogen bonds (H-bonds), water displacement etc.^{94,95} The binding pocket must therefore be compatible with the small molecule ligand with respect to volume, topology, and physicochemical environment.⁹⁴ A competitive inhibitor of GALE would target the UDP-Sugar binding site (**Figure 8A-C**). No X-ray structure of GALE without NAD has even been determined, suggesting NAD may be structurally integral and would be more challenging to target. The structure of GALE with only NAD and no substrate bound has been solved (PDB:1EK5).⁹⁶ In the UDP-GalNAc binding site, the UDP moiety is likely to be much more tightly bound than the hexose sugar moiety, owing to the need for rotation of the hexose to enable hydride insertion from the opposing side it was abstracted from. Thus, a small molecule inhibitor would ideally target this region.

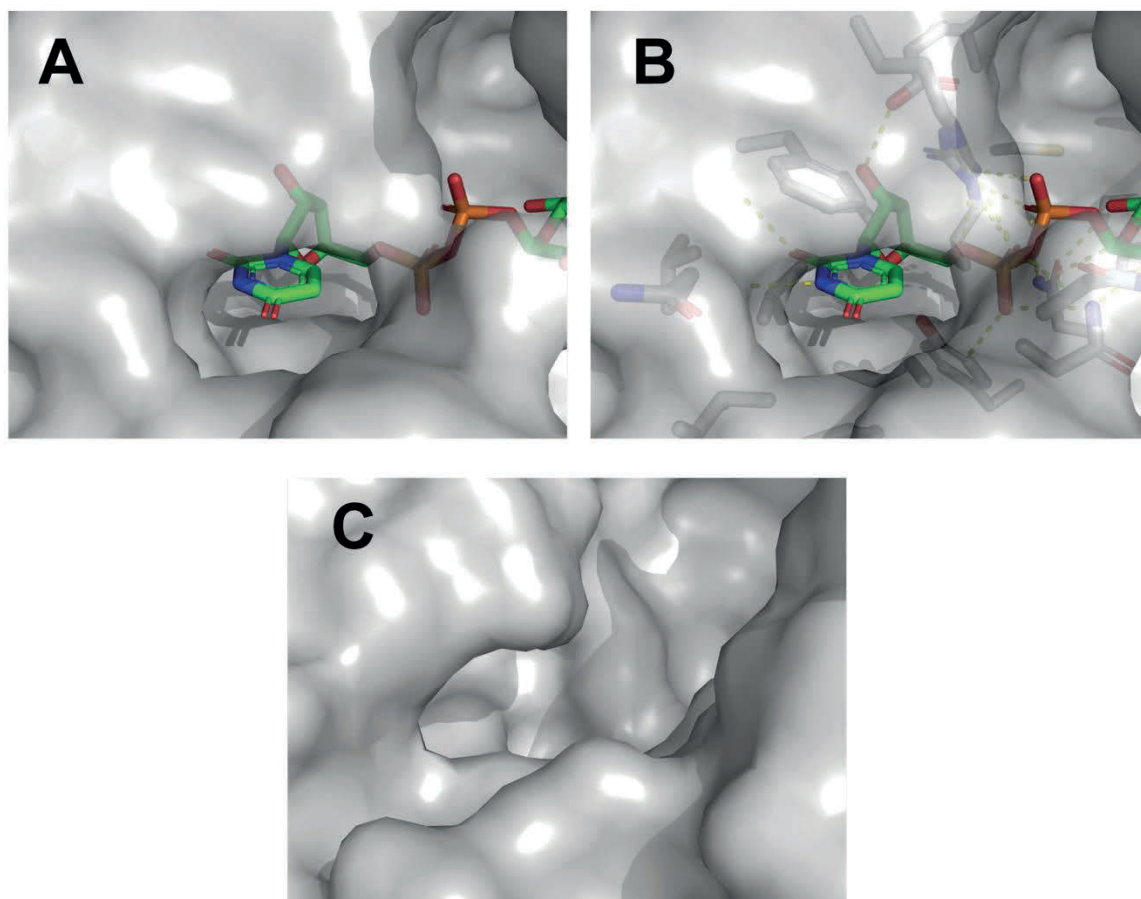


Figure 8: GALE UDP-Sugar binding site structure.

A) Crystal structure of GALE (1EK6) in complex with UDP-Glc (green) with modelled surface. B) As in panel A but with UDP-Glc bonding residues shown. C) crystal structure with surface of GALE without substrate (1EK5).⁹⁶

One of the most frequently targeted classes of enzyme are kinases. Responsible for phosphorylating other proteins, they use the substrate adenosine triphosphate, a nucleotide much like UDP. Kinase inhibition has seen a lot of success by targeting the ATP binding domain.⁹⁷ Similarly, the presence of a UDP binding pocket in GALE would suggest GALE is targetable by small molecules. Enzyme accessibility is another factor governing druggability. An enzyme expressed in the nucleus will be harder to inhibit than the same enzyme in the cytosol. Fortunately, GALE is a cytosolic enzyme and thus accessibility due to enzyme localisation is not predicted to be a major limiting factor.

Finally, GALE has also been partially validated for small molecule inhibition by two prior efforts, namely those by the Bertozzi and Ferguson labs.^{98,99} The Bertozzi group synthesised a library (~1000) of uridine nucleoside analogues as substrate mimetic inhibitors. Screening of

this library identified **compound 1 (Figure 9)** with a mid to low micromolar IC_{50} in the presence of 100 μ M UDP-Gal against recombinant GALE. However, it was not elaborated upon in any further study, likely due to cell permeability issues⁹⁸. Meanwhile, the Ferguson lab explored GALE inhibition as a therapeutic target for *Trypanosoma brucei*. In this case, the aim was to identify an inhibitor with selectivity for *T. brucei* GALE over human GALE as a treatment for African sleeping sickness. The Ferguson lab's approach was to screen a library (~800) of known drug-like compounds in an inhibition assay against *T. brucei* GALE. There was limited success in targeting human GALE over *T. brucei* GALE, however, a nanomolar inhibitor of GALE was identified (**compound 2, Figure 9**). Unfortunately, **compound 2** is a known drug called ebselen with many other known targets and thus would have off target effects.¹⁰⁰ Regardless, both studies demonstrate that GALE is targetable by small molecules.

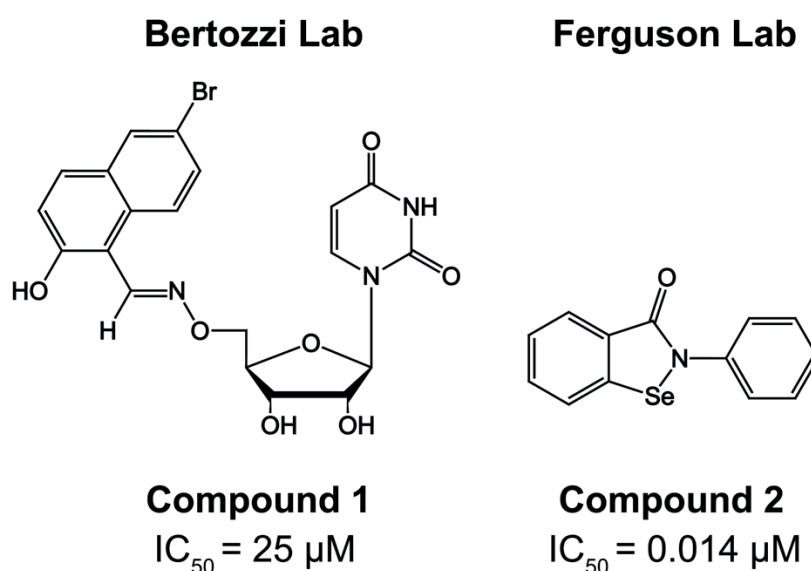


Figure 9: Structures and IC_{50} values for the inhibitors of GALE identified by the Bertozzi and Ferguson labs.

1.4 Inhibitor Development

1.4.1 Phenotypic and Target-Based Screening

The development of a small molecule inhibitor for an enzyme is a well-established process and begins with identifying compounds that produce a desired effect e.g. inhibition of enzymatic activity, cell death, etc. Typically, these compounds are identified by screening large libraries to identify initial “hits”, which are then chemically elaborated or modified to improve pharmacodynamic (PD) and pharmacokinetic (PK) properties.¹⁰¹ Screening is usually

performed in one of two ways; 1) screening in cells to find hits that produce a phenotypic change (phenotypic screening), 2) screening against recombinant enzymes using biophysical or biochemical assays (target-based screening).¹⁰¹

Phenotypic screening is most useful when the enzymes involved in the pathway to be disrupted are unknown. The main benefits include uncovering new targets/mechanisms of action and the identified hits are known to work in the complex cellular environment.¹⁰² The challenges are that the target (or targets) of hits are unknown which can complicate the development of improved leads with better on-target activity.¹⁰² Additionally, complex cellular models and assays may be required to accurately identify compounds that produce the desired phenotype.

In target-based screening, libraries are screened against a well-validated target (usually a single gene or gene product) by either biophysical or biochemical assays.¹⁰³ This method is typically used when the exact target is established, and structural and functional insights are already available. The advantages lie in the ease of designing more selective ligands, especially with structures of related off-target proteins.¹⁰⁴ A major limitation with this approach is that ligands optimised *in vitro* may not be active or permeable *in cellulo*. Therefore, careful consideration of lipophilicity and other PK properties must be considered during lead generation.

The early development of a small molecule inhibitor by a target-based approach follows several key initial steps. 1) *in vitro* biophysical/biochemical library screening to identify hits. 2) Hit validation by orthogonal assays and testing for enzymatic activity. 3) hit to lead generation to develop properties such as potency, ligand efficiency (LE), solubility, and PK. 4) testing and optimisation *in cellulo*.¹⁰¹ Step 1 requires a suitable expression system that can produce stable and active enzyme in the high yields that are required for screening small molecule libraries. Furthermore, key is improving target engagement of any identified hits, and since improving inhibitor potency is driven initially through improving binding, crystallographic information can greatly accelerate this process.

1.4.2 Fragment-Based Drug Discovery

As discussed target-based screening requires hit identification achieved through screening large libraries ($>10^6$) of drug-like molecules (“rule of 5” compounds) using biophysical, biochemical, or phenotypic screening.¹⁰⁵ However, in recent decades this approach has been challenged by so-called fragment-based drug discovery (FBDD).¹⁰⁶ Here, smaller libraries ($\sim 10^3$) of low molecular weight and chemically “simple” compounds are screened in biophysical assays at high concentrations.¹⁰⁶ Most fragments today will typically adhere to the “rule of 3”, analogous to the rule of 5 used for drug-like compounds (**Figure 10A**). This rule was first suggested by *Congreve et al* after assessing a range of fragment-like hits identified by Astex.¹⁰⁷ They noted that on average useful hits could be categorised in orders of three based on their physiochemical properties: molecular weight (MW) <3 , CLogP ≤ 3 , hydrogen bond acceptors ≤ 3 and hydrogen bond donors ≤ 3 . As such, most fragment libraries today will contain compounds that follow this principle. Additional considerations for fragment libraries include balancing 3D and 2D structures with molecular weight and ensuring fragments have easily expandable vectors for fragment expansion.¹⁰⁸

Due to their small size fragments are typically identified with weak affinities (mid-high micromolar K_D) for the protein target, due to the few atoms available for interaction. This necessitates fragment elaboration through medicinal chemistry techniques to develop potent inhibitors.¹⁰⁹ Due to their simplicity fragment hits can also be expected to be fairly promiscuous binders, thus elaboration is also required to generate specificity for the target. Consequently, to fully utilise FBDD, structural information on hits is often essential. The expansion/elaboration of fragments is typically achieved by one of two methods (**Figure 10B**): 1) Fragment merging, where identified hits that occupy different pockets are “merged” by either linkers or functionalisation of a scaffold to form a more diverse and potent binder.¹¹⁰ This method is difficult in practice owing to stereochemical changes induced by linking. 2) Fragment growing, where an initial fragment hit is elaborated based on structural insights gained from crystal structures (i.e. structure-based drug design, SBDD).¹¹⁰

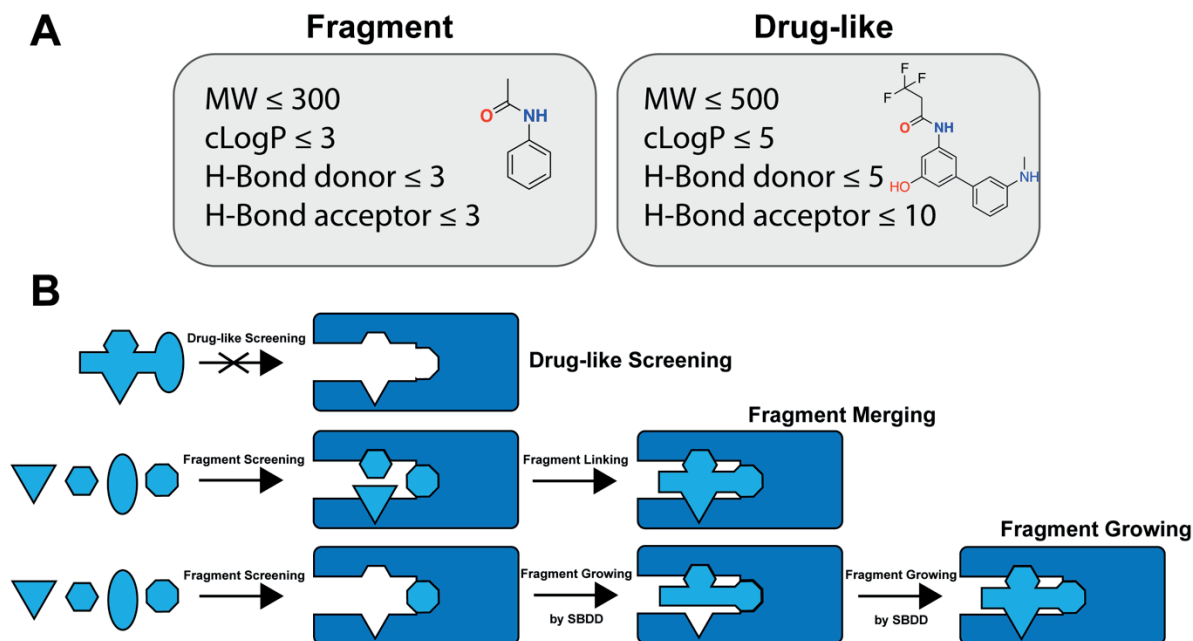


Figure 10: FBDD, rule of 3 and elaboration strategies

A) Comparison of rule of 3 vs rule of 5 compounds. B) Illustration of drug-like compound screening vs fragment screening followed by fragment growing and merging.

There are several major advantages to FBDD. Firstly, the small size and simplicity of the fragments allows for more efficient probing of a diverse chemical space, thus enabling smaller library sizes that are essential for academic drug discovery projects.¹¹¹ Fragments are often more hydrophilic than drug-like molecules.¹¹² Consequently, hits usually have more specific interactions such as hydrogen bonding, increased solubility, and increased ligand efficiency (LE).^{111,112} LE is a useful metric for selecting favourable fragment analogues.¹¹³ It is a measure of the average binding contribution of each heavy atom in a molecule. The simplest definition of LE is shown in **Equation 1**, using ΔG (assuming standard conditions) and dividing this by the heavy atom count (HAC).¹¹³ This can also be expressed using $RT\ln(K_D)$, which can be further simplified to part three of **Equation 1**, giving LE in units $\text{kcal}\cdot\text{mol}^{-1}\text{HA}^{-1}$ (although LE is usually given without units). Consideration of LE is useful for avoiding inefficient fragment elaboration.

$$LE = \frac{-\Delta G}{HAC} = \frac{-RT\ln(K_D)}{HAC} = \frac{-1.37}{HAC} \times \text{Log}(K_D)$$

Equation 1: Equation for determining ligand efficiency (LE)

Where ΔG is the Gibbs free energy under standard conditions.

HAC is the heavy atom count.

R is the ideal gas constant in $\text{Jmol}^{-1}\text{K}^{-1}$

T is the temperature in K.

K_D is the dissociation constant of the ligand in M.

The fragment hits are also more likely to be identified across different regions of the protein due to their small size. Together, this enables a better understanding of the interactions driving binding and facilitates elaboration into more potent compounds with better ligand efficiencies. The applicability of FBDD to developing drugs is now well validated, with over forty FBDD derived compounds in clinical trials and six Food and Drug Administrations (FDA) approved drugs.¹¹⁴ With these benefits and prior validation, FBDD was deemed an appropriate method for developing a GALE inhibitor.

1.5 Project Aims

1.5.1 Aims

The main aim of this project is to develop a cell-permeable small molecule inhibitor of GALE using a FBDD approach. This would provide an O-GalNAc glycosylation inhibitor that would serve two purposes. Firstly, such an inhibitor would provide a useful chemical tool for studying O-GalNAc glycosylation. Secondly, it would enable further exploration and validation of GALE as a possible therapeutic target for mucin expressing cancers.

Chapter 2 Protein Expression, Characterisation and Crystallography

2.1 Chapter Introduction

As has been discussed in the thesis introduction and aims, a FBDD approach will be adopted to develop a small molecule inhibitor of GALE. Essential to this will be a suitable protein expression system (for assays) and crystallography conditions (to determine ligand binding modes). Thus, in this chapter, the efforts are described to express and characterise human GALE, and to optimise crystallography conditions that can be used to determine protein/ligand complexes.

2.1.1 Protein Expression, Purification and Characterisation

2.1.1.1 Protein Expression and Prior Work

GALE has been expressed in the past using both *E.coli* and *pichia pastoris* in prior publications.^{96,115} Whilst both methods produced active GALE in high yields, we chose to express GALE in a *Spodoptera frugiperda* cell line (Sf21 cells). The advantages include high yielding expression, simple cell growth and expression induction.¹¹⁶ Insect cell expression is typically performed using viral vectors. The baculovirus is specific to insect cells, by preparing a recombinant virus with the protein of interest (POI) inserted into the viral genome, infected cells will overexpress viral proteins as well as the POI. This is a convenient approach since the virus stocks generated can be stored over long periods (>12 months). Since we expected repeat expression would be essential, the insect cells/baculovirus expression system was used to produce recombinant GALE.

The expression and purification were first optimised as part of an MRes project (W. M. Browne, MRes Thesis, Imperial College London, 2020) conducted by me (the author) as part of this project.¹¹⁷ In summary, a baculovirus based insect cell expression system was chosen to express recombinant GALE. The GALE DNA coding-sequence (DNASU, Clone ID: 40708) was cloned into a pTriEx based, in-house vector. This is a shuttle vector designed to enable protein expression from both *E.coli* and Sf21 insect cells. The fusion protein construct consisted of a 6xHis and glutathione S-transferase (GST) C-terminal tags to enable purification, separated from GALE by a human rhino virus (HRV) 3C protease cleavage sequence for tag removal

(**Figure 11A**). This vector could be combined with baculovirus DNA (via homologous recombination) and transfected into cells to generate recombinant virus. The recombinant virus was then used to infect Sf21 insect cells (usually 1 L) as described in the methods of prior publications written by myself.¹¹⁸ Once infected, cells express both viral proteins and our GALE construct, cells were then lysed, and the fusion protein extracted. Purification was performed by incubation of cell lysates with glutathione beads followed by washing and subsequent cleavage of GALE from the beads with the HRV 3C protease (**Figure 11B**). Washing of the beads after incubation with protease yielded elution fractions E1-E3 (**Figure 11B**). The enzyme was purified further by size exclusion chromatography (**Figure 11C**). The expressed enzyme was obtained in high yield (approximately 9 mg/L of culture) and purity (**Figure 11B-D**).

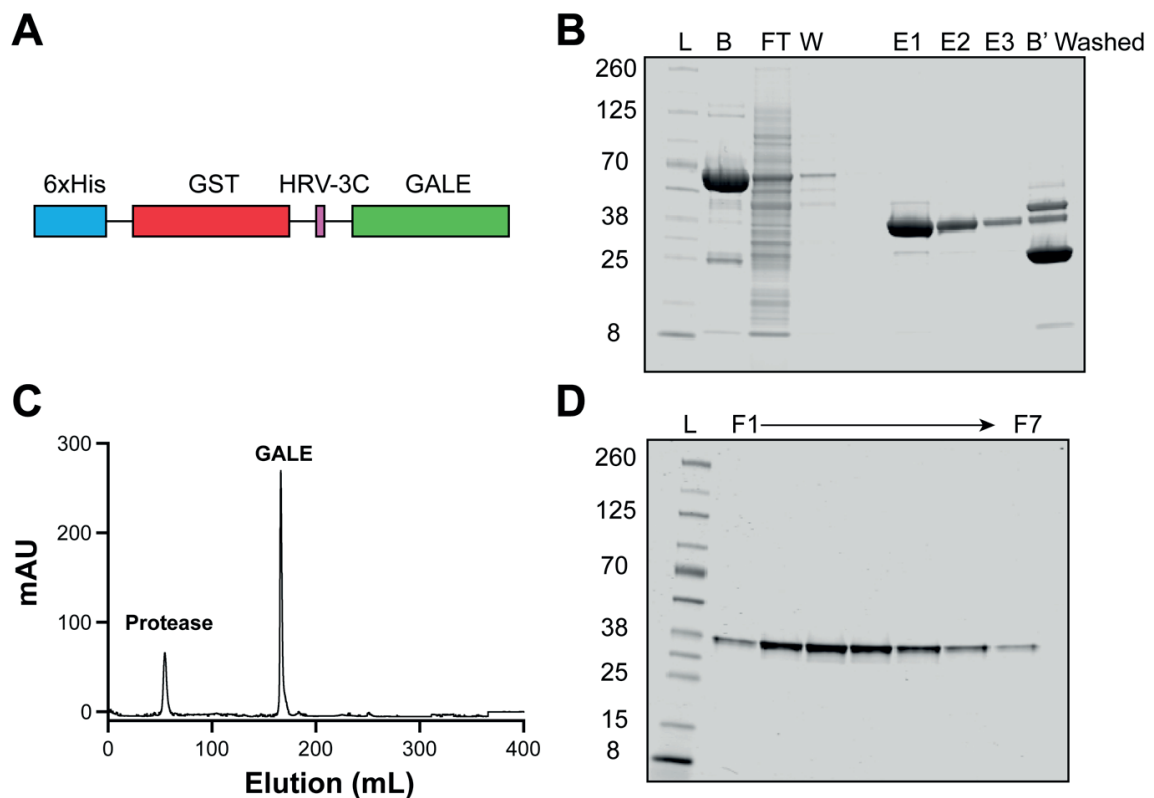


Figure 11: GALE fusion construct and purification

A) Gene construct in the pTriEx vector used for GALE expression. B) SDS-PAGE gel stained with safe blue showing the purification of GALE. Left to Right, Ladder, GST-beads after incubation with pelleted cell lysate supernatant and washing, supernatant after bead incubation, wash, Elution fractions from beads after incubation with HRV 3C protease (E1-E3). C) Size exclusion chromatogram of GALE after elution from beads. D) SDS-PAGE gel of fractions collected from size exclusion chromatography (F1-F7), major band at 38 kDa is GALE. It should be noted the panels B-D have been used in an earlier MRes project as part of this PhD degree. (W. M. Browne, MRes Thesis, Imperial College London, 2020.)

2.1.1.2 Michaelis-Menten

After protein expression and purification, the activity of the enzyme is typically assessed with a Michaelis-Menten plot.¹¹⁹ The initial rate (v_0) is measured at varying substrate concentration ($[S]$), from this insight into the enzyme efficiency and substrate binding is gained, and can be used to optimise enzyme reaction conditions.¹²⁰ The most useful information obtained are the kinetic parameters K_m and k_{cat} (**Equation 2**). The K_m (given in units of concentration) is the Michaelis-Menten constant and is defined as the concentration at which half the maximal enzymatic rate is achieved, it is therefore a correlate of the affinity of a protein for the substrate. k_{cat} (given in units of time^{-1}) is the turnover number and defines the number of molecules of substrate converted per unit time. Comparison of these parameters to those published for other protein preparations of the same enzyme are often used to validate an expression system.

$$v_0 = \frac{[E]_t \cdot k_{cat} \cdot [S]}{K_m + [S]}$$

Equation 2: Michaelis-Menten Equation to determine k_{cat} .

Where v_0 is the initial rate in $\mu\text{M/s}$.

K_m is the Michaelis-Menten constant in μM .

$[E]_t$ is the concentration of enzyme in μM in the assay.

k_{cat} is the turnover number in s^{-1} .

$[S]$ is the concentration of substrate at the start of the reaction in μM .

2.1.1.3 Isothermal Titration Calorimetry (ITC)

Also useful for characterising expressed enzymes is determination of the affinity of the substrate (or ligands) for the enzyme and the proportion of available binding sites (N). The measure of affinity between substrate/ligand for a protein is given by the K_D and is defined as the concentration at which half of the available substrate/ligand is bound to the protein (**Figure 12**). ITC is a biophysical assay used to measure the binding and thermodynamics of molecules in solution. It can determine useful parameters such as K_D , enthalpy change (ΔH), Gibbs free energy change (ΔG), and binding site occupancy (N).¹¹⁰ In general, the assay works by titrating in a ligand to a sample of protein. After each ligand addition the release or intake of thermal energy is recorded to determine the parameters mentioned above. ITC is relatively

low throughput, so it is not typically used for compound screening, however, it has proved useful as a secondary assay to quantitatively assess hits.¹²¹

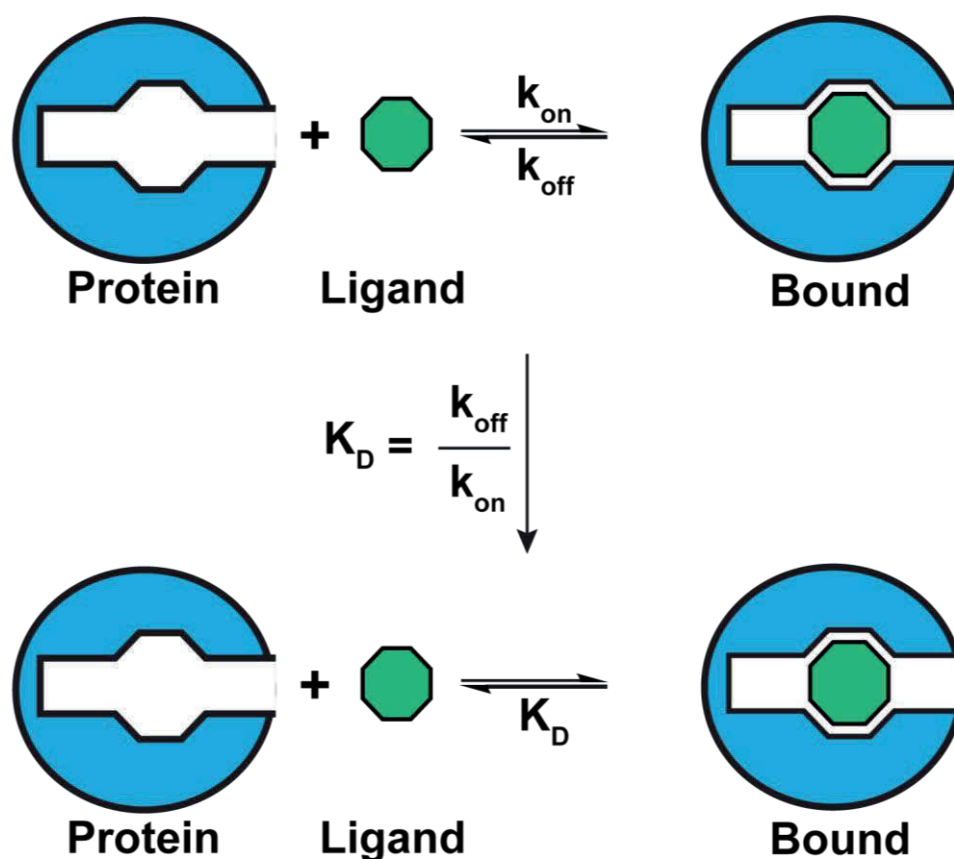


Figure 12: Illustration depicting reversible ligand binding with K_D

2.1.2 Crystallography in Drug Discovery

After validation that the expressed target enzyme is active, it is advantageous to develop crystallography conditions for the protein.¹²² As discussed above, structural information on hit and lead binding can be invaluable to drive increases in potency and guide functional group modification (e.g. to improve PK). Proteins that cannot be crystallised can often be much harder to develop ligands for.

Crystallisation itself has two main phases: nucleation, and growth.¹²³ Nucleation is the development of small initial clusters of molecules that can then grow into larger lattices. Nucleation as a limiting factor can be overcome with “seeding”. Here, small fragments of

crystal (possibly from earlier obtained protein crystals) are used as a template for further growth. This can significantly improve crystal hit rates.¹²⁴

Experiments to identify ligand/protein complexes are typically carried out using two main strategies; 1) soaking experiments, where protein crystals are first grown and the fragment added directly to solutions containing the crystals, or 2) using co-crystallisation, where protein crystals are grown from solutions of both protein and ligand.¹²⁵ Which approach is used is often dependent on the stability of the protein/protein crystals to solvents required for small molecule solubility. Soaking experiments for ligands binding to a substrate binding pocket usually require this pocket to be empty in the crystal system (apo structures), otherwise the ligand must compete with the substrate, which can be difficult for tight binding substrates. Fortunately, GALE had been crystallised in the past and several structures and starting conditions were available.⁹⁶ The PDB entry 1EK5 is an “apo” structure of GALE, with only NAD bound and no UDP-Sugar. The entry 1EK6 contains substrate bound GALE in its native homodimeric form.⁹⁶ For this project the apo structure of GALE was likely more useful, therefore in this chapter, the conditions used to produce PDB 1EK5 were used for initial scouting of crystallography conditions.

In summary, this chapter details the efforts to characterise the insect cell expressed GALE, and establish crystallography conditions suitable for small molecule screening.

2.2 Results and Discussion

2.2.1 Activity Characterisation

To demonstrate that the expression system previously produce active enzyme, we used a biochemical high performance liquid chromatography (HPLC) assay. In brief, GALE was incubated with UDP-Gal in the presence of NAD at 37 °C for 15 minutes. The relative conversion of UDP-Gal to UDP-Glc was determined using analytical HPLC and UV absorption peaks for each substrate (**Figure 13A**). GALE will epimerase both UDP-Gal/Glc and UDP-GalNAc/GlcNAc, and although the aim is to inhibit the conversion of UDP-GlcNAc to UDP-GalNAc, the reaction for UDP-Gal to UDP-Glc was used for several reasons. 1) UDP-Gal and UDP-Glc have more complete separation by HPLC than UDP-GalNAc and UDP-GlcNAc, meaning the obtained results are more consistent and reliable. 2) Prior efforts to develop inhibitors of GALE also use this reaction, enabling direct comparison of identified inhibitors to those already published.^{98,99} 3) Since all the substrates bind to the same pocket, an inhibitor of the UDP-Gal/Glc reaction will also inhibit the UDP-GalNAc/GlcNAc. The first experiment performed with this assay involved varying enzyme concentration to determine the amount of enzyme required to convert approx. 20 % of UDP-Gal to UDP-Glc. This ensures the enzyme is under initial rate conditions, which are a requisite of Michaelis-Menten kinetics. The appropriate concentration was found to be 5 nM GALE under the time frame of the assay (15 minutes) (**Figure 13B**). This concentration was then used to produce a Michaelis-Menten plot to calculate kinetic parameters (**Figure 13C** and **Table 1**). The determined K_m value of $87 \pm 16 \mu\text{M}$ is comparable to the previously published K_m of $69 \pm 12 \mu\text{M}$, the slight variation is likely due to the different assays used (HPLC vs coupled enzyme assay) and differences in pH (7.5 vs 8.8) that the assays were performed at.¹²⁶ All activity assays in this report were performed at pH 7.5, since this is close to the pH of the cytosol, where GALE is active.¹²⁷ The comparable K_m and k_{cat} results validate the insect cells expression system for producing active recombinant GALE.

GALE	K_m (μM)	k_{cat} (s^{-1})
Insect Cell Expressed	87 ± 16	19 ± 1
Published Values ¹²⁶	69 ± 12	36 ± 1

Table 1: GALE kinetics parameters.

GALE K_m and k_{cat} values from insect cells and literature. Where K_m is the Michaelis-Menten Constant and k_{cat} is the turnover number. Error is determined from technical triplicates.

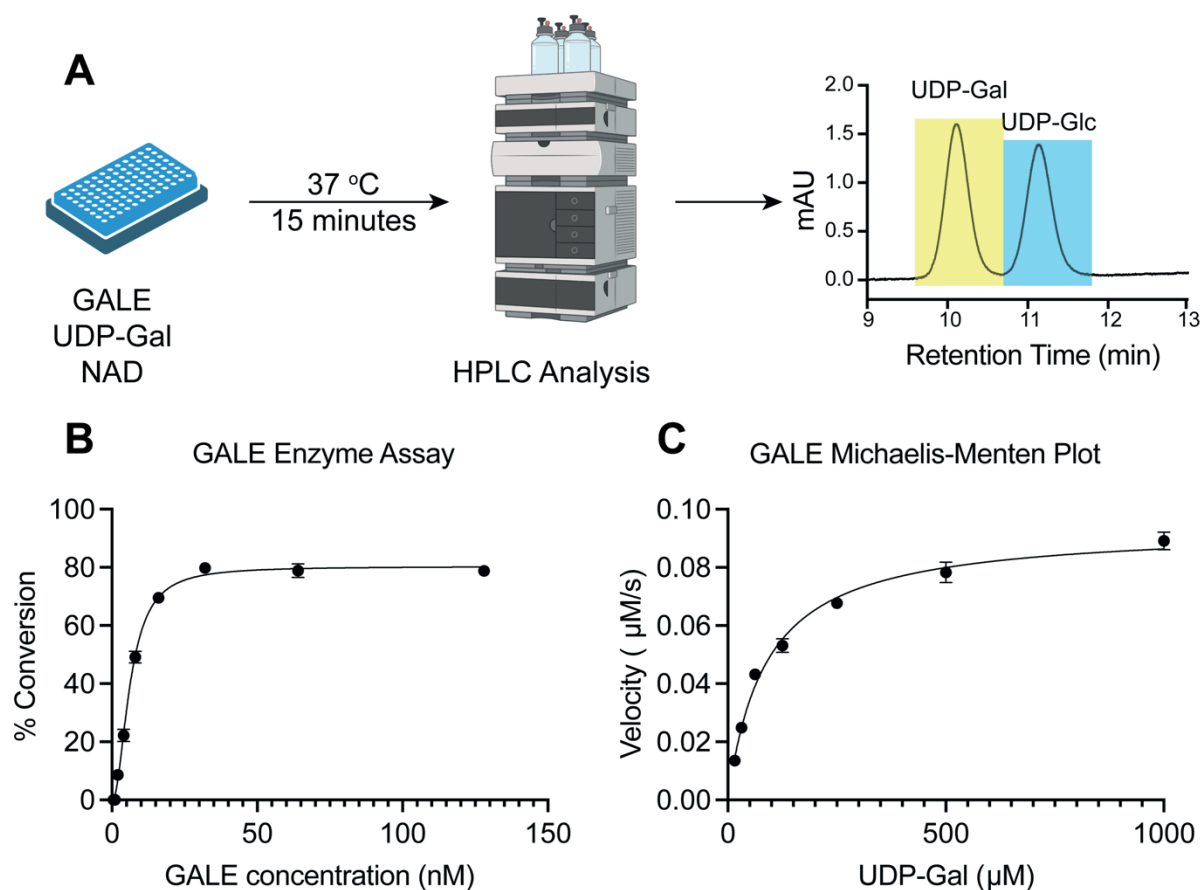


Figure 13: GALE Michaelis-Menten activity validation.

A) Illustration of the GALE UDP-Gal HPLC activity assay. B) Enzyme dependence where the enzyme concentration is varied at 100 μM UDP-Gal and the % conversion measured. C) Michaelis-Menten plot at 5 nM GALE. Assay results are from technical triplicates.

2.2.2 GALE Crystallography – Initial Screen

Having developed an expression system that expressed high yielding active GALE the next aim was to optimise crystallography conditions suitable for fragment soaking experiments. Initial experiments involved an optimisation screen based around the crystallography conditions used to produce 1EK5, the apo structure of GALE obtained by *J. B. Thoden et al.*⁹⁶ This is described in full in the methods section but in brief, the screen covered the conditions EPPS 0.1 M, 17-22% PEG 8K, 0.05-0.25 M KCl (**Figure 14A**). The screen was assessed with protein in both 25 mM HEPES, pH 7.5, 500 mM NaCl and protein dialysed into a 100 mM potassium phosphate buffer pH 7.5, 150 mM NaCl buffer (herein HEPES buffered, and Phosphate buffered, respectively). The HEPES buffered protein was used since the initial purification of GALE was performed in this buffer, whilst the Phosphate buffer was used in purifying GALE in *J. B. Thoden et al.*'s publication. The hit rates and hit well numbers for all plates are summarised in **Figure 14B-E**. The hit rate was 12% for both screens with no clear trend between crystallisation and condition, suggesting crystal growth is driven more by random nucleation events than specific conditions.

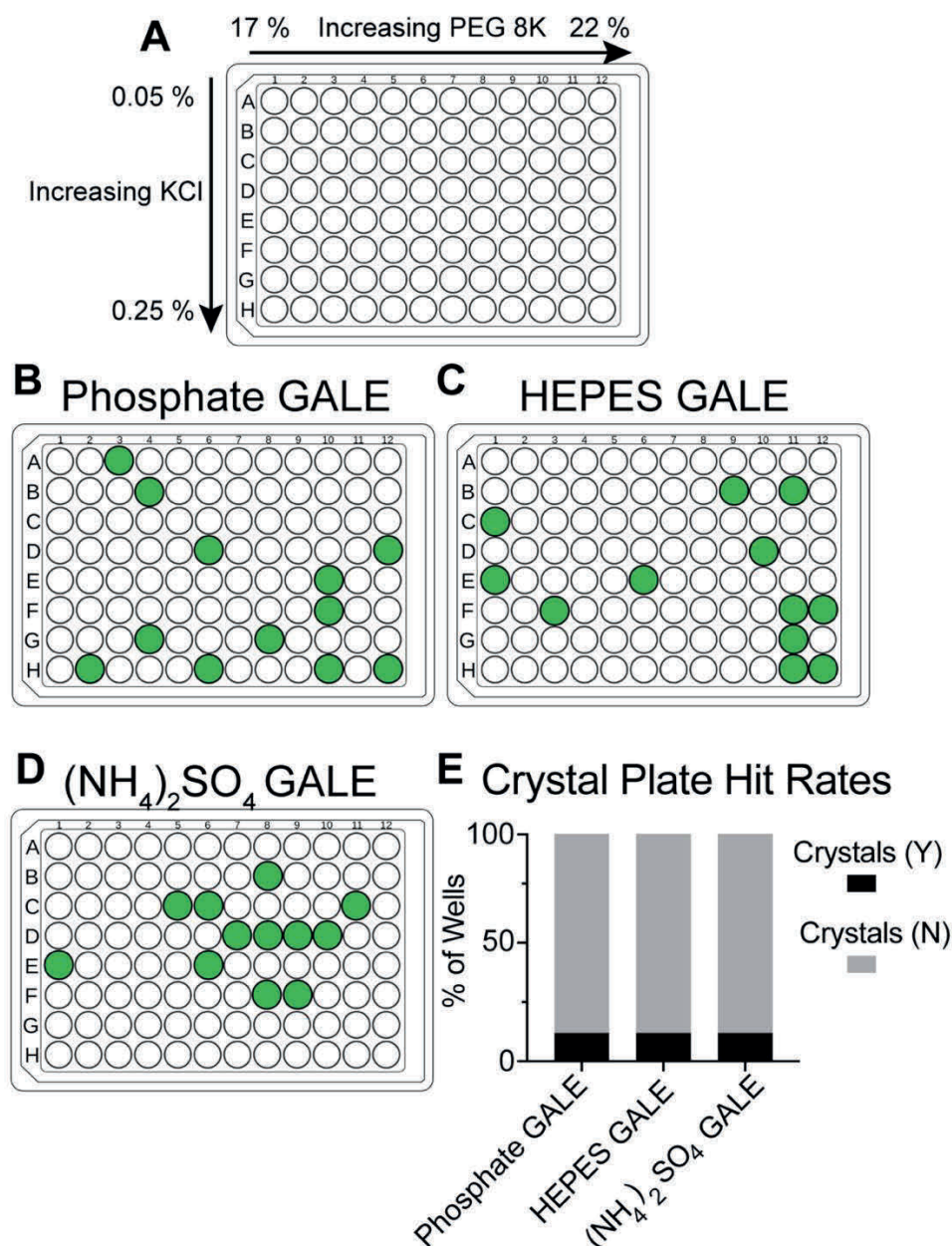


Figure 14: Initial crystallography condition screening results.

A) Conditions and layout used for the 1EK5 based condition screen. Increasing PEG concentration horizontally, increasing KCl concentration vertically. B) Hit locations for the phosphate buffered GALE plate. C) Hit locations for the HEPES buffered GALE. D) Hit locations for GALE precipitated with ammonium sulfate. Green = crystal growth, White = no crystal growth E) % crystal growth across crystallography trays. Black = % with crystal growth, grey = % of wells with no crystal growth.

Several representative wells containing crystals from each screen are shown in **Figure 15**. In general crystals from HEPES buffered GALE were slightly longer (mean average of 209 μm) than those Phosphate buffered GALE samples (mean average of 150 μm). HEPES buffered crystals were also generally longer and thinner. Both conditions typically only had 1-2 crystals growing of between 100-200 μm lengths, rather than multiple small crystals.

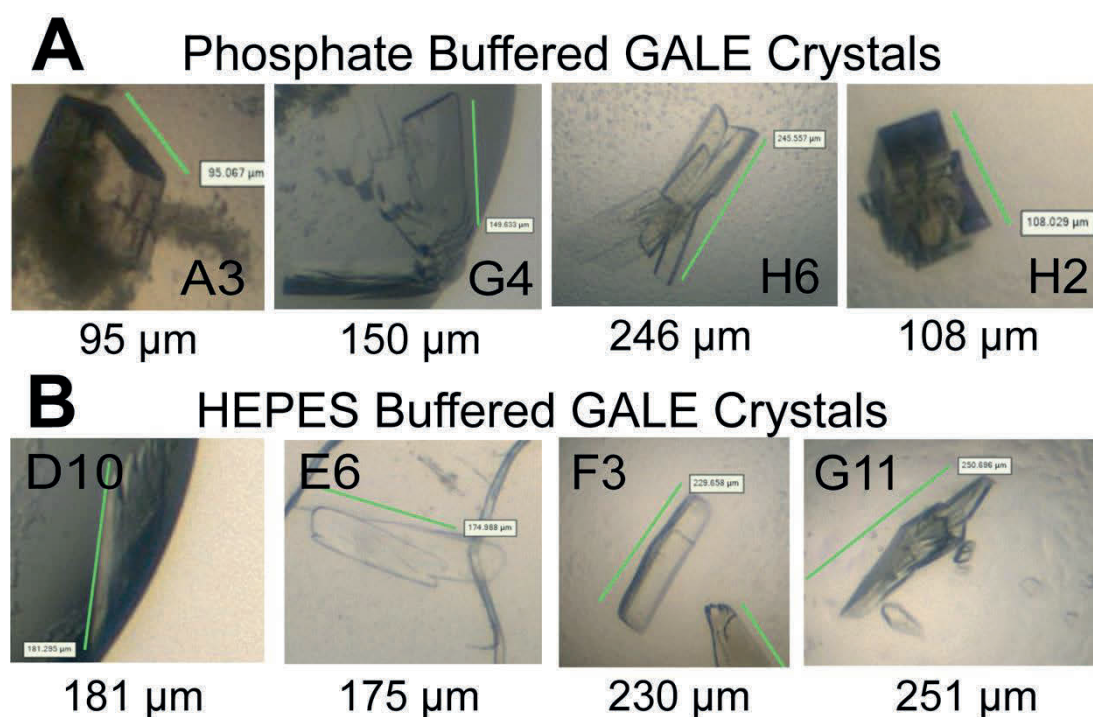


Figure 15: Crystal morphology summary from initial crystallography condition screening.

Crystal images for representative wells from crystallography trays set up using HEPES buffered, and Phosphate buffered samples of GALE. Length of green scale shown below images. All images taken on day 5.

From each condition (HEPES and Phosphate buffered) several crystals were sent to the Diamond Light Source (DLS) Synchrotron for data collection (A3, H6, D10). Structures were solved for some of the best diffracting crystals (all were 1.7-1.9 Å resolution) from each protein buffer sample (conditions detailed in **Table 2** and structure validation in **Table 3**).

Well	Protein Buffer	KCl (M)	PEG 8K (%)	EPPS (M)
A3	Phosphate	0.05	17.9	0.1
H6	Phosphate	0.25	19.3	0.1
D10	HEPES	0.136	21.1	0.1

Table 2: Diffracting crystal crystallisation conditions from initial condition screen (1EK5 based).

Well	A3	H6	D10
Wavelength (Å)	0.98	0.98	0.98
Resolution range (Å)	55.05 – 1.88	67.97 – 1.73	57.75 – 1.80
Space group	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁
Cell dimension: a, b, c (Å)	45.9, 110.1, 135.5	45.8, 109.8, 135.9	45.7, 109.8, 135.8
α , β , γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
R _{work}	0.20	0.23	0.21
R _{free}	0.24	0.26	0.25

Table 3: Structure validation and information for the crystals analysed from wells A3, H6 and D10. ("Table 1" values).

Full statistical validation in section 2.5.

The search model used for molecular replacement for the first of these crystals (A3, Phosphate buffered) was the published structure 1EK5.⁹⁶ Processing of the data using the model 1EK5 followed by rounds of manual model building with Coot and refinement with Refmac found the structure to be a "mixed homodimer", where each monomer differed based on the presence or absence of the natural substrate UDP-Glc (**Figure 16A-D**).⁹⁶ The monomer containing only NAD and no UDP-Glc (herein the "apo structure") has well resolved density around the NAD cofactor (**Figure 16C-D**). Surprisingly, the monomer containing NAD and UDP-Glc has high ligand occupancy where density is observed for these two ligands (1.0 and 1.0 respectively), but no density is observed around C4 of the nicotinamide of NAD (**Figure 16B**), suggesting some degradation of the ligand. Additionally, the presence of UDP-Glc in the substrate binding pocket of one of the ligands was unexpected since, unlike NAD, UDP-Glc was not added to the protein prior to crystallisation. This would suggest that GALE was binding endogenous UDP-Glc during expression and carrying the substrate through the purification process. However, the affinity of GALE for UDP-Glc is not particularly strong ($K_D \sim 10 \mu\text{M}$), therefore it is surprising that this would not have been removed during the purification.¹¹⁵ How GALE is able to retain the UDP-sugar is uncertain.

The structure solved for condition A3 was then used in molecular replacement of crystals in condition H6 (Phosphate buffered protein). Model building with Coot followed by refinement found the structure to be a similar "mixed homodimer" to that observed for condition A3 with

NAD only and NAD plus substrate chains, again, with a “degraded” NAD missing density around C4 (**Figure 16E**). However, the structure obtained from condition H6 is missing significantly more density around the Glc moiety of UDP-Glc. The density displays a fully occupied UDP, but a total absence of Glc. Whether the hydrolysis occurred before, during, or after crystallisation is unclear. Molecular replacement was once again used to solve the structure for crystals from D10 (HEPES buffered protein). This time, almost complete density was observed for UDP-Glc, but again NAD displayed apparent “degradation” (**Figure 16F**) in the ligand bound chain. (Structural information and validation of all three crystals shown in **Table 3**, full details in section 2.5)

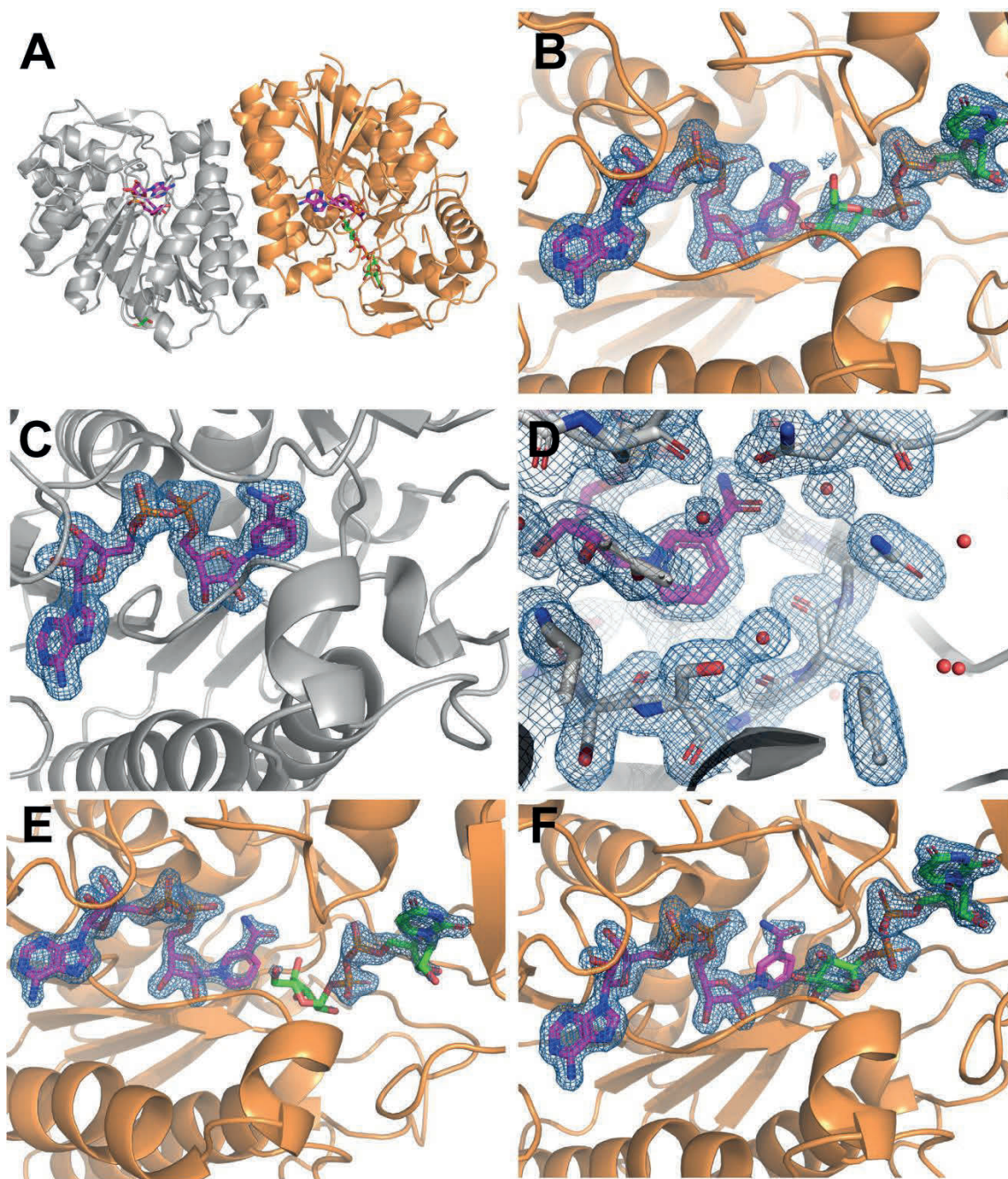


Figure 16: Structures obtained for crystals A3, H6 and D10.

A) Dimer obtained from crystal A3, phosphate buffered GALE 1EK5 screen. B) A3, Chain B ligand binding sites showing both UDP-Glc and NAD with ligand 2Fo-Fc maps. C) A3, "Apo" chain A with only NAD bound and NAD 2Fo-Fc maps, observed for crystals A3, H6 and D10. D) A3, "Apo" chain A with NAD and neighbouring residue 2Fo-Fc maps showing no substrate density, observed for crystals A3, H6 and D10. E) crystal H6, chain B showing both ligands bound and ligand 2Fo-Fc maps. F) Crystal D10, chain B showing both ligands bound and 2Fo-Fc maps. All maps are displayed at 1 Å σ level. Grey: apo chain, orange: ligand bound chain B, pink: NAD, green: UDP-Glc, blue mesh: 2Fo-Fc maps

2.2.3 GALE ITC Characterisation

Regardless of the differences observed for structures around the ligand bound protein chain in A3, H6, and D10 the GALE monomer containing only NAD is structurally similar to that obtained in 1EK5 (RMSD 0.38). Therefore, this apo monomer likely reflects the active apo structure of GALE in cells. The solved structure should be suitable for ligand soaking experiments to determine future inhibitor/protein interactions. It is not clear if the mixed dimer observed (one monomer substrate bound, one apo monomer) exists in solution as a mixed dimer, or if this only forms in crystallography.

The presence of UDP-Glc/UDP in the structure was unexpected. Prior work by *A. Pey et al.* has examined the effects of clinically relevant mutations on substrate binding to human GALE expressed from *E. coli*.¹¹⁵ To investigate this, the authors used isothermal titration calorimetry (ITC). ITC experiments determined that the expressed protein contained N values (where N is the proportion of active site available for binding, 0-1) of 0.20 and 0.23 with respect to UDP-Glc and NAD, respectively. They also noted several unsuccessful attempts to restore the blocked UDP-sugar substrate binding site.

To determine if the GALE expressed from insect cells has similar occupancy, ITC assays were performed (**Table 4, Figure 17A-B**). For both UDP-Glc and NAD, similar K_D and N values were obtained (**Table 4**). This would suggest that the occupancy of the active site is unaffected by both expression method and purification protocols.

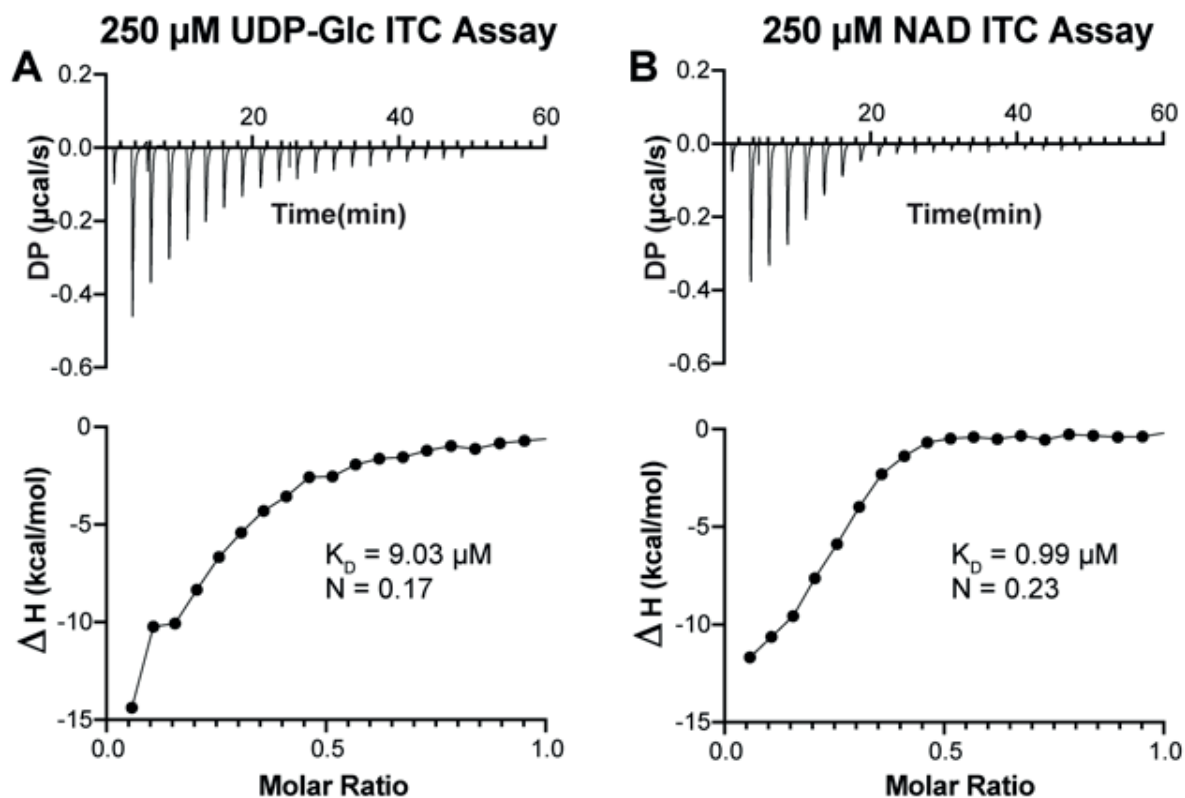


Figure 17: Insect cell expressed GALE ITC characterisation.

A) ITC assay thermogram (top) and Integral (bottom) plots with UDP-Glc at 250 μM . B) As in A, but with NAD instead of UDP-Glc. Conditions: 25 μM GALE, 25 mM HEPES pH 7.5, 50 mM NaCl

GALE source	UDP-Glc K_D (μM)	UDP-Glc N Value	NAD K_D (μM)	NAD N Value
Insect Cell Expressed	9.0 ± 4.1	0.17 ± 0.04	1.0 ± 0.1	0.24 ± 0.04
Published Values ¹¹⁵	7.9 ± 1.1	0.20 ± 0.05	1.0 ± 0.1	0.23 ± 0.01
(NH ₄) ₂ SO ₄ Precipitated	7.8 ± 2.1	0.17 ± 0.05	n.d.	n.d.

Table 4: ITC results for different GALE preparations.

Where N is the occupancy of the enzyme and K_D is the dissociation constant

Additionally, the occupancy of the active site by UDP-Glc has also been noted for the *E. coli* homologue of GALE by J. L. Vanhooke and P.A. Frey.¹²⁸ They identified a so called “abortive complex” that occurs when UDP-Glc binds to the enzyme that contains NADH, the reduced

form of NAD. The hypothesis is that after the abstraction of a hydride forming the 4-keto-pyranose intermediate and NADH, dissociation of the 4-keto sugar followed by binding of UDP-Glc (or other substrate) causes a conformational change that traps the substrate in the enzyme, inactivating it in the process. They also highlight that this occurs during expression.

2.2.4 GALE Ammonium Sulfate Precipitation

From the structures and ITC data obtained, one monomer likely represents an abortive complex (approximately 85 % of GALE purified) whilst the other monomer is active GALE (15% of GALE purified). However, because of the poor density around the “NAD” in the obtained structures, it is difficult to confirm if this is NAD or NADH present, i.e. it is not obvious if these structures show the abortive complex (NADH and UDP-Glc), or just GALE in complex with UDP-Glc and NAD. Again, despite the uncertainty around the substrate bound chain, the apo chain was still suitable for future soaking experiments, but the consistency of crystal growth still needed optimisation and a monomer or dimer with only apo GALE (similar to that obtained in 1EK5) may be more reproducible. To achieve this, the purified enzyme was precipitated with $(\text{NH}_4)_2\text{SO}_4$ to remove the bound UDP-sugar, since $(\text{NH}_4)_2\text{SO}_4$ was used in the purification of GALE that produced the apo monomer 1EK5. $(\text{NH}_4)_2\text{SO}_4$ decreases the solubility of the protein by disrupting the hydration layers (water molecules will instead form around the NH_4^+ and SO_4^{2-} ions) causing precipitation, a process known as “salting-out”. In principle, the bound ligands (NAD and UDP-Glc) could remain soluble or else be released from the protein due to a shift in ionic conditions. The K_D and N values of the precipitated and repurified enzyme were then measured by ITC for UDP-Glc as before, the results summarised in **Table 4** and shown in **Figure 18**. Despite precipitation and dialysis, the same N value (0.17 vs 0.17) was obtained suggesting the precipitation had failed to remove the abortive complex/bound UDP-Glc and reactivate the protein.

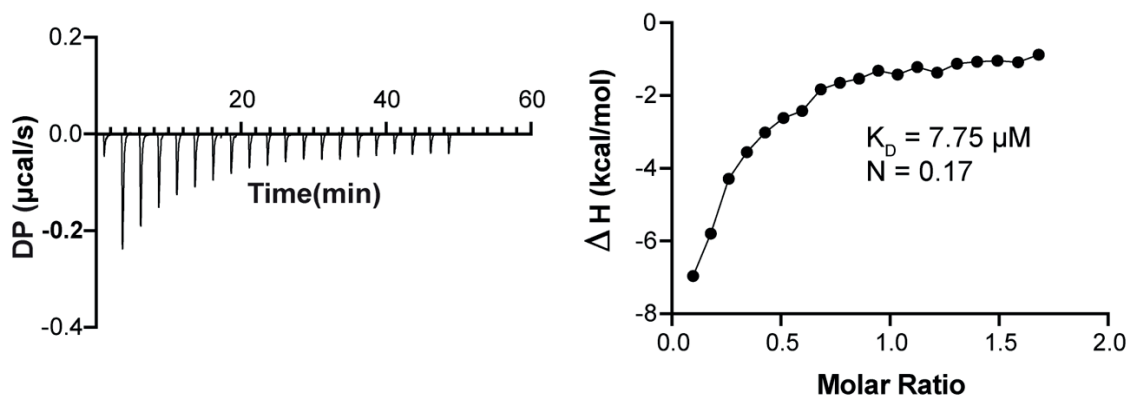


Figure 18: $(\text{NH}_4)_2\text{SO}_4$ precipitated GALE ITC assay thermogram (left) and Integral plot (right) with UDP-Glc at 250 μM

Next, crystallography trays using the $(\text{NH}_4)_2\text{SO}_4$ precipitated GALE preparation were set up to establish if there was any effect on the frequency of crystal formation or structure. Trays were set up with the same condition screen based around the 1EK5 conditions described previously. As shown in **Figure 14D and E**, there was no change in the crystal hit rate, however crystals were observed in different conditions to previous screens. Collectively, the spread of crystal growth across the three 1EK5 screens suggest that crystal formation is less dependent on KCl and PEG 8K concentrations and are instead determined by random nucleation events. Crystals from this plate were submitted for analysis at the DLS. It should be noted that at this stage structures were only solved using Phaser molecular replacement or Dimple (see methods 2.4.4.2), since it was clear from Dimple (using the previously solved structures as the input model) and the 2Fo-Fc maps, whether UDP-Glc was bound or not. The determined structure was the same as that observed previously, a mixed homodimer of apo and substrate-bound GALE, notably with well resolved complete NAD density in the apo chain A (**Figure 19A-D**). The state of NAD (degraded or complete) appears indicative of whether UDP-Sugar was bound.

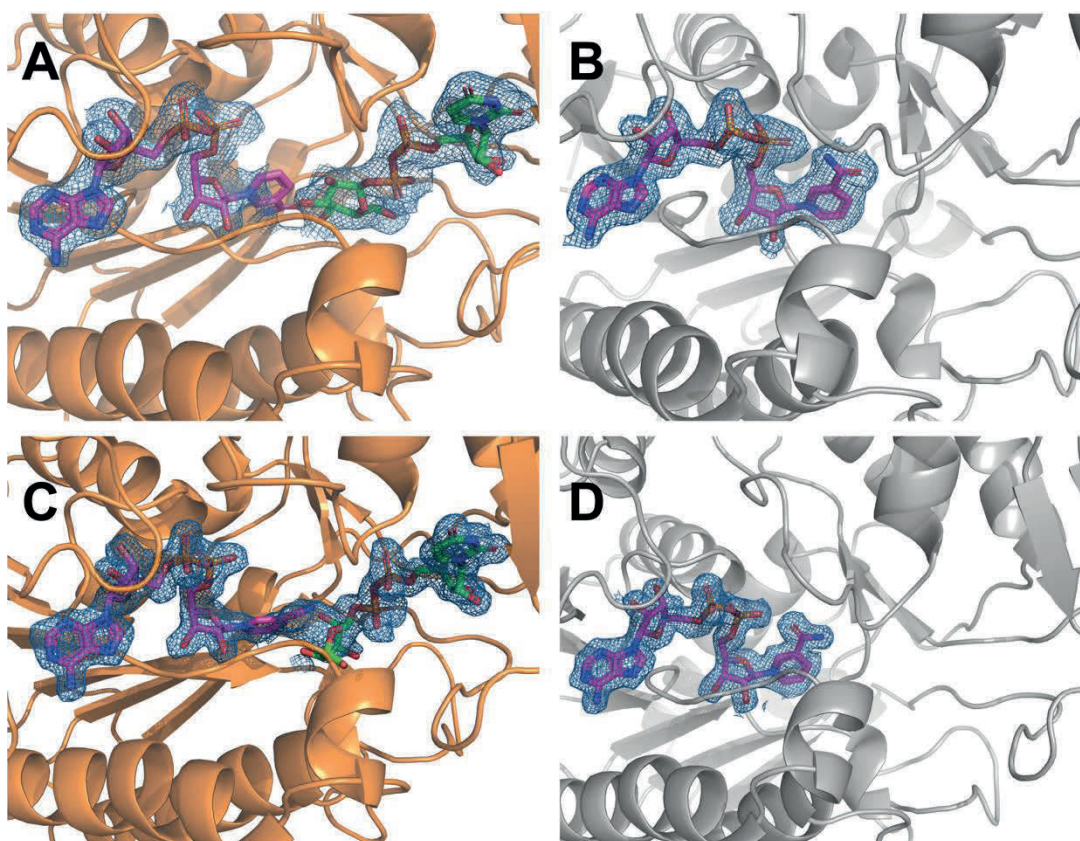


Figure 19: Crystal structures obtained for crystals C11 and F8.

A) C11 ligand bound chain B with NAD and UDP-Glc bound with ligand 2Fo-Fc maps. B) C11 apo chain A with only NAD bound and ligand 2Fo-Fc maps. C) As for panel A but for crystal from well F8. D) As for panel B but for crystal from well F8. Grey: apo chain A, orange: ligand bound chain B, pink: NAD, green: UDP-Glc, blue mesh: ligand 2Fo-Fc maps. All maps are displayed at 1 Å σ level.

Well	F8	C11
Wavelength (Å)	0.98	0.98
Resolution range (Å)	85.63 - 1.62	45.91 - 2.04
Space group	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁
Cell dimension:		
a, b, c (Å)	45.9, 110.0, 136.3	45.9, 109.9, 136.0
α , β , γ (°)	90, 90, 90	90, 90, 90
R _{work}	0.30	0.27
R _{free}	0.33	0.33

Table 5: Structure validation and information for the ammonium sulfate precipitated GALE crystals analysed from wells F8 and C11.

Full statistical ("Table 1" values) validation in section 2.5.

2.2.5 GALE Crystallography – Seeding

Given that the protein is active, expressed in good yields and crystallises to give a reproducible structure with a free substrate binding pocket, the protein was considered suitable for use in fragment screening and inhibitor development. However, improvements in the crystallisation frequency and reliability were still needed. Previous screens indicated nucleation rather than precise conditions were essential for crystal growth. As discussed in the introduction, seeding is a crystallographic method used to drive crystallisation by providing a template of small fragments of protein crystals. Since our previous screens showed poor frequency and consistency of crystal growth, a seed stock of crushed GALE crystals was made to use in seeding. A full plate with 0.1 M EPPS, pH 7.5, 136 mM KCl, 19.3 % w/v PEG 8K was set up with a 1.5x serial seed dilution series, where each row contained one seed concentration (**Figure 20A**). As expected, significantly improved hit rates were observed (**Figure 20A and 20E**). High seed concentrations gave smaller crystals but larger quantities, whilst more dilute seed concentrations produced the inverse (**Figure 21A**). The seed concentration which produced the best combination of size and quantity (row C, 2.25-fold dilution from stock) was used to set up three crystal trays with one seed concentration and condition (**Figure 20B-D**). This produced more consistent crystal morphology (**Figure 21B**). Several crystals from the plate in **Figure 20B** were analysed by X-ray diffraction to confirm an apo structure was still present. Surprisingly, both crystals for which data was collected (E11 and D9, plate B in **Figure 20**) showed no evidence of UDP-Glc or any other ligand in either binding site, only density for NAD was observed (**Figure 22A-D**). This may explain the improved crystal hit rate observed. Importantly even if there is a mixture of mixed homodimer and apo dimer within the crystals obtained, both are suitable for soaking with future hit ligands since both contain at least one apo monomer.

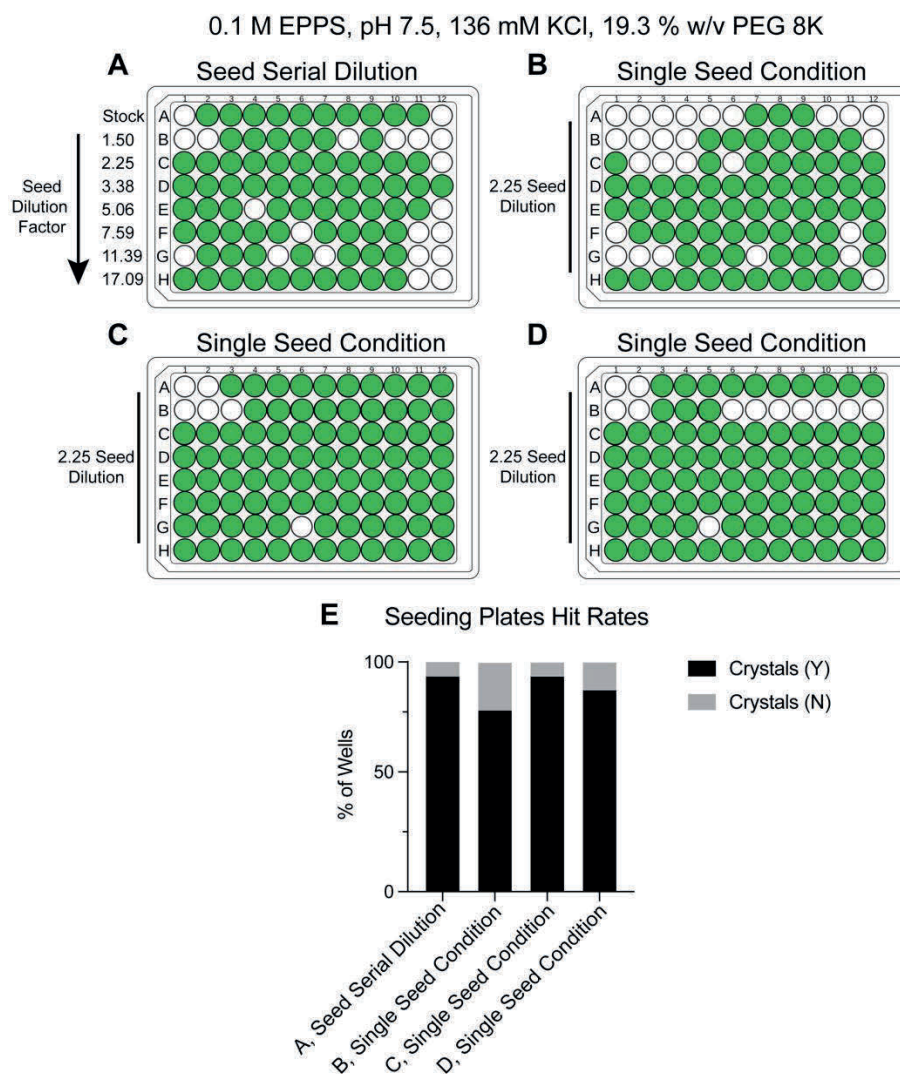


Figure 20: Hit rate summary for crystallography plates set up using crystal seeding.

All seeding trays used HEPES buffered GALE. A) Hit location for a crystallography tray using a single precipitant condition but with different concentrations of seed stock in each row (seed dilution factor shown vertically). B-D) Hit location for crystallography plates using single precipitant and single seed (2.25x) concentration. Green = crystal growth, White = no crystal growth E) Summarised hit rates for plates A-D. Black = % with crystal growth, grey = % of wells with no crystal growth.

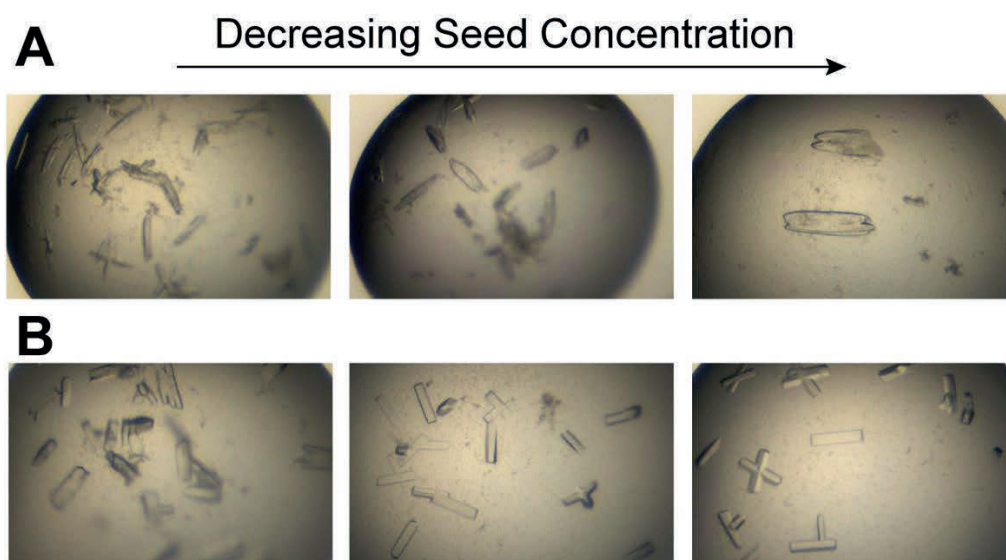


Figure 21: Representative images showing crystal morphology from crystal seeding experiments.

A) Serial seed dilution crystallography drops showing the effect of decreasing seed concentration on crystal growth. B) Representative wells showing single seed concentration drops (2.25x seed dilution from stock). All images taken on day 5.

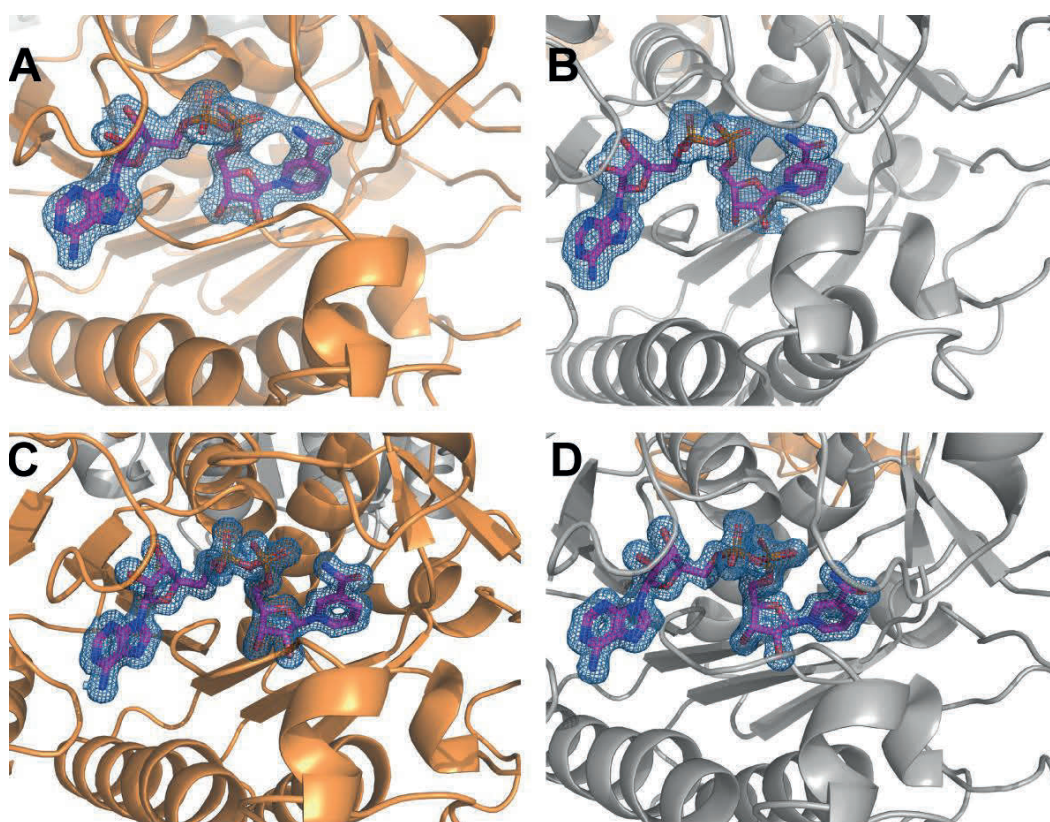


Figure 22: Crystal structures obtained for crystals E11 and D9.

E11 apo chain B with NAD bound with ligand 2Fo-Fc maps. B) E11 apo chain A with only NAD bound and ligand 2Fo-Fc maps. C) As for panel A but for crystal from well D9. D) As for panel B but for crystal from well D9. Grey: apo chain A, orange: ligand bound chain B, pink: NAD, green: UDP-Glc, blue mesh: ligand 2Fo-Fc maps. All maps are displayed at 1 Å σ level.

Well	D9	E11
Wavelength (Å)	0.98	0.98
Resolution range (Å)	67.66 - 1.43	46.38 - 2.00
Space group	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁
Cell dimension: a, b, c (Å)	46.1, 110.8, 135.3	45.3, 109.6, 137.1
α , β , γ (°)	90, 90, 90	90, 90, 90
R _{work}	0.25	0.22
R _{free}	0.26	0.28

Table 6: Structure validation summary for the GALE crystals grown using seed stock from wells D9 and E11.

Full statistical validation ("Table 1" values) in section 2.5

2.3 Chapter 2 Summary

In summary, human GALE was expressed and purified from Insect cells using the baculovirus expression system. The enzyme was folded and active with kinetic parameters similar to those determined in prior publications. ITC experiments determined that approximately 80-85 % of the enzyme is in a so called “abortive complex” potentially formed through the binding of UDP-Glc to GALE containing NADH rather than the cofactor NAD, however this was not clear from crystallography due to apparently degradation of NAD, indicated by an absence of density around the nicotinamide ring. Precipitation and dialysis of the protein had no effect on the amount occupancy of the UDP-Sugar substrate binding pocket. Multiple crystallography conditions produced a GALE structure with a “mixed homodimer” with one monomer containing only NAD, and the other containing apparently degraded UDP-Glc and NADH/NAD. Optimisation of the conditions and the use of crystal seeding increased the consistency of crystal growth from 12 % to 76 %. X-ray diffraction experiments using several crystals grown with seeding revealed a structure containing only NAD in both monomers. From these results, we believed we had a suitably active and stable protein for use in biophysical screening assays and a crystal system that would be applicable for determining hit fragment/protein complexes.

2.4 Methods

2.4.1 Cloning and Protein Expression

The method of protein cloning, expression, and purification has been published in prior work by myself (William Browne) and has been described in an earlier MRes report and therefore these methods have not been included here.¹²⁹

For later crystallography experiments requiring potassium phosphate buffered GALE, ~5 mg of protein was dialysed overnight against 2 L of 150 mM potassium phosphate pH 7.5, 150 mM NaCl buffer. The sample was then reconcentrated to 20 mg/mL.

2.4.1.1 Ammonium Sulfate precipitation

4.1 M $(\text{NH}_4)_2\text{SO}_4$ was used to precipitate ~ 6 mg of GALE (added to 55% $(\text{NH}_4)_2\text{SO}_4$ saturation). The $(\text{NH}_4)_2\text{SO}_4$ was added fast. The sample was spun at 2000g for 10 minutes and the supernatant discarded. The precipitate was resuspended using a 25 mM HEPES, pH 7.5, 500 mM NaCl buffer, dialysing overnight against 2 L of the same buffer. The sample was then reconcentrated to 20 mg/mL (termed $(\text{NH}_4)_2\text{SO}_4$ precipitated GALE).

2.4.2 Epimerisation Assay

Epimerisation assays were 20 μL reactions performed in 25 mM HEPES, pH 7.5, 50 mM NaCl, 200 μM NAD with variable concentrations of UDP-Gal and GALE. Reactions were run at 37 °C for 15 minutes. Quenching was carried out by heat inactivation at 95 °C for 1 minute. Samples were cooled to 4 °C and analysed by high performance liquid chromatography (HPLC) using a ArcPremier Waters instrument with a Dionex CarboPac PA1 IC (2mm x 150 mM) column (Waters, Wilmslow, UK). Injections were 5 μL . Eluent A was 1 M NaOAc/1 mM NaOH (Sigma-Aldrich). Eluent B was 1 mM NaOH. Eluent change detailed below.

Time (min)	% Eluent A
0	60
4	64
8	68
10	70
12	72
14	72
16	76
22	82
27	82

Table 7: HPLC assay elution conditions.

The % conversion for each assay was determined by the relative difference in peak area between UDP-Gal and UDP-Glc peaks. Kinetics factors (K_m and k_{cat}) were determined by non-linear curve fitting using Graphpad Prism 9.0 (GraphPad Software). Calculations used a modified version of the Michaelis-Menten equation, shown below.

$$v(GALE) = \frac{[GALE]_t \cdot k_{cat} \cdot [UDP - Gal]}{K_m + [UDP - Gal]}$$

Equation 3: Michaelis-Menten Equation as used in GraphPad.

Where v is the initial rate in $\mu\text{M/s}$.

K_m is the Michaelis-Menten constant in μM .

$[GALE]_t$ is the concentration of GALE in μM in the assay.

k_{cat} is the turnover number in s^{-1} .

$[UDP - Gal]$ is the concentration of UDP-Gal at the start of the reaction in μM .

2.4.3 Protein ITC Characterisation

ITC characterisation was performed using a MicroCal PEAQ ITC (Malvern Panalytical, Malvern, United Kingdom) system at 25 °C. 300 μL purified GALE at 50 μM (30 μM for precipitated GALE experiment) in 25 mM HEPES, pH 7.5, 50 mM NaCl buffer, was loaded into the sample cell.

Ligands were prepared to match protein buffer composition. Ligands were titrated into the cell with 1x0.4 μ L injection followed by 19x2 μ L injections, 4 s injection durations with 150 s spacing. Heat release per titration was measured and integrated using the MicroCal PEAQ ITC software. All kinetic parameters such as K_d , Enthalpy (ΔH), Gibbs free energy (ΔG), and stoichiometry (N) were determined using a single site occupancy model.

2.4.4 Protein Crystallography

In general, GALE expressed and purified as described above was concentrated to 20 mg/mL and NAD added to a final concentration of 2 mM. The protein was crystallised in MRC 2 drop Swisssci plates (Swisssci, Buckinghamshire, UK) by sitting drop vapour diffusion. Drops consisted of 100 nL GALE/NAD with 100 nL of precipitant mixture equilibrating with 40 μ L of precipitant. In seeding experiments, 10 nL of seed stock and 90 nL of precipitant was used. Drops were dispensed by NT8 (Formulatrix, Frankfurt, Germany). Plates were then sealed with a clear adhesive film. Wells were monitored by an automated Rock Imager 1000 with RockMaker software (Formulatrix). All images shown in the results section are taken on day 5 after plate setup.

2.4.4.1 Condition Screening

Several custom condition screens were set up to optimise for crystallisation conditions:

- A condition screen based on PDB:1EK5⁹⁶ was set up with the following concentration ranges: EPPS 0.1 M, 17-22% PEG 8K, 0.05-0.25 M KCl. Crystals were fished and frozen using cryoprotectant described as follows: 0.1M EPPS, 0.25M KCl, 20% PEG 8K, 20% ethylene glycol). This screen was performed with samples of GALE in 25 mM HEPES, pH 7.5, 500 mM NaCl, 2 mM NAD and 10 mM K_3PO_4 , pH 7.5, 150 mM NaCl, 2 mM NAD. This screen was repeated with GALE precipitated with $(NH_4)_2SO_4$

- A seeding screen performed using seed stock generated by vortexing crystals from plate 1 described above and making a 1.5-fold serial dilution (A-H) in 100 mM EPPS, pH 7.5, 136 mM KCl, 19.3 % w/v PEG 8K. Each plate row contained a different dilution.
- A full plate using the single condition 100 mM EPPS, pH 7.5, 136 mM KCl, 19.3 % w/v PEG 8K, seed dilution C and MRC 3 drop Swissci plates.

2.4.4.2 Data Collection

Data was collected at the DLS remotely and performed by Dr. Andy Purkiss.

Data for all other crystals was collected on Beamline IO4 (DLS). Datasets consisted of 1800 images with an oscillation of 0.1 °, exposure time of 40 ms per image per crystal. Data sets were indexed, scaled and merged with xia2.¹³⁰ In general, data was imported into CCP4 cloud and molecular replacement performed using either MOLREP or PHASER.¹³¹ For crystal A3 (detailed in the results and discussion) this was performed using published structure 1EK5 as the search model (PDB: 1EK5⁹⁶). Subsequent structures were solved by DIMPLE using the “mixed homodimer” determined for crystal A3. Structures were refined with rounds of Refmac and manual model building in COOT.^{132,133}

2.5 Chapter 2 Crystal Structure “Table 1” Validation

Well	A3	H6	D10
Wavelength (Å)	0.98	0.98	0.98
Resolution range (Å)	55.05 – 1.88	67.97 – 1.73	57.75 – 1.80
Space group	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁
Cell dimension: a, b, c (Å)	45.9, 110.1, 135.5	45.8, 109.8, 135.9	45.7, 109.8, 135.8
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Total reflections	1 020 303 (51 052)	972 967 (40 283)	64426
Unique reflections	78 572 (3889)	72 637 (3535)	64426
Multiplicity	13.0 (13.0)	13.4 (11.4)	12.7 (7.3)
Completeness (%)	100 (100)	100 (97.8)	100 (99.3)
Mean I/sigma(I)	10.4 (0.7)	7.9 (0.3)	8.8 (0.5)
Wilson B-factor	31.6	25.3	23.3
Reflections used in refinement	51 052	72 534	64 374
Reflections used for R_{free}	2 773	3 572	3 179
R_{meas}	0.36 (4.65)	0.06 (5.06)	
CC1/2	1 (0.30)	1 (0.41)	1 (0.3)
R_{work}	0.20	0.23	0.21
R_{free}	0.24	0.26	0.25
RMSD - bond lengths	0.01	0.01	0.01
RMSD - bond angles	1.59	1.65	1.46
Ramachandran favored (%)	97.2	96.7	97.1
Ramachandran allowed (%)	2.8	3.3	2.9
Ramachandran outliers (%)	0.0	0.0	0.0
Rotamer outliers (%)	4.5	4.3	4.5
Clash score	7.6	5.8	6.7

Table 8: Full statistical validation for crystals A3, H6 and D10.

Well	F8	C11
Wavelength (Å)	0.98	0.98
Resolution range (Å)	85.63 - 1.62	45.91 - 2.04
Space group	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁
Cell dimension: a, b, c (Å)	45.9, 110.0, 136.3	45.9, 109.9, 136.0
α, β, γ (°)	90, 90, 90	90, 90, 90
Total reflections	1 140 586 (5 5846)	500 888 (19 006)
Unique reflections	88 670 (4 317)	44 507 (2 130)
Multiplicity	12.9 (12.9)	11.3 (8.9)
Completeness (%)	100 (98.0)	99.4 (95.4)
Mean I/sigma(I)	2.4 (0.5)	6.7 (0.9)
Wilson B-factor	15.42	28.44
Reflections used in refinement	55 846	44 430
Reflections used for R _{free}	2 773	2 207
R _{meas}	0.11 (2.18)	0.09 (2.58)
CC1/2	1 (0.3)	1 (0.5)
R _{work}	0.30	0.27
R _{free}	0.33	0.33
RMSD - bond lengths	0.01	0.01
RMSD - bond angles	1.43	1.53
Ramachandran favored (%)	95.2	95.5
Ramachandran allowed (%)	4.5	4.1
Ramachandran outliers (%)	0.3	0.4
Rotamer outliers (%)	2.4	5
Clash score	14.1	10.3

Table 9: Full statistical validation for crystals F8 and C11.

Well	D9	E11
Wavelength (Å)	0.98	0.98
Resolution range (Å)	67.66 - 1.43	46.38 - 2.00
Space group	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁
Cell dimension: a, b, c (Å)	46.1, 110.8, 135.3	45.3, 109.6, 137.1
α, β, γ (°)	90, 90, 90	90, 90, 90
Total reflections	3 263 471 (128 147)	668 546 (32 886)
Unique reflections	128 880 (6361)	51 174 (2535)
Multiplicity	25.3 (20.1)	13.1 (13.0)
Completeness (%)	100 (100)	100 (100)
Mean I/sigma(I)	10.5 (0.8)	10.4 (0.7)
Wilson B-factor	11.6	38.2
Reflections used in refinement	128 870	51 103
Reflections used for R _{free}	6 351	2 487
R _{meas}	0.09 (5.62)	0.05 (3.9)
CC1/2	1 (0.3)	1 (0.3)
R _{work}	0.25	0.22
R _{free}	0.26	0.28
RMSD - bond lengths	0.009	0.008
RMSD - bond angles	1.64	1.56
Ramachandran favored (%)	98.1	97.7
Ramachandran allowed (%)	1.9	2.3
Ramachandran outliers (%)	0	0
Rotamer outliers (%)	5.2	3.6
Clash score	2.9	3.3

Table 10: Full statistical validation for crystals D9 and E11

3. Chapter 3 Non-Covalent Fragment Screening and Elaboration

3.1 Chapter Introduction

3.1.1 Chapter Overview

The previous chapter described the expression of active and stable GALE, as well as the optimisation of robust crystallography conditions. In this chapter the aim was to combine two orthogonal screening methods to identify fragment binders of GALE. Thermal shift screening would provide a cost and time effective initial screen, whilst crystallographic screening (with the same library) would identify hits and provide immediate structural information. Comparison of each hit list would help to validate and identify which fragments should be focused on for elaboration. To further aid follow up elaboration screening was performed with the Enamine “DSI-Poised library”.¹³⁴ This library contains fragments (< 300 MW) that are designed to consist of two distinct functional groups linked by an easily synthesisable chemical bond, such as an amide, imine or aryl-aryl bond. This enables rapid analogue synthesis by combinatorial chemistry from commercially available analogues of each functional group. The design of follow up compounds would then be further guided by virtual docking/screening of potential analogues.

3.1.2 Fragment Screening Methods

3.1.2.1 Screening Overview

Focus was placed first on developing a suitable screening method to identify fragments that bind to GALE. Typically employed biophysical binding assays are thermal shift (TS), isothermal titration calorimetry (ITC), NMR, and surface plasmon resonance (SPR) assays.¹³⁵ It is also not uncommon for fragments to be screened virtually with docking software.¹³⁵

3.1.2.2 Thermal Shift Assays

Thermal shift screening has been used to identify fragment hits that have been developed into more potent binders.^{136,137} Thermal shift assays operate based on the principle that a protein in complex with a ligand has a greater degree of the thermal stability than a protein in the

absence of the ligand. Additionally, tighter binders make stronger interactions with the protein and therefore provide more resistance to thermal denaturation than weaker binders.¹³⁸ For thermal shift assays, the stability of the enzyme in solution is described by its T_m , the temperature at which 50% of the protein is denatured. The denaturing of the enzyme can be measured using fluorescent dyes which bind to the enzyme in a non-specific manner as the enzyme unfolds during heating. One of the most frequently used dyes is SYPROTM orange. As more dye molecules bind to the protein, the fluorescence intensity of the dye increases and can be measured.¹³⁹ A major benefit of thermal shift assays is that they are both inexpensive and high throughput, making them an ideal option for academic drug discovery projects. However, the thermal shift assay does suffer from high false-negative and false-positive rates, this is largely since protein thermal stability is not only affected by ligand binding but also a range of other factors such as solubility, pH, and salt concentration.¹⁴⁰ TS assays therefore require extensive validation to minimise the false positive/negatives.

3.1.2.3 Crystallographic Screening

A more novel approach has been to screen libraries of fragments using X-ray crystallography.¹⁴¹ Where typically fragment libraries are screened first by a biophysical technique and the identified hits used in crystal fragment soaking experiments to produce fragment-protein complexes, crystallographic screening performs this process with every fragment in a library. This ensures the maximum amount of structural information regarding fragment-protein binding is obtained, even for fragments that are very weak binders since soaking can be performed in the 10s of mM range.¹⁴¹ The protein/ligand obtained enables elaboration guided by structural insights, an approach known as structure-based drug design (SBDD).¹⁴² In summary, these features make crystallographic screening an extremely powerful technique for identifying hits for difficult-to-target proteins, and the wealth of structural information helps to accelerate elaboration. The limitation of this method is the significant barrier posed by requiring ~1000 reproducible crystals due to library size, resolutions < 2.5 Å, and crystals stable to either DMSO or ethylene glycol.

3.1.2.4 Virtual Screening with ICM

SBDD is a powerful approach for developing small molecules ligands for proteins and is used throughout the drug discovery process. Ligands can be elaborated from initial hits using protein/ligand structures by rationalising favourable interactions between the protein and ligand. However, due to advances in computational power over the last 30 years and the now vast number of protein structures in the protein data bank (PDB), virtual docking of ligands to predict binding strength has become routine.¹⁴³ Virtual docking software uses computer algorithms to model proteins and ligands such that favourable interactions can be predicted. A score can then be produced for each ligand to enable ranking based on theoretical binding strength, thus enabling screening of entire small molecules libraries.¹⁴³ Virtual screening is an extremely effective method for assessing the effects of a variety of ligand structural changes, so that fewer compounds need testing experimentally.

In this chapter, virtual screening was effectively used to facilitate fragment elaboration. The docking software was produced by Molsoft using a technique called Internal Coordinate Mechanics (ICM).^{144,145} In ICM, docking is performed using a biased probability Monte Carlo approach to position ligand coordinates in a grid of potential maps calculated for a defined protein pocket/surface. The receptor maps of the protein are calculated using grid spacings of 0.5 Å and five potentials; van der Waals potentials for both a hydrogen and a non-hydrogen probe, hydrogen bonding, electrostatic and hydrophobic potentials. The force field (FF) used for this task is the Empirical Conformational Energy Program for Peptides (ECEPP).¹⁴⁶ Notably, this FF maintains fixed values for bond length and angle, with only the dihedral torsion angle being variable. In this context, a biased probability Monte Carlo simulation enables random sampling of the ligand within the receptor maps.¹⁴⁷ This is necessary since it is inefficient to sample every position and pose of a group of ligands within a protein pocket. Instead, a random sample of positions within the receptor and conformations of the ligand and are assessed, this is then repeated with a new position/conformation until all ligand conformations are assessed. The scores are weighted based on 1) The internal energy of the ligand. 2) Entropic penalty of binding. 3) hydrogen bonding interactions. 4) polar and non-polar solvation energy change upon binding. 5) Electrostatic energy. 6) Hydrophobic energy.¹⁴⁵

3.2 Results and Discussion

3.2.1 DSI-Poised Library Thermal Shift Fragment Screening

Thermal shift (TS) serves as a suitable primary binding screening assay to find putative binders of the protein of interest. Thus, a TS assay for GALE was first validated with the objective to screen the Enamine DSI poised library, by first assessing the effects of pH and DMSO. The pH was optimised to ensure maximum enzyme stability. The results show a steady increase in thermal stability between pH 5.5 and 7.5 (**Figure 23A**), with a maximum of 45.2 °C. The stability then decreases slightly at pH 8.0 to 44.9 °C. Given the apparent decrease in stability beyond pH 7.5, and GALE being a cytosolic enzyme and therefore likely to be at physiological pH (7.4) in cells, no higher pH was assessed. The effect of DMSO on enzyme stability was also assessed. A limitation of fragment screening is its lack of sensitivity, requiring screening to be performed at high concentrations (mM). Consequentially, DMSO concentrations are often between 0% and ~5% (v/v) depending on stock concentrations. The effect on thermal stability is a near linear correlation with a total decrease in the T_m of 1.6 °C between 0% and 4.5% (v/v) DMSO (**Figure 23B**). These results demonstrate that DMSO is well-tolerated by GALE at pH 7.5.

The thermal shift assay was then validated with known binders of GALE. The assays were run with a 2-fold dilution series of both UDP-GalNAc and NAD (**Figure 23C-D**). The UDP-GalNAc series produced a typical dose response with plateauing above 200 μ M, whilst NAD displayed a saturation curve up to 800 μ M. Both results indicate that the thermal shift assay is suitable for detecting small molecule binders.

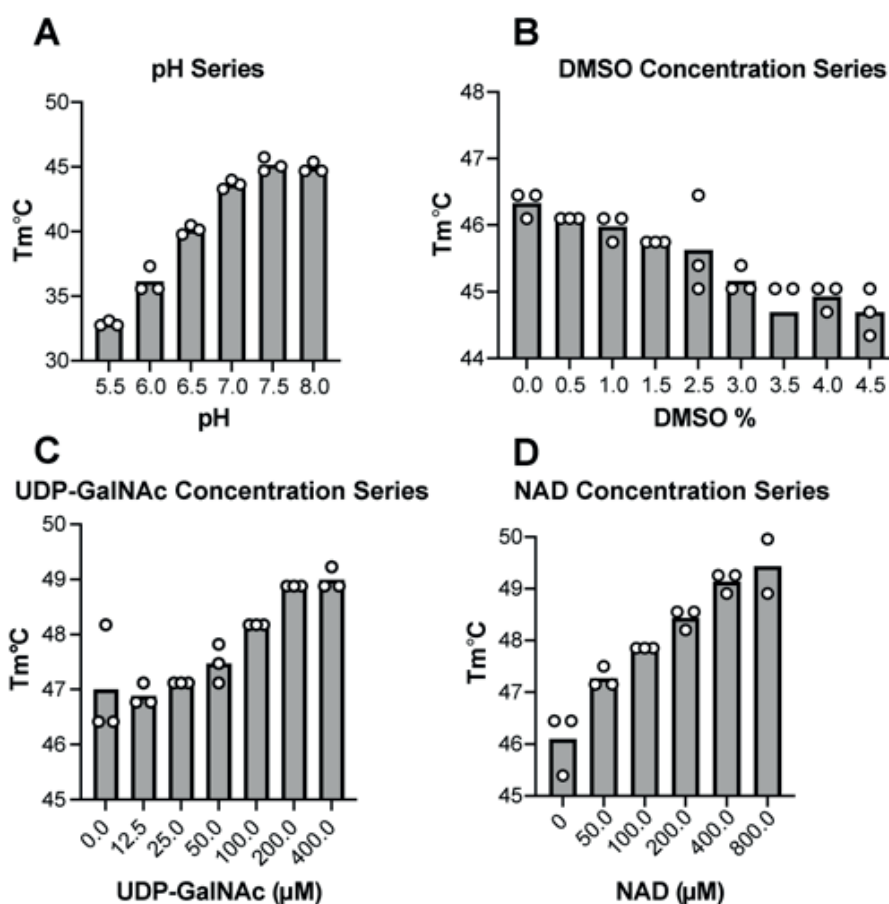


Figure 23: Thermal shift assay condition optimisation.

A) pH series performed from pH 5.5-8.0. B) Thermal shift assay performed at different DMSO concentrations (% V/V). C) Results of a thermal shift assay performed with varying concentrations of UDP-GalNAc. D) Thermal shift assay performed with varying concentrations of NAD. All experiments were performed in technical triplicate with mean averages. 2 μM GALE was used in all assays.

With these conditions optimised and valid positive controls, the Enamine DSI-Poised library was selected for screening. To overcome some of the limitations of thermal shift assays, namely their false positive rate and low sensitivity, screening was performed in technical duplicates at both 1 mM and 4 mM fragment concentrations. Compounds that showed a positive dose-response relationship by increasing stability from 1 to 4 mM are more likely to be real hits, providing an extra filter to differentiate fragments. An overview of the workflow for this screen is shown in **Figure 24**.

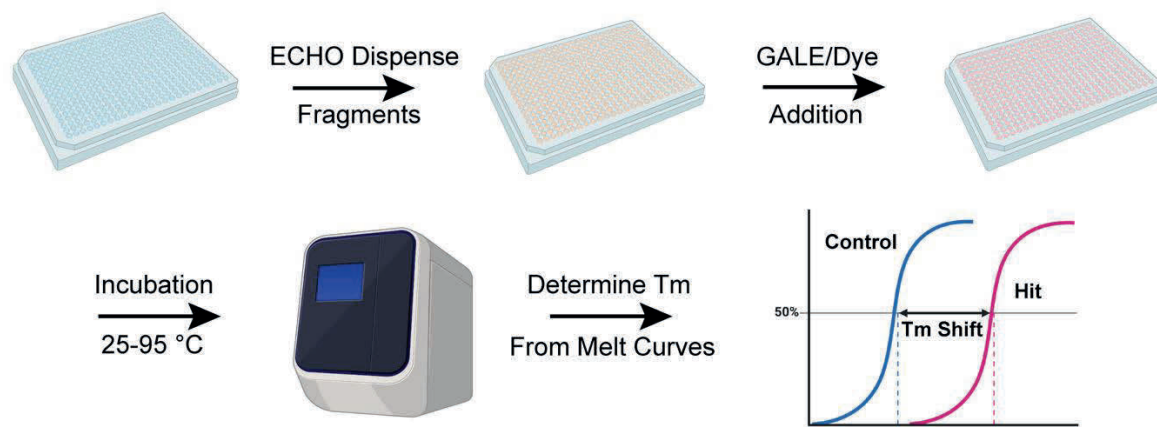


Figure 24: Thermal shift fragment screening workflow.

1) fragments are dispensed by Echo into 384 well plates. 2) GALE/Dye mixtures dispensed into each well. 3) samples are heated from 25 °C to 95 °C and fluorescence measured by RT-PCR instrument. 4) melt curves are recorded and the T_m determined.

The results of the fragment screen are summarised in **Figure 25**. Hits were identified as compounds that stabilised the protein with a T_m greater than the maximum T_m recorded from the DMSO controls (56 controls per plate). The spread of values was greater for fragments at 4 mM compared to 1 mM, likely due to secondary effects such as compound solubility and protein aggregation at higher concentrations. To further refine the hit list, fragment hits were separated into two tiers. Tier 1 fragments were hits at both 1 mM and 4 mM and showed an increase in the ΔT_m from 1 mM to 4 mM, suggesting a positive dose-response as seen with positive controls. Tier 2 hits were hits at 4 mM only. 4 mM was chosen to identify not only binders, but also compounds with high hydrophilicity. Destabilising compounds were not considered since it was rationalised that compounds binding to the substrate binding pocket would be likely to stabilise the protein, as seen with the natural substrate. The ΔT_m values for hits from both lists are shown in **Figure 26A-B**. The ΔT_m values ranged from 0.4 – 1.5 °C, which is in agreement with published values from thermal shift fragment screens.¹³⁶ Six compounds gave a concentration dependent change in ΔT_m (**Figure 26C**).

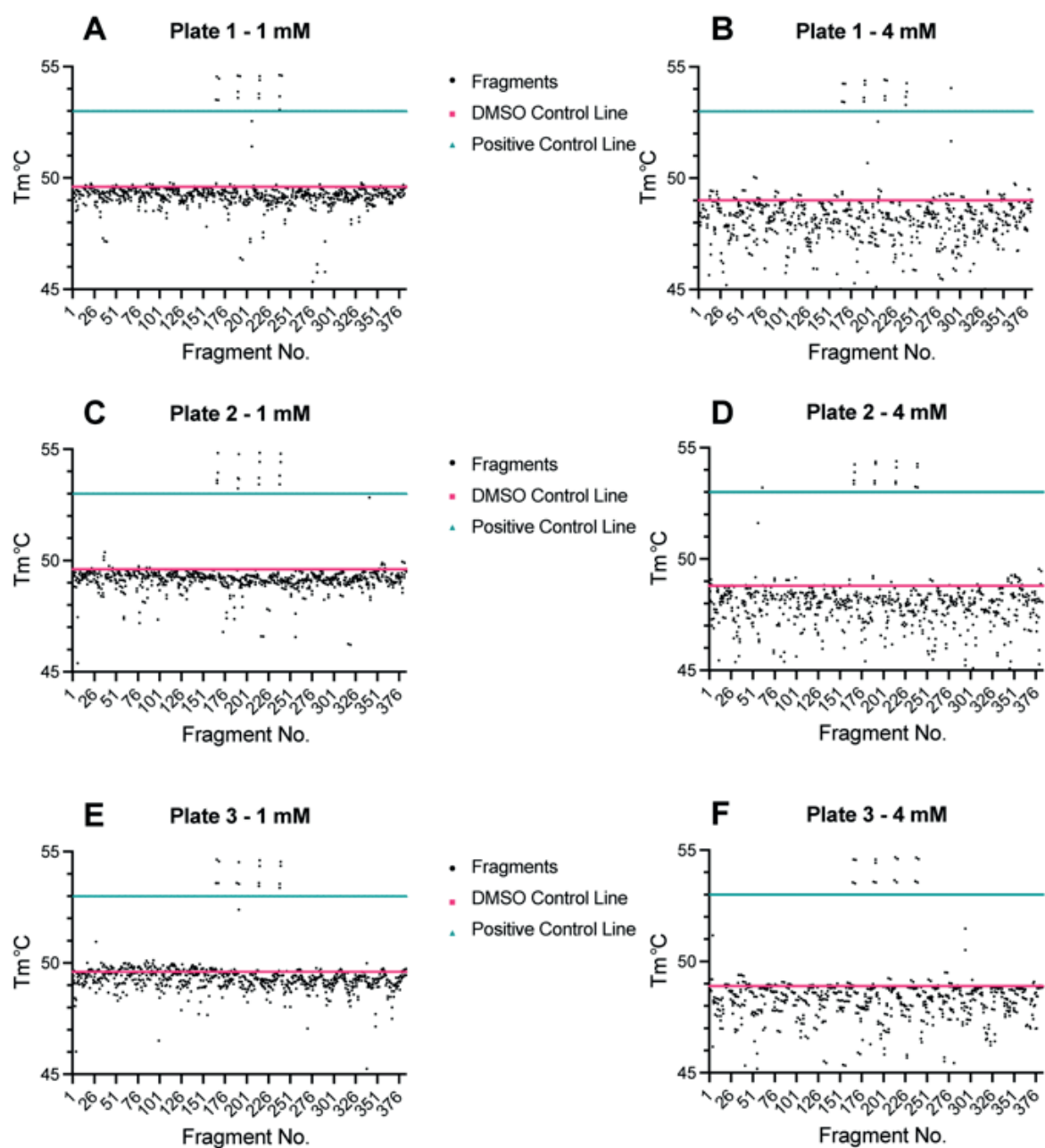


Figure 25: Thermal shift fragment screening results.

A-F) The DSI library covers three 384-well plates. The results for each plate are shown for assays at 1 mM (A, C, E) and 4 mM (B, D, F) fragment concentration. The DMSO control line (pink) shows the maximum T_m value obtained from the DMSO control samples. Anything above the positive control line (turquoise) is a result from one of the positive controls. Fragments in-between these two lines can be considered hits.

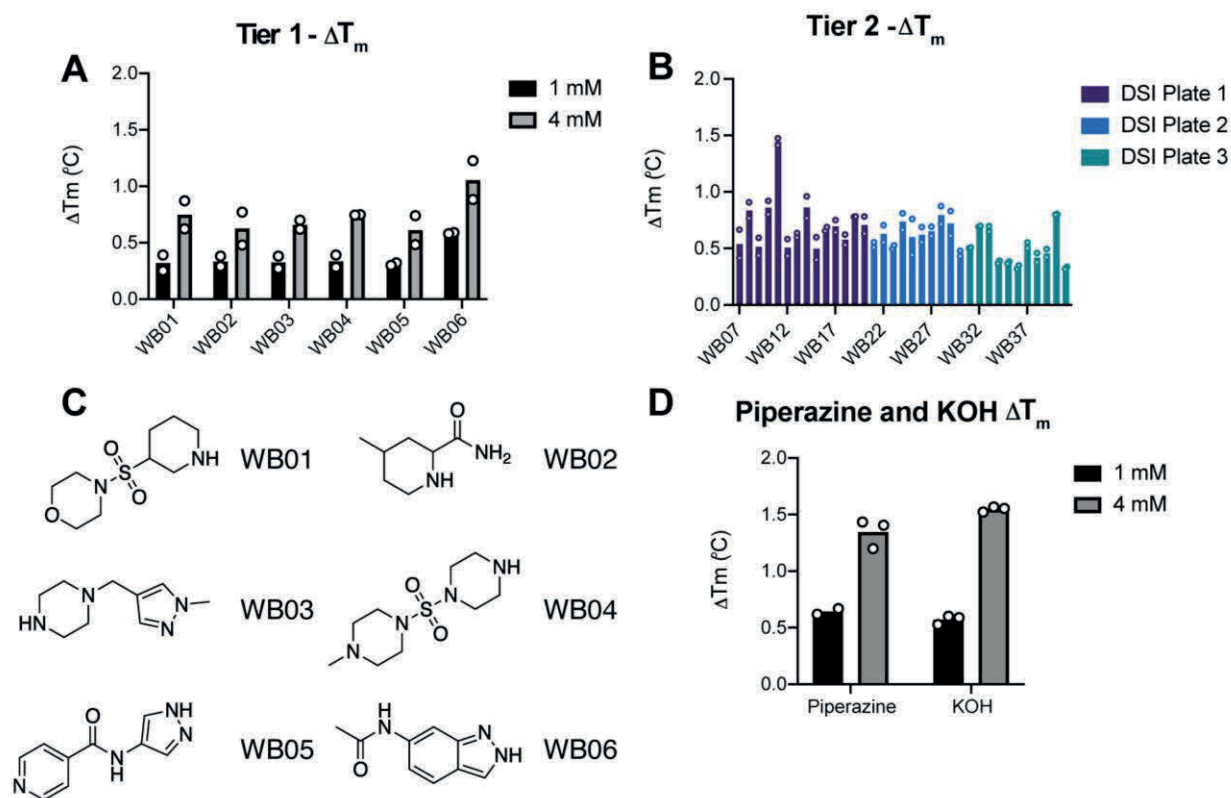


Figure 26: Fragment screening hit results, ranking, and structures.

A) Tier 1 thermal shift screening hit ΔT_m values at both 1- and 4-mM fragment. Result show technical duplicates. B) Tier 2 thermal shift screening hit ΔT_m values at 4 mM fragment. Hits are separated by the 384-well plate in which they were screened. Result show technical duplicates. C) Structures of the tier 1 hits. D) Thermal shift assay ΔT_m results for piperazine and KOH at 1 and 4 mM. Results display technical triplicates.

In these structures, the common denominator is the presence of basic, cyclic functional groups such as piperazines. Common functional groups appearing in hits are expected because structurally related groups should bind to the same regions of a protein if specific interactions are being made. However, to rule out a false-positive result, it was also investigated if piperazine alone was enough to induce stabilisation of GALE (**Figure 26D**). Piperazine induces a ΔT_m at both 1 mM and 4 mM concentrations similar to those observed for many of the hits **WB01-WB-06**. While it was possible that piperazine was binding to GALE and stabilising the protein, the next aim was to establish if small changes in pH could be affecting the results (at 25 mM HEPES the pH shifts from 7.50 to 7.72 with 4 mM piperazine). To assess this, GALE in the presence of 1 mM and 4 mM KOH was assayed by thermal shift. Similar thermal stabilisation of the protein was observed for KOH as with piperazine. Thus, it appears basic conditions (as found in SP³ hybridised, nitrogen-containing functional groups)

are sufficient to induce stabilisation of GALE. Unfortunately, this created significant uncertainty around compounds containing these groups since a significant proportion of the compounds in the library contained basic nitrogen. However, several hits with less basic groups that produced positive ΔT_m values were still identified, such as **WB05** and **WB06** (see below).

Despite these promising results, the structure-activity relationship of screened fragments was still uncertain. It was thus important to perform an additional, orthogonal screen with the same DSI-poised library to rule out false-positive hits.

3.2.2 Crystallographic Screening and Hit Validation

3.2.2.1 XChem Screening

As discussed above, crystallographic screening has emerged as a powerful method for finding fragment hits. Full fragment libraries are screened by soaking each fragment into pre-grown protein crystals. Diffraction data is then collected for each crystal to identify additional electron density from fragment binding. Screening this way enables the identification of weak binders, because screening can be carried out at >10 mM. Identified hits also inherently provide structural information, key to fragment elaboration.

An optimised reproducible and consistent crystallography system for producing apo GALE structures has been described in Chapter 2. A crystallographic screen was therefore feasible, and several crystallography trays were set up to produce the approximately 1100 wells containing 200 nL drops with GALE crystals required.

Screening of the DSI-Poised library was performed at the XChem facility. A simplified workflow is shown in **Figure 27**. 250 fragments were screened at 20 mM (20 % ethylene glycol) soaking concentration for three hours, to assess how well the crystals tolerated the fragment and ethylene glycol. It appeared that after three hours the crystals were beginning to crack and

dissolve. Therefore, the remaining 607 compounds were screened at 10 mM (10 % ethylene glycol) concentration to prevent damage to the crystals.

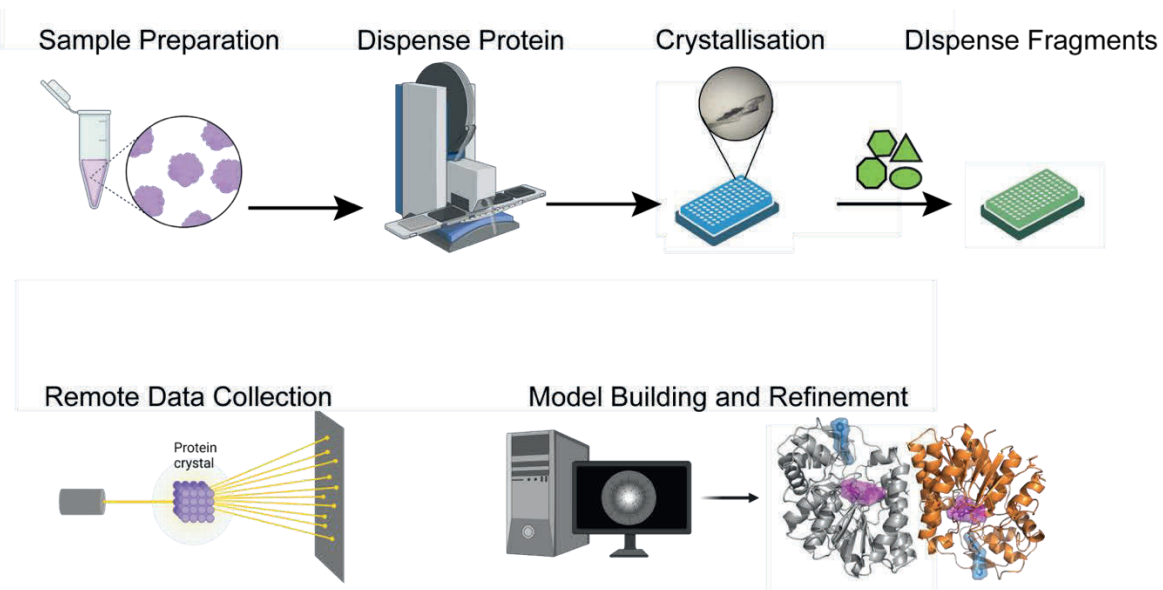


Figure 27: Illustration of the XChem screening process.

Including protein sample preparation, protein dispensing, crystallisation, fragment addition, data collection, followed by structure determination to identify bound fragments.

From this screen, three fragment hits were identified (**Figure 28A**) from crystals diffracting to between 1.7 and 1.9 Å resolution (full statistical validation for XChem hits can be found in section 3.5). Two fragments (**X0121** and **X0213**) were identified in the same binding pocket (**Figure 28B**), the UDP binding part of the protein as illustrated for **X0213** in **Figure 28C**.

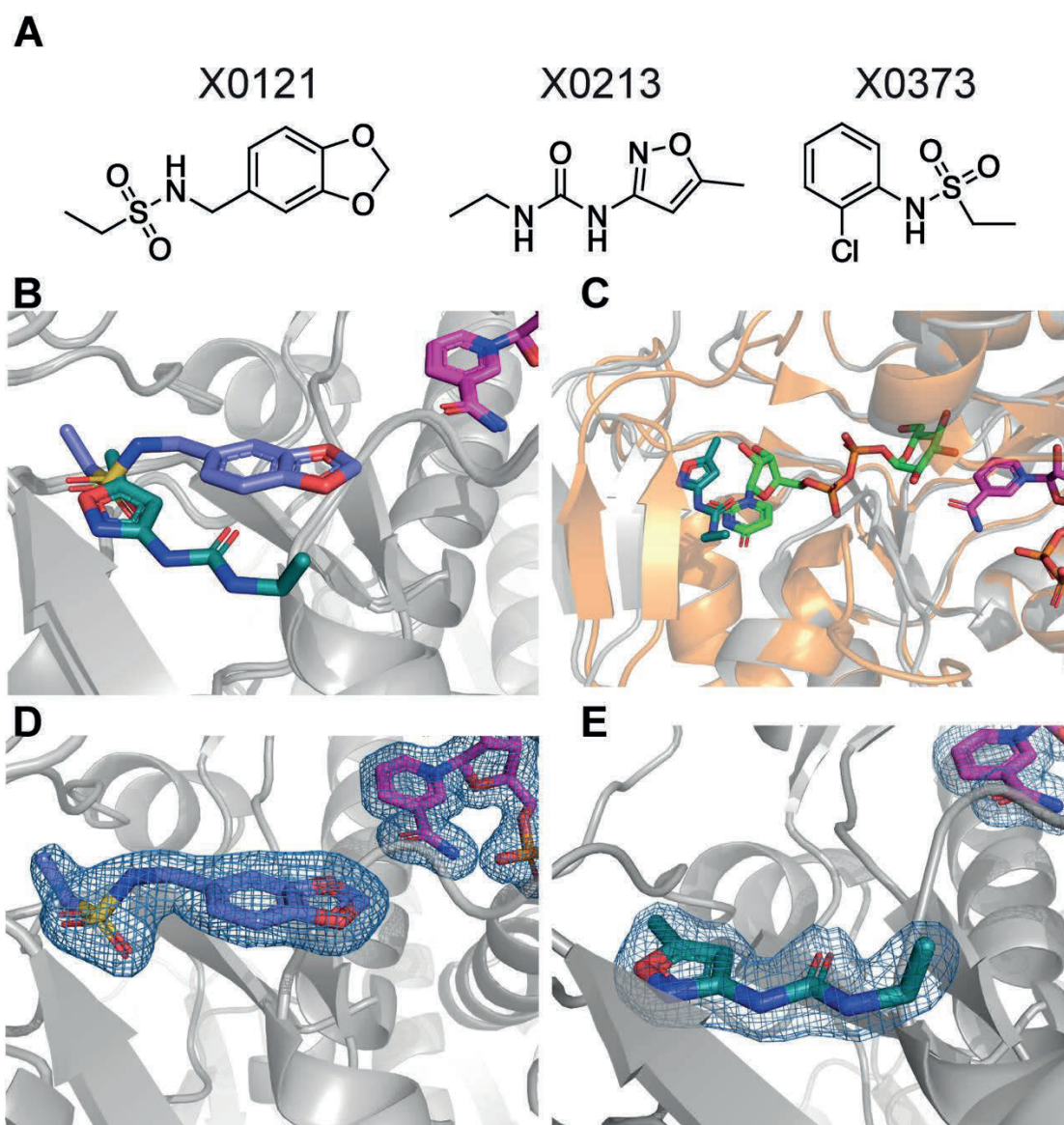


Figure 28: XChem screening fragment structures and **X0121** and **X0213** binding sites.

A) Structures of the three XChem hits. B) Alignment of crystal structures solved for **X0121** and **X0213**. C) Alignment of **X0213** structure (grey) with PDB: 1EK6 structure (orange). D) **X0121** and NAD with 2Fo-Fc density maps. E) **X0213** and NAD with 2Fo-Fc density maps. All maps are at σ -level 1. (NAD: pink, UDP-Glc: green)

Whilst both substrates occupy the same pocket, the network of interactions are different, as expected given the significant differences in structures of the two fragments. Fragment **X0213** has an occupancy of 1, meaning 100 % of the protein molecules in the crystal contain the ligand. **X0213** makes two key hydrogen bonds through the isoxazole nitrogen and the urea N-H moiety with residues Phe226 and Asn224, respectively, in the protein backbone (**Figure 29A and C**). A surface model of the protein displays a cleft/fold around fragment **X0213** made by the neighbouring residues (**Figure 29B**). The bonding pattern to Phe226 and Asn224 is of note

because these are the same two residues that are predicted to be hydrogen bonding the uracil of UDP (**Figure 29D**). The urea and nitrogen of the isoxazole are acting as isosteres for the hydrogen bonding regions of the uracil ring (**Figure 29E**). Such clear similarities between the binding interactions of ligand and substrate are evidence that this ligand is likely to be a true hit. Additionally, several hydrogen bonds are made to adjacent water molecules via the urea carbonyl and isoxazole oxygen lone pairs.

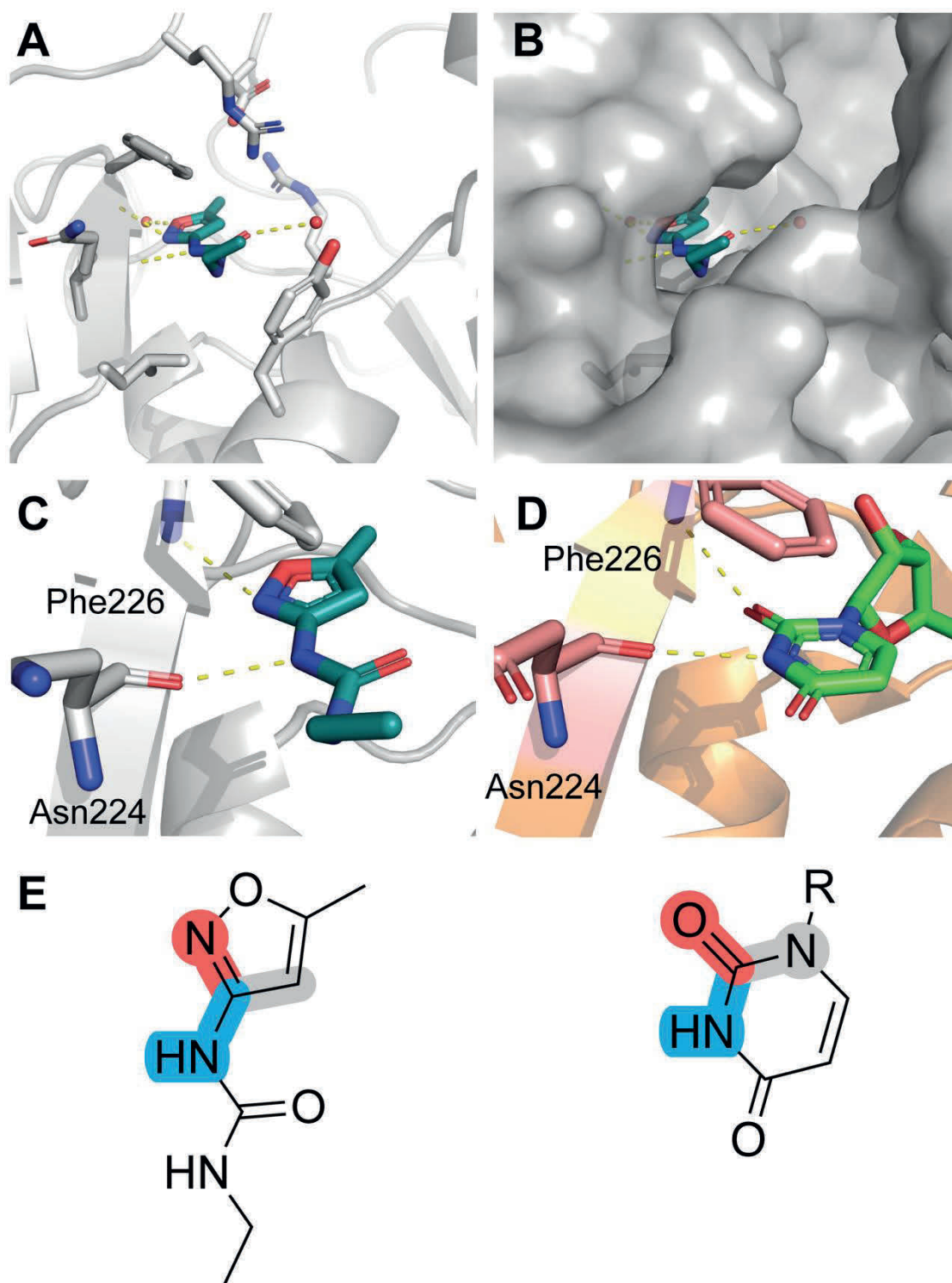


Figure 29: XChem hit **X0213**-protein interaction analysis.

A) **X0213** structure showing residues that form the binding pocket as sticks. Dashed yellow lines show H-bond. Interacting water molecules are shown as red spheres. B) **X0213** binding pocket surface calculated in Pymol. C) Image of the key H-bonds formed between Asn224 (bottom) and Phe226 (top) to **X0213**. D) Image showing key H-bonds made to the uracil of UDP-Glc by Asn224 and Phe226. E) Diagram illustrating the region of **X0213** acting as an isostere for uracil. (Red: hydrogen bond acceptor, blue: hydrogen bond donor).

Fragment **X0121** (Occupancy 1) has an extensive system of hydrogen bonding to water molecules within the pocket, centred mainly around the sulfonamide, whilst a π -cation interaction is formed through the benzodioxol ring and the Arg300 side chain (**Figure 30A-B**). A hydrogen bond is also present through N-H of the sulfonamide with Asp303. In general, there are fewer binding similarities between **X0121** and UDP than for **X0213**, which mimics the natural substrate. It should be noted that for both ligands, the strength of the ligand-water hydrogen bonds is unknown and may only be relevant in crystallisation. However, they do provide information about the putative binding pocket that will be valuable for ligand elaboration.

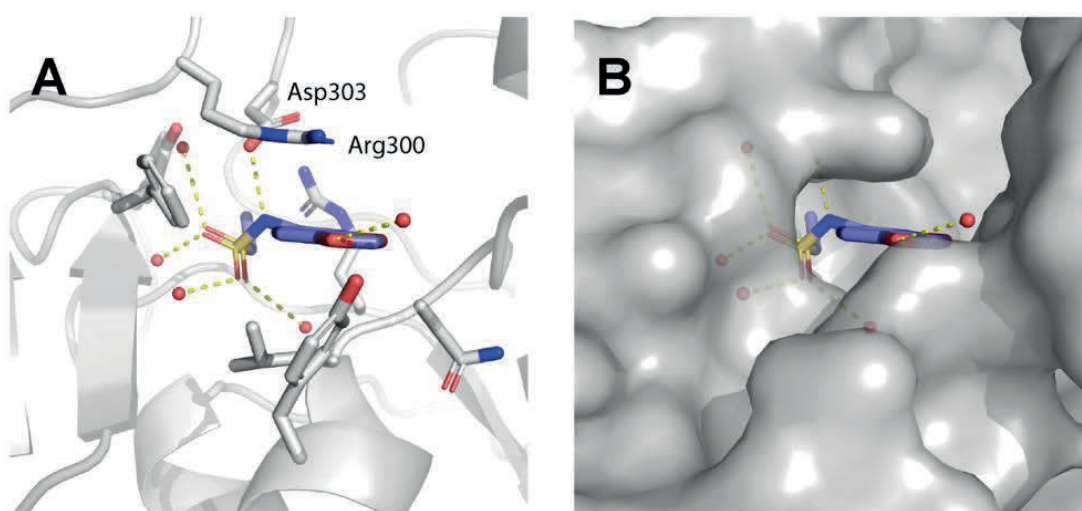


Figure 30: XChem hit **X0121**-protein interaction analysis.

A) **X0121** structure showing residues that form the binding pocket as sticks. Dashed yellow lines show H-bond. Interacting water molecules are shown as red spheres. B) **X0121** binding pocket surface calculated in Pymol.

The final compound, **X0373**, occupies a completely different area of the structure (**Figure 31A-B**). **X0373** is found at an interface of the two GALE monomers and has the lowest occupancy of all the hits (0.60). Additionally, this compound occupies an interface that is unlikely to exist in solution and is likely only present in crystallographic packing, making it much less useful as a starting point for elaboration.

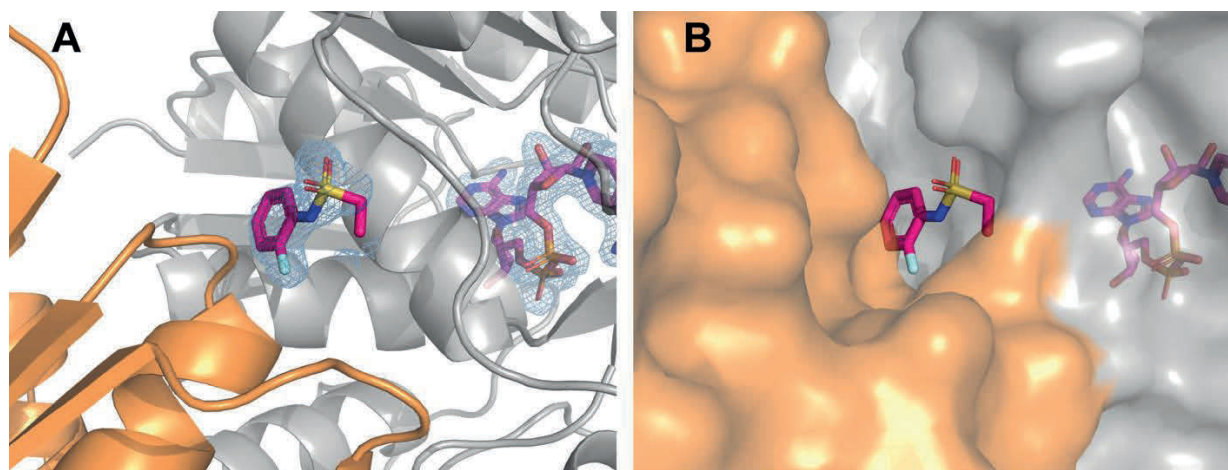


Figure 31: XChem hit **X0373**-GALE crystal structures.

*A) Structure a 2Fo-Fc map of **X0373** (red). B) Pymol calculated protein surface showing the pocket occupied by **X0373**. Pink: NAD. All maps are at σ -level 1.*

The hit rate for the XChem fragment screen was low (0.35 %). All three identified hits were found from the two hundred and fifty fragments that were screened at 20 mM, whilst none were identified from the crystals soaked at 10 mM fragment concentration. The results indicate that, although optically it appeared as though the crystals were unstable at 20 mM, they were still diffracting to a reasonable resolution of $< 2.5 \text{ \AA}$. Additionally, even 20 mM is a low fragment concentration compared with other crystallographic screens, which are often conducted up to 50 mM concentration.¹⁴⁸ Further optimisation around the crystal tolerance to higher fragment and ethylene glycol concentrations may have produced a better hit rate. Regardless, two of the identified hits are in the ideal location for developing GALE inhibitors, since they occupy the UDP binding moiety of the protein.

3.2.2 XChem Hit Validation

Although the hits **X0121**, **X0213** and **X0373** were observed in the crystal structures, validation that these are “true positives” and interact with the protein in solution requires orthogonal assays. The thermal shift data (Section 3.2.1) were considered, shown in **Figure 32A**. **X0121** has minimal effect on protein stability, whilst **X0373** showed significant destabilisation. **X0213** produced a ΔT_m of $\sim 0.5 \text{ }^\circ\text{C}$ at both 1 and 4 mM, therefore it would have been considered a tier 2 hit. The differences in stabilisation of **X0121** and **X0213** is unexpected, given the lower

occupancy of **X0213** in the crystallographic screen. However, occupancy in crystallography is not a quantitative measure of binding in solution and may explain the observed differences.

To confirm and quantify that **X0213** binds not only when soaked into crystals but also in solution, we purchased fragments **X0121** and **X0213** from a commercial vendor (Enamine) and determined the K_D for **X0213** by isothermal titration calorimetry (ITC) (**Table 11, Figure 32B**). In brief, the ITC experiment involved titrating from a 4 mM stock of fragment into 25 μ M GALE solution until saturation was achieved. The K_D was significantly weaker than the natural substrate (>30-fold weaker). However, the occupancy (N value) was 0.16, a value comparable to that of the natural substrate and commensurate with the existence of an abortive complex (see chapter 2). Given that both UDP-Glc and **X0213** bind to the same pocket, this is not an unsurprising result. This data also enables calculation of the LE using the K_D as 0.39. The ITC results clearly confirm both the binding of **X0213** in solution and binding to the substrate binding pocket of GALE. For Fragment **X0121**, suitable data could not be obtained to determine the K_D at the fragment concentrations used (4 mM in syringe) indicating **X0121** as a weaker binding ligand.

Ligand	K_D (μ M)	N Value
X0213	370.0 ± 44.6	0.16 ± 0.05
UDP-Glc	9.0 ± 4.1	0.17 ± 0.04

Table 11: ITC assay results for **X0213** binding to recombinant GALE.

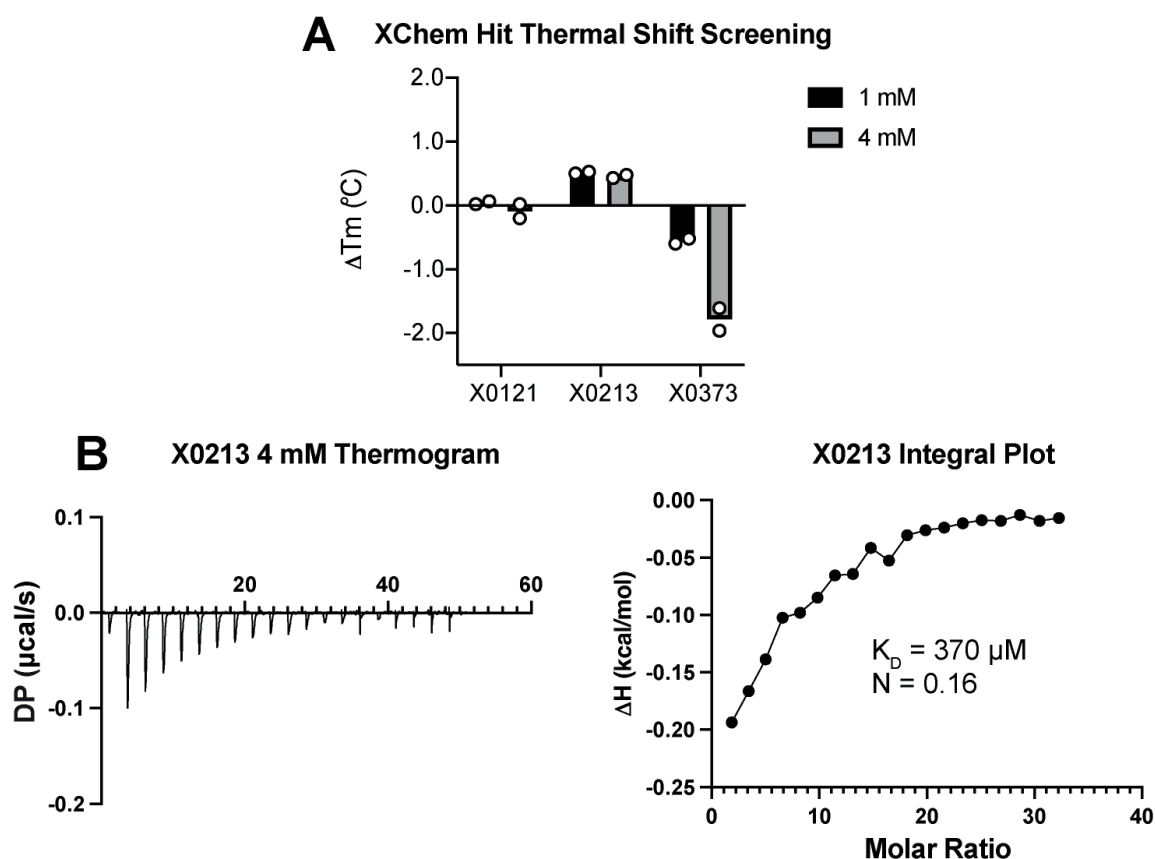


Figure 32: X0213 biophysical characterisation

A) Thermal shift screening ΔT_m values (relative to DMSO controls) for the XChem hit fragments at both 1 and 4 mM. B) X0213 ITC assay thermogram and integral plots.

Despite the weak binding of the XChem hits **X0213** and **X0121** to GALE, the fragments were next assessed for their effect on enzyme activity. In brief, inhibition assays were performed at 5 nM GALE, 100 μ M UDP-Gal for 15 minutes at 37 °C (initial velocity conditions) in the presence of varying concentrations of fragment (1.6 mM – 12.5 μ M) in technical duplicate. The relative conversion of UDP-Gal to UDP-Glc was then determined by HPLC as described for the epimerisation assays described in Chapter 2. The results were normalised relative to DMSO controls. Both fragments showed minimal effects on enzyme activity, even at 1.6 mM Fragment concentrations (**Figure 33**). Surprisingly, both produced almost identical curves, suggesting that either the fragments have similar affinities, despite previous results, or that the inhibition at high concentrations is the result of aggregation or other non-affinity related effects. Regardless, these results demonstrate the need for fragment elaboration to develop a useful inhibitor. Based on results of the thermal shift screen, ITC experiments, activity assays,

and the isosteric binding, compound **X0213** was chosen as the preferred candidate for further elaboration and characterisation.

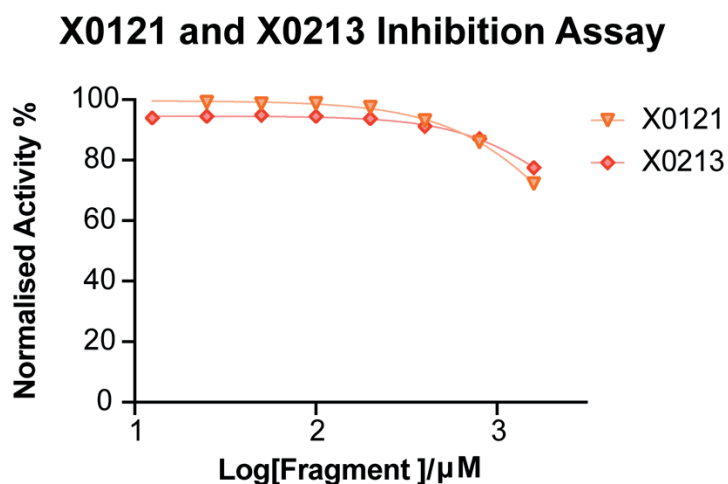


Figure 33: X0121 and X0213 inhibition assay results.

Activity is given as a % of DMSO controls. Results show mean and range from two technical replicates.

3.2.3 Screening Summary

To summarise, non-covalent fragment screening of the Enamine DSI-Poised library was performed using both thermal shift assays and X-ray crystallography. Thermal shift identified several hits that displayed concentration-dependent protein stabilisation. However, the results are complicated by protein stabilisation caused by basic functional groups such as piperazine. Fragments that possess less basic groups (e.g. pyridine) may still prove to be useful hits. Further validation of these compounds (e.g. **WB05** and **WB06**) would require secondary thermal shift assays with higher buffer concentrations and evaluation by ITC.

Crystallographic screening identified three hits: **X0121**, **X0213**, and **X0373**. **X0373** was found at a crystallographic interface between the GALE monomers. This interface is unlikely to occur in solution and therefore, **X0373** has limited potential for follow-up. **X0121** and **X0213** both occupy the uracil binding region of the protein substrate binding site. **X0121** does not appear to mimic the natural substrate, whilst in **X0213** the urea and isoxazole are acting as clear isosteres to the uracil. ITC confirmed **X0213** to binding with a K_D of $370.0 \pm 44.6 \mu\text{M}$. No

binding could be determined for **X0121** using ITC. Additionally, **X0213** was the only crystallographic hit also identified as a hit in the thermal shift screen. Neither fragment showed any significant enzyme inhibition in the presence of 100 μM substrate. Given **X0213** has been validated by thermal shift, ITC, crystallography, and its apparent isosteric binding, the subsequent sections focus on **X0213** for elaboration.

3.2.3 Elaboration of the fragment **X0213** toward a GALE lead compound

3.2.3.1 Analogue Series 1 Virtual Screening

Fragment **X0213** displays only weak binding to GALE ($K_D = 370.0 \pm 44.6 \mu\text{M}$). The natural substrate UDP-Glc has a much stronger K_D of $9.0 \pm 4.1 \mu\text{M}$ and cellular substrate concentrations are likely in the mid-high μM range in cells.^{149,150} To effectively inhibit GALE, the affinity of **X0213** for GALE would need to be improved by several orders of magnitude. Fragment “growing” is therefore essential to expand the low molecular weight hit into larger compounds that make additional interactions to the protein. This is typically done by screening as many analogues of the initial hit as possible. Designing, making, and testing tens or hundreds of analogues based solely on chemical intuition from crystal structures requires considerable time and often, luck. To aid in the design process, virtual docking software has emerged as a routine tool. Computational screening enables vast libraries of analogues to be ranked by predicted binding, the results can then be evaluated, and the most promising analogues shortlisted for synthesis and testing. The process can be further optimised by initially screening libraries of commercially available analogues virtually, the best predicted binders can then be ordered from vendors and tested experimentally.

A "substructure" search of a chemical library looking for compounds that are structurally related to an input of the Enamine chemical stock using **X0213** as the input identified 1404 available analogues. These compounds were then docked into a protein structure derived from the GALE-**X0213** complex in which the **X0213** ligand was removed but the protein conformation preserved. The ICM software was used as described above. At this stage, little was known about the environment around **X0213**, aside from what was obvious from the

crystal structure. The compounds selected for ordering and testing were therefore selected based on several criteria: 1) Fragments should contain only a single modification of the fragment. i.e. if the isoxazole or its substituents are altered, the rest of the fragment should remain the same. This should enable a clearer understanding of the SAR if a particular modification improves binding. 2) The selected fragments should be structurally distinct enough to probe a range of regions within the pocket. 3) The fragments should have docking scores high enough to place them in the top 10% of predicted binders.

From the docking results, eight analogues of **X0213** were selected and ordered (**Figure 34A**). It should be noted that in ICM's scoring function, a more negative value indicates better predicted binding. From the docking results, compounds in which the urea is formed from a benzylamine or a substituted analogue thereof were clearly favoured. This is due to predicted π -stacking interactions between the phenyl ring of the fragment and phenol ring of Tyr211 (**Figure 34B**). All eight compounds contained modification around the ethylamine of the parent fragment **X0213**. This is because compounds with modifications at this position occurred much more frequently in the top 100 docked compounds. Due to the clear preference of the docking software for benzyl substitution on the ethylamine side of the urea, six of the compounds were substituted benzyl analogues. The remaining two compounds still contained aromatic groups but were slightly more structurally diverse.

3.2.3.2 Analogue Series 1 Characterisation

It is important to note that from this section of the chapter onwards, the majority of thermal shift and activity assays were performed by an MRes student, Sing Hei Lok, with supervision from myself (the thesis author, William Browne). Additionally, some of the results are included in the student's MRes report (S. H. Lok, MRes Thesis, Imperial College London, 2023.)

Assaying all eight compounds (plus future analogues) by ITC was considered impractical due to the low throughput of the assay and the quantity of enzyme required per assay (~0.3 mg per assay). Therefore, an assay pipeline was developed (**Figure 34C**) consisting of several assays in sequence; 1) thermal shift. 2) enzyme inhibition assays. 3) ITC analysis of the best

mild improvements in thermal stabilisation (0.4, 0.9, and 0.5 °C respectively) compared to the parent compound **X0213** (0.3 °C). **WBX4** was a clear standout with a ΔT_m of 3.1 °C. Docking of this compound predicts that it forms an additional H-bond to NH in the backbone of Asn224 (**Figure 35B**). Comparison of the docking score vs ΔT_m shows no correlation (**Figure 35C**). However, **WBX4** was predicted to have the highest affinity, and this was confirmed in the thermal shift assay.

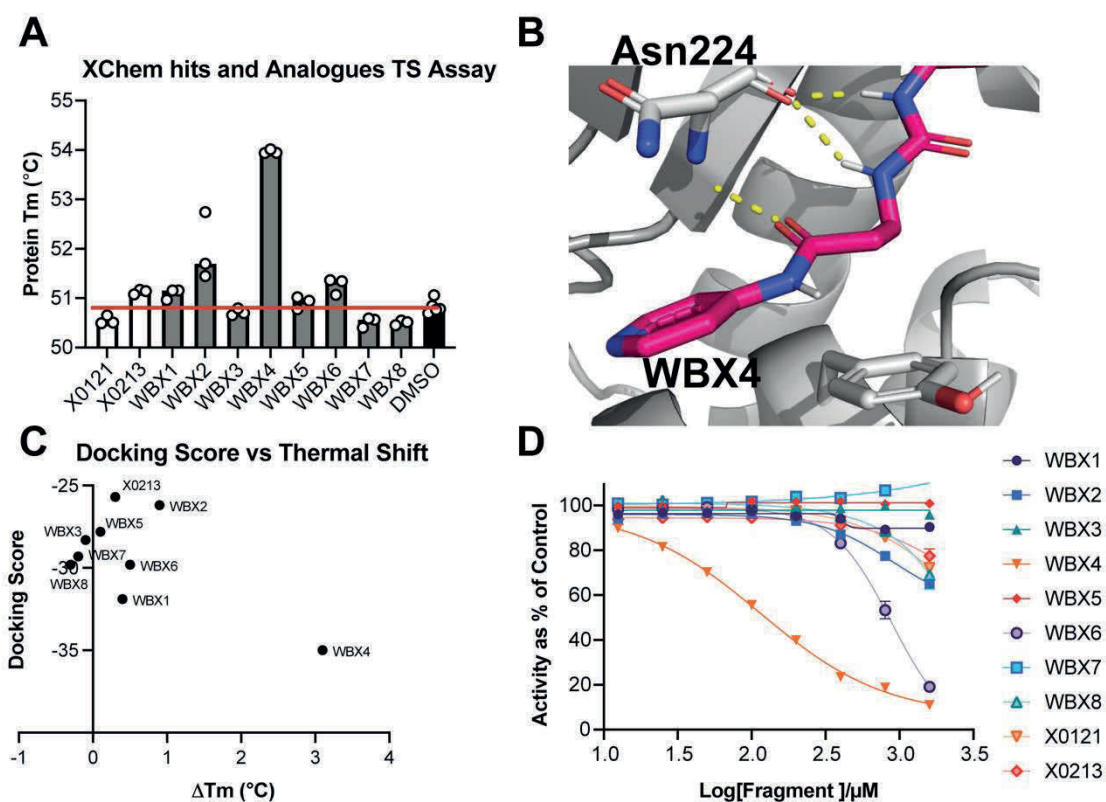


Figure 35: Analogue series one docking, thermal shift, and inhibition assays results.

A) Thermal shift screening results of the **X0213** analogues (including **X0213** and **X0121**). The red line indicates the mean value of the six DMSO control samples. All other samples were run in technical triplicate. B) Docking structure predicted by ICM showing the additional H-bond made by **WBX4** to Asn224. C) Plot showing docking score predicted by ICM vs ΔT_m . D) Inhibition assay results for **X0121**, **X0213** and analogues. Results are from technical duplicates with mean and range displayed

Fragment	ΔT_m °C
X0213	0.3
WBX1	0.4
WBX2	0.9
WBX3	-0.1
WBX4	3.1
WBX5	0.1
WBX6	0.5
WBX7	-0.2
WBX8	-0.3

Table 12: X0213 and analogues ΔT_m values calculated from DMSO controls.

All eight compounds were then assayed in inhibition assays against GALE using the same method described above for fragments **X0121** and **X0213**. **WBX4** was the only fragment that displayed a measurable IC_{50} of 115 μM (**Figure 35D**). **WBX6** also displayed inhibition with an IC_{50} of 882 μM . The steep dose response-curve of **WBX6** and the high fragment concentration at which it occurs may indicate that protein aggregation or changes in protein solubility are responsible for the drop in activity. Regardless, the thermal shift data and inhibition assays clearly indicate **WBX4** as the fragment with a significant improvement in protein binding.

WBX4 was further validated by ITC to determine K_D (**Figure 36A**). Due to the expected tighter binding, the concentration of **WBX4** stock solution used for the titrations was decreased to 250 μM , aside from this, the experiment was identical to that performed for **X0213**. The K_D was measured to $8.87 \pm 5.85 \mu M$ (~100-fold improvement over parent fragment), which gives a LE of approximately 0.33, only a slight decrease from the initial hit **X0213**. This K_D is approximately the same binding strength as the natural substrate UDP-Glc (also ~9 μM). This result is supported by the inhibition assay where the IC_{50} in the presence of 100 μM UDP-Glc was ~100 μM , which would be expected for ligands with equivalent binding strength. Thus far, NAD has not been included in the ITC assays. To confirm that the presence of NAD does not interfere with ligand binding, the ITC assay was repeated with **WBX4** in the presence of 200 μM NAD. The K_D measured for the fragment in this experiment was $3.79 \pm 0.75 \mu M$ (**Figure 36B**). Whilst this result did suggest tighter binding, the data was also of better quality

compared to the ITC experiment where NAD was not included. Therefore, NAD likely does not improve the binding but rather stabilises the protein so that more consistent data is obtained producing smaller error. In any case, The K_D of **WBX4** lies between approximately 3-9 μM , close to a 100-fold improvement over the $\sim 390 \mu\text{M}$ determined for the parent fragment **X0213**.

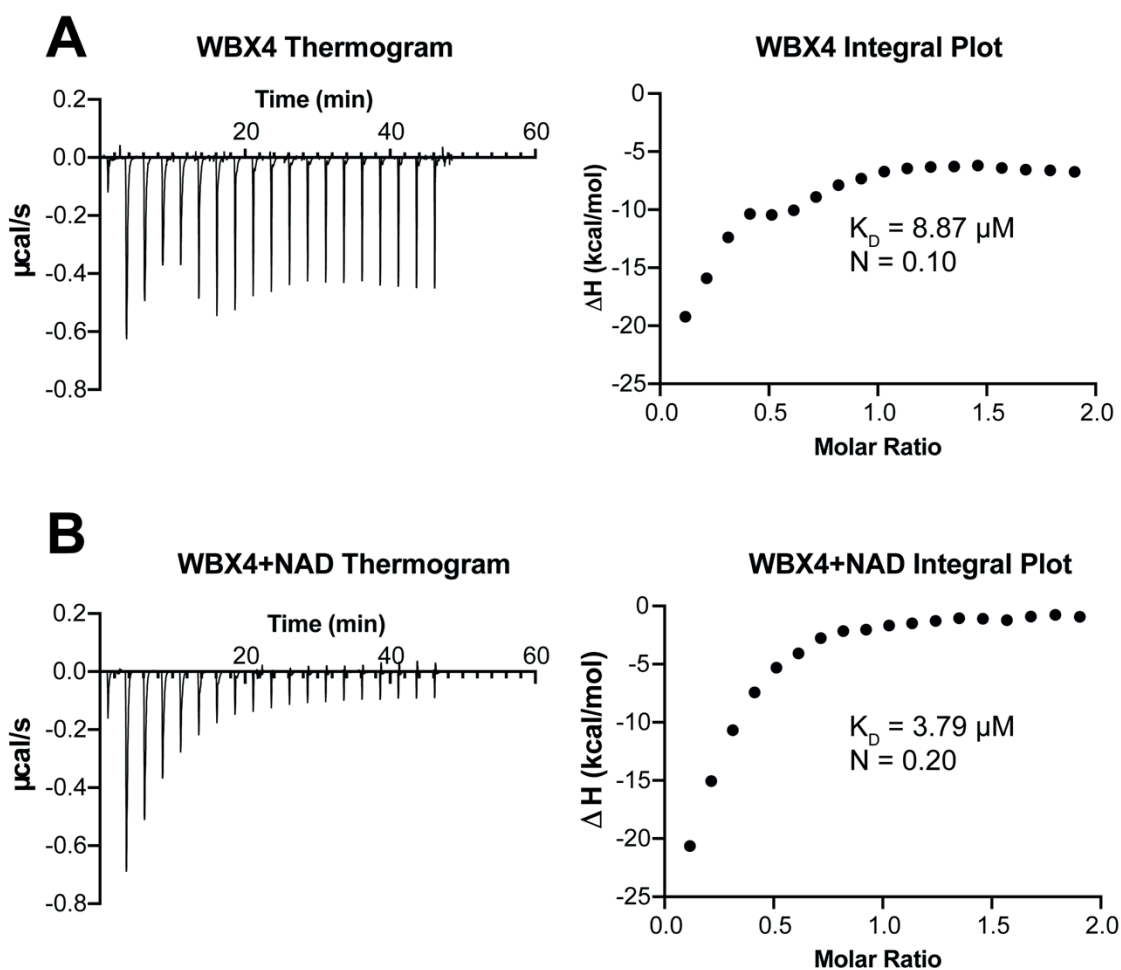


Figure 36: WBX4 ITC assay characterisation.

A) **WBX4** ITC assay results with both thermogram and integral plots. B) As in panel A but in the presence of NAD.

3.2.3.3 Analogue Series 1 Crystallography

The final step in the hit assay workflow (**Figure 34C**) requires obtaining crystal structures for selected analogues. Whilst the crystal form and conditions used for screening at XChem produced reasonable structures and consistent crystal growth with seeding, the XChem

screen revealed the susceptibility of the crystals to DMSO. A collaborator, Dr. Stephane Moulleron at the Francis Crick Institute re-expressed GALE from *E. coli*, this is described in full in the methods section of this chapter. Fortunately, the pTriEx-6 vector discussed in Chapter 2 is also suitable for expression of the same GST-3C-GALE fusion protein. The aim was to determine if this affected the amount of UDP-Glc bound to GALE or if the crystal structure obtained would now be free of UDP-Glc and contain only NAD. To assess this, GALE was expressed and purified from *E. coli* and assayed by ITC, titrating in UDP-Glc to determine the active site occupancy (N) (**Figure 37**). The ITC assay with UDP-Glc produced approximately the same K_D as was observed for the insect cell-expressed GALE ($\sim 9 \mu\text{M}$ in both cases). However, there was an increase in the N value from 0.16 to 0.24, an 8 % increase in the available substrate binding site. However, the effect was still modest and could have been the result of batch variability rather than expression system. To fully compare this, the several replicates would be required with protein expressed from multiple batches. In any case, since the expression of GALE in *E. coli* was easier to scale up, the remaining crystallography in this chapter was performed by Dr. Moulleron using *E. coli*-expressed GALE.

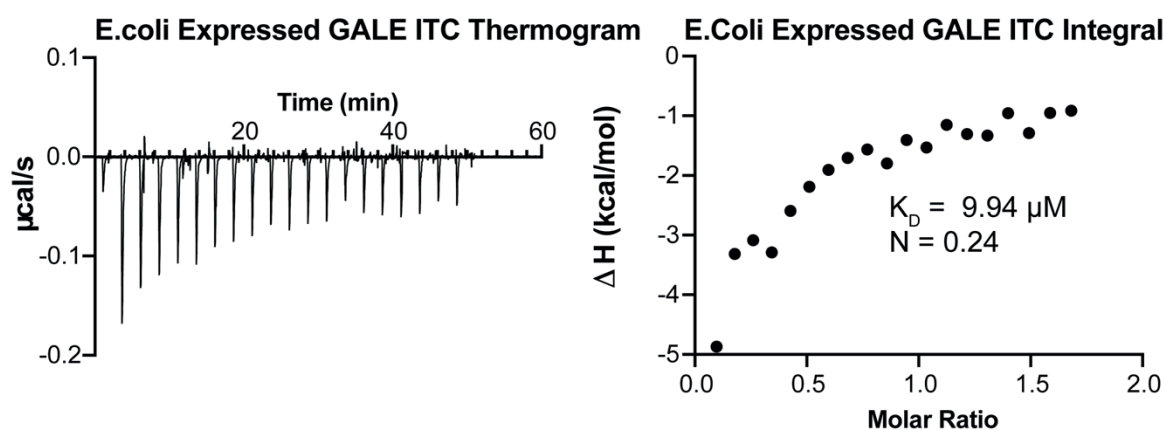


Figure 37: ITC assay results for *E. coli* expressed GALE with UDP-Glc showing both thermogram and integral plots.

Fragment **WBX4** was soaked at 1 mM into crystals of GALE (expressed from *E. coli*) optimised by Dr. Stephane Moulleron. Data were collected at the Diamond Light Source (DLS) synchrotron. Diffracting crystals containing **WBX4** were obtained at 1.2 Å resolution (validation in section 3.5). The structure consisted of two GALE monomers as has been observed previously (**Figure 38A**). Both monomers contain one NAD molecule (**Figure 38A**). In contrast to previously obtained protein/ligand structures, both monomers also contain one ligand (**WBX4**) with very well-defined density (**Figure 38B**). No density suggesting bound UDP-

Sugar was observed. ITC suggested 8 % more free binding site in E. coli-expressed GALE compared to insect cell-expressed GALE. This slight difference is unlikely to be the reason for an absence of UDP-Sugar in the protein structure. Rather, the most likely explanation is that **WBX4** has a much higher affinity for GALE compared to the XChem hits (~100-fold tighter). Thus, **WBX4** can displace the bound UDP-sugar from the substrate binding pocket of GALE during soaking.

The binding mode of **WBX4** is identical to the parent compound **X0213** where the two fragments are structurally equivalent (isoxazole and urea) with hydrogen bonds to both Asn224 and Phe226 via the urea and isoxazole, respectively (**Figure 38C-D**). The additional arylamide in **WBX4** makes two key interactions absent in the parent fragment **X0213**. 1) The amide carbonyl makes an additional hydrogen bond to the N-terminus of Asn224, accounting for three hydrogen bonds from fragment to Asn224 alone (**Figure 38D**). 2) The Tyr211 side chain has shifted, potentially forming a π -stacking interaction with the conjugated system of the arylamide (**Figure 38D**). However, this was not detected in the interaction prediction using Pymol. Comparison of the crystal structure binding pose of **WBX4** to that predicted by ICM reveals identical ligand conformations (**Figure 38E**). However, the position of Tyr211 is notably different, occupying a position parallel to the fragment. This highlights the limitations of ICM where the receptor (protein) is not flexible and thus cannot consider these changes.

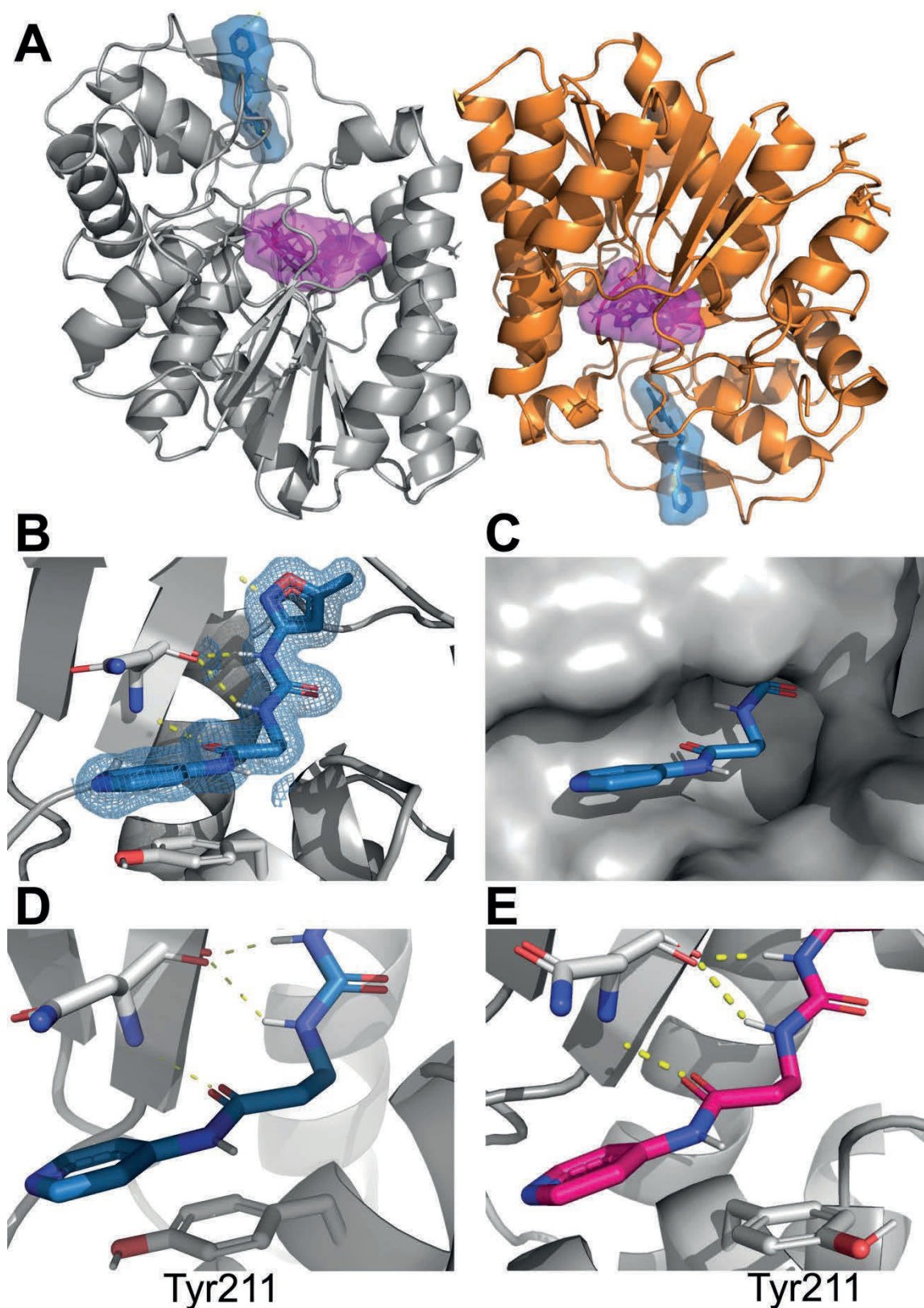


Figure 38: Crystal and docking structures for fragment **WBX4**.

A) Crystal structure GALE dimer in complex with **WBX4** (blue) and NAD (pink) with calculated ligand surfaces. B) **WBX4** with 2Fo-Fc map and Tyr211 (bottom) and Asn224 (top) residues. C) **WBX4** displayed with calculated protein surface. D) Enhanced view of panel B showing Tyr211 position. E) Docking structure of **WBX4** showing the same ligand position (as in crystal panel D). Tyr211 is notably in a different position. Dashed yellow lines indicate H-bonds.

From this first series of analogues, the structure for **WBX2** in complex with GALE was also successfully solved (**Figure 39**) to 1.2 Å (validation in section 3.5). Since this fragment produced the second highest ΔT_m in the series (0.9 °C), it is unsurprising that a structure was determined for this fragment, given that tighter binding fragments are more likely to bind to the crystallised protein. Consistent with the lower ΔT_m of **WBX2** compared to **WBX4** (0.9 vs 3.1 °C respectively), the solved structure contained two different conformations for **WBX2**. In one, the phenol ring of the fragment is parallel to the phenol ring of Tyr211. In the other, the fragment phenol ring is perpendicular to Tyr211. This data would suggest that the π -stacking interaction between fragment and Tyr211 is not particularly strong, otherwise only one conformation would be observed. It may be that for fragment **WBX4** this interaction is equally weak and that most of the improvement in binding comes from the additional H-bond.

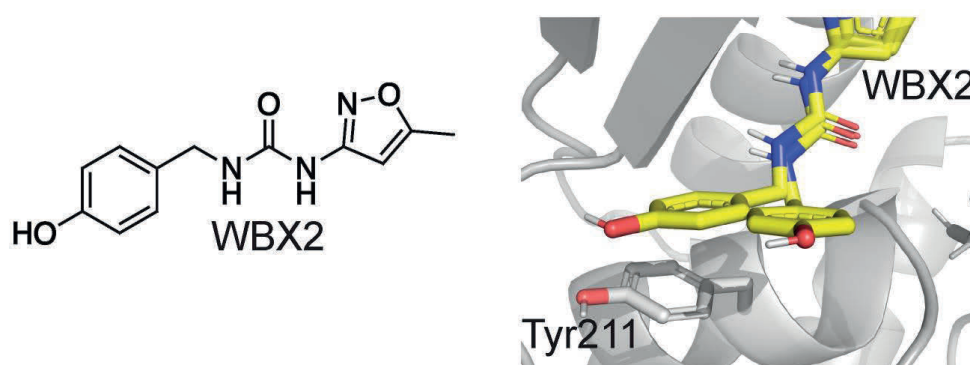


Figure 39: WBX2 chemical and crystal structure

3.2.4 WBX4 Elaboration

3.2.4.1 Analogue Series 2 Virtual Screening

WBX4 provided a clear improvement in binding (~100-fold) over the initial parent fragment and produced a measurable IC_{50} of 115 μM . Despite this success, an effective inhibitor in cells would still require a significantly lower K_D and IC_{50} against recombinant enzyme, due to factors such as metabolism, cell membrane penetration, and efflux that work against the inhibitor. Thus, a second round of analogues was designed using the existing hits and existing structural information. Contrary to the previous round, the initial docking process to identify potential

hits was performed from a library of 64 *in silico*-designed fragments, rather than from commercially available analogues.

The fragments were docked into the structure obtained for fragment **WBX4**. From the docking results, nine compounds were selected to be synthesised by the chemical supplier Enamine (**Figure 40**). Several of the chosen compounds contained changes to the methyl group of the isoxazole (**WBX9-12**) to probe the binding properties of the pocket filled by this group. **WBX13-15** contained alterations to the pyridine ring of the arylamide moiety. **WBX16** contained an oxazole to assess if the oxygen might mimic the natural substrate better than the nitrogen of the isoxazole. **WBX17** is a cyclised version of **WBX4**, potentially improving binding due to a reduced entropic penalty compared with **WBX4**. It should be noted that whilst there were other compounds that produced higher docking scores, many of these were too expensive to order as they were synthetically challenging. Future work will require in-house synthesis and testing of these compounds.

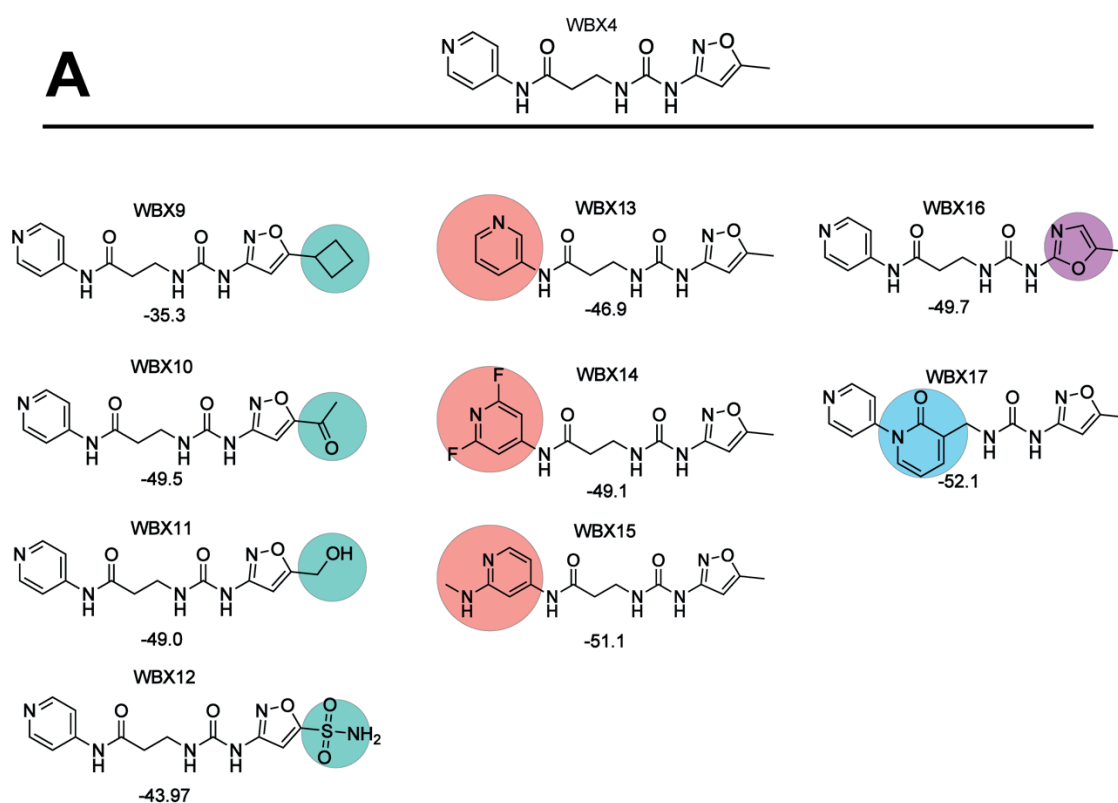


Figure 40: Structures of **WBX4** and its analogues separated by modification. Modifications are highlighted with circles. Docking scores are also shown.

3.2.4.2 Analogue Series 2 Characterisation

The **WBX4** analogues were characterised with the same assay workflow used to for the first series of analogues (**Figure 34C**). The thermal shift screening was carried out with the same conditions used for both the DSI library and analogue series one screening (25 mM HEPES buffer). The T_m of the protein in the presence of each fragment is shown in **Figure 41A**, whilst the ΔT_m relative to a DMSO control is shown in **Table 13**. From the isoxazole analogues (**WBX9-WBX12**), two compounds (**WBX9** and **WBX12**) produced a T_m greater than the parent compound (**Figure 41A**, **Table 13**). **WBX12** displayed only a marginal increase in protein stabilisation compared to the parent compound (0.2 °C increase). In contrast, **WBX9** produced a significant increase in thermal stability (2.2 °C) compared to **WBX4**. This was a surprising result given that of all the analogues tested from analogue series 2, **WBX9** had the worst docking score (-35.3). Of the isoxazole analogues (**WBX9-12**), fragment **WBX9** was the only one to contain a lipophilic modification, suggesting the presence of a lipophilic pocket. This is best illustrated in a 2D representation of the **WBX9** docking result (**Figure 41B**). The green areas of the 2D diagram are lipophilic regions, and the docking software clearly identifies that the cyclobutyl group of **WBX9** can fill a lipophilic region. However, the orientation of the isoxazole has also been flipped to accommodate this, removing the crucial hydrogen bond made between Phe226 and the isoxazole nitrogen, thus resulting in a poor score.

Of the pyridine analogues (**WBX13-15**), no significant increase in protein stabilisation was observed. The only other notable compound was **WBX17**, the cyclised version of **WBX4**, which produced a T_m 0.8 °C higher than the parent compound. This potentially demonstrates the benefits of reducing ligand flexibility.

As an additional validation to confirm that none of the compounds were inducing protein stabilisation due to fragment basicity (seen in the DSI library screening), the thermal shift assay was also run at 100 mM HEPES (rather than the 25 mM used previously) and with a piperazine control (**Figure 41C**). The same trend was observed in both assays (25 mM and 100 mM HEPES) validating that the results discussed above were not affected by pH shift induced by the fragments. Comparison of the docking score with thermal shift values was once again

poor (**Figure 41D**). Again, this highlights the limitations of a static receptor for docking into a pocket that has considerable flexibility in solution. Future work will likely involve utilising the multiple crystal structures of GALE obtained to perform more sophisticated docking that accounts for protein flexibility.

In continuation of the workflow, the **WBX4** analogues were then run in an enzyme inhibition assay as used for previous fragments (**Figure 41E**). It should be noted here that at the time of submission, biochemical data could not be obtained for **WBX15** and **WBX16** due to time limitations. From the other analogues, every compound displayed some level of enzyme inhibition over the fragment concentration used (12.5-1600 μ M). **WBX4** (the parent compound) and **WBX12** display almost identical inhibitory activity, consistent with their almost identical T_m values from the thermal shift assay. It was surprising that such a big modification (methyl to sulfonamide) has almost no effect on IC_{50} . Most notably, **WBX9** displayed the best predicted IC_{50} (9 μ M) of all the compounds tested, consistent with thermal shift results. Due to the better inhibitory profile of **WBX9**, a full dose-response curve was not observed across the concentration range used. To determine a more reliable IC_{50} and to confirm that shifts in pH from the fragments were not affecting activity, the assay was performed with **WBX9** across a concentration range of 1.6-1600 μ M fragment concentration with 100 mM HEPES buffer. The duplicates for this experiment displayed some significant variation (**Figure 41F**) and the curve fitted to the data had an R^2 of only 0.84. The IC_{50} was measured as 1.3 μ M. However, given the variation in the data, this result is likely quite different from the true value and a repeat would be necessary in future experiments. Regardless, extrapolation of the data in **Figure 41E** provides a reasonable estimate of **WBX9**'s IC_{50} at 9 μ M.

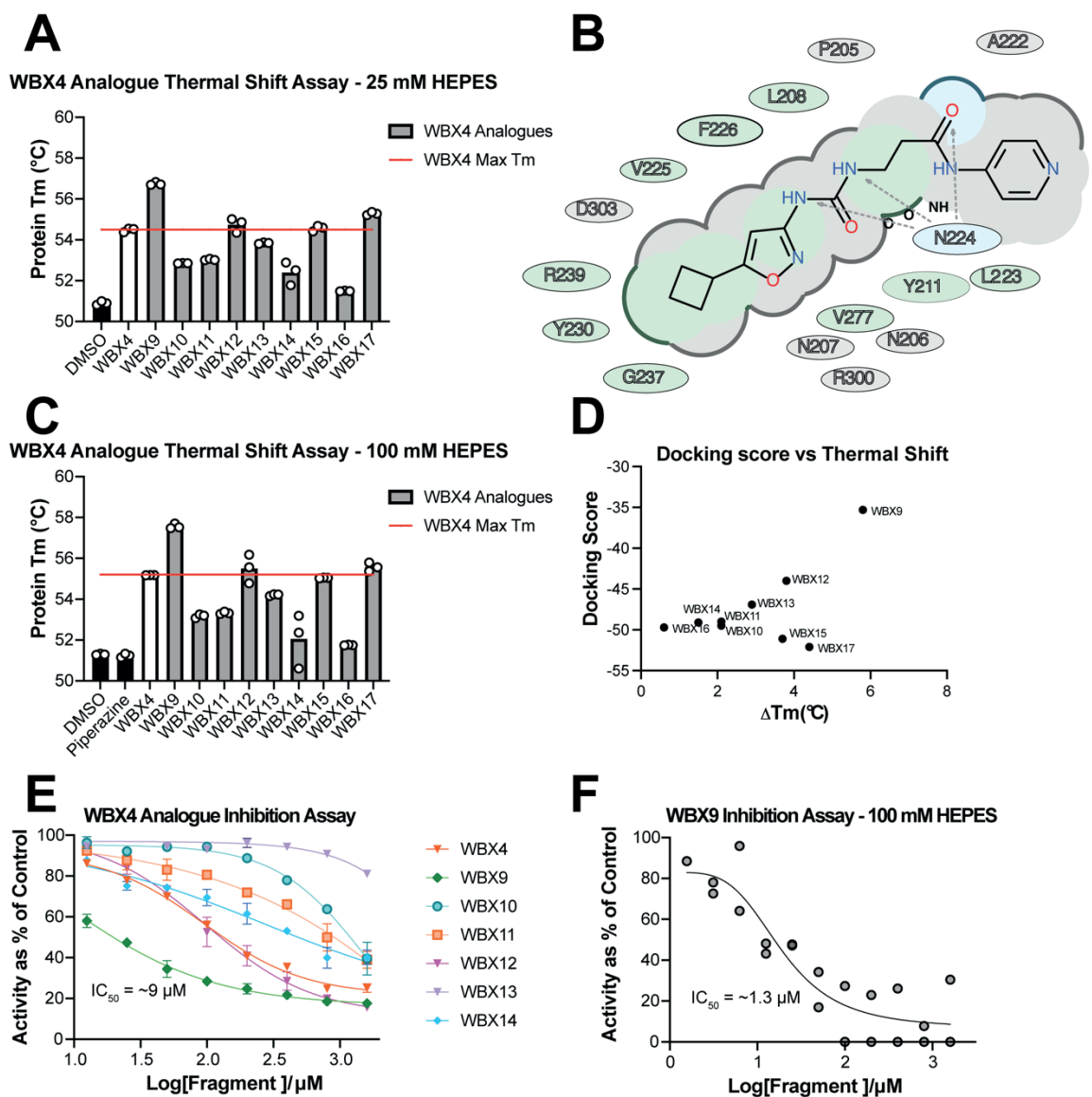


Figure 41: Analogues series two characterisation.

A) Thermal shift screening results for **WBX4** analogues performed at 25 mM HEPES buffer (results show technical triplicates). Red line shows the maximum value obtain for **WBX4**. B) 2D illustration of the **WBX9** docking result. Green areas shown lipophilic regions. Blue regions are polar. Larger circles showing residues indicates more significant interactions to the ligand. C) As displayed in panel A but with 100 mM HEPES buffer. D) A plot of docking score vs ΔT_m . E) Inhibition assay results for **WBX4** and analogues **WBX9-14** (results show mean and range from technical duplicates). F) inhibition assay results for **WBX9** repeated at 100 mM HEPES. Experiment performed in technical duplicate.

Fragment	ΔT_m °C
WBX4	3.6
WBX9	5.8
WBX10	2.1
WBX11	2.1
WBX12	3.8
WBX13	2.9
WBX14	1.5
WBX15	3.7
WBX16	0.6
WBX17	4.4

Table 13: WBX4 and analogues ΔT_m values calculated from DMSO controls.

To further characterise **WBX9**, ITC was used to determine the K_D . This experiment produced excellent data (**Figure 42**) and a K_D of 496 ± 37 nM (LE 0.36). The occupancy (N) was calculated at 0.51, a substantial increase in available binding sites compared to any previous ITC experiment. It is speculated that the much tighter affinity of **WBX9** enables it to dislodge some of the bound substrate blocking the active site in the abortive complex, thus providing a lower N value.

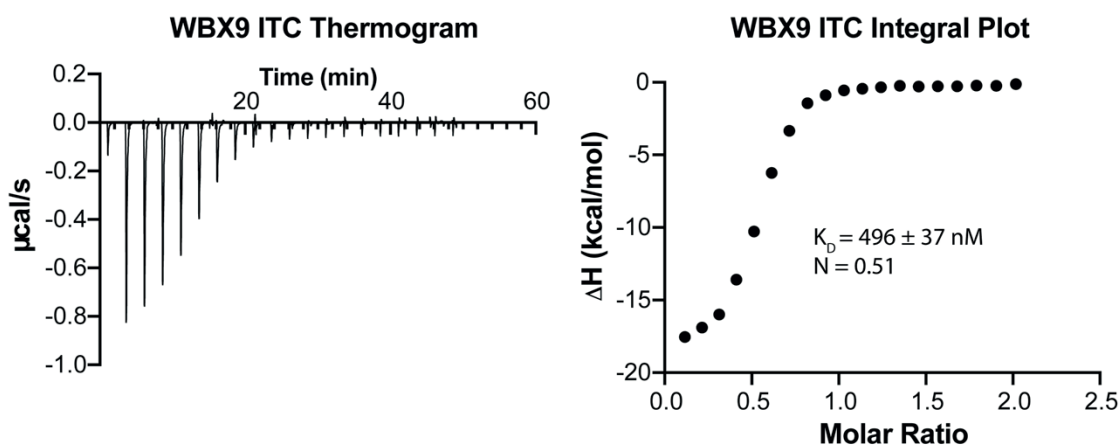


Figure 42: WBX9 ITC assay results with thermogram and integral plots.

Data collected by MRes student Sing Hei Lok.

3.2.4.3 Analogue Series 2 Crystallography

Finally, several of the **WBX4** analogues were soaked with crystals of GALE at 1 mM to determine protein/ligand complexes and confirm the binding modes, again by Dr. Stephane Mouilleron. GALE structures containing ligands were obtained for **WBX9**, **12**, and **15** and **17** at excellent resolutions (≤ 1 Å). Fragment **WBX17** binds in a very similar orientation as the parent fragment **WBX4** (**Figure 43A**), with key hydrogen bonds to Asn224 and Phe226. Crucially, the cyclisation around the amide does not disrupt the hydrogen bond made to Asn224. The main difference between this fragment and the parent compound is the shift in angle of the pyridine ring (**Figure 43A**), from in-plane with the amide in **WBX4**, to a $\sim 45^\circ$ angle in the structure for **WBX17**. From the thermal shift results, **WBX17** displays a clear increase in protein stability compared to **WBX4** ($\Delta T_m = 0.8$ °C). Based on the crystal structure obtained, there are two likely explanations. First, the cyclised fragment can make stronger π -stacking interactions with Tyr211 than the uncyclized fragment (**WBX4**). Second, the cyclised fragment (**WBX17**) has a reduced entropic penalty of binding due to reduced conformational freedom. Of note, the differences in pyridine orientation did not have a negative effect on binding, suggesting that this ring is not a driver of potency. Removal of the pyridine may help to significantly improve ligand efficiency in future analogues.

As was the case for the other fragments, **WBX15** also binds in much the same way as previous compounds. **WBX15** had the second-best docking score of all the fragments (-51.1) due to an additional hydrogen bond predicted to the carbonyl of Ala222. The crystal structure reveals that this interaction was accurately predicted (**Figure 43B**). Paradoxically, this did not translate into increased thermal stability (only a 0.1 °C over **WBX4**). Further characterisation of this compound in an inhibition assay will be required to confirm that the additional H-bond has no effect on protein binding.

WBX12 substituted the methyl group of the methyl isoxazole for a sulfonamide. Surprisingly given the size difference of the two groups, the inhibition assayed determined an IC_{50} identical to that of **WBX4** (~ 9 μ M). From the crystal structure, the sulfonamide neatly fills a pocket

(**Figure 43C**). However, no additional interactions (H-bond or otherwise) appear to be made to the surrounding residues. This is best illustrated in the 2D interaction diagram generated by ICM (**Figure 43D**).

Finally, the crystal structure obtained for **WBX9** confirms that the isoxazole-urea portion of the fragment binds in the same manner as previous fragments (the isoxazole is not flipped as predicted by ICM). Crucially, this allows the cyclobutyl group to adopt a conformation that fills the adjacent lipophilic pocket, without losing the network of H-bonds to Asn224 and Phe226 (**Figure 43E-F**).

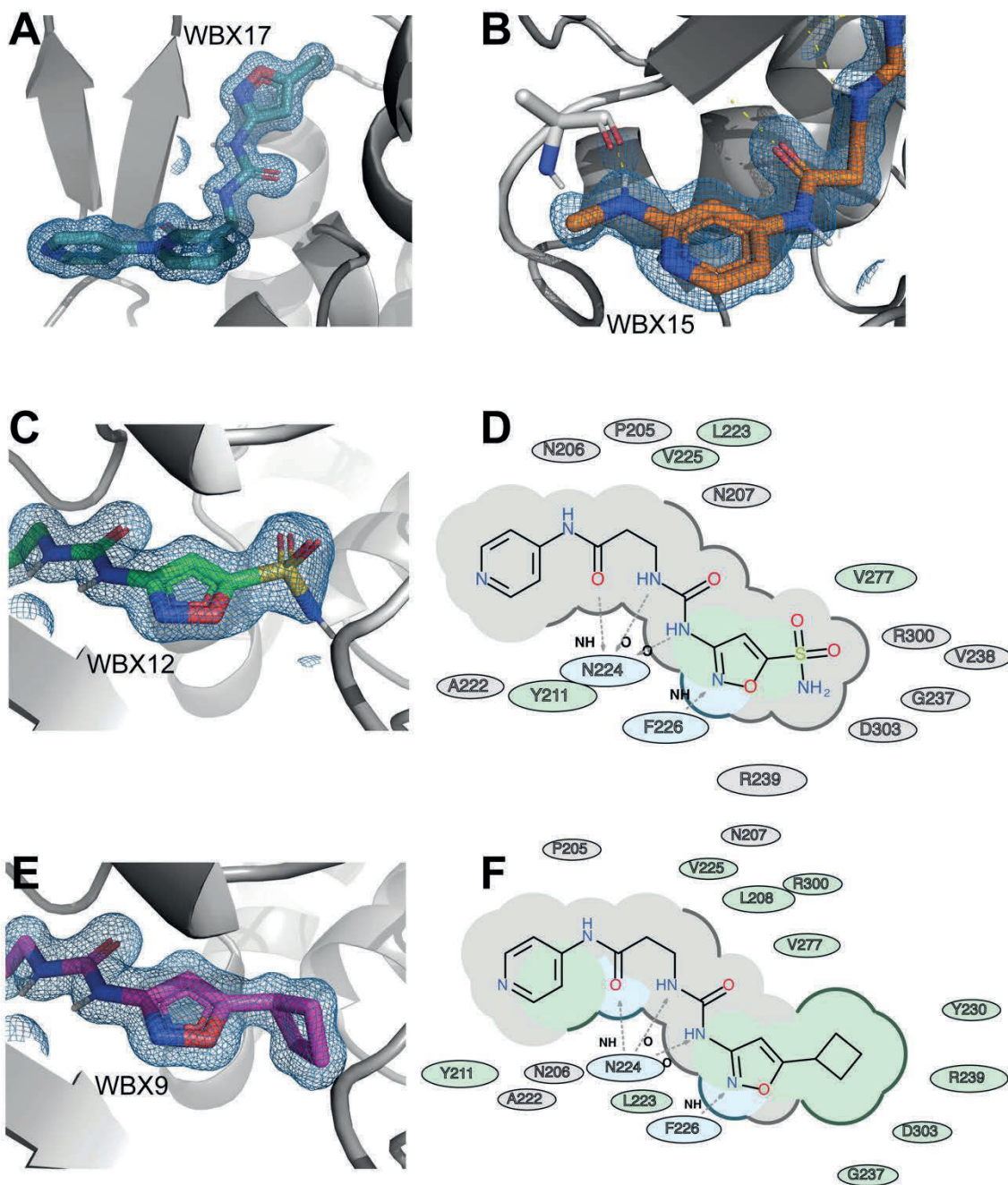


Figure 43: Analogue series 2 crystal structures and binding interaction analysis.

A) **WBX17** in complex with GALE. B) WBX15 in complex with GALE showing H-bond made to Ala222. C) WBX12 in complex with GALE. D) 2D interaction diagram generated in ICM from WBX12 crystal structure. E) Structure of **WBX9** in complex with GALE. F) 2D interaction diagram generated in ICM from **WBX9** crystal structure. All blue meshes show 2Fo-Fc maps at σ level 1.

3.3 Chapter 3 Summary

To summarise, this chapter describes the efforts to screen and elaborate fragment hits using a cascade of biophysical, biochemical, and crystallographic assays. Screening was first performed with the Enamine DSI-Poised library by thermal shift assay. The screen was performed at both 1 and 4 mM to identify hits that displayed dose dependent shifts in protein stability (T_m). Unfortunately, subtle shifts in pH (0.1-0.2) resulted in positive protein stabilisation, particularly at 4 mM, resulting in a high false positive result. Nevertheless, hits were still identified that stabilised the protein and did not contain highly basic groups. These compounds may prove useful starting points for parallel lead series.

As an alternative approach, the same library was screened by crystallography at the XChem Facility. The aim was to identify compounds that were hits in both screens to provide additional validation. The Screen identified three hit fragments: **X0121**, **X0213**, and **X0373**. Of these only **X0213** was also a hit in the thermal shift screen. Both **X0121** and **X0213** bound in the uracil binding region of the GALE substrate binding site, whilst **X0373** bound in a crystallographic interface between GALE monomers. Validation of fragment binding by ITC could only be obtained for **X0213** (K_D 370.0 ± 44.6 μ M). **X0213** was thus used in a structure search to identify commercially available analogues, based on virtual screening of these compounds, eight analogues were ordered for experimental validation. An assay workflow using thermal shift, enzyme inhibition assays, ITC, and crystallography was used to characterise the fragments. This workflow identified **WBX4** as having a clear increase in binding affinity (~ 100 -fold) in addition to a measurable IC_{50} of 115 μ M. Crystallography determined the structure of the **WBX4**/GALE complex and found it in good agreement with the predicted docking. The increase in binding was likely driven by an additional H-bond made by **WBX4** to Asn224.

After this success a second series of analogues were designed (guided by virtual docking results) to further increase potency. Nine compounds were selected and custom synthesised by Enamine. Testing of these compounds in the assay workflow identified **WBX9** as having a significant increase in binding (K_D 496 nM) and inhibition ($IC_{50} \sim 9$ μ M). Crystallography

determined that this increase was driven through filling of a lipophilic pocket by a cyclobutyl group, which had been substituted for a methyl group present in the parent compound (**WBX4**). In all, an initial hit with weak protein binding from an XChem crystallographic screen has been elaborated using virtual docking and structure guided design to a nM binder.

3.4 Methods

3.4.1 DSF Thermal Shift Assay Optimisation

Thermal shift assays were prepared at 4 °C with 2 µM GALE at variable concentrations of, substrate, dye and DMSO. Reactions were 20 µL and prepared in 384-well PCR plates (Applied Biosystems). Plates were then spun at 1000 rpm for 1 minute and incubated at 4 °C for 10 minutes. Melt curves were acquired using a QuantStudio 12K Flex Real-Time PCR System (ThermoFisher). A temperature gradient was run from 25-95 °C (3 °C/min). Acquired melt curves were analysed using the Protein Thermal Shift™ Software (Applied Biosystems).

3.4.2 DSF Thermal Shift Screening

The methods below were adapted from an earlier early stage assessment (ESA) report.

A premixed protein-dye sample was prepared at 4 °C as follows: 2 µM GALE, 5X SYPRO orange dye (ThermoFisher), 2 mM NAD, 25 mM HEPES, pH 7.5, 50 mM NaCl. Fragments (Enamine, DSI-Poised and Essentials library) were dispensed with an Echo 550 (Labcyte) into 384-well PCR plates (Applied Biosystems). 20 µL of protein-dye mix was dispensed into each well using a Viafil dispenser (Integra). Each plate contained 56 DMSO only controls and 8 positive controls (UDP-GlcNAc 1 mM). Plates were then spun at 1000 rpm for 1 minute and incubated at 4 °C for 10 minutes. Each library screen was run at both 1 mM and 4 mM fragment concentration in technical duplicates.

Melt curves were acquired using a QuantStudio 12K Flex Real-Time PCR System (ThermoFisher). A temperature gradient was run from 25-95 °C (3 °C/min). Acquired melt curves were analysed using the Protein Thermal Shift™ Software (Applied Biosystems).

3.4.3 XChem Fragment Screening

GALE was concentrated to 20 mg/mL in 25 mM HEPES, pH 7.5, 500 mM NaCl, 2 mM NAD. The protein was crystallised in MRC 3 drop Swisssci plates by sitting drop vapour diffusion. Drops consisted of 100 nL GALE/NAD with 90 nL of 100 mM EPPS, pH 7.5, 136 mM KCl, 19.3 % w/v PEG 8K precipitant, 10 µL of seed stock in precipitant. Equilibrating with 40 µL of precipitant. Drops were dispensed by NT8 (Formulatrix). In total 6 plates with crystals were grown.

Wells containing appropriate crystals were selected using a crystal ranking system and fragments in ethylene glycol dispensed into these wells using an Echo 550 (Labcyte) to a final concentration of either 10 or 20 mM. Crystals were left to soak with fragments for ~3 hours at 4 °C. Crystals were fished using a Shifter (Oxford Lab Technologies, Oxford, UK) and frozen in liquid nitrogen and stored in pucks for data collection.¹⁵¹

3.4.4 XChem Crystal Data Collection

Data for all crystals was collected on Beamline IO4 (DLS) by unattended data collection at 100 K. Datasets consisted of 1800 images with an oscillation of 0.1 °, exposure time of 40 ms per image per crystal. Data sets were indexed, scaled and merged with xia2.¹³⁰ Data was processed using the XChem pipeline by Dr. Andy Purkiss and the full method is described: <https://www.diamond.ac.uk/Instruments/Mx/Fragment-Screening/The-XChem-Pipeline.html>.^{152,153}

However, in summary, Xia2 auto-processed files were imported in the XChem explorer (XCE) pipeline where batch molecular replacement was first performed on protein crystal only controls using DIMPLE and the “mixed homodimer” search model solved for GALE previously.¹⁵⁴ A ground-state model was then determined for GALE from these structures

using the Pan-Dataset Density Analysis (PanDDA) method.¹⁵⁵ Ligands models and geometry restraints were prepared from the Enamine DSI-Poised library using AceDRG.¹⁵⁶ Batch molecular replacement using DIMPLE is then performed on the protein/fragment samples using the ground state structure. Hits are identified from unfilled density “blobs” using PanDDA and refinement performed with XCE.

3.4.5 ICM Virtual Screening

Virtual ligand screening was conducted using ICM-Pro (version 3.9-3a) with the Virtual Ligand Screening (VLS) package. The methods below were adapted from an earlier MRes dissertation. (W. M. Browne, MRes Thesis, Imperial College London, 2020.)

First, ICM receptor maps for a specified binding site (in this case the UDP binding pocket) are made employing five types of interaction potentials: 1) Van der Waals interactions for hydrogen probes, 2) Van der Waals interactions for heavy atoms, 3) hydrophobic terms, 4) lone-pair-dependent H-bonding potentials, and 5) electrostatic terms.¹⁴⁵ The force field (FF) used for this task is the Empirical Conformational Energy Program for Peptides (ECEPP).¹⁴⁶ Notably, this FF maintains fixed values for bond length and angle, with only the dihedral torsion angle being variable. Ligand conformational docking within the receptor maps uses the biased probability Monte Carlo method.¹⁴⁷ This approach randomly situates one of the low-energy ligand conformations within the receptor map before shifting to another random position. Following each move, local energy minimization is applied, and desolvation energies are calculated. By combining large-scale positional adjustments with local energy minimization for the flexible ligand, a global energy optimization of both the ligand and receptor is achieved. The number of moves is determined by an algorithm influenced by the chosen "Thoroughness" value (set to 3 for all screening iterations). The ligands are modelled based on the Merck Molecular Force Field (MMFF).¹⁵⁷

In general docking was performed in several steps:

1. Import of the crystal structure to be docked into.
2. Conversion of the PDB file to an ICM object, keeping tight waters in the structure.
3. Identification of the binding pocket (UDP pocket) and building of receptor maps.
4. Docking of libraries from SDF files

All docking parameters kept as default except for thoroughness (set to 3).

3.4.6 ITC Fragment Assays

ITC experiments were performed as described in chapter 1 with minor adjustments as described below.

ITC characterisation was performed using a MicroCal PEAQ ITC system at 25 °C. 300 µL purified GALE at 25 µM in 25 mM HEPES, pH 7.5, 50 mM NaCl, 4 % DMSO buffer, was loaded into the sample cell. Ligands were prepared at 4 mM in buffer matching the protein. Ligands were titrated into the cell with 1x0.4 µL injection followed by 19x2 µL injections, 4 s injection durations with 150 s spacing. Heat release per titration was measured and integrated using the MicroCal PEAQ ITC software. All kinetic parameters such as K_D , Enthalpy (ΔH), Gibbs free energy (ΔG), and stoichiometry (N) were determined using a single site occupancy model.

3.4.7 Fragment Inhibition Assays

The assays detailed in this section were performed by and MRes student (Sing Hei Lok) under supervision.

Fragments were dispensed in a 2-fold dilution series from 1600 µM to 12.5 µM by Echo 550 (Labcyte) into 384-well PCR plates (Applied Biosystems) from 100 mM DMSO fragment stocks. DMSO volume was adjusted to 4 % final DMSO concentration. 20 µL of GALE 5 nM, 25 mM HEPES, pH 7.5, 50 mM NaCl, 2 mM NAD and 100 µM UDP-Gal was dispensed to each fragment

containing well. Reactions were prepared in technical duplicates with four DMSO only controls. Reactions were run at 37 °C for 15 minutes. Quenching was carried out by heat inactivation at 95 °C for 1 minute. Samples were analysed by HPLC as described in methods section 2.4.2.

3.4.8 Expression and Purification of GALE from *E.coli*

Expression and purification of GALE was performed by Dr. Stephane Mouilleron and the methods below provided by him.

Protein expression in BL21 (DE3) *E. coli* cells (Invitrogen). 1L of ZYM5052 autoinduced media was inoculated with one colony, incubated for 10h at 37°C followed by 24h at 20°C. Cells were harvested and resuspended in lysis buffer (50 mM HEPES, pH 7.5, 10 mM imidazole, 500 mM NaCl, 0.2% v/v Tween 20, 1 mM TCEP, 0.5 mM AEBSF, 15 µg/ml benzamidine, 10% glycerol and complete EDTA-free protease inhibitor tablets), lysed by French press, clarified, and stored at -80°C. An HisTrap column (5ml) was equilibrated in 5CV of Buffer A (50 mM HEPES pH 7.5, 500 mM NaCl, 10 mM Imidazole). GALE lysate was loaded into the column followed. By a 20CV wash step with buffer A. The protein was eluted with 5CV of buffer A supplemented with 250 mM Imidazole. GALE was then concentrated to 5 ml and loaded onto a superdex 200 26/60 equilibrated in buffer B (20 mM HEPES pH 7.5, 100 mM NaCl, 0.5 mM TCEP). The fractions with active GALE are pooled together and cleaved with His-3C protease. The sample was then loaded again on a HisTrap column in buffer A. The cleaved product was recovered in the flowthrough, using a Vivaspin20 10,000 MWCO and loaded onto a superdex 75 26/60 equilibrated in buffer B. Purified GALE was then concentrated at 15 mg/ml and stores at -80°C.

3.4.9 Data Collection for X0213 Analogue Structures

Data collection and structure refinement was performed by Dr. Stephane Mouilleron and the methods below provided by him.

GALE was crystallised at 20 °C using sitting-drop vapour diffusion. Sitting drops of 1 µl consisted of a 1:1:0.2 (vol:vol:vol) mixture of protein, well solution and GALE seed stock. Seeds stock were obtained from crystals grown in the same precipitant as below but without seeds, crystals were not as nice. Seeding improved the quality of crystals. Well solutions were as follows 16% PEG 3350, 0.2 M Na Malonate pH 7. Crystals appeared within a few hours and reached their maximum size after 2 to 3 days.

All crystals were cryoprotected in well solution supplemented with 30% ethylene glycol and then soaked for 30 min in the cryoprotectant solution and 5 mM compounds. Crystals were then flash-frozen in liquid nitrogen and X-ray data were collected by unattended data collection at 100 K at beamlines I04 1 and I03 of the Diamond Light Source Synchrotron (Oxford, UK). Data collection and refinement statistics are summarised in section 3.5. Data sets were indexed, scaled and merged with xia2.¹³⁰ Molecular replacement used the atomic coordinates of human GALE from PDB 1EK5 in PHASER (McCoy et al., 2007).¹³¹ Refinement used Phenix.¹⁵⁸ Model building used COOT with validation by PROCHECK.^{132,159}

3.5 Chapter 3 Crystal Structure “Table 1” Validation

Fragment	X0121	X0213	X0373
Wavelength (Å)	0.98	0.98	0.98
Resolution range (Å)	85.91 - 1.73	85.50 – 1.86	86.07-1.88
Space group	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁
Cell dimension:			
a, b, c (Å)	46.1, 110.6, 136.4	46.1, 110.8, 135.6	46.1, 110.4, 137.4
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Total reflections	500 127 (24 978)	399 747 (18 003)	394 044 (18230)
Unique reflections	73 814 (3 607)	59 377 (2 880)	58 125 (2820)
Multiplicity	6.9 (6.8)	6.7 (6.3)	6.8 (6.5)
Completeness (%)	100 (98.9)	100 (99.2)	100 (98)
Mean I/sigma(I)	7.9 (0.6)	6.5 (0.7)	6.1 (0.5)
Wilson B-factor	28.7	15.5	22.3
Reflections used in refinement	73 607 (3 602)	59 184 (2 865)	52 725
Reflections used for R _{free}	3 738	2 999	2 655
R _{meas}	0.07 (2.50)	0.07 (2.29)	0.08 (2.39)
CC1/2	1 (0.30)	1 (0.2)	1 (0.3)
R _{work}	0.20	0.20	0.18
R _{free}	0.25	0.24	0.24
RMSD - bond lengths	0.01	0.01	0.01
RMSD - bond angles	1.74	1.67	2.25
Ramachandran favored (%)	97.2	97.8	97.4
Ramachandran allowed (%)	2.3	2.2	2.2
Ramachandran outliers (%)	0.4	0.0	0.0
Rotamer outliers (%)	3.1	2.0	4.0
Clash score	8.7	3.3	1.93

Table 14: Full “Table 1” validation for the XChem hit crystal structures.

Fragment	WBX4	WBX2
Wavelength (Å)	0.98	0.98
Resolution range (Å)	43.52 - 1.22	42.53 - 1.17
Space group	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁
Cell dimension:		
a, b, c (Å)	45.9, 109.6, 134.9	46.1, 108.6, 136.7
α, β, γ (°)	90, 90, 90	90, 90, 90
Total reflections	2 622 698 (179 275)	2 777 320 (125 277)
Unique reflections	200 142 (16 666)	224 283 (17 247)
Multiplicity	13.1 (9.8)	12.4 (7.1)
Completeness (%)	97.9 (83.1)	96.59 (75.1)
Mean I/sigma(I)	7.6 (0.4)	13.45 (0.7)
Wilson B-factor	14.6	15.45
Reflections used in refinement	198 477 (16 666)	223 563 (17 223)
Reflections used for R _{free}	1 980 (200)	11 304 (880)
R _{merge}	0.13 (2.67)	0.07(2.12)
R _{meas}	0.14 (2.82)	0.07 (2.29)
CC1/2	0.99 (0.35)	1 (0.41)
R _{work}	0.15	0.15
R _{free}	0.19	0.18
RMSD - bond lengths	0.23	0.02
RMSD - bond angles	1.34	1.47
Ramachandran favored (%)	97.37	97.96
Ramachandran allowed (%)	2.48	2.04
Ramachandran outliers (%)	0.15	0
Rotamer outliers (%)	0	0.36
Clash score	3.77	3.68

Table 15: Full "Table 1" validation for the structures of WBX4 and WBX2.

Fragment	WBX17	WBX15
Wavelength (Å)	0.98	0.98
Resolution range (Å)	38.12 - 0.96	43.66 - 0.98
Space group	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁
Cell dimension:		
a, b, c (Å)	46.1, 110.3, 135.3	46.1, 110, 135.4
α, β, γ (°)	90, 90, 90	90, 90, 90
Total reflections	16 758 419 (1 452 195)	15 904 693 (1 554 733)
Unique reflections	417 044 (38 525)	390 833 (33 403)
Multiplicity	40.2 (35.9)	40.7 (40.1)
Completeness (%)	99.12 (92.90)	98.1 (86.1)
Mean I/sigma(I)	12.84 (0.48)	13.65 (0.41)
Wilson B-factor	10.37	11.11
Reflections used in refinement	414 908 (38 521)	385 001 (33 401)
Reflections used for R _{free}	2000 (186)	1982 (167)
R _{merge}	0.13 (2.67)	0.14 (5.25)
R _{meas}	0.14 (2.82)	0.14 (5.32)
CC1/2	1 (0.46)	1 (0.41)
R _{work}	0.14	0.15
R _{free}	0.16	0.18
RMSD - bond lengths	0.02	0.015
RMSD - bond angles	1.51	1.43
Ramachandran favored (%)	98.10	97.95
Ramachandran allowed (%)	1.90	2.05
Ramachandran outliers (%)	0.00	0.00
Rotamer outliers (%)	0.87	1.22
Clash score	4.38	3.41

Table 16: Full "Table 1" validation for the structures of **WBX17** and **WBX15**.

Fragment	WBX12	WBX9
Wavelength (Å)	0.98	0.98
Resolution range (Å)	38.11 - 1.00	43.68 - 0.95
Space group	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁
Cell dimension:		
a, b, c (Å)	46.1, 110.3, 135.3	46.1, 110.2, 135.4
α, β, γ (°)	90, 90, 90	90, 90, 90
Total reflections	15 040 642 (1 493 141)	17 546 262 (1 657 185)
Unique reflections	370 351 (36 718)	415 324 (33 175)
Multiplicity	40.6 (40.7)	42.2 (41.4)
Completeness (%)	99.52 (96.13)	94.51 (77.38)
Mean I/sigma(I)	10.13 (0.34)	11.87 (0.36)
Wilson B-factor	11.77	11.63
Reflections used in refinement	368 740 (35 297)	407 994 (33 170)
Reflections used for R _{free}	2011 (196)	1999 (162)
R _{merge}	0.17 (6.12)	0.13 (10.64)
R _{meas}	0.17 (6.20)	0.13 (10.77)
CC1/2	1 (0.39)	1 (0.41)
R _{work}	0.14	0.15
R _{free}	0.16	0.18
RMSD - bond lengths	0.02	0.01
RMSD - bond angles	1.44	1.12
Ramachandran favored (%)	97.80	97.51
Ramachandran allowed (%)	2.20	2.49
Ramachandran outliers (%)	0.00	0.00
Rotamer outliers (%)	1.05	1.05
Clash score	3.05	3.32

Table 17: Full "Table 1" validation for the structures of WBX12 and **WBX9**.

Chapter 4 Covalent Fragment Screening and Characterisation

4.1 Chapter Introduction

In this chapter are discussed experiments concerning covalent fragments. It should be noted that the libraries used here were the property of GSK, co-funders of this project. Details where collaborators performed experiments are included in the methods section of the report.

4.1.1 Covalent Fragments

The topics discussed in Chapter 3 have been regarding non-covalent fragments, where binding (or inhibition) is achieved through intermolecular fragment-protein interactions (hydrogen bonding, Van Der Waals, etc). Covalent fragments and drugs contain an electrophilic “warhead” that is designed to form covalent bonds with amino acids with nucleophilic side chains.¹⁶⁰ This mode of action results in proteins covalently modified by the fragment. The reaction between covalent fragment and protein is designed to occur in two distinct steps. 1) The fragment makes reversible non-covalent interactions with the protein. 2) This orientates the warhead so that it can react with a nucleophilic residue to form the covalent bond (**Figure 44**).¹⁶¹ Crucially, the efficiency of each of these two steps is defined by two separate kinetic parameters, step 1 by K_i and step 2 by k_{inact} (analogous to K_m and k_{cat} respectively). The ratio between these two (k_{inact}/K_i) is used to determine labelling efficiency and compare covalent compounds.¹⁶² The design of the warhead must be such that it is sufficiently electrophilic to react with the desired amino acid side chain, but not so reactive as to label the protein without first forming specific non-covalent interaction.¹⁶⁰

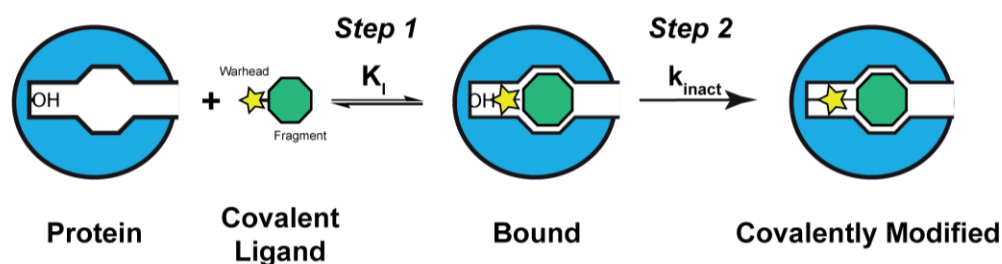


Figure 44: Covalent fragment kinetics.

Diagram showing how a covalent fragment (green) with an electrophilic warhead (yellow) first binds to a protein (blue) non-covalently before forming a covalent bond with nearby nucleophilic residues. The kinetic parameters K_I and k_{inact} are also displayed.

The design of the warhead is therefore a major consideration in the development of these fragments, as it enables tuning of the libraries to target specific amino acids.¹⁶³ The most nucleophilic amino acid is cysteine and the acrylamide and chloroacetamide functional groups have proven to be potent warheads for selectively targeting them.¹⁶⁴ More recently, sulfonyl fluoride exchange (SuFEx) chemistry has been exploited to design warheads selective for Ser, Tyr and Lys residues.¹⁶⁵

4.1.2 Covalent Inhibitors

Historically many covalent inhibitors/drugs have been identified through serendipity and contained a variety of warheads. One of the earliest covalent drugs marketed was Aspirin (**Figure 45**) in 1899, a non-steroidal anti-inflammatory drug (NSAID) that inhibits cyclooxygenase enzymes (COX-1 and COX-2) through covalent modification of a serine.¹⁶⁶ Penicillin (**Figure 45**), another breakthrough drug, contains the β -lactam functional groups which reacts with a serine residue in a bacterial transpeptidase.¹⁶⁷ In both cases, the precise mechanism of action was not elucidated until years after their use. Despite these successes, covalent drugs were left underexplored due to toxicity concerns around introducing reactive compounds into biological systems and possible interference with assays.¹⁶⁴ However, there are several key benefits provided by covalent fragments which encouraged further study; 1) The covalent modification of the protein is irreversible; this can overcome competition with natural substrates leading to increased potency. 2) Modification is dependent on nucleophilic residues in specific environments, improving the specificity of the ligand often giving superior

safety profiles over non-covalent analogues. 3) Targeting nucleophilic residues on difficult to target proteins can expand the “druggable” proteome 4) Covalent fragments can be conveniently screened by intact denaturing LCMS (intact-LCMS).¹⁶⁸

Identifying and developing covalent inhibitors follows two main approaches: electrophile first, and ligand first. In “electrophile first” approaches, the aim is to identify covalent hits from the start, this usually means screening libraries of compounds designed specifically to contain covalent warheads.¹⁶⁴ A prime example of this strategy that illustrates the benefits described above was the development of the KRAS(G12C) inhibitor sotorasib (**Figure 45**).¹⁶⁹ KRAS is a well validated oncoprotein that is mutated in approximately 25% of all cancers, with a major mutant being the G12C variant.¹⁷⁰ Ever since the validation of KRAS as a promising target, multiple efforts were made using conventional small molecule inhibitors to target it with no success and several factors suggested covalent compounds may provide a solution.¹⁷¹ First, the presence of a mutation to cysteine in the substrate binding pocket meant that a covalent compound labelling this residue would likely inactivate the enzyme. Additionally, the fact that this mutation is not present in healthy cells provides a level of selectivity unachievable with non-covalent compounds.¹⁷² Finally, screening against KRAS(G12C) could be performed using a disulfide fragment library (containing acrylamide warheads) with a convenient intact-LCMS assay as the readout.¹⁷³ A limitation of screening in this way is that, if the target contains multiple nucleophilic residues that the library is designed to target (e.g. Cys), screening can identify compounds that label residues with no effect on activity. Additionally, in this case the site of labelling needs to be confirmed for each hit to guide which compounds should be further developed.

The “ligand first” approach involves first identifying non-covalent ligands for a protein and then modifying these with a covalent warhead to neighbouring residues.¹⁶⁴ The benefit of this approach is that structure-based design can be used to target specific residues close to the ligand, limiting the chance of developing covalent compounds that bind to functionally irrelevant residues. This is well illustrated in the case of afatinib (**Figure 45**), an acrylamide-containing covalent inhibitor of the oncoprotein receptor tyrosine kinase (RTK) epidermal growth factor receptor (EGFR).¹⁷⁴ The non-covalent inhibitor erlotinib (**Figure 45**) displayed potent inhibition of EGFRs containing specific mutations that promoted growth of non-small

cell lung cancer (NSCLC).¹⁷⁵ However, much of this efficacy was lost by a T790M gatekeeper residue mutation.¹⁷⁵ Design of covalent analogues such as Afatinib were able to target Cys797 located adjacent to the substrate binding site, partially restoring the lost activity.¹⁷⁶ This is an excellent example of how rational structure-based design can be used to improve the potency of non-covalent compounds with low activity.

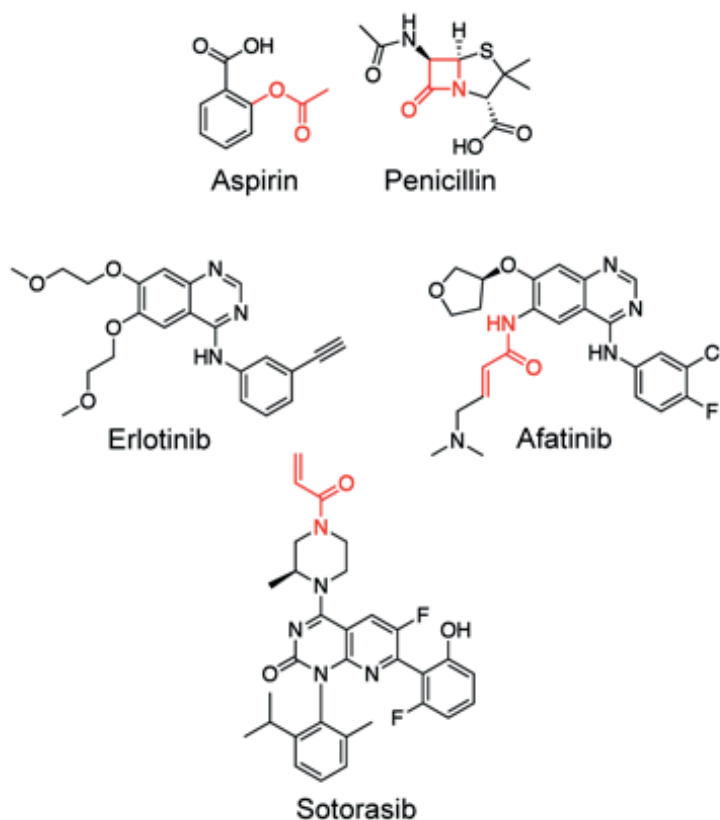


Figure 45: FDA approved covalent drugs.

Warheads shown in red.

4.1.3 Activity-Based Protein Profiling (ABPP)

Another key use of covalent protein labelling compounds is in the field of chemoproteomics, specifically through activity-based protein profiling (ABPP).¹⁷⁷ In general, ABPP works by first using a probe that covalently labels and inactivates targets that are enzymatically active (i.e. have a free and functional active site), and subsequent determination of the target protein. The ABPP probe will typically consist of two main parts, a reactive group with a warhead to covalently label and inactivate the protein and a reporter tag to enable protein identification.¹⁷⁸ This approach pioneered by Ben Cravatt was developed for two main

purposes: 1) comparative profiling, where the aim is to assess the relative activities of an enzyme using probes designed to react with and inactivate a specific class of enzyme (e.g. serine hydrolases) within a complex proteome (**Figure 46A**) and 2) competitive profiling, where the aim is to perform competition experiments with covalent probes to identify inhibitors (**Figure 46B**).⁸⁹ Isotopic Tandem Orthogonal Proteolysis-Activity-Based Protein Profiling (isoTOP) ABPP has expanded these experiments to perform covalent fragment screening in complex systems to identify ligandable cysteines.¹⁷⁹

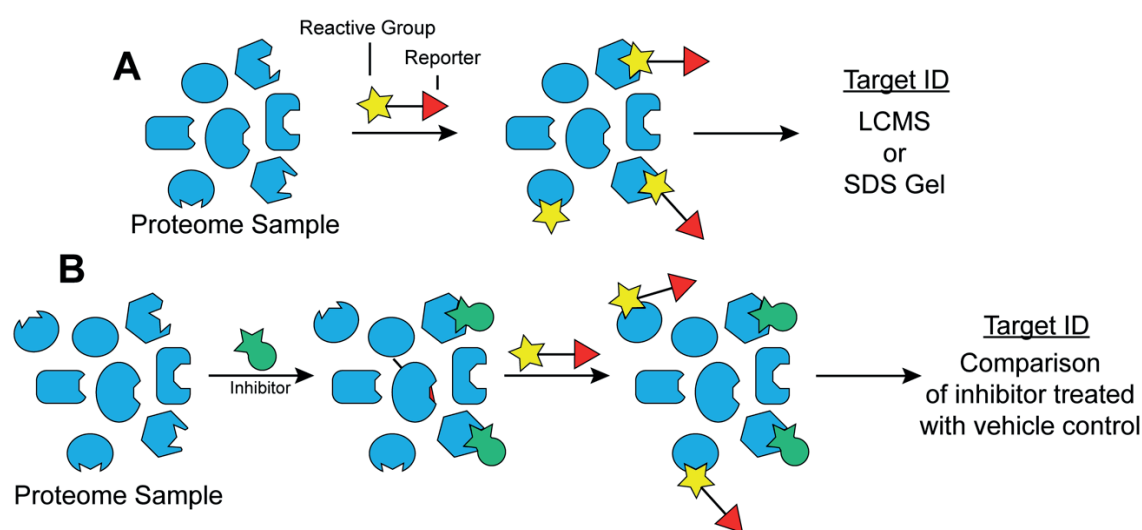


Figure 46: Activity based protein profiling.

A) Illustration of comparative profiling. Proteomic samples are reacted with probes designed to covalently modify and inactivate a class of proteins. A reporter tag is then used for target identification and quantification. B) Illustration of competitive profiling. Proteomic samples are first treated with an inhibitor, a covalent probe is then added specific for the inhibitor target (and related enzymes) Probe labelled compounds are then identified and quantified for comparison with no inhibitor control to confirm inhibitor binding.

4.1.4 GALE as a Covalent Fragment Target

As mentioned above, a major target of covalent fragments are cysteines due to their high nucleophilicity. However, other functional groups such as Ser, Lys, and Tyr have been targeted successfully with SuFEx chemistry. GSK have published previously on the use of aryl sulfonyl fluorides as warheads selective for Lys and Tyr residues.^{180,181} GALE contains an exposed Tyr (Tyr211) adjacent to the UDP binding site of GALE (**Figure 47A-B**), particularly close to where the uridine of UDP binds. Targeting Tyr211 with a covalent fragment screen seemed a

promising approach to identify hit compounds that could be elaborated into covalent inhibitors of GALE.

This chapter describes the efforts to develop covalent inhibitors of GALE using both “electrophile first” and “ligand first” approaches and discussed how the identified ligands could be used in future work.

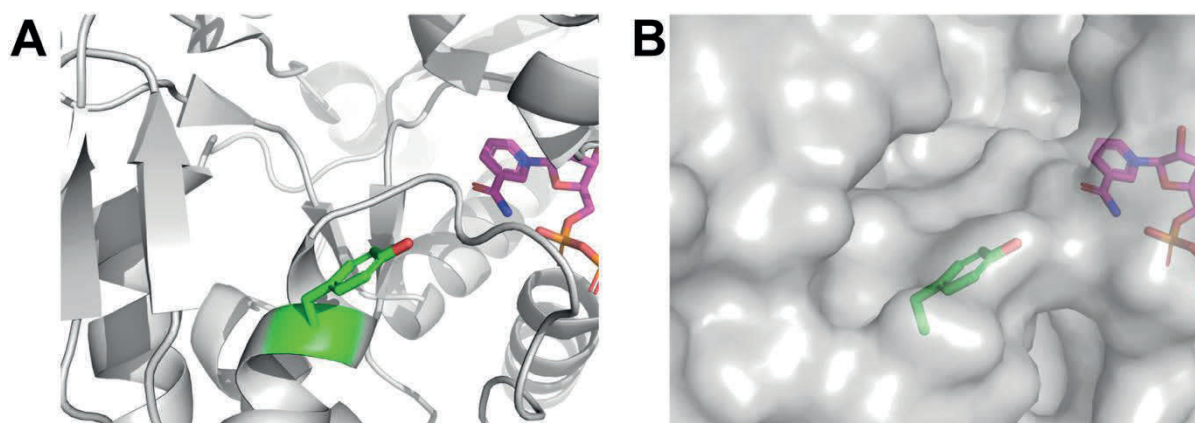


Figure 47: Location of covalent fragment target Tyr211 in GALE.

A) Crystal structure o GALE with Tyr211 (green) and protein cartoon (grey). B) Tyr211 (green) with modelled protein surface (grey). NAD (purple).

4.2 Results and Discussion

4.2.1 Aryl Sulfonyl Fluoride Library Screening

4.2.1.1 Screening Introduction and Overview

To identify covalent fragments that modify GALE a 347-member sulfonyl fluoride library was screened against GALE using denaturing intact mass spectrometry (Intact-LCMS) (**Figure 48A**), detailed in methods section 4.4.1). In Brief, fragments (100 μ M) were incubated with GALE (0.5 μ M, 25 mM HEPES, pH 7.5, 50 mM NaCl) for 24 hours at 4 $^{\circ}$ C. Samples were then injected onto an intact-LCMS system. Spectra were deconvoluted and the relative intensities of peaks from native protein and the protein-fragment reaction product compared to determine % labelling (**Figure 48B**).

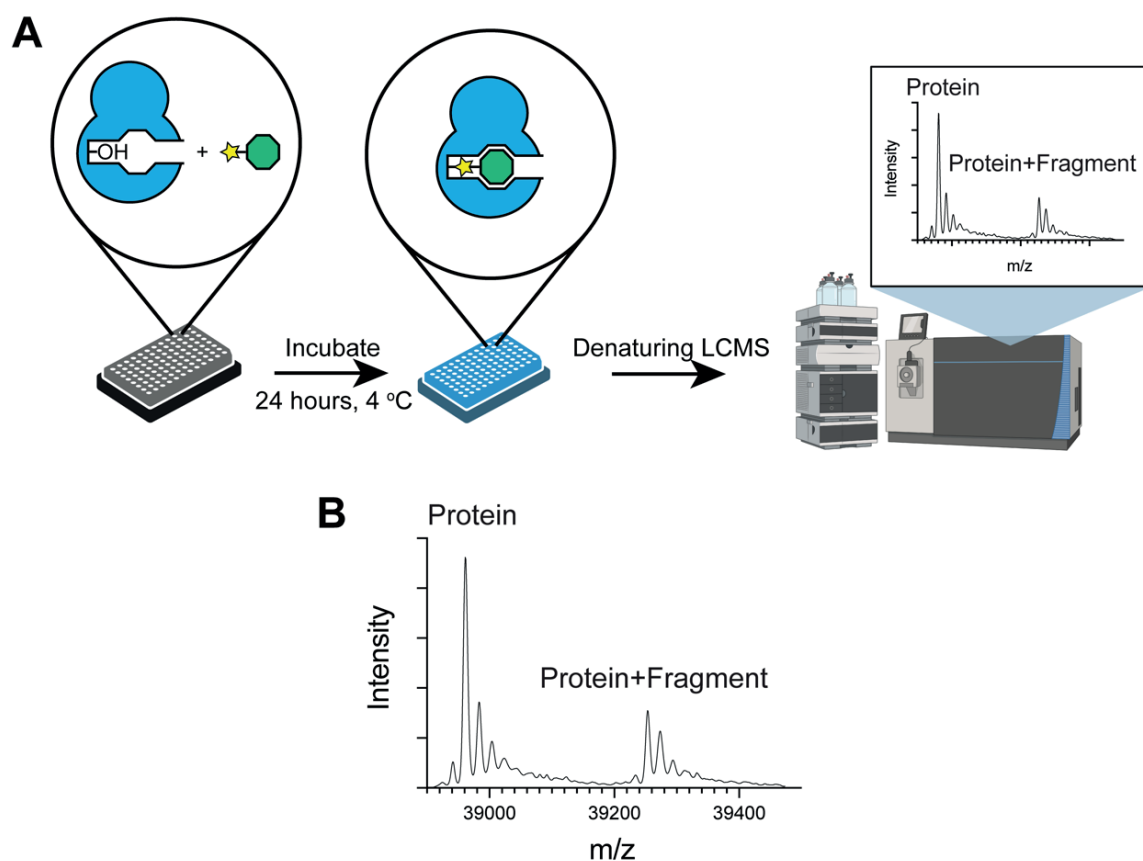


Figure 48: A) Intact protein LCMS assay.

Protein is incubated with fragments and the % protein labelling determined by LCMS. B) Example deconvoluted protein LCMS spectra showing protein only peaks and protein+fragment peaks.

Screening of the library yielded 24 hit compounds (6.9% hit rate) that showed significant (>15%) single labelling of GALE (**Figure 49A and C**). Fragments that gave multiple protein labelling peaks were excluded. These compounds displayed peaks with a molecular weight of the protein plus two or more fragment molecules indicative of high intrinsic reactivity and therefore lacking protein specificity. The structures of these compounds and their labelling data are shown in **Figure 49B**. Of note, no fragment hit exceeded 32% labelling of the protein.

4.2.1.2 Hydroxyquinoline Hits

From these results, investigation of the structure activity relationship (SAR) found several common functional groups. **GSK01** and **GSK08** are both hydroxyquinolines, the higher labelling of **GSK01** (21%) vs **GSK08** (16%), however, is unexpected. The reactivity of aryl sulfonyl fluorides is governed by factors that affect the electrophilicity of the sulfonyl moiety.¹⁸⁰ *Meta*-substituted electron withdrawing groups (such as quinolines) are less reactive than their *para*-analogues.¹⁸⁰ Therefore, assuming protein modification driven purely by electrophile reactivity, **GSK08** would be expected to have increased protein labelling. Therefore, the higher labelling of **GSK01** (21%) would suggest some degree of specific interaction driven labelling by this fragment.

4.2.1.3 Nitroaryl Sulfonyl Fluoride Hits

Other common functional groups were also observed; **GSK03**, **GSK04** and **GSK12** are all nitroaryl sulfonyl fluorides. However, this class of sulfonyl fluorides poses a high degree of intrinsic reactivity due to the strong electron withdrawing effect of the nitro group.¹⁸⁰ Of note, **GSK18** contains a methyl-isoxazole, a functional group identified in the XChem fragment hit **X0213**.

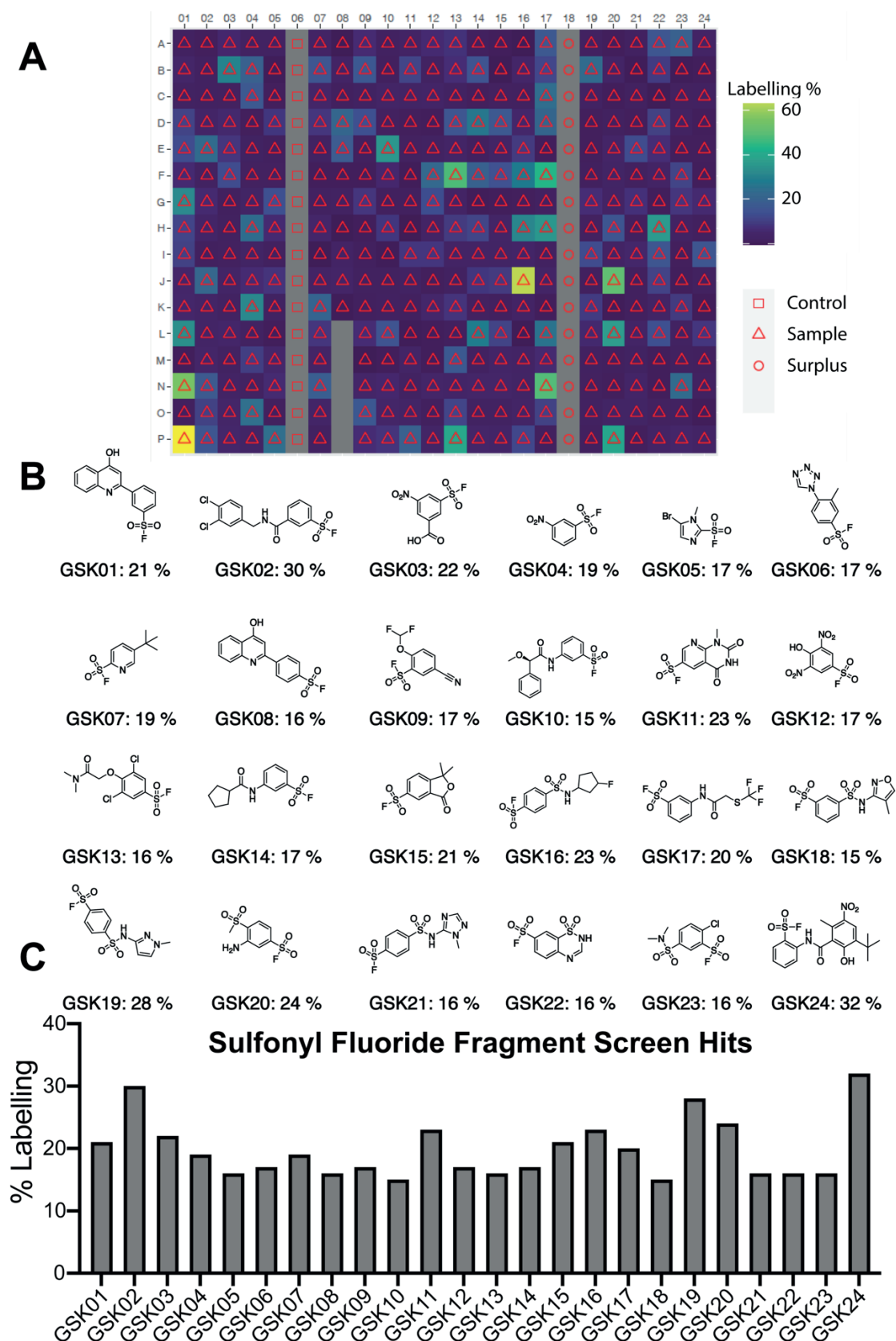


Figure 49: Covalent fragment screening results.

A) Heat map showing % protein labelling results from intact-LCMS screening of an aryl sulfonyl fluoride library. B) structures of the hits identified from LCMS screening with ID and % labelling. C) Plot showing % protein Labelling of covalent hits.

4.2.1.4 Covalent Fragment Hit Virtual Docking

To help identify hits that had the highest probability of labelling Tyr211, the compounds **GSK01-GSK32** were docked using ICM, the docking software described in Chapter 3. It should be noted that when docking covalent fragments with ICM, the side chain being labelled is first defined, and the covalent bond made between ligand and side chain before fitting of the reversible binding portion of the fragment. This order of steps contrasts with how covalent inhibitors/ligands should bind, first with reversible binding and then covalent modification. However, this docking approach still provides useful information to assess compounds, since the best binding pose when covalently bound to the protein is likely to be the binding pose required for formation of the covalent bond. If ligands provide poor docking scores when covalently bound, it is unlikely that they label this residue, because this pose is needed to form the covalent bond. From crystal structures obtained for the non-covalent fragments discussed in Chapter 3, Tyr211 has some flexibility. Therefore, for the docking with covalent fragments, the Tyr side chain was treated as flexible to enable the evaluation of a wider range of conformations. A limitation of the docking process is that binding of ligands or substrates is known to affect the degree to which the binding site is “open” or “closed” (crystal structure 1EK6 vs 1EK5, **Figure 50A-B**). However, the software treats the protein as an inflexible receptor, which is untrue of GALE in solution. Regardless, docking is a useful tool for assessing hits, as seen for the non-covalent fragment binders in Chapter 3.

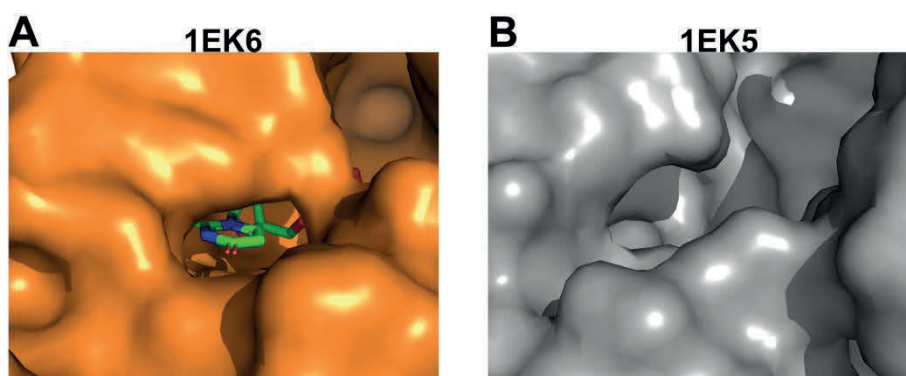


Figure 50: Change in GALE binding site structure upon substrate binding.

A) 1EK6 PDB, GALE in complex with UDP-GlcNAc (green) with modelled surface (orange). B) 1EK5 PDB, GALE in apo structure with modelled surface (grey)

The scores from the docking process are shown in **Figure 51A-B**. From these results, **GSK01**, **GSK02**, **GSK08**, and **GSK21** had the best scores (more negative = better binding). The docking score was then plotted against protein labelling to identify compounds that not only displayed the best labelling (from intact-LCMS), but also have a higher likelihood of labelling Tyr211 (from docking score). The standout fragments from this plot were **GSK01**, **GSK02**, **GSK08**, **GSK21**, and **GSK24**, with **GSK03** and **GSK04** giving reasonable values for both score and labelling. From these compounds, four fragments were selected for further validation, **GSK01-04**. The docking poses for these fragments can be seen in **Figure 51C-F**. **GSK02**, **03**, and **04** are all predicted to bind in a similar pose, with the aryl ring of the warhead in each aligning in the same place. The dichlorophenyl ring of **GSK02** fills the sugar moiety binding site (**Figure 51D**). Interestingly, **GSK01** is predicted to bind in the pocket of the protein that binds the UDP motif (**Figure 51C**). These fragments were selected for further investigation based on their labelling behaviour, docking score, the existence of structurally related hits suggesting structure-based binding (**GSK03** and **GSK04**), and the availability of solid compound stocks at GSK (a limiting factor for **GSK08** and **GSK21**).

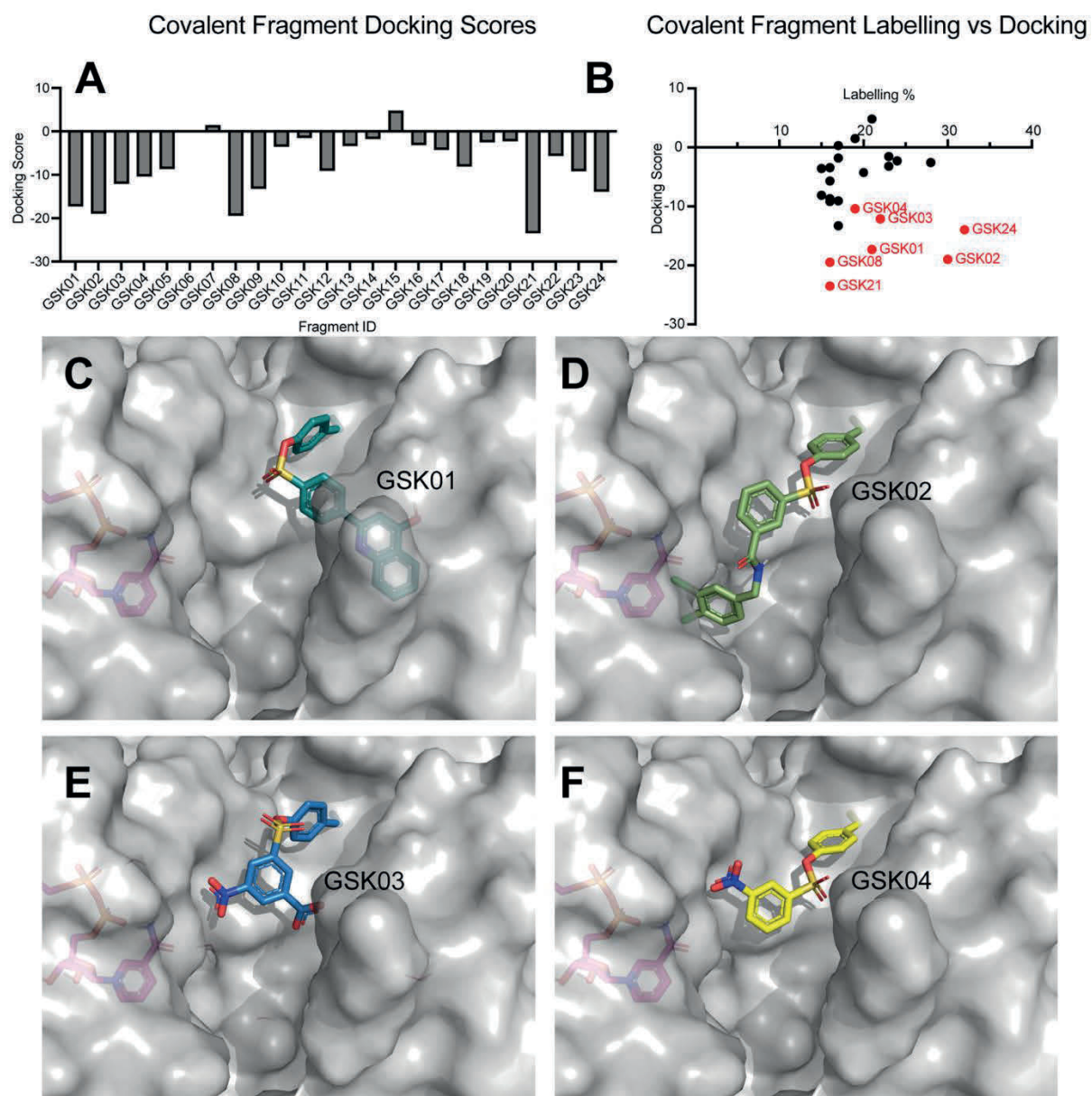


Figure 51: Covalent fragment hit docking results and structures.

A) Plot of covalent docking scores for the intact LCMS hits. B) Plot of docking score vs % labelling. Notable hits shown in red. C-F) Docking poses of hit **GSK01**, O2, O3, and O4. NAD shown in purple.

4.2.2 GSK Covalent Fragment Hit Validation

4.2.2.1 Concentration Series

As initial validation of the hit compounds **GSK01-GSK04** (**Figure 52A**), a labelling experiment was performed with a fragment concentration series to rule out fragments that do not display dose-dependent protein modification (suggesting non-specific, high reactivity compounds) and to assess concentration dependent labelling (**Figure 52 B**). The fragments were incubated for 24 hours at 4 °C at fragment concentrations between 6 µM and 100 µM and analysed by intact-LCMS. **GSK01** displayed moderate concentration dependent labelling from 6 to 12 µM but then plateaued from 12 to 100 µM, not exceeding 18% protein modification. Similarly, **GSK02** gave some concentration-dependent labelling from 6 to 50 µM, with an apparently anomalous result at 25 µM. Like **GSK01**, the labelling plateaued around 21%. **GSK03** had the most obvious concentration dependent labelling of GALE, however, it too plateaued above 50 µM at 38%. This fragment also displayed significant peaks corresponding to double protein labelling, above 50 µM. **GSK04** had the poorest results with no clear labelling below 100 µM.

From this experiment, **GSK03** and **GSK04** produced unreliable data, with either little to no labelling or significant secondary labelling of the protein. These results suggest poor specificity and indicate that labelling is driven by intrinsic high reactivity rather than the desired non-covalent reversible binding, followed by covalent modification. **GSK01** and **GSK02** produced more consistent data with single labelling. However, both fragments also plateau at around 15-25% protein labelling, even at high concentrations (100 µM). This is potentially indicative of non-specific binding, where increasing fragment concentration should result in increased protein labelling for fragments where protein modification is driven by specific interactions.

Alternatively, as has been noted previously (Chapter 2), the recombinant GALE expressed has a UDP-Sugar binding site that is approximately 80% occupied by UDP-Glc (abortive complex). It is possible that the plateauing observed for fragments **GSK01** and **GSK02** is caused by these fragments labelling Tyr211, a residue close to the UDP binding pocket, but that only ~20% is

available for labelling due to the presence of the abortive complex. Regardless, the obtained data from this validation assay were questionable (and from singlets), warranting further validation with a more thorough assay.

4.2.2.2 Kinetics Assay

To probe the reactivities of **GSK01** and **GSK02** in more detail, a kinetics assay was performed. Covalent fragments are dependent on two kinetics parameters, the K_D describing the reversible binding of the fragment to the protein and the k_{inact} describing the efficiency of covalent modification. Covalent fragment labelling efficiency thus has a time-dependent factor that needs to be assessed with a time course assay. In this experiment, fragments were incubated with GALE at a range of fragment concentrations and labelling analysed by intact-LCMS at multiple time points. The full details of this experiment can be found in the methods section 4.4.4 but in brief, fragments were incubated with GALE from 12.5 μM to 200 μM and the labelling determined at seven time points from 0 – 380 minutes (6.5 hours). The results of this experiment are shown in **Figure 52C-D**.

Both **GSK01** and **GSK02** display a clear time-dependent increase in protein labelling as expected. However, the trend in labelling does not reveal a clear dependence on fragment concentration as would be expected for residue-specific fragments. Like the earlier validation assay, **GSK02** gave slightly higher average labelling than **GSK01** (25.6 vs 21.8% respectively). Both fragments also see their labelling plateau at around 20-25% as had been observed previously. This may provide further validation to the hypothesis that the fragments are labelling Tyr211, but modification above ~25% is prevented by the abundance of the abortive complex. Alternatively, the plateauing observed could be due to degradation of the sulfonyl fluoride warhead through hydrolysis.¹⁸⁰ The data obtained thus far for **GSK01** and **GSK02** were inconclusive, and it was not clear if the obtained hits would be useful for follow-up and fragment elaboration. Repeats of the kinetics experiments may have provided more information as well as analysis of some of the other hit fragments.

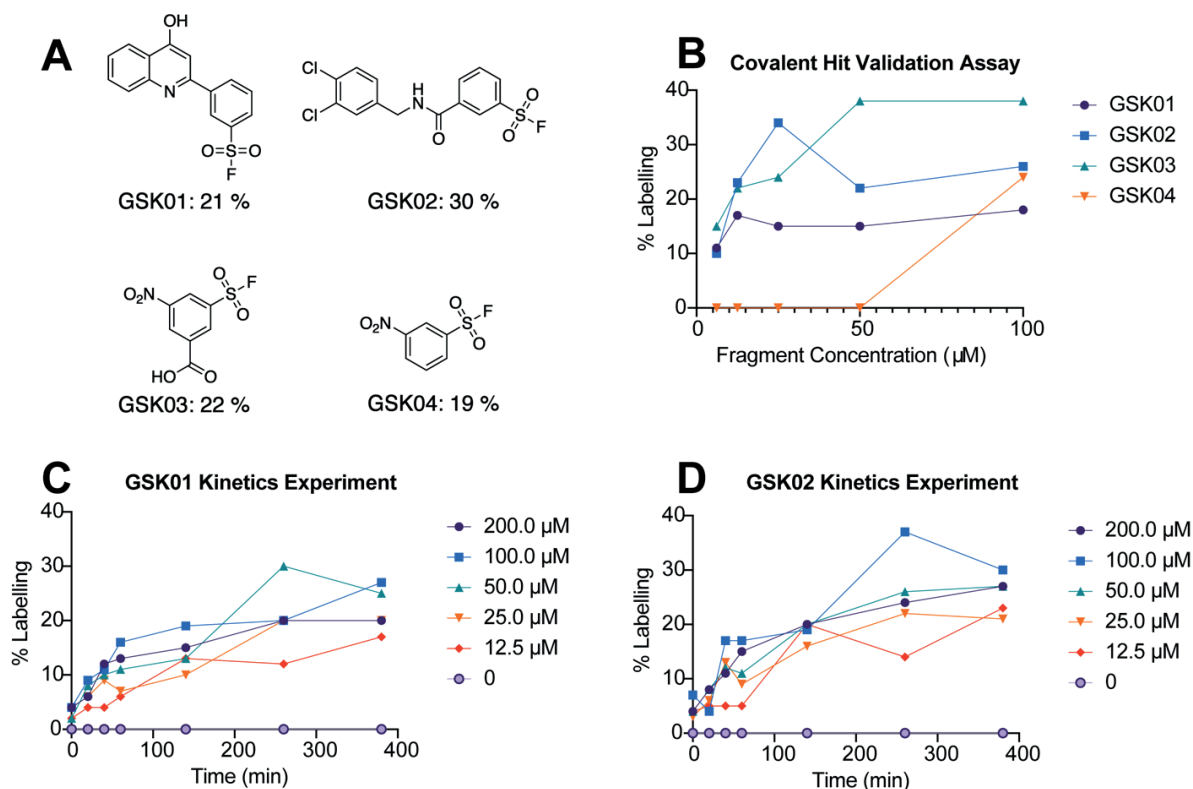


Figure 52: Covalent fragment hit validation and kinetics assay results.

A) Structures of hits **GSK01**, 02, 03, and 04 with % protein labelling from the initial screen. B) hit validation assay results. C-D) kinetics assay results for **GSK01** and **GSK02**. (All experiments run in singlet)

Instead, hits **GSK01** or **GSK02** were next evaluated for the effect their labelling had on enzymatic activity. This would provide the clearest evidence to support the usefulness of these compounds for further development and justify necessary repeats. To assess the effects of **GSK01** and **GSK02** on activity, the first experiment involved incubating GALE with either 100 or 50 μM fragment for 24 hours. A HPLC activity assay was then run with these samples with GALE at varying concentrations. The results of this experiment are seen in **Figure 53A-B**. At GALE concentrations above 25 nM, all reactions (fragments and controls) reached completion (75% is the point of equilibrium). At 3 nM to 6 nM GALE, the inhibitory effect of each fragment on activity can be seen. **GSK01** has little to no effect on the activity of GALE (<10% inhibition) at both 50 and 100 μM , in contrast **GSK02** reduces the enzymatic activity by ~50%. This can be seen most clearly in **Figure 53B** where the activity of each fragment is displayed as a % of the DMSO controls. From this data it is difficult to relate this to the labelling observed previously. If the 20-30% labelling observed at these concentrations reflects 100% labelling of the active protein (non-abortive complex) GALE in solution, then the inhibition of the enzyme

should be closer to 100% if targeting this pocket. However, it should also be noted that it is not guaranteed that complete modification of Tyr211 from the active GALE proportion would result in complete loss of activity. Alternatively, the ligands may be labelling a Tyr or other nucleophilic amino side chain in an allosteric site. Regardless, these results suggested that **GSK02** had a 2-fold inhibitory effect on the enzymatic activity of GALE compared to a protein DMSO only control . Further characterisation of inhibition would require a kinetics experiment similar to the experiment in **Figure 52C** and **D**, except with an activity assay performed at each time point in addition to intact-LCMS. This would provide evidence that the inhibition and labelling correlate and are directly related. However, this would have been a technically challenging experiment and at this stage it was still uncertain how useful **GSK02** would be for follow up. First confirming that the ligands modify a residue in a defined pocket amenable to elaboration was of higher priority.

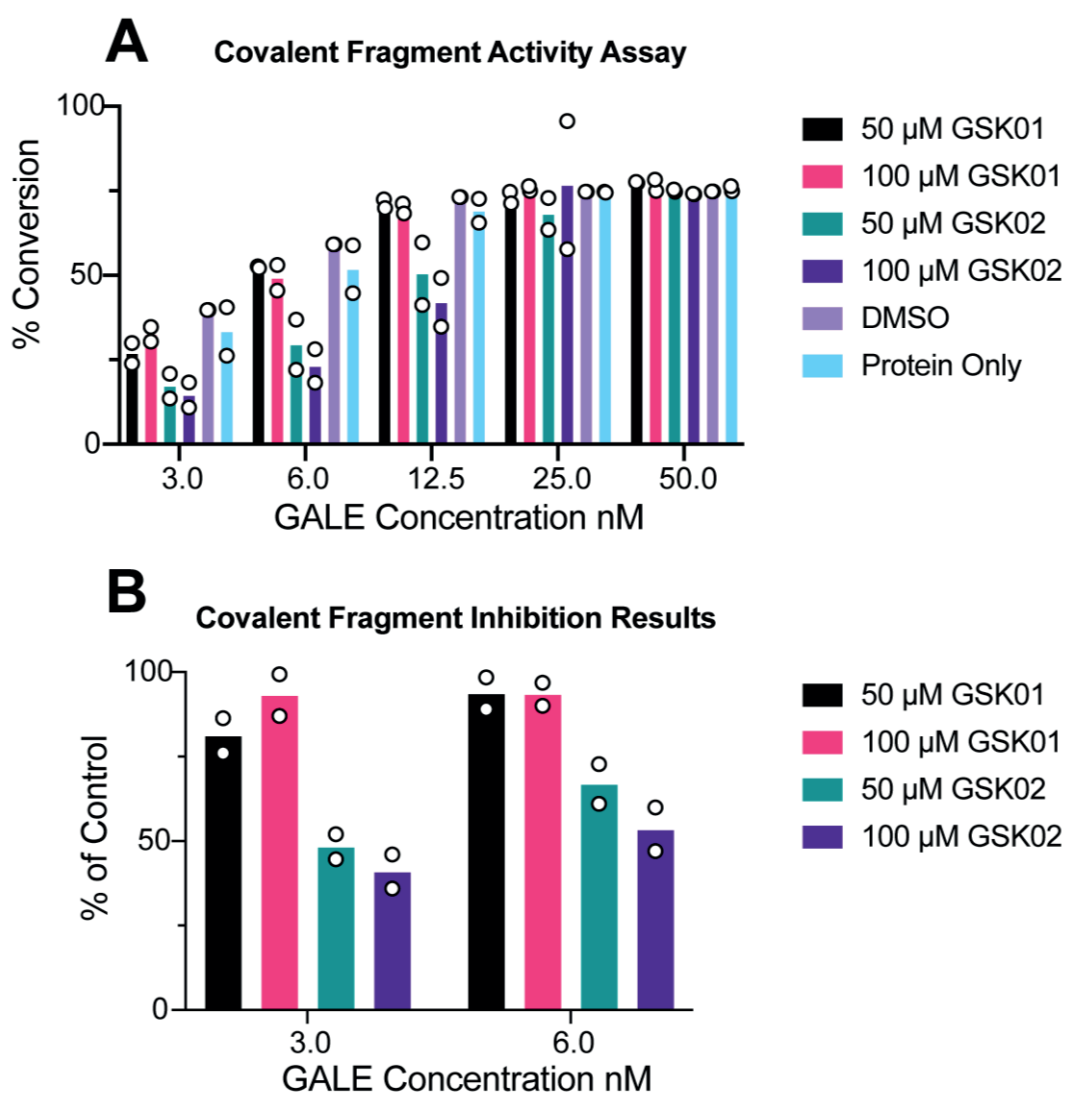


Figure 53: GSK Covalent fragment hit inhibition assays.

A) % Conversion for different concentrations of GALE reacted with covalent fragments plus DMSO and protein only controls.

B) Activity assay results for 3 and 6 nM GALE concentrations reacted with covalent fragments shown as a% of a DMSO control. (Plots show the results from experiments performed in independent duplicate)

4.2.3 GSK Covalent Fragment Modification Site ID Experiments

Much of the decision of whether to follow up and elaborate **GSK01** or **GSK02** is dependent on the site that it is labelling. No data clearly indicates if the fragment labels a residue in the substrate binding site, a Tyr, or any other amino acid on the protein. This information is essential since labelling of a region of the protein with no well-defined pocket would render elaboration into a potent inhibitor difficult. There are several methods which can be used to

confirm which residue is primarily being labelled in GALE. The most useful is direct structural information of the protein in complex with the fragment after reaction with the residue being modified. A structure also provides information on the interactions made between fragment and protein that could be used for structure guided design. Alternatively, mutation of the suspected residue being labelled to a phenylalanine could also be used, since this would result in a loss of protein modification in an intact-LCMS assay due to the removal of the nucleophilic residue. The problem with this method, however, is that GALE contains sixteen Tyr residues and would require significant effort to sequentially mutate and express the mutant proteins for assaying. Mutagenesis would, however, be suitable to rule out Tyr211 specifically. The final option is tryptic digest LCMS/MS (herein just LCMS/MS) experiments. Here, the protein is first labelled, and the protein digested with trypsin to yield a peptide preparation. LCMS/MS analysis of these peptides can then be used to identify the labelled residue (**Figure 54A**). The advantage of this technique is that it is relatively easy to perform, and the results are obtained quickly. Dr. Steven Howell (a collaborator at the Francis Crick Institute) has experience in these experiments and thus the LCMS method was attempted first. It should be noted here that protein digestion and data collection for the intact-LCMS experiments described below were performed by Dr. Steven Howell.

For this experiment, both **GSK01** and **GSK02** were reacted with GALE at either 25, 50 or 100 μ M (plus a 1% DMSO control) for 24 hours at 4 °C. Tryptic digest LCMS/MS analysis (detailed in 4.4.6) identified the two Tyr residues (Tyr104 and Tyr211, **Figure 54B and C**) with the correct mass modification for fragment **GSK01**. Tyr104 was the only residue identified with the **GSK02** mass modification. The results of this experiment are displayed in **Figure 54D-F**. These graphs represent the area of the mass peak for each peptide, which has the expected fragment mass modification (plus isotopes, M+1, M+2). The peptides detected plus the modified residue are shown in **Table 18**.

There is considerable variation in the area detected for each peptide, however, given different peptides and modification are being compared, this is expected. Different peptides and modification ionise to different degrees, which directly influences the amount detected (peak intensity). The essential takeaway from this data is that: 1) Only two residues were detected with any modification (Tyr104 and Tyr211). 2) Peptides with this mass modification were not detected in the controls (negligible peptide+**GSK01** peaks were detected in the control for Tyr104, **Figure 54E**).

The identification of two labelling sites (when in the intact-LCMS screening only one labelling peak is observed) was not unexpected. Mass spectrometry is a highly sensitive technique. As such, even at very low percentages of labelled protein, modified peptides are still observed in the spectra. Since each peptide is different, the intensities of two different peptides cannot be compared for quantification. Additionally, comparison of the unlabelled and labelled peptides cannot be compared because the modification also affects detection. Therefore, from the LCMS/MS data obtained, it can only be stated that both Tyr211 and Tyr104 are modified by **GSK01**, and only Tyr104 was labelled by **GSK02**. Lack of labelling on Tyr211 by **GSK02** was unexpected since this compound displayed inhibition of the enzyme and Tyr211 is much closer to the substrate binding site than Tyr104. Tyr104 may be an allosteric site.

Peptide No.	Fragment/Residue	Peptide (Charge)	m/z
1	GSK01 Tyr211 Peptide	YFNPTGAHASG <u>C</u> IGEDPQGIPNNLMP <u>Y</u> VSQVAIGR (+++)	1338.6
2	GSK01 Tyr104 Peptide	AVGESVQKPLD <u>Y</u> YR (++)	954.4
3	GSK02 Tyr104 Peptide	AVGESVQKPLD <u>Y</u> YR (++)	983.4

Table 18: GSK covalent fragment site ID experiment fragment modified peptides.

Table showing peptides identified with the correct fragment mass modification from tryptic digest LCMS/MS experiment.

The charge of each peptide is also shown as well as the m/z of that peptide. Modified Tyr shown in bold and underlined.

Tyr211 peptide also contains a Cys modified by acetamide (part of sample preparation).

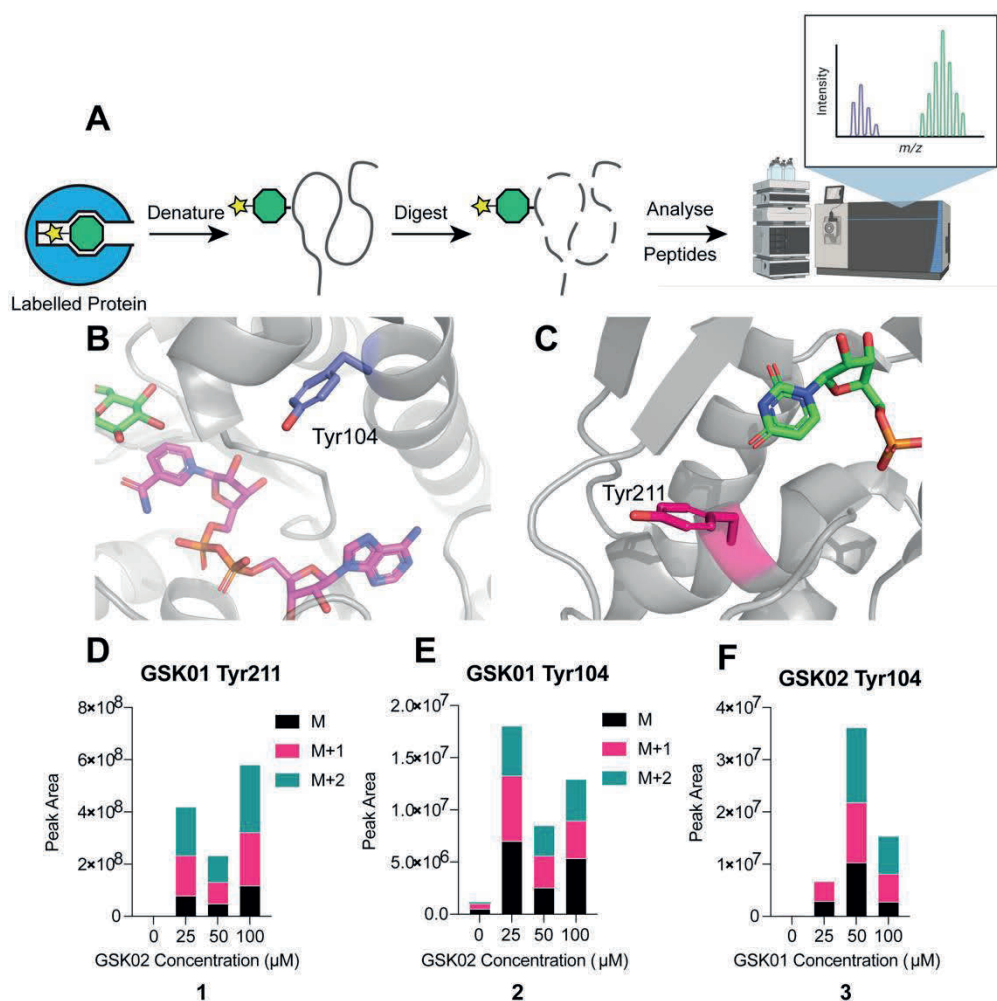


Figure 54: GSK covalent fragment protein modification site ID experiment results.

A) Diagram showing the process of tryptic digest LCMS/MS. Protein is first denatured, digested by trypsin, and finally peptides analysed by mass spectrometry. B) Crystal structure showing Tyr104 (purple) and NAD (pink). C) Crystal structure showing Tyr211 (pink) and UDP-Glc (green). D-E) Graphs displaying peak areas of fragment modified peptides 1-3 detected by mass spectrometry. M is parent ion, the most common isotope. Whilst M+1 and M+2 are the parent ion one 1 and 2 mass units heavier, respectively.

Regardless, due to the uncertainty of these results, further characterisation would be needed to confirm these results likely by mutagenesis or crystallography. There is also no well-defined pocket around Tyr104 making **GSK02** unattractive for elaboration and analogue design.

4.2.4 XChem Covalent Fragment Analogues

Earlier in this chapter, Tyr211 was highlighted as being close to the UDP-Sugar binding site of GALE. Fortunately, Tyr211 is also adjacent to the binding site of the crystallographic hit

compounds **X0213** and **X0121** as well as their analogues (**Figure 55A and B**). Within the analogues of **X0213**, a crystal structure was also obtained of **WBX2** (**Figure 55C**) in complex with GALE (**Figure 55D**). The structure shows two alternate conformations of **WBX2**, with the phenol ring either forming a π -stacking interaction with Tyr211 or situated adjacent to Tyr211. These two alternative confirmations are not unexpected since the addition of the phenol ring did not improve binding over the parent compound **X0213** (both at ~400-500 μ M as assessed by ITC, see Chapter 3).

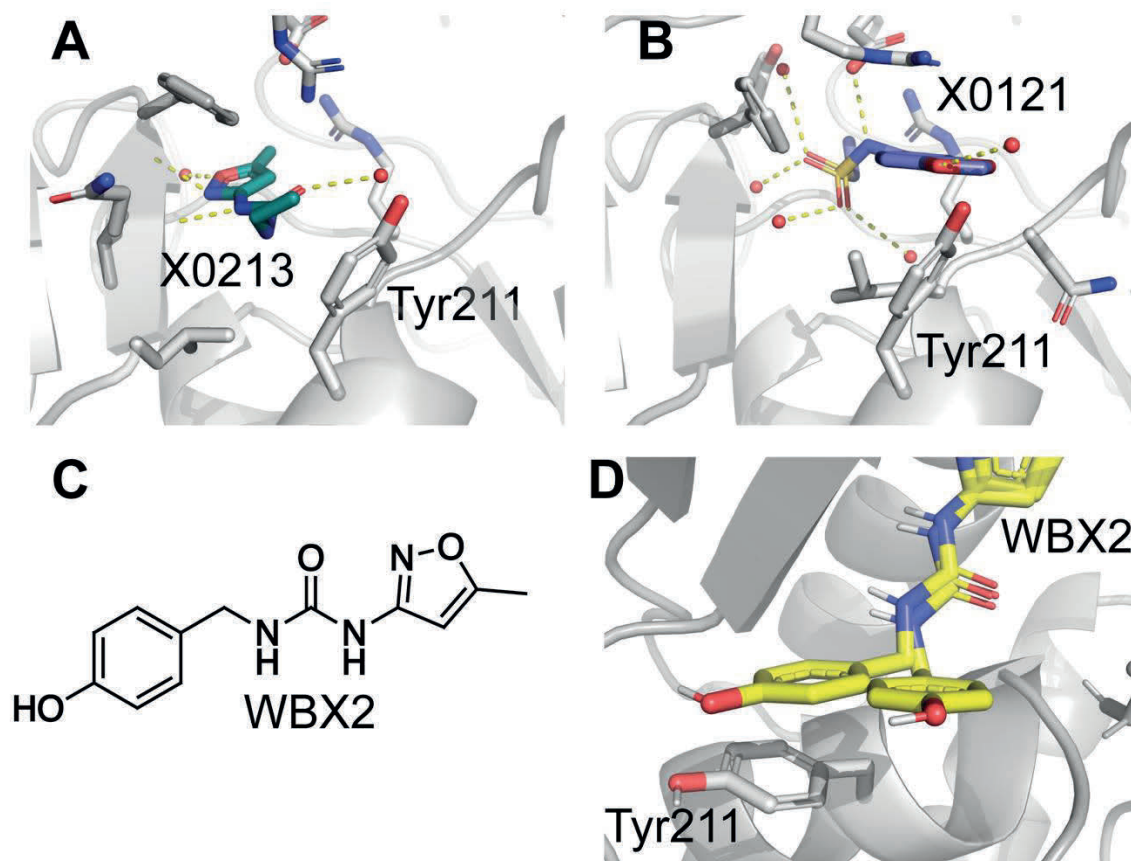


Figure 55: Non-covalent to covalent fragment candidates and structures.

A) Crystal structure of XChem hit **X0213** (cyan) in complex with GALE with Tyr211 labelled. B) As shown in A but with XChem hit **X0121** (purple). C) Structure of non-covalent fragment **WBX2** (yellow). D) As in panel A except with **WBX2**.

Thus far, an “electrophile first” approach has been adopted to identify covalent inhibitors of GALE. However, with the validation of Tyr211 as ligandable, a “ligand first” approach was an exciting possible option. Additionally, the tolerance of **X0213** to the addition of a substituted aryl group had already been confirmed with the non-covalent analogues WBX1-WBX8 which displayed similar thermal shifts, inhibition and/or K_D values to the parent compound **X0213** (Chapter 3, **Table 12**). The aryl ring flexibility displayed in the crystal structure of compound **WBX2** may also prove beneficial, since specific geometries are required for nucleophilic attack

on the sulfonyl fluoride. The geometry required for nucleophile trajectory may be aided by some flexibility of the aryl group. It should be noted that **X0121**, whilst also adjacent to Tyr211, was avoided for modification to a covalent fragment due to the synthetic challenges posed by possible analogues.

4.2.5 XChem Covalent Analogue Design and Docking

To first assess if the binding pocket of GALE would accommodate the addition of an aryl sulfonyl fluoride on analogues of **X0213**, several analogues (**Figure 56A**) were designed and docked into the crystal structures of the parent compound (i.e. analogues of **X0213** docked into **X0213** structure) using ICM Pro. When designing these analogues, the aim was to change as little of the initial hit (**X0213**) as possible, whilst still incorporating the aryl sulfonyl fluoride warhead. Additionally, ligands were designed to incorporate different linker lengths between the warhead and the urea moiety. Finally, the sulfonyl fluoride moiety was placed on either meta- or para-positions to assess how this may change the binding modes. However, it should be noted that the docking method of ICM requiring a pre-formed covalent bond, the docking algorithm did not consider the differences in electrophilicities of the *meta*- and *para*-warheads.

From the docking scores shown in **Figure 56B**, ligands **WB-C3** and **WB-C4** have the best predicted binding scores (-28.7 and -29.4 respectively). Comparison of the crystal structure for **X0213** (**Figure 56C**) to the docked ligands **WB-C3** and **WB-C4** (**Figure 56D-G**) revealed that these are predicted to bind the fragment portion of the molecules in the exact same position as the parent compound **X0213**. Given that docking was performed into the **X0213** structure, it is unsurprising that the compound fragments that mimic this interaction have particularly good docking scores. Interestingly, for these fragments, the flexibility of Tyr211 appears to enable each to bind in similar poses, despite the structural differences of the *meta*- and *para*-substitutions. **WB-C5** and **WB-C6** (**Figure 56A**, grey box) have slightly lower scores due to the lack of the H-bonding nitrogen in the urea moiety, but displayed the same binding position as **WB-C3** and **-C4** (not included in the **Figure 56**). **WB-C1** and **WB-C2** have the poorest scores - this is because the docking of these two fragments is not predicted to fit into the **X0213** binding site. Instead, fragments **WB-C1** and **WB-C2** fill a groove adjacent to where the other fragments bind (**Figure 56C**).

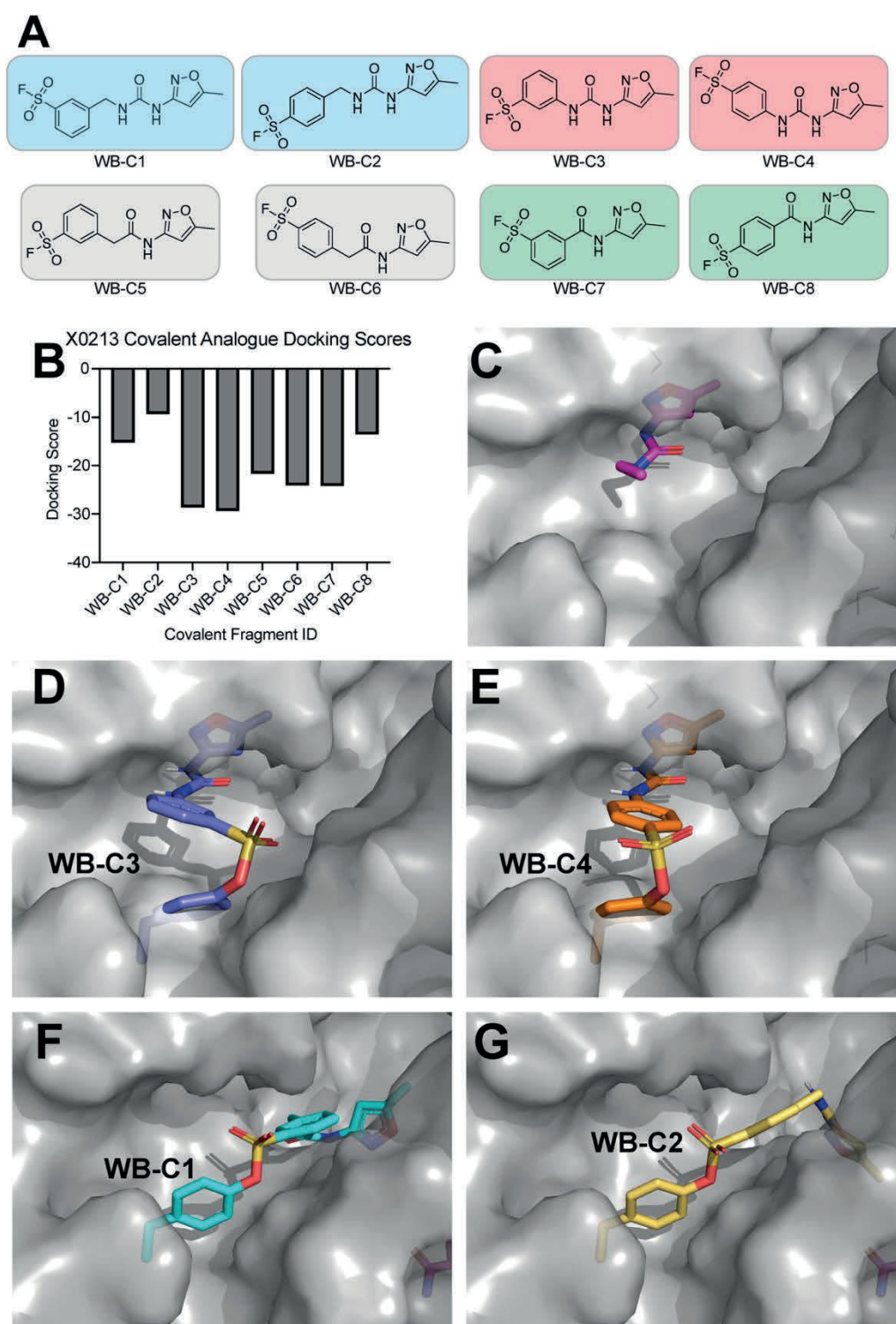


Figure S6: Covalent compounds **WB-C1** – **WB-C8** structures and virtual docking results.

A) Structures of the covalent analogues of **X0213**. Para/meta- analogue pairs are displayed in the same box colour. B) Docking scores for covalent analogues from docking into **X0213** crystal structure. C) crystal structure of **X0213** with modelled protein surface. D-E) Docking pose adopted by **WB-C3** and **WB-C4** (purple and orange respectively). F-G) Docking pose adopted by **WB-C1** and **WB-C2** (blue and yellow respectively). It should be noted that the aromaticity was removed from the Tyr211 phenol ring when transferring from ICM to Pymol, but this did not affect docking.

4.2.6 XChem Covalent Analogue Intact-LCMS Screening and SAR

Each fragment had reasonable docking scores and predicted binding poses. All the docked compounds were ordered from custom synthesis provider Enamine, excluding **WB-C2** and **WB-C3** due to cost limitations. Each of these compounds was then assayed by intact-LCMS as described for the GSK covalent fragment screening, with 100 μ M fragment for 24 hours at 4 °C. The deconvoluted LCMS spectra for each of these compounds is shown in **Figure 57A-F** and summarised in **Figure 57G**. From these data, **WB-C1** clearly gives the highest protein modification, with 100% protein labelled under the assay conditions. No peak for non-modified protein was observed. The fragment **WB-C6** displayed no labelling of the protein.

4.2.6.1 *WB-C7 and WB-C8*

From the data interesting structure activity relationship (SAR) factors can be observed. The most reactive compounds are **WB-C7/8**, as would be expected as the only compounds with electron withdrawing amides substituted to the aryl sulfonyl fluoride. However, with comparison of *meta*- and *para*- ligand pairs **WB-C7/8** similar protein labelling was observed (both give multiple protein modification, approximately 35%) contrary to what would be expected based on warhead reactivity. Based on this the *para*- substituted amide **WB-C8** is predicted to be more reactive, since aryl sulfonyl fluorides are activated by electron withdrawing effects.

4.2.6.2 *WB-C5 and WB-C6*

Based on reactivity, **WB-C5/6** should produce similar labelling, however **WB-C5** labels 32% of the protein whilst **WB-C6** shows no labelling. This clearly indicates that for these two compounds the conformations adopted by the *meta*- substituted compound (**WB-C5**) is significantly preferred, which was not predicted in the docking software (Both having almost identical docking scores).

4.2.6.2 WB-C4 and WB-C6

Comparison of **WB-C4** and **WB-C6** highlights the importance of the nitrogen forming the urea in **WB-C4**, since absence of this nitrogen completely abrogates labelling (**WB-C6**). The loss in labelling was not unexpected since both nitrogen atoms in the urea of **X0213** make key hydrogen bonds to the protein (discussed in detail in Chapter 3, **Figure 27**).

4.2.6.2 WB-C1, WB-C2 and WB-C3

Complete SAR to assess the effect of linker length and sulfonyl fluoride position was not possible due to not having fragments **WB-C2** and **WB-C3** (cost limitation). Comparison of **WB-C1/C3** (both with a *meta*-substituted aryl ring) would confirm the benefits of having the extra carbon between urea and aryl sulfonyl fluoride, whilst comparison of **WB-C1/C2** (structural isomers) would have provided insight into whether *meta*- or *para*- is preferred with the additional carbon between warhead and urea.

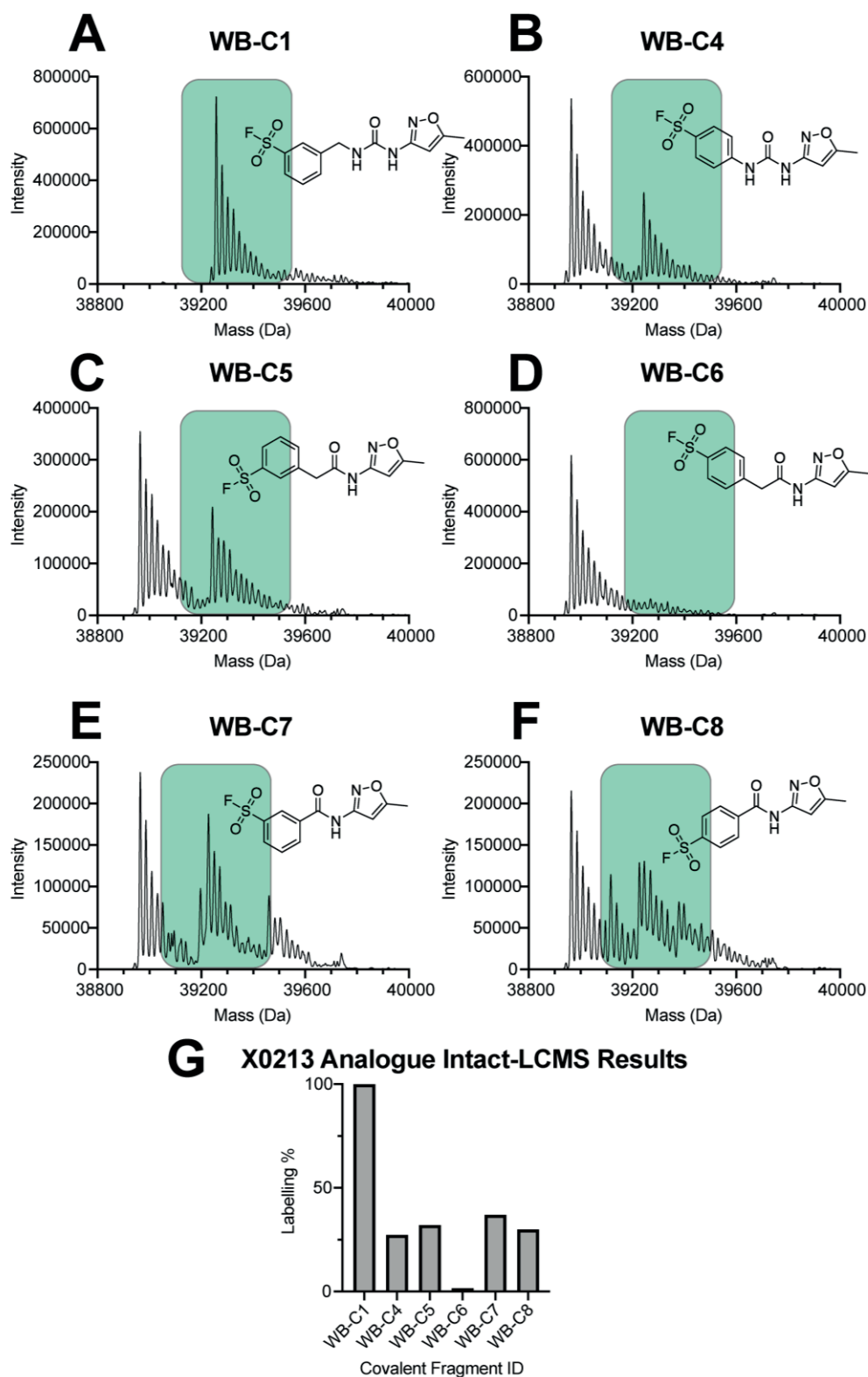


Figure 57: Covalent fragment intact-LCMS results.

A-F) Intact LCMS traces showing for GALE reacted with covalent fragments. Protein+Fragment peaks are highlighted with green boxes. G) Summary of the protein labelling for each fragment.

4.2.7 XChem Covalent Analogue Docking Evaluation

Comparison of the docking scores and the labelling revealed no trend (**Figure 58A**). This highlights some of the potential limitations with docking into crystal structures. The inflexible receptor is unable to account for conformational changes that may benefit one fragment over another. To assess the extent to which such conformational changes may have impacted docking scores, the fragments **WB-C1** to **C8** were docked into the crystal structure obtained for non-covalent hit **WBX2** (herein the alternate structure).

4.2.7.1 Docking into WBX2 Crystal Structure (Alternate Structure)

In contrast to fragment **X0213**, **WBX2** is structurally identical to **WB-C1**, except with a para-substituted hydroxyl group in place of the meta-substituted sulfonyl fluoride (**Figure 59B**). The slight variation in the two crystal structures may produce docking results that correlate better with the assay data and validate this docking approach for future analogue design.

The results for docking into the alternate GALE structure are shown in **Figure 58C**. The most significant difference between the two docking attempts is seen for fragment **WB-C1**, where the score changed from -15.3 to -27.4, a significant improvement. In contrast, the *para*-analogue of this compound (**WB-C2**) produced a similar score to that observed previously (-13.2). Notably, **WB-C1**'s improved score is a result of the docking now placing the fragment into the same pocket as the analogue **WBX2** (**Figure 58D**).

WB-C3 and **WB-C4** gave the best predicted binding as seen for the docking performed into the structure from fragment **X0213**. However, no correlation to the labelling data was observed (**Figure 58E**). These results highlight the limitations of covalent docking for predicting modifications that alter the linker length between fragment and warhead, and the limitations posed by not being able to consider changes in warhead reactivity caused by different functional groups on the warhead. If docking is to be used for future fragment elaboration more sophisticated software that considers receptor flexibility will be required.

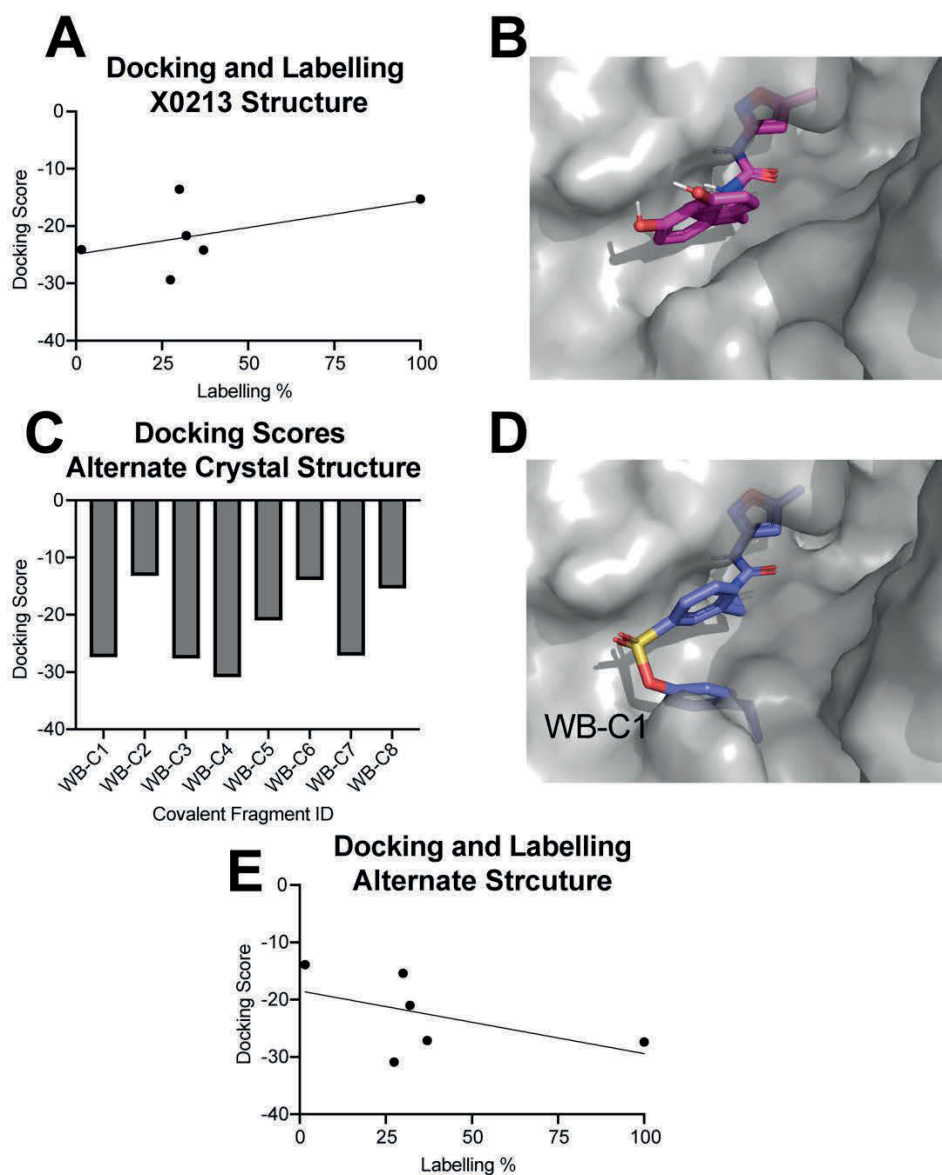


Figure 58: Covalent fragment docking evaluation.

A) Plot of virtual docking scores into the structure obtained for **X0213** vs intact LCMS assay protein labelling%. B) Crystal structure of **WBX2** (pink) with modelled protein surface (grey). C) Docking scores for fragments docked into the crystal structure obtained for **WBX2**. D) Docking pose predicted for **WB-C1**(violet). E) Plot of docking score from **WBX2** structure vs labelling%.

4.2.8 Effect of WB-C1 on GALE Activity

WB-C1 shows complete labelling of the protein, with no detectable unmodified protein. Based on ligand design, this compound is also structurally most similar to the non-covalent analogue **WBX2**. It is therefore extremely likely that **WB-C1** was labelling Tyr211. Therefore, the next question was whether labelling has any inhibitory effect on GALE activity.

An activity assay was performed to assess the inhibitory properties of the most promising covalent compound **WB-C1**. At 5 nM GALE concentration, a 90% inhibition was observed for GALE reacted with 100 μ M **WB-C1** compared to a control reaction containing DMSO (**Figure 59A**). Reaction of GALE with 25 μ M fragment yielded an ~66% decrease in activity. A 10 nM GALE concentration yielded lower inhibitory capacity at both concentrations of inhibitor **WB-C1** (**Figure 59B**)

These data indicate that **WB-C1** is a GALE inhibitor that can be subverted by increasing the concentration of GALE enzyme, in support of a covalent inhibitor. Additionally, 100 μ M **WB-C1** should yield 100% protein modification, but 100 % inhibition was not observed. This is likely because the residue being targeted (Tyr211) is adjacent to the active site, not a residue located directly in the pocket. Therefore, inhibition is still dependent on the strength of reversible non-covalent binding of the fragment portion of the compound.

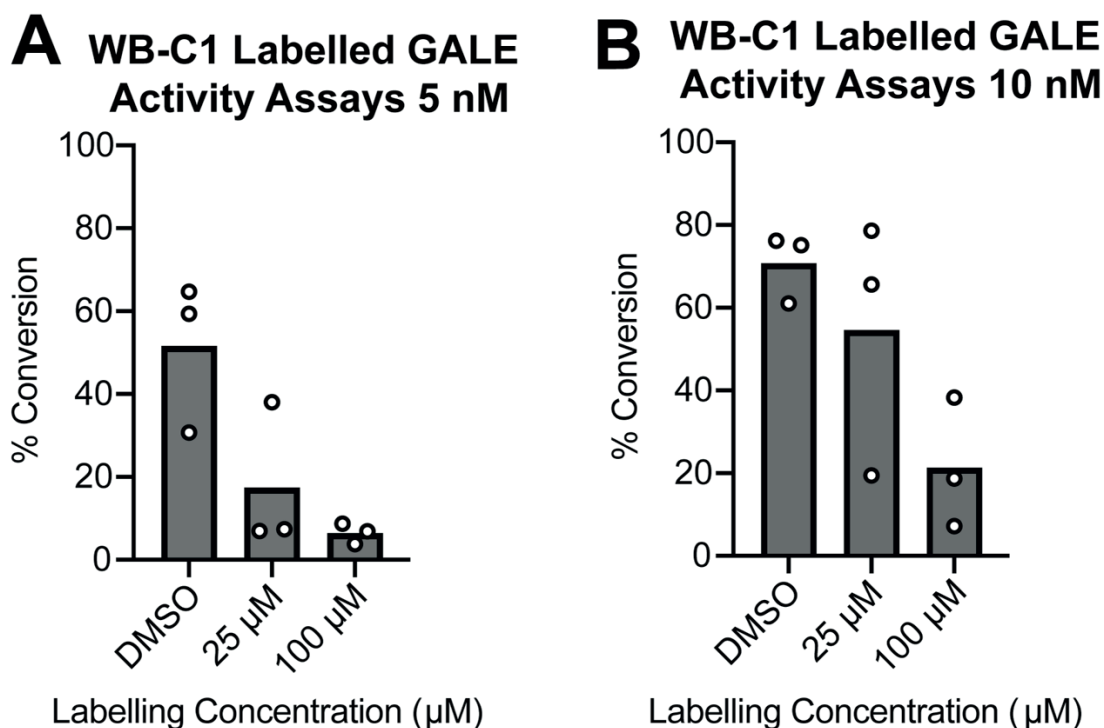


*Figure 59: Covalent fragment **WB-C1** effect on GALE activity.*

*A) Activity assays performed with GALE at 5 nM reacted with **WB-C1** or with a DMSO control. B) Activity assays performed with GALE at 10 nM reacted with **WB-C1** or with a DMSO control.*

4.3 Chapter 4 Summary

In summary, this chapter describes the use of covalent fragment screening by intact-LCMS to identify compounds that covalently modify GALE. Hits were validated using virtual docking, tryptic digest LCMS/MS, kinetic and enzyme activity assays. Of these two compounds (**GSK01** and **GSK02**) showed reproducible protein labelling. Site ID experiments confirmed **GSK01** modified Tyr104 and Tyr211 but had minimal effect on enzyme activity. In contrast, modification by **GSK02** was only observed on Tyr104 but had a 2-fold decrease in enzyme activity from approximately 30% protein labelling. Despite these promising results, the hits were not elaborated in favour of a “ligand first” approach. The screen results validate Tyr211 as tractable and non-covalent fragments have previously been validated by crystallography as binding close to this residue (Chapter 3). Therefore, a series of analogues based on **WBX2** were designed and screened for protein labelling. Fragment **WB-C1** produced 100% protein labelling under the assay conditions and inhibited the enzyme ~10-fold. This result

demonstrates an elegant example of targeted development of a covalent inhibitor from non-covalent hits, using structure guided design against a validated nucleophilic residue.

4.4 Methods

4.4.1 Intact LCMS

Intact-LCMS was performed as described by *A. Aatkar et al* with some modification, the following methods are adapted from this publication.¹⁸¹

5 µL of sample was injected onto an intact protein mass spectrometer with the following set up; an Agilent G230B time-of-flight (ToF) Accurate Mass Series spectrometer, an Agilent Bio-HPLC PLRP-S (1000 Å, 5 µM x 50 mm x 1 mm) reverse phase column, and an Agilent Dual AJS ESI ion source (Agilent Technologies, Cheshire, UK).¹⁸¹

The flow rate and temperature was set at 0.5 mL/min and 70 °C, respectively. The solvent system used was as follows; solvent A (water, 0.2% (v/v) formic acid), solvent B (acetonitrile, 0.2% (v/v) formic acid). The applied gradient is described in Table 1. For data acquisition the positive ion scanning was performed between 600 – 3200 Da with a scan rate of 1.2s. The obtained spectra were analysed and deconvoluted with the MassHunter BioConfirm software (Agilent Technologies, Cheshire, UK). The expected mass range used was 30000 – 50000 Da. The spectra were then exported and analysed using software developed internally at GSK to determine relative% labelling. The equation used to determine labelling is shown in **Equation 4**.

Time/min	Solvent B (%)
0.00	20
0.60	20
0.61	50
1.00	100
1.20	100
1.21	20

Table 19: Solvent B% change with time

$$\% \text{ Labelling} = \frac{\text{Intensity of modified protein}}{\text{Intensity of protein only} + \text{Intensity of modified protein}} \times 100$$

Equation 4: Calculation used to determine% protein labelling from intact-LCMS assays.

4.4.2 Covalent Fragment Screening

In summary, fragments (100 µM Final) were dispensed by ECHO 550 into a Greiner 384 well plate (Sigma Aldrich) followed by the addition of 10 µL GALE (0.5 µM, 25 mM HEPES, pH 7.5, 50 mM NaCl,) and incubation for 24 hours at 4 °C. DMSO only controls were also included. The total reaction volume was 10 µL. Samples were then analysed by intact-LCMS as described above and the relative % labelling determined. Full data output and all deconvoluted LCMS spectra are available on request.

4.4.3 Covalent Fragment Validation

The selected fragments (**GSK01-06**) were first dispensed by ECHO 550 into a Greiner 384 well plate to provide a final concentration of 100, 50, 25, 12.5, and 6.125 µM in 10 µL (1% DMSO). Followed by incubation with 10 µL GALE (0.5 µM, 25 mM HEPES, pH 7.5, 50 mM NaCl,) for 24 hours at 4 °C as previously described (See Chapter X). DMSO controls were also included. Samples were then analysed by intact-LCMS to determine% labelling.

4.4.4 Covalent Fragment Kinetics Assays

Kinetics experiments were performed with fragments **GSK01** and **GSK02**. Fragments were first dispensed by ECHO 550 into a Greiner 384 well plate to provide a final concentration of 200, 100, 50, 25, and 12.5 µM in 10 µL, with 7 replicates per concentration. Followed by incubation with 10 µL GALE (0.5 µM, 25 mM HEPES, pH 7.5, 50 mM NaCl,) at 4 °C. One sample per fragment concentration was then analysed by intact-LCMS at the following time points; 0, 20, 40, 60, 140, 260, and 380 minutes.

4.4.5 Covalent Fragment Activity Assays

4.4.5.1 *GSK01 and GSK02*

GSK01 and **GSK02** were dispensed to a final concentration of 100 and 50 μM . 20 μL of GALE (0.5 μM protein, 25 mM HEPES, pH 7.5, 50 mM NaCl,) was then dispensed to each sample and incubated at 4 °C for 24 hours. After this time a HPLC epimerisation assay was performed with GALE from each reaction at concentrations 50, 25, 12.5, 6, and 3 nM GALE, using 300 μM UDP-Gal and 40 min incubation at 37 °C. Quenching was carried out by heat inactivation at 95 °C for 1 minute.

4.4.5.2 *WB-C1*

WB-C1 was dispensed to a final concentration of 100 and 25 μM . 20 μL of GALE (1 μM , 25 mM HEPES, pH 7.5, 50 mM NaCl, 2mM NAD) was then dispensed to each sample and incubated at 4 °C for 24 hours. After this time, a HPLC epimerisation assay was performed with GALE from each reaction at concentrations 10 and 5 nM GALE, using 100 μM UDP-Gal and 15 minute incubation at 37 °C. Quenching was carried out by heat inactivation at 95 °C for 1 minute. HPLC was run as described in section 2.4.2.

4.4.6 Tryptic Digest LCMS/MS

GALE was first incubated with either **GSK01** or **GSK02** (25 μM , 50 or 100 μM) or DMSO (1%) for 24 hours at 4 °C before treating with iodoacetamide (IAA). Samples were purified by SDS-PAGE gel and digested in-gel with 100 ng trypsin (modified sequencing grade, Promega, Wisconsin, USA) overnight at 37 °C, then acidified with Trifluoroacetic acid (TFA).

LCMS/MS experiments were performed by Dr. Steven Howell and the methods below provided by him.

On an Ultimate 3000 nanoRSLC HPLC (Thermo Scientific, Massachusetts, USA), 1-10ul was loaded on a 20mm x 75µm Pepmap C18 trap column (Thermo Scientific) prior to elution via a 50cm x 75µm EasySpray C18 column into a Lumos Tribrid Orbitrap mass spectrometer (Thermo Scientific). A 45' gradient of 6% - 40% B was used followed by washing and re-equilibration (A= 2% ACN, 5% DMSO, 0.1% formic acid; B= 80% ACN, 5% DMSO, 0.1% formic acid).

The Orbitrap was operated in "Data Dependent Acquisition" mode followed by MS/MS in "TopS" mode, with Orbitrap accumulation at R=30K of higher-energy C-trap dissociation (HCD) fragmented parent ions.

Raw files were processed using Maxquant (maxquant.org) and Perseus (maxquant.net/perseus) with a recent download of the uniprot homo sapiens reference proteome database and a common contaminants database. A decoy database of reversed sequences was used to filter false positives, at a peptide false detection rate of 1%.

Visualisation and integration of MS1 peptide signals was performed in Skyline (<https://skyline.ms/project/home/begin.view>).

4.4.7 Covalent Fragment Docking

In general docking was performed as described in Chapter 3 section 3.4.5 with the following steps:

1. Import of the crystal structure to be docked into.
2. Conversion of the PDB file to an ICM object, keeping tight waters in the structure.
3. Identification of the binding pocket (UDP pocket) and building of receptor maps.
4. Selection of Tyr211 as the site for covalent modification.
5. Defining the sulfonyl fluoride as the warhead.
6. Docking of libraries from SDFiles

All docking parameters kept as default except for thoroughness (set to 3).

Chapter 5 Conclusions and Future Work

5.1 Conclusions and Summary

5.1.1 Aims

The aim of this project was to develop a cell-permeable small molecule inhibitor of the human UDP-Galactose-4-epimerase (GALE). Inhibition of GALE will result in a loss of O-GalNAc glycosylation and provide both a useful chemical tool for the study of O-GalNAc glycosylated biomolecules. Additionally, such an inhibitor could be used to validate GALE as a therapeutic target for epithelial cancers by downregulating mucin expression. To achieve this aim, a FBDD approach was employed.

5.1.2 Protein Characterisation and Crystallography

First, a robust protein expression and purification protocol was established using insect cells that produced active GALE in high yield. Characterisation of this protein did identify a so called “abortive complex” where much of the expressed protein (~80%) contained trapped UDP-Glc. Attempts to remove the bound substrate by protein precipitation were unsuccessful. The abortive complex had also been noted previously from human GALE expressed in *E. coli*.¹¹⁵ Given that the enzyme was active and expressed in high yields, further efforts were not employed to remove the bound substrate. However, future work will involve additional purification steps to try reactivating the GALE complex. Protein washing steps including buffers containing high phosphate or malonate may dislodge the bound UDP-Sugar by competing for the UDP phosphate binding region of GALE. Regardless, robust crystallography conditions were then optimised for GALE that produced a “mixed-homodimer”, where one monomer contained both UDP-Glc and NAD and the other contained only NAD (Apo). The reproducibility of the dimer was validated (>90 % crystal growth hit rate) and due to the presence of the Apo monomer, these conditions were considered acceptable for future experiments to determine protein/fragment structures.

5.1.3 Non-Covalent Fragments

Non-covalent fragment screening was performed with two assays. First, fragments from the Enamine DSI-poised library were screened by thermal shift assay, followed by X-ray crystallographic screening of the same library. The results of the thermal shift screen were

compromised by shifts in pH of the buffer solution induced by fragments containing basic functional groups (particularly amines). Fortunately, crystallographic screening performed with the same fragment library yielded three hit fragments: **X0121**, **X0213**, and **X0373**. Fragments **X0121** and **X0213** bound to the uridine binding region of the UDP binding pocket. **X0373** was identified at a crystallographic interface between the two GALE monomers. **X0213** was subsequently validated as the best candidate for elaboration by comparison with the earlier thermal shift screen (only **X0213** was identified as a hit in this screen) and by ITC assay.

Initially, eight analogues of **X0213** were selected for experimental testing from a virtual screen of >1000 commercially available analogues. These compounds (**WBX1-8**) were then characterised using a screening workflow of thermal shift assay, enzyme inhibition assay, ITC, and finally crystallography. This workflow identified fragment **WBX4** as a 100-fold improvement in binding ($K_D \sim 9 \mu\text{M}$) and inhibition ($IC_{50} \sim 100 \mu\text{M}$) over the parent fragment (**Figure 60**). The structure of **WBX4** in complex with GALE was subsequently solved and confirmed that the increase in potency was likely driven by an additional H-bond made to Asn224 via an amide carbonyl.

Motivated by the success of the first round of elaboration, a second series of compounds were designed (guided by the available structural information) and virtually docked into GALE. From the docked fragments, nine (**WBX9-17**) were ordered for custom synthesis from Enamine and screened in the workflow described above. Again, one fragment (**WBX9**) was identified with a clear improvement (~ 10 -fold) in binding ($K_D \sim 500 \text{ nM}$) and inhibition ($IC_{50} \sim 10 \mu\text{M}$). Solving the structure for **WBX9** in complex with GALE confirmed the substitution of a cyclobutyl group in place of the methyl isoxazole (which filled a lipophilic pocket) was the cause of the 10-fold increase in binding over the parent compound, **WBX4**. It should be noted, however, that the correlation between inhibition/binding with the score predicted in the docking was poor.

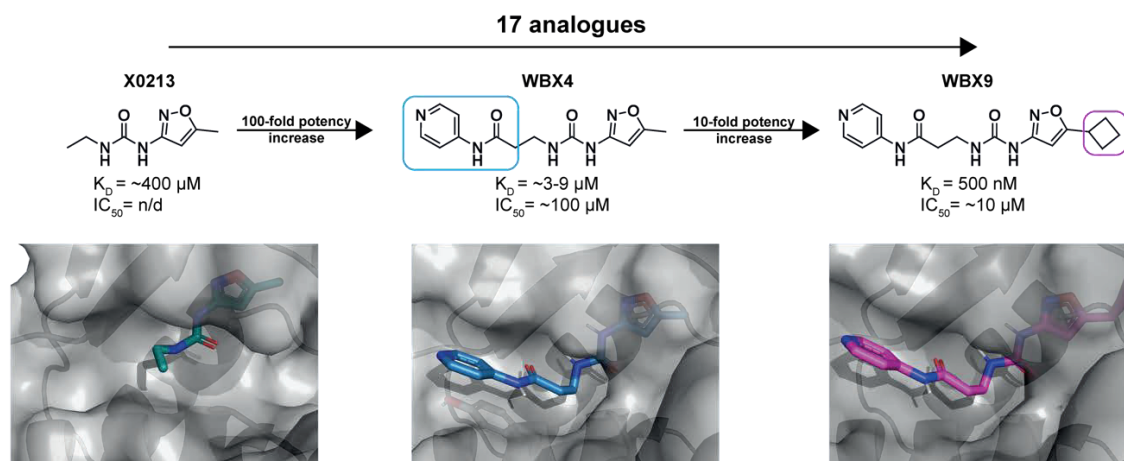


Figure 60: Non-covalent inhibitor development summary.

5.1.4

In parallel to the non-covalent fragments, covalent fragments (from a GSK library) were also screened against GALE by intact-LCMS, with the aim of targeting a binding site Tyr (Tyr211). This identified several hit fragments that produced a single mass modification. Validation of these compounds confirmed two fragments (**GSK01** and **GSK02**) as giving reproducible labelling. Only one fragment (**GSK02**) displayed any inhibitory effect upon GALE modification and only **GSK01** could be confirmed to modify Tyr211 by mass spectrometry site ID experiments. The success of the screen and the usefulness of the identified hits for elaboration was uncertain. However, the results did validate that covalent fragments could modify GALE with specificity and that this modification could modulate activity.

With GALE validated as targetable by covalent fragments, several covalent analogues of non-covalent fragment **WBX2** were designed (**WB-C1-C8**) and screened by intact-LCMS. Fragment **WB-C1** produced 100 % protein modification under the screening conditions, confirming the ligand first approach had worked. Modification with **WB-C1** also decreased the activity of GALE by ~ 10 -fold compared to a DMSO only control.

5.2 Future Work

5.2.1 Non-Covalent Fragments

Whilst fragment **WBX9** shows strong GALE binding and inhibition, to date, all assays and characterisation have been conducted with recombinant enzyme. To achieve our aim of a small molecule O-GalNAc glycosylation inhibitor, a compound active in cells is required. It is possible that fragment **WBX9** may show some cellular activity, however, it is likely that further increases in potency towards GALE will first be required. Thus, future work will involve a more thorough medicinal chemistry campaign. Initially, the aim of this campaign will be to further optimise potency against recombinant GALE ($K_D \sim 50$ nM). Given the already 1000-fold increase in potency obtained with just seventeen analogues, this should be an obtainable goal. Following the initial potency optimisation, analogue activity in cellular assays will be assessed. As such, a cellular assay will also need to be developed and could take two forms 1) an O-GalNAc assay analogous to a cellular enzyme-linked immunosorbent assay (ELISA) where antibodies are exchanged for O-GalNAc binding lectins conjugated to a reported enzyme (e.g. horse radish peroxidase). 2) A cellular assay that is based on mucin expression such as previously developed by *K. Pedram et al.*⁸² As with most drug discovery projects, optimisation will be a cyclical process, with numerous iterations of compound design, *in vitro* testing, crystallography, and *in cellulo* testing. Additional considerations that will be factored into future analogue design will be considerations of metabolically susceptible.

5.2.2 Covalent Fragments

The covalent fragment **WB-C1** displayed good inhibitory activity and protein labelling, however, further work is required to fully characterise fragment **WB-C1**. Firstly, the efficiency of protein labelling is still uncertain. A full kinetics experiment (as performed with **GSK01** and **GSK02**) is required to assess how fast covalent modification is reached and what the relationship between concentration and labelling is (k_{inact}/K_I). Site ID experiments are required to confirm that Tyr211 is the only residue being modified, both with mass spectrometry and potentially through mutants. Finally, crystal structures of the covalently modified protein would provide invaluable information to confirm the binding pose adopted after reaction with the protein.

To improve on **WB-C1**, there are two factors which could be further developed. 1) The minimal concentration required for 100% protein labelling is still unknown, a more efficient inhibitor will require 100% modification from as lower a fragment concentration as possible. 2) The binding strength of the fragment portion of the compound needs to be improved to enable better inhibition of the enzyme after covalent modification of Tyr211. The non-covalent compound **WBX2** (that **WB-C1** was designed from) is a weak binder (350-450 μ M) that displays no effect on enzyme activity even at 800 μ M. In contrast **WB-C1**, which has an identical fragment structure, modifies Tyr211 and the effect on activity is dramatic.

As has been discussed, a 10-fold increase in inhibition and binding was observed by the addition of a cyclobutyl group to fragment **WBX4** to yield fragment **WBX9** (Figure X). It is possible that analogous addition of a cyclobutyl group to **WB-C1** (**Figure 61**, **WB-C9**) will yield a significant improvement in both inhibition and labelling efficiency (**Figure 61**).

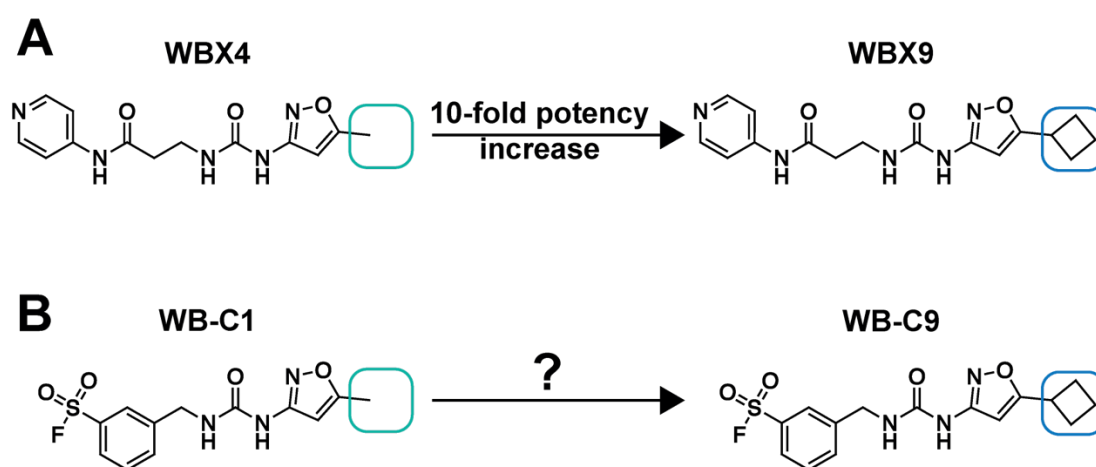


Figure 61: Future elaboration strategy for fragment **WB-C1**

A) schematic showing the addition of a cyclobutyl group to the non-covalent fragment **WBX4** that produced a 10-fold increase in potency. B) Schematic showing the proposed addition of a cyclobutyl group to **WB-C1**.

Full characterisation of **WB-C9** will be required to confirm the hypothesis that the inhibitory effect of the covalent compound will be increased by addition of a cyclobutyl group, as was seen for the non-covalent fragments (see Chapter 3). If this hypothesis is correct, it will suggest that potency increasing changes made to non-covalent fragments can be incorporated

into the covalent analogues to a similar effect. Eventually, these compounds will also require testing for GALE modification and inhibition in cells.

Development of a covalent inhibitor should be underpinned by appropriate applicability. Sulfonyl fluorides are yet untested as warheads for therapeutics. Therefore, a compound that covalently modifies GALE in cells would likely be used as a chemical tool. The first tool application would be as an O-GalNAc glycosylation inhibitor used for research purposes. Alternatively, if the covalent inhibitor could be modified to contain an alkyne, this would enable the use of “click” chemistry to attach on tags such as fluorophores. This tool compound could then be used to confirm cellular engagement with GALE of non-covalent lead compounds in experiments similar to the competitive profiling ABPP experiments described in the section 4.1.3.

References

- 1 C. Reily, T. J. Stewart, M. B. Renfrow and J. Novak, *Nat. Rev. Nephrol.*, 2019, **15**, 346–366.
- 2 A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar and P. H. Seeberger, *Essentials of glycobiology, third edition*, Cold spring harbour laboratory press, New York, Third., 2017.
- 3 P. Stanley, *Cold Spring Harb. Perspect. Biol.*, 2011, **3**, 1–13.
- 4 H. J. Joshi, L. Hansen, Y. Narimatsu, H. H. Freeze, B. Henrissat, E. Bennett, H. H. Wandall, H. Clausen and K. T. Schjoldager, *Glycobiology*, 2018, **28**, 284–294.
- 5 K. Ohtsubo and J. D. Marth, *Cell*, 2006, **126**, 855–867.
- 6 C. Reily, T. J. Stewart, M. B. Renfrow and J. Novak, *Nat. Rev. Nephrol.*, 2019, **15**, 346–366.
- 7 P. Gagneux, T. Hennet and A. Varki, in *Essentials of Glycobiology*, ed. and M. E. E. Ajit Varki, Richard D Cummings, Jeffrey D Esko, Hudson H Freeze, Pamela Stanley, Carolyn R Bertozzi, Gerald W Hart, Cold Spring Harbor Laboratory Press, New York, 4th edn., 2022.
- 8 P. Van Den Steen, P. M. Rudd, R. A. Dwek and G. Opdenakker, *Crit. Rev. Biochem. Mol. Biol.*, 1998, **33**, 151–208.
- 9 R. Amon, E. M. Reuven, S. Leviatan Ben-Arye and V. Padler-Karavani, *Carbohydr. Res.*, 2014, **389**, 115–122.
- 10 H. H. Freeze, H. Schachter and T. Kinoshita, in *Essentials of Glycobiology*, 2015.
- 11 S. Mikkola, *Molecules*, , DOI:10.3390/molecules25235755.
- 12 N. M. Akella, L. Ciraku and M. J. Reginato, *BMC Biol.*, 2019, **17**, 1–14.
- 13 M. de las Rivas, E. Lira-Navarrete, T. A. Gerken and R. Hurtado-Guerrero, *Curr. Opin. Struct. Biol.*, 2019, **56**, 87–96.
- 14 M. Muraoka, M. Kawakita and N. Ishida, *FEBS Lett.*, 2001, **495**, 87–93.
- 15 C. Fu, H. Zhao, Y. Wang, H. Cai, Y. Xiao, Y. Zeng and H. Chen, *HLA*, 2016, **88**, 275–286.
- 16 S. Schnitzler, O. Prokop and M. v. Ardenne, *Naturwissenschaften*, 1969, **56**, 285.
- 17 E. P. Bennett, U. Mandel, H. Clausen, T. A. Gerken, T. A. Fritz and L. A. Tabak, *Glycobiology*, 2012, **22**, 736–756.

- 18 D. T. Tran and K. G. Ten Hagen, *J. Biol. Chem.*, 2013, **288**, 6921.
- 19 I. Sanz-Martinez, S. Pereira, P. Merino, F. Corzana and R. Hurtado-Guerrero, *Acc. Chem. Res.*, 2023, **56**, 548.
- 20 T. Ju, V. I. Otto and R. D. Cummings, *Angew. Chemie - Int. Ed.*, 2011, **50**, 1770–1791.
- 21 G. F. Springer, P. R. Desai and I. Banatwala, *Naturwissenschaften*, 1974, **61**, 457–458.
- 22 G. F. Springer, *Science (80-.)*, 1984, **224**, 1198–1206.
- 23 T. B. Rømer, M. K. M. Aasted, S. Dabelsteen, A. Groen, J. Schnabel, E. Tan, J. W. Pedersen, A. D. Haue and H. H. Wandall, *Br. J. Cancer* 2021 1259, 2021, **125**, 1239–1250.
- 24 T. Ju, K. Brewer, A. D. Souza, R. D. Cummings and W. M. Canfield, *J. Biol. Chem.*, 2002, **277**, 178–186.
- 25 T. Ju and R. D. Cummings, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 16613–16618.
- 26 T. Ju, G. S. Lanneau, T. Gautam, Y. Wang, B. Xia, S. R. Stowell, M. T. Willard, W. Wang, J. Y. Xia, R. E. Zuna, Z. Laszik, D. M. Benbrook, M. H. Hanigan and R. D. Cummings, *Cancer Res.*, 2008, **68**, 1636–1646.
- 27 S. Terashima, Y. Takano, T. Otori, T. Kanno, T. Kimura, R. Motoki and T. Kawaguchi, *Surg. Today*, 1998, **28**, 682–686.
- 28 A. Konno, Y. Hoshino, S. Terashima, R. Motoki and T. Kawaguchi, *Clin. Exp. Metastasis*, 2002, **19**, 61–70.
- 29 D. W. Kufe, *Nat. Rev. Cancer*, 2009, **9**, 874–885.
- 30 S. S. Dhanisha, C. Guruvayoorappan, S. Drishya and P. Abeesh, *Crit. Rev. Oncol. Hematol.*, 2018, **122**, 98–122.
- 31 M. Kesimer and J. K. Sheehan, *Methods Mol. Biol.*, 2012, **842**, 67.
- 32 K. Ganguly, S. Rauth, S. Marimuthu, S. Kumar and S. K. Batra, *Cancer Metastasis Rev.*, 2020, **39**, 647.
- 33 J. P. M. Van Putten and K. Strijbis, *J. Innate Immun.*, 2017, **9**, 281.
- 34 M. C. Rose, *Am. J. Physiol. - Lung Cell. Mol. Physiol.*, , DOI:10.1152/ajplung.1992.263.4.l413.
- 35 N. Moniaux, M. Andrianifahanana, R. E. Brand and S. K. Batra, *Br. J. Cancer* 2004 919, 2004, **91**, 1633–1638.
- 36 N. MONIAUX, S. NOLLET, N. PORCHET, P. DEGAND, A. LAINE and J.-P. AUBERT, *Biochem. J.*, 1999, **338**, 325–333.

- 37 B. Macao, D. G. A. Johansson, G. C. Hansson and T. Härd, *Nat. Struct. Mol. Biol.* 2005 **131**, 2005, **13**, 71–76.
- 38 D. H. Wreschner, M. A. McGuckin, S. J. Williams, A. Baruch, M. Yoeli, R. Ziv, L. Okun, J. Zaretsky, N. Smorodinsky, I. Keydar, P. Neophytou, M. Stacey, H.-H. Lin and S. Gordon, *Protein Sci.*, 2002, **11**, 698–706.
- 39 J. R. Gum, J. W. Hicks, N. W. Toribara, B. Siddiki and Y. S. Kim, *J. Biol. Chem.*, 1994, **269**, 2440–2446.
- 40 D. Li, M. Gallup, N. Fan, D. E. Szymkowski and C. B. Basbaum, *J. Biol. Chem.*, 1998, **273**, 6812–6820.
- 41 J. L. Desseyn, J. P. Aubert, I. Van Seuning, N. Porchet and A. Laine, *J. Biol. Chem.*, 1997, **272**, 16873–16883.
- 42 N. W. Toribara, A. M. Robertson, S. B. Ho, W. L. Kuo, E. Gum, J. W. Hicks, J. R. Gum, J. C. Byrd, B. Siddiki and Y. S. Kim, *J. Biol. Chem.*, 1993, **268**, 5879–5885.
- 43 L. A. Bobek, H. Tsai, A. R. Biesbrock and M. J. Levine, *J. Biol. Chem.*, 1993, **268**, 20563–20569.
- 44 B. J. W. Van Klinken, A. W. C. Einerhand, H. A. Büller and J. Dekker, *Glycobiology*, 1998, **8**, 67–75.
- 45 F. Levitin, O. Stern, M. Weiss, C. Gil-Henn, R. Ziv, Z. Prokocimer, N. I. Smorodinsky, D. B. Rubinstein and D. H. Wreschner, *J. Biol. Chem.*, 2005, **280**, 33374–33386.
- 46 D. Ambort, M. E. V. Johansson, J. K. Gustafsson, H. E. Nilsson, A. Ermund, B. R. Johansson, P. J. B. Koeck, H. Hebert and G. C. Hansson, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 5645–5650.
- 47 P. Dhar and J. McAuley, *Front. Cell. Infect. Microbiol.*, 2019, **9**, 452801.
- 48 J. L. McAuley, S. K. Linden, W. P. Chin, R. M. King, H. L. Pennington, S. J. Gendler, T. H. Florin, G. R. Hill, V. Korolik and M. A. McGuckin, *J. Clin. Invest.*, 2007, **117**, 2313–2324.
- 49 S. K. Lindén, Y. H. Sheng, A. L. Every, K. M. Miles, E. C. Skoog, T. H. J. Florin, P. Sutton and M. A. McGuckin, *PLoS Pathog.*, , DOI:10.1371/JOURNAL.PPAT.1000617.
- 50 G. Z. Ng and P. Sutton, *Genes Immun.* 2016 **173**, 2016, **17**, 203–206.
- 51 E. P. Lillehoj, S. W. Hyun, B. T. Kim, X. G. Zhang, D. I. Lee, S. Rowland and K. C. Kim, *Am. J. Physiol. - Lung Cell. Mol. Physiol.*, , DOI:10.1152/AJPLUNG.2001.280.1.L181/ASSET/IMAGES/LARGE/H50110233006.JPEG.
- 52 E. P. Lillehoj, H. Kim, E. Y. Chun and K. C. Kim, *Am. J. Physiol. - Lung Cell. Mol. Physiol.*,

- 2004, **287**, 809–815.
- 53 J. J. Rahn, J. W. Chow, G. J. Horne, B. K. Mah, J. T. Emerman, P. Hoffman and J. C. Hugh, *Clin. Exp. Metastasis*, 2005, **22**, 475–483.
 - 54 Y. Li, H. Kuwahara, J. Ren, G. Wen and D. Kufe, *J. Biol. Chem.*, 2001, **276**, 6061–6064.
 - 55 K. Kato, E. P. Lillehoj, Y. S. Park, T. Umehara, N. E. Hoffman, M. Madesh and K. C. Kim, *J. Immunol.*, 2012, **188**, 2014–2022.
 - 56 K. C. Kim and E. P. Lillehoj, <https://doi.org/10.1165/rcmb.2008-0169TR>, 2012, **39**, 644–647.
 - 57 P. K. Singh and M. A. Hollingsworth, *Trends Cell Biol.*, 2006, **16**, 467–476.
 - 58 J. H. Yen, S. Xu, Y. S. Park, D. Ganea and K. C. Kim, *Brain. Behav. Immun.*, 2013, **29**, 70–81.
 - 59 T. Kawai and S. Akira, *Trends Mol. Med.*, 2007, **13**, 460–469.
 - 60 W. Lu, A. Hisatsune, T. Koga, K. Kato, I. Kuwahara, E. P. Lillehoj, W. Chen, A. S. Cross, S. J. Gendler, A. T. Gewirtz and K. C. Kim, *J. Immunol.*, 2006, **176**, 3890–3894.
 - 61 Y. H. Sheng, S. Triyana, R. Wang, I. Das, K. Gerloff, T. H. Florin, P. Sutton and M. A. McGuckin, *Mucosal Immunol.* 2013 63, 2012, **6**, 557–568.
 - 62 E. C. Woods, F. Kai, J. M. Barnes, K. Pedram, M. W. Pickup, M. J. Hollander, V. M. Weaver and C. R. Bertozzi, *Elife*, , DOI:10.7554/eLife.25752.
 - 63 S. Nath and P. Mukherjee, *Trends Mol. Med.*, 2014, **20**, 332.
 - 64 S. K. Gautam, P. Khan, G. Natarajan, P. Atri, A. Aithal, A. K. Ganti, S. K. Batra, M. W. Nasser and M. Jain, *Cancers (Basel)*, , DOI:10.3390/CANCERS15061640.
 - 65 W. Chen, Z. Zhang, S. Zhang, P. Zhu, J. K. S. Ko and K. K. L. Yung, *Int. J. Mol. Sci.*, , DOI:10.3390/IJMS22126567.
 - 66 X. Cheng and K. Cheng, *Cancer Metastasis Rev.*, 2021, **40**, 71.
 - 67 F. Van Roy and G. Berx, *Cell. Mol. Life Sci.*, 2008, **65**, 3756–3788.
 - 68 J. P. Their, *Nat. Rev. Cancer* 2002 26, 2002, **2**, 442–454.
 - 69 L. D. Roy, M. Sahraei, D. B. Subramani, D. Besmer, S. Nath, T. L. Tinder, E. Bajaj, K. Shanmugam, Y. Y. Lee, S. I. L. Hwang, S. J. Gendler and P. Mukherjee, *Oncogene* 2011 3012, 2010, **30**, 1449–1459.
 - 70 D. Vestweber, *Semin. Cell Biol.*, 1992, **3**, 211–220.
 - 71 I. P. Witz, *Immunol. Lett.*, 2006, **104**, 89–93.
 - 72 S. R. Barthel, J. D. Gavino, L. Descheny and C. J. Dimitroff, *Expert Opin. Ther. Targets*,

- 2007, **11**, 1473.
- 73 M. J. Paszek, C. C. Dufort, O. Rossier, R. Bainer, J. K. Mouw, K. Godula, J. E. Hudak, J. N. Lakins, A. C. Wijekoon, L. Cassereau, M. G. Rubashkin, M. J. Magbanua, K. S. Thorn, M. W. Davidson, H. S. Rugo, J. W. Park, D. A. Hammer, G. Giannone, C. R. Bertozzi and V. M. Weaver, *Nature*, 2014, **511**, 319–325.
 - 74 S. Cascio and O. J. Finn, *Biomol. 2016, Vol. 6, Page 39*, 2016, **6**, 39.
 - 75 S. Wisnovsky, L. Möckl, S. A. Malaker, K. Pedram, G. T. Hess, N. M. Riley, M. A. Gray, B. A. H. Smith, M. C. Bassik, W. E. Moerner and C. R. Bertozzi, *Proc. Natl. Acad. Sci. U. S. A.*, 2021, **118**, e2015024118.
 - 76 E. C. M. Brinkman-Van Der Linden and A. Varki, *J. Biol. Chem.*, 2000, **275**, 8625–8632.
 - 77 D. Raina, P. Agarwal, J. Lee, A. Bharti, C. J. McKnight, P. Sharma, S. Kharbanda and D. Kufe, *PLoS One*, 2015, **10**, e0135156.
 - 78 Y. Leng, C. Cao, J. Ren, L. Huang, D. Chen, M. Ito and D. Kufe, *J. Biol. Chem.*, 2007, **282**, 19321–19330.
 - 79 O. Engebraaten, G. Sivam, S. Juell and Ø. Fodstad, *J. Cancer*, 2000, **88**, 970–976.
 - 80 R. Gutierrez, P. D. Shah, O. Hamid, A. L. Garfall, A. Posey, M. R. Bishop, G. R. Blumenschein, M. L. Johnson, S. Lee, J. J. Luke, D. Morgensztern, T. J. Fountaine, R. Dryer-Minnerly, S. Najmi, P. Hufner and K. Chagin, https://doi.org/10.1200/JCO.2021.39.15_suppl.e14513, 2021, **39**, e14513–e14513.
 - 81 T. Gao, Q. Cen and H. Lei, *Biomed. Pharmacother.*, 2020, **132**, 110888.
 - 82 K. Pedram, D. J. Shon, G. S. Tender, N. R. Mantuano, J. J. Northey, K. J. Metcalf, S. P. Wisnovsky, N. M. Riley, G. C. Forcina, S. A. Malaker, A. Kuo, B. M. George, C. L. Miller, K. M. Casey, J. G. Vilches-Moure, M. J. Ferracane, V. M. Weaver, H. Läubli and C. R. Bertozzi, *Nat. Biotechnol.*, 2023, 1–11.
 - 83 G. Huet, S. Hennebicq-Reig, C. De Bolos, F. Ulloa, T. Lesuffleur, A. Barbat, V. Carrière, I. Kim, F. X. Real, P. Delannoy and A. Zweibaum, *J. Cell Biol.*, 1998, **141**, 1311.
 - 84 S. Satoh, Y. Hinoda, T. Hayashi, M. D. Burdick, K. Imai and M. A. Hollingsworth, *J. Cancer*, 2000, **88**, 507–518.
 - 85 T. J. McCorvie and D. J. Timson, *Handb. Glycosyltransferases Relat. Genes, Second Ed.*, 2014, **2**, 1449–1464.
 - 86 P. A. Frey, *FASEB J.*, 1996, **10**, 461–470.
 - 87 A. Broussard, A. Florwick, C. Desbiens, N. Nischan, C. Robertson, Z. Guan, X. J. J.

- Kohler, L. Wells and M. Boyce, *J. Biol. Chem.*, 2020, **295**, 1225.
- 88 J. M. Termini, Z. A. Silver, B. Connor, A. Antonopoulos, S. M. Haslam, A. Dell and R. C. Desrosiers, *PLoS One*, , DOI:10.1371/journal.pone.0179949.
- 89 B. F. Cravatt, A. T. Wright and J. W. Kozarich, *Annu. Rev. Biochem.*, 2008, **77**, 383–414.
- 90 C. R. Shurer, M. J. Colville, V. K. Gupta, S. E. Head, F. Kai, J. N. Lakins and M. J. Paszek, *ACS Biomater. Sci. Eng.*, 2018, **4**, 388–399.
- 91 D. J. Timson, *IUBMB Life*, 2006, **58**, 83–89.
- 92 J. Fridovich-Keil, L. Bean, M. He and R. Schroer, *GeneReviews*[®].
- 93 Z. Zhao, A. Ukidve, J. Kim and S. Mitragotri, *Cell*, 2020, **181**, 151–167.
- 94 A. L. Hopkins, *Pract. Med. Chem. Fourth Ed.*, 2008, 395–408.
- 95 G. Klebe and H. J. Böhm, *J. Recept. Signal Transduct. Res.*, 1997, **17**, 459–473.
- 96 J. B. Thoden, T. M. Wohlers, J. L. Fridovich-Keil and H. M. Holden, *Biochemistry*, 2000, **39**, 5691–5701.
- 97 P. Cohen, D. Cross and P. A. Jänne, *Nat. Rev. Drug Discov.* 2021 207, 2021, **20**, 551–569.
- 98 K. A. Winans and C. R. Bertozzi, *Chem. Biol.*, 2002, **9**, 113–129.
- 99 M. D. Urbaniak, J. N. Tabudravu, A. Msaki, K. M. Matera, R. Brenk, M. Jaspars and M. A. J. Ferguson, *Bioorganic Med. Chem. Lett.*, 2006, **16**, 5744–5747.
- 100 C. Santi, C. Scimmi and L. Sancineto, *Molecules*, , DOI:10.3390/MOLECULES26144230.
- 101 J. P. Hughes, S. S. Rees, S. B. Kalindjian and K. L. Philpott, *Br. J. Pharmacol.*, 2011, **162**, 1239–1249.
- 102 J. G. Moffat, F. Vincent, J. A. Lee, J. Eder and M. Prunotto, *Nat. Rev. Drug Discov.*, 2017, **16**, 531–543.
- 103 S. Wang, Z. Wang, L. Fang, Y. Lv and G. Du, *Int. J. Drug Discov. Pharmacol.*, 2022, **2**.
- 104 G. E. Croston, *Expert Opin. Drug Discov.*, 2017.
- 105 M. Q. Zhang and B. Wilkinson, *Curr. Opin. Biotechnol.*, 2007.
- 106 D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke and H. Jhoti, *Nat. Rev. Drug Discov.*, 2016, **15**, 605–619.
- 107 M. Congreve, R. Carr, C. Murray and H. Jhoti, *Drug Discov. Today*, 2003, **8**, 876–877.
- 108 J. D. St. Denis, R. J. Hall, C. W. Murray, T. D. Heightman and D. C. Rees, *RSC Med. Chem.*, 2021, **12**, 321.
- 109 P. Kirsch, A. M. Hartman, A. K. H. Hirsch and M. Empting, *Molecules*, 2019, **24**.

- 110 D. E. Scott, A. G. Coyne, S. A. Hudson and C. Abell, *Biochemistry*, 2012, **51**, 4990–5003.
- 111 A. Carbery, R. Skyner, F. Von Delft and C. M. Deane, *J. Med. Chem.*, 2022, **65**, 11404–11413.
- 112 M. Bon, A. Bilsland, J. Bower and K. McAulay, *Mol. Oncol.*, 2022, **16**, 3761–3777.
- 113 C. W. Murray, D. A. Erlanson, A. L. Hopkins, G. M. Keserü, P. D. Leeson, D. C. Rees, C. H. Reynolds and N. J. Richmond, *ACS Med. Chem. Lett.*, , DOI:10.1021/ml500146d.
- 114 L. Wang, J. Gao, R. Ma, Y. Liu, M. Liu, F. Zhong, J. Hu, S. Li, J. Wu, H. Jiang, J. Zhang and K. Ruan, *Magn. Reson. Lett.*, 2022, **2**, 107–118.
- 115 A. L. Pey, E. Padín-Gonzalez, N. Mesa-Torres and D. J. Timson, *Arch. Biochem. Biophys.*, , DOI:10.1016/j.abb.2014.07.030.
- 116 D. L. Jarvis, in *Methods in Enzymology*, 2009.
- 117 W. M. Browne, Imperial College London, 2020.
- 118 M. F. Debets, O. Y. Tastan, S. P. Wisnovsky, S. A. Malaker, N. Angelis, L. K. R. Moeckl, J. Choi, H. Flynn, L. J. S. Wagner, G. Bineva-Todd, A. Antonopoulos, A. Cioce, W. M. Browne, Z. Li, D. C. Briggs, H. L. Douglas, G. T. Hess, A. J. Agbay, C. Roustan, S. Kjaer, S. M. Haslam, A. P. Snijders, M. C. Bassik, W. E. Moerner, V. S. W. Li, C. R. Bertozzi and B. Schumann, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 25293–25301.
- 119 J. E. Scott and K. P. Williams, *Validating Identity, Mass Purity and Enzymatic Purity of Enzyme Preparations*, 2004.
- 120 K. A. Johnson and R. S. Goody, *Biochemistry*, 2011, **50**, 8264–8269.
- 121 N. Drinkwater, H. Vu, K. M. Lovell, K. R. Criscione, B. M. Collins, T. E. Prisinzano, S. A. Poulsen, M. J. Mcleish, G. L. Grunewald and J. L. Martin, *Biochem. J.*, 2010, **431**, 51–61.
- 122 H. Zheng, J. Hou, M. D. Zimmerman, A. Wlodawer and W. Minor, *Expert Opin. Drug Discov.*, 2014, **9**, 125.
- 123 A. M. Kierzek and P. Zielenkiewicz, *Biophys. Chem.*, 2001, **91**, 1–20.
- 124 T. Bergfors, *J. Struct. Biol.*, 2003, **142**, 66–76.
- 125 B. Wienen-Schmidt, M. Oebbeke, K. Ngo, A. Heine and G. Klebe, *ChemMedChem*, 2021, **16**, 292–300.
- 126 D. J. Timson, *FEBS J.*, 2005, **272**, 6170–6177.
- 127 A. Odunewu-Aderibigbe and L. Fliegel, *IUBMB Life*, 2014, **66**, 679–685.
- 128 P. A. Frey and A. D. Hegeman, *Acc. Chem. Res.*, , DOI:10.1021/ar300246k.

- 129 M. F. Debets, O. Y. Tastan, S. P. Wisnovsky, S. A. Malaker, N. Angelis, L. K. R. Moeckl, J. Choi, H. Flynn, L. J. S. Wagner, G. Bineva-Todd, A. Antonopoulos, A. Cioce, W. M. Browne, Z. Li, D. C. Briggs, H. L. Douglas, G. T. Hess, A. J. Agbay, C. Roustan, S. Kjaer, S. M. Haslam, A. P. Snijders, M. C. Bassik, W. E. Moerner, V. S. W. Li, C. R. Bertozzi and B. Schumann, *Proc. Natl. Acad. Sci. U. S. A.*, , DOI:10.1073/pnas.2007297117.
- 130 G. Winter, C. M. C. Lobley and S. M. Prince, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, , DOI:10.1107/S0907444913015308.
- 131 A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J. Appl. Crystallogr.*, , DOI:10.1107/S0021889807021206.
- 132 P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, , DOI:10.1107/S0907444910007493.
- 133 G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin, *urn:issn:0907-4449*, 2011, **67**, 355–367.
- 134 O. B. Cox, T. Krojer, P. Collins, O. Monteiro, R. Talon, A. Bradley, O. Fedorov, J. Amin, B. D. Marsden, J. Spencer, F. Von Delft and P. E. Brennan, *Chem. Sci.*, 2016, **7**, 2322–2330.
- 135 Q. Li, *Front. Mol. Biosci.*, 2020, **7**.
- 136 H. L. Silvestre, T. L. Blundell, C. Abell and A. Ciulli, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 12984–12989.
- 137 M. C. Silva-Santisteban, I. M. Westwood, K. Boxall, N. Brown, S. Peacock, C. McAndrew, E. Barrie, M. Richards, A. Mirza, A. W. Oliver, R. Burke, S. Hoelder, K. Jones, G. W. Aherne, J. Blagg, I. Collins, M. D. Garrett and R. L. M. van Montfort, *PLoS One*, , DOI:10.1371/journal.pone.0065689.
- 138 M. Hermodson, *Proteins Struct. Funct. Genet.*, 1996, **24**, 409.
- 139 K. Gao, R. Oerlemans and M. R. Groves, *Biophys. Rev.*, 2020, **12**, 85–104.
- 140 M. C. Deller, L. Kong and B. Rupp, *Acta Crystallogr. Sect. Struct. Biol. Commun.*, 2016, **72**, 72–95.
- 141 T. G. Davies and I. J. Tickle, *Top. Curr. Chem.*, 2012, **317**, 33–59.
- 142 M. Batool, B. Ahmad and S. Choi, *Int. J. Mol. Sci.*, 2019, **20**.
- 143 E. H. B. Maia, L. C. Assis, T. A. de Oliveira, A. M. da Silva and A. G. Taranto, *Front. Chem.*, 2020, **8**.
- 144 R. Abagyan, M. Totrov and D. Kuznetsov, *J. Comput. Chem.*, 1994, **15**, 488–506.

- 145 M. A. C. Neves, M. Totrov and R. Abagyan, *J. Comput. Aided. Mol. Des.*, 2012, **26**, 675–686.
- 146 Y. A. Arnautova, A. Jagielska and H. A. Scheraga, *J. Phys. Chem. B*, 2006, **110**, 5025–5044.
- 147 R. Abagyan and M. Totrov, *J. Mol. Biol.*, 1994, **235**, 983–1002.
- 148 A. Douangamath, D. Fearon, P. Gehrtz, T. Krojer, P. Lukacik, C. D. Owen, E. Resnick, C. Strain-Damerell, A. Aimon, P. Ábrányi-Balogh, J. Brandão-Neto, A. Carbery, G. Davison, A. Dias, T. D. Downes, L. Dunnett, M. Fairhead, J. D. Firth, S. P. Jones, A. Keeley, G. M. Keserü, H. F. Klein, M. P. Martin, M. E. M. Noble, P. O’Brien, A. Powell, R. N. Reddi, R. Skyner, M. Snee, M. J. Waring, C. Wild, N. London, F. von Delft and M. A. Walsh, *Nat. Commun.*, , DOI:10.1038/s41467-020-18709-w.
- 149 M. Sunden, D. Upadhyay, R. Banerjee, N. Sipari, V. Fellman, J. Kallijärvi and J. Purhonen, *Cell Reports Methods*, 2023, **3**, 100518.
- 150 E. Chaves-Olarte, I. Florin, P. Boquet, M. Popoff, C. Von Eichel-Streiber and M. Thelestam, *J. Biol. Chem.*, 1996, **271**, 6925–6932.
- 151 N. D. Wright, P. Collins, L. Koekemoer, T. Krojer, R. Talon, E. Nelson, M. Ye, R. Nowak, J. Newman, J. T. Ng, N. Mitrovich, H. Wiggers and F. Von Delft, *Acta Crystallogr. Sect. D Struct. Biol.*, 2021, **77**, 62–74.
- 152 The XChem Pipeline - - Diamond Light Source,
<https://www.diamond.ac.uk/Instruments/Mx/Fragment-Screening/The-XChem-Pipeline.html>, (accessed 10 November 2023).
- 153 T. Krojer, R. Talon, N. Pearce, P. Collins, A. Douangamath, J. Brandao-Neto, A. Dias, B. Marsden and F. Von Delft, *Acta Crystallogr. Sect. D Struct. Biol.*, 2017, **73**, 267–278.
- 154 M. Wojdyr, R. Keegan, G. Winter and A. Ashton, *Acta Crystallogr. Sect. A*, 2013, **69**, s299–s299.
- 155 N. M. Pearce, A. R. Bradley, T. Krojer, B. D. Marsden, C. M. Deane and F. Von Delft, *Struct. Dyn. (Melville, N.Y.)*, , DOI:10.1063/1.4974176.
- 156 F. Long, R. A. Nicholls, P. Emsley, S. Gražulis, A. Merkys, A. Vaitkus and G. N. Murshudov, *Acta Crystallogr. Sect. D Struct. Biol.*, 2017, **73**, 112–122.
- 157 T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 490–519.
- 158 P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R.

- Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, , DOI:10.1107/S0907444909052925.
- 159 A. A. Vaguine, J. Richelle and S. J. Wodak, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, , DOI:10.1107/S0907444998006684.
- 160 A. Keeley, L. Petri, P. Ábrányi-Balogh and G. M. Keserű, *Drug Discov. Today*, 2020, 25, 983–996.
- 161 J. M. Strelow, *J. Biomol. Screen.*, 2017, 22, 3–20.
- 162 C. McWhirter, *Annu. Rep. Med. Chem.*, 2021, **56**, 1–31.
- 163 J. S. Martin, C. J. MacKenzie, D. Fletcher and I. H. Gilbert, *Bioorganic Med. Chem.*, 2019, **27**, 2066–2074.
- 164 L. Boike, N. J. Henning and D. K. Nomura, *Nat. Rev. Drug Discov.*, 2022, 21, 881–898.
- 165 H. S. Hahm, E. K. Toroitich, A. L. Borne, J. W. Brulet, A. H. Libby, K. Yuan, T. B. Ware, R. L. McCloud, A. M. Ciancone and K. L. Hsu, *Nat. Chem. Biol.*, 2020, **16**, 150–159.
- 166 J. R. Vane and R. M. Botting, in *Thrombosis Research*, 2003, vol. 110, pp. 255–258.
- 167 R. R. Yocum, J. R. Rasmussen and J. L. Strominger, *J. Biol. Chem.*, 1980, **255**, 3977–3986.
- 168 J. Singh, R. C. Petter, T. A. Baillie and A. Whitty, *Nat. Rev. Drug Discov.*, 2011, 10, 307–317.
- 169 H. A. Blair, *Drugs*, 2021, 81, 1573–1579.
- 170 L. Goebel, M. P. Müller, R. S. Goody and D. Rauh, *RSC Med. Chem.*, 2020, 11, 760–770.
- 171 L. Huang, Z. Guo, F. Wang and L. Fu, *Signal Transduct. Target. Ther.*, 2021, 6.
- 172 A. R. Moore, S. C. Rosenberg, F. McCormick and S. Malek, *Nat. Rev. Drug Discov.*, 2020, 19, 533–552.
- 173 J. M. Ostrem, U. Peters, M. L. Sos, J. A. Wells and K. M. Shokat, *Nature*, 2013, **503**, 548–551.
- 174 G. D. Goss, M. Cobo, S. Lu, K. Syrigos, K. H. Lee, E. Göker, V. Georgoulas, D. Isla, A. Morabito, Y. J. Min, A. Ardizzoni, S. Bender, A. Cseh and E. Felip, *EClinicalMedicine*, , DOI:10.1016/j.eclinm.2021.100940.
- 175 G. Recondo, F. Facchinetti, K. A. Olaussen, B. Besse and L. Friboulet, *Nat. Rev. Clin. Oncol.*, 2018, 15, 694–708.
- 176 S. H. I. Ou, *Crit. Rev. Oncol. Hematol.*, 2012, 83, 407–421.
- 177 N. Jessani and B. F. Cravatt, *Curr. Opin. Chem. Biol.*, 2004, 8, 54–59.

- 178 A. E. Speers and B. F. Cravatt, *Curr. Protoc. Chem. Biol.*, 2009, **1**, 29–41.
- 179 S. Wang, Y. Tian, M. Wang, M. Wang, G. B. Sun and X. B. Sun, *Front. Pharmacol.*, 2018, 9.
- 180 K. E. Gilbert, A. Vuorinen, A. Aatkar, P. Pogány, J. Pettinger, E. K. Grant, J. M. Kirkpatrick, K. Rittinger, D. House, G. A. Burley and J. T. Bush, *ACS Chem. Biol.*, , DOI:10.1021/acschembio.2c00633.
- 181 A. Aatkar, A. Vuorinen, O. E. Longfield, K. Gilbert, R. Peltier-Heap, C. D. Wagner, F. Zappacosta, K. Rittinger, C. W. Chung, D. House, N. C. O. Tomkinson and J. T. Bush, *ACS Chem. Biol.*, , DOI:10.1021/acschembio.3c00034.