Structure-based inhibitor design for substrate reduction therapy towards primary hyperoxaluria and galactosaemia

---

Sabrina Mackinnon Green Templeton College



Supervisors: Dr Wyatt Yue, Dr Paul Brennan

A thesis submitted for the degree of *Doctor of Philosophy* in Clinical Medicine at the Structural Genomics Consortium

Nuffield Department of Medicine, University of Oxford

Michaelmas term 2019

Declaration

I declare that no parts of this research or thesis have been submitted for another award, degree or diploma at any other university or learning institution. This thesis contains no other person's work, except where stated in text.

Acknowledgements

Thank you to my supervisors, Wyatt Yue and Paul Brennan, for bringing their collective and considerable expertise to bear in supporting and guiding me through this extremely rewarding project. In particular to Wyatt, for his seemingly endless supply of calm logic, kind patience and sage advice.

I would like to thank the many wonderful people I have met at the SGC these past four years. Above all, I truly appreciate the ready friendship, sound advice and all-round encouragement I have received from past and present members of the Yue group, both in and out of the lab. I owe particular thanks to Jola Kopec, Elzbieta Rembeza, Romain Talon and Michael Fairhead for teaching me everything I needed to know in the lab, and more besides. And to the SGC MOB, Biotechnology and Protein Crystallography groups and the Diamond Light Source i04-1 beamline staff, giants whose shoulders permitted me to see this project through to a satisfying conclusion. Thank you also to the EPSRC and Nuffield Department of Medicine, whose funding allowed me to spend four years doing exactly what I have always dreamed of doing – studying science at Oxford.

And finally, to my fiancé Drew, who deserves a medal for providing four years of continuous love, support and remedial chocolate, and my foster parents Bill and Shona, who encouraged me to follow my dreams.

Abstract

Inborn errors of metabolism are a group of more than 1000 inherited single gene disorders for which few effective treatments are available. In almost all of these disorders, mutations cause a loss of function in the encoded protein that is essential to normal metabolism, resulting in accumulation of metabolites upstream of the variant and/or deficiency of downstream metabolites. Where the disease pathogenesis results from toxicity of accumulated metabolites, inhibition of proteins preceding the metabolic block has been shown to be an effective therapeutic strategy. For example, proof-of-concept, genetic knockdown studies support the inhibition of galactokinase 1 (hGALK1), which precedes the defective variant in galactose metabolism, to treat classic galactosaemia; and inhibition of hydroxyacid oxidase 1 (hHAO1), which precedes the defect in glycolate metabolism, to treat primary hyperoxaluria. However, existing inhibitors for these two targets have limited scaffold diversity and their poor pharmacokinetic and toxicity profiles preclude therapeutic use. In this work, fragment screening by X-ray crystallography was combined with solution binding and activity assays to generate novel chemical starting points for the development of hGALK1 and hHAO1 inhibitors. Two fragments weakly inhibited hGALK1 upon binding to novel allosteric pockets away from the active site. Merging of fragments in one such pocket yielded two inhibitor hits of good potency and efficacy with potential for future drug development. For hHAO1, analogues of fragments binding to the active site, a nearby novel pocket and at the subunit-subunit interface yielded a total of 9 high efficacy inhibitors, spanning all three sites and diverse scaffolds. Altogether, the compounds described in this work hold promise for addressing the unmet therapeutic need in classic galactosaemia and primary hyperoxaluria, and the methodology developed suggests a broader application to the field of inborn errors of metabolism.

Abbreviations

AEX – Anion exchange chromatography

AGXT – Alanine-glycine aminotransferase 1 AI-

TB – Auto-induction terrific broth (medium)

ALDH4A1 - Δ-1-pyrroline-5-carboxylate dehydrogenase (also aldehyde dehydrogenase 4A1)

AMPPNP – Adenylyl-imidodiphosphate

AR – Aldose reductase

ASKHA – Acetate/sugar kinase/hsc70/actin (protein superfamily)

ATP/ADP – Adenosine triphosphate/ diphosphate

AUC – Analytical ultracentrifugation

BCS – Basic Chem Space (crystallisation screen)

B_{max} – Maximal specific binding response (RU)

CCPST – 4-carboxy-5-[(4-chlorophenyl)sulfanyl]-1,2,3-thiadiazole

CDST – 4-carboxy-5-dodecylsulfanyl-1,2,3-triazole

CHES – *N*-Cyclohexyl-2-aminoethanesulfonic acid

CHO – Chinese hamster ovary (cells)

CID – Compound ID number from the PubChem database

cLogD – logarithm of calculated distribution coefficient between lipid and aqueous phases

cLogP – logarithm of calculated partition coefficient between n-octanol and water (type of cLogD)

CRISPR/Cas9 – Clustered regularly interspaced short palindromic repeats/ CRISPR-associated protein 9

CV – Column volume

CYB2 – Cytochrome b2, mitochondrial (or lactate dehydrogenase [Cytochrome])

DAO – D-amino acid oxidase

DIPEA – *N*,*N*-Diisopropylethylamine

DLS – Diamond Light Source (synchrotron)

DMSO – Dimethyl sulfoxide

DNA – Deoxyribonucleic acid

DSF – Differential scanning fluorimetry EC_{50} –

Half maximal effective concentration

EDTA – Ethylenediaminetetraacetic acid

eIF2α – Eukaryotic translation initiation factor 2 subunit alpha

ER – Endoplasmic reticulum

FAD – Flavin adenine dinucleotide

FMN – Flavin mononucleotide GALE

– UDP-glucose 4-epimerase GALK1

– Galactokinase 1

GALK2 – *N*-acetylgalactosamine kinase (or galactokinase 2)

GALM – Galactose mutarotase

GALT – Galactose-1-phosphate uridylyltransferase

GHMP – galactokinase, homoserine kinase, mevalonate kinase and phosphomevalonate kinase (protein family)

GOT2 – Aspartate aminotransferase (or glutamate oxaloacetate transaminase, mitochondrial)

GOX – Glycolate oxidase

GPCRs – G-protein-coupled receptors

GRPHR – Glyoxylate reductase – hydroxypyruvate reductase

HAO1 – Hydroxyacid oxidase 1 (or glycolate oxidase)

HAO2 – Hydroxyacid oxidase 2

HEPES – 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HIN – Hampton Research Index (crystallisation screen)

HOGA – 4-hydroxy-2-oxoglutarate aldolase (or dihydrodipicolinate synthase-like, mitochondrial)

HPLC – High performance liquid chromatography

HTS – High throughput screening

IC_{50} – Half maximal inhibitory concentration IEM

– Inborn error of metabolism

IUPAC – International Union of Pure and Applied Chemistry (naming system)

JCSG – Joint Centre for Structural Genomics (crystallisation screen)

K_D – equilibrium dissociation constant

Km – Michaelis-Menten constant

LB – Lysogeny broth/Luria-Bertani (medium)

LCHAO – Long chain hydroxyacid oxidase

LC-MS – Liquid chromatography-mass spectrometry

LDH – Lactate dehydrogenase

LFS – Ligand Friendly Screen (crystallisation screen)

LLE_{AT} – Lipophilic ligand efficiency score, developed by Astex Therapeutics

MDH/MDHC1 – Mandelate dehydrogenase

MIB (buffer) – Sodium malonate dibasic monohydrate, Imidazole, Boric acid: pH 4.0-10.0

MOPS – 3-(*N*-morpholino)propanesulfonic acid

MPD – 2-methyl-2,4-pentanediol

MVDD/MVD1 – Mevalonate 5-diphosphate decarboxylase

MVK/KIME – Mevalonate kinase

NAD – Nicotinamide adenine dinucleotide

NADP(H) – Nicotinamide adenine dinucleotide phosphate (reduced)

NCT – National clinical trial

NIH – National Institute of Health Ni-

NTA – Nickel-nitrilotriacetic acid NMP

– *N*-Methyl-2-pyrrolidone NMR –

Nuclear magnetic resonance

OMIM – Online Mendelian Inheritance in Man (database)

P – Probability value

PAH – Phenylalanine hydroxylase

PanDDA – Pan-dataset density analysis

PCR – Polymerase chain reaction

PDB – Protein databank PEG

– Polyethylene glycol

PH1/2/3 – Primary hyperoxaluria type 1/2/3

PKU – Phenylketonuria

PLP – Pyridoxal 5'-phosphate

PREPX – Parallel Rapid Expression & Purification for X-ray crystallography

PRODH2 – Hydroxyproline dehydrogenase (or proline dehydrogenase 2)

Q-TOF – Quadrupole time-of-flight (mass spectrometry)

Rg – Radius of gyration, Å

RISC – RNA-induced silencing complex

RMSD – Root mean square deviation RNA

– Ribonucleic acid

RNAi – RNA interference

SAR – Structure activity relationship

SAXS – Small angle X-ray scattering

SDS-PAGE – Sodium dodecyl sulphate-polyacrylamide gel electrophoresis

SEC – Size exclusion chromatography

SGC – Structural Genomics Consortium

S_NAr – Nucleophilic aromatic substitution

SPR – Surface plasmon resonance

SRT – Substrate reduction therapy

SSMD – Strictly standardised mean difference

TACA – 5-(undecan-6-ylthio)-2H-1,2,3-triazole-4-carboxylic acid

TCEP – Tris(2-carboxyethyl)phosphine

TEV – Tobacco etch virus

TIM (barrel) – Triosephosphateisomerase barrel (protein fold)

TKP – 3-decyl-2,5-dioxo-4-hydroxy-3-pyrroline

Tm – Melting temperature, °C

UDP – Uridine diphosphate

Vmax – Maximal velocity

Contents

Declaration 2

Acknowledgements 2

Abstract 3

Abbreviations 4

Contents 8

List of Figures 13

List of tables 18

1. Introduction 19

1.1 Inborn errors of metabolism perturb function by blocking a specific metabolic pathway 19

1.1.1 Galactose metabolism primarily occurs in the liver via the Leloir pathway 20

1.1.2 Classic galactosemia is a disorder of galactose metabolism 22

1.1.3 Glyoxylate metabolism primarily occurs in the liver via peroxisomal and mitochondrial pathways 24

1.1.4 Primary hyperoxaluria – disorder of glyoxylate metabolism 25

1.2 Treatment of inborn errors of metabolism 27

1.2.1 Dietary therapy aims to restore steady state balance of the affected pathway 28

1.2.2 Organ transplantation aims to replace the defective pathway 29

1.2.3 Alternative methods to restore defective protein function 30

1.2.4 Evolution from dietary supplementation to other substrate reduction therapies 31

1.3 Structure-guided hit discovery 33

1.3.1 High Throughput Screening (HTS) 34

1.3.2 Fragment screening 35

1.3.3 Crystallography-based fragment screening 38

1.4 Structural biology of hGALK1 39

1.4.1 GALK1 is a sugar kinase belonging to the GHMP kinase superfamily 39

1.4.2 GHMP family of kinases 41

1.4.3 Structure of human hGALK1 44

1.4.4 Small molecule inhibitors of hGALK1 46

1.5 Structural biology of hHAO1 48
1.5.1 HAO1 belongs to the (*S*)-2-hydroxyacid oxidase family 48
1.5.2 The human 2-hydroxyacid oxidases, hHAO1 and hHAO2, are FMN-dependent 51
1.5.3 Catalytic mechanism of hHAO1 revealed by crystal structures 53
1.5.4 Small molecule inhibitors of hHAO1 55
1.6 Thesis Overview 61
2. Methods 62
2.1 Recombinant production of hHAO1 and hGALK1 protein 62
2.1.1 HAO1 PREPX protein production 62
2.1.2 HAO1 large-scale protein production 65
2.1.3 GALK1 standard protein production 66
2.1.4 GALK1 optimised protein production 66
2.2 X-ray crystallography studies of hHAO1 and hGALK1 67
2.2.1 Crystallisation trials 67
2.2.2 Crystallisation follow-up screen 68
2.2.3 Crystal soaking and mounting 70
2.2.4 Data collection, processing and refinement 71
2.3 Amplex Red fluorescence-based activity assays 71
2.3.1 HAO1 Amplex Red fluorescence-based activity assay 72
2.3.2 GALK1 Amplex Red fluorescence-based activity assay 73
2.4 Additional solution assays used to characterise hGALK1 and hHAO1 74
2.4.1 Kinase Glo luminescence-based assay to measure hGALK1 activity 74
2.4.2 Small-angle X-ray scattering (SAXS) analysis of hHAO1 76
2.4.3 Surface plasmon resonance to measure fragment binding to hHAO1 76
2.4.4 Native mass spectrometry to measure compound binding to hHAO1 76
2.5 Synthesis of hHAO1 follow-up compounds derived from fragment **6** 77
3. Structural biology of hHAO1 inhibition 78
3.1 Optimisation of hHAO1 protein production and crystallisation for drug discovery 79

3.1.1 Comparison of crystallisability of hHAO1 constructs 79
3.1.2 Scale-up of protein production and crystallisation for fragment screening 83
3.2 Expanding the structural repertoire of hHAO1 84
3.2.1 Small angle X-ray scattering (SAXS) data of purified hHAO1 85
3.2.2 Structure of hHAO1 in complex with CCPST 86
3.2.3 Structure of hHAO1 in complex with substrate 89
3.3 Fragment screening campaign: from 409 structures to 3 binding pockets 91
3.3.1 Preparing hHAO1 crystals for fragment soaking and data collection 92
3.3.2 Processing and overall analysis of fragment datasets 93
3.3.3 Fragments **1** and **2** bound in the glycolate binding pocket 96
3.3.4 Fragments **3** - **5** bound near to the gating loop 102
3.3.5 Fragments **6** and **7** bound at the subunit-subunit interface 106
4. Characterisation and optimisation of fragment hits for hHAO1 109
4.1 Characterisation of hHAO1 fragment hits in solution 109
4.1.1 Application of Amplex Red assay to hHAO1 fragment characterisation 111
4.1.2 Application of SPR assay to hHAO1 fragment characterisation 113
4.2 Follow-up compounds from active site fragments **1** and **2** 115
4.2.1 Single point assay of follow-up compounds from fragment **1** 112
4.2.2 Single point assay of follow-up compounds from fragment **2** 114
4.2.3 Concentration-response curves for selected inhibitors 117
4.2.4 Crystal structures of hHAO1 in complex with follow-up compounds 120
4.3 Follow-up compounds from fragment **5** at the gating loop 123
4.3.1 Single point assay of follow-up compounds from fragment **5** 126
4.3.2 Concentration-response curves for selected inhibitors 128
4.3.3 Crystal structure of hHAO1 after soaking follow-up compounds 130
4.4 Follow-up compounds from fragment **6** at oligomeric interface 131
4.4.1 Synthesis of follow-up compounds by poised chemistry 131
4.4.2 Single point assay of follow-up compounds from fragment **6** 134

4.4.3 Concentration-response curves for selected inhibitors 135

4.4.4 Investigation of compound effect on hHAO1 tetramer by native mass spectrometry 136

4.4.5 Crystal structure of hHAO1 after soaking follow-up compounds 140

5. Structural biology of hGALK1 inhibition 143

5.1 Optimisation of hGALK1 protein production and crystallisation for drug discovery 144

5.1.1 Expression and purification of recombinant hGALK1 from *E. coli* 144

5.1.2 Initial crystal trials for hGALK1 in HEPES buffer 145

5.1.3 Optimisation of buffer conditions to improve diffraction of hGALK1 crystals 146

5.1.4 Crystal trials of hGALK1 in phosphate buffer 149

5.2 Crystal structures of hGALK1 liganded complexes 153

5.2.1 Overall description of datasets collected with known ligands of hGALK1 153

5.2.2 Structure of hGALK1 in complex with galactose and ADP 154

5.2.3 Structures of hGALK1 in complex with tool compounds **T1**, **T2**, **T3** and **T7** 157

5.3 Fragment screening campaign: from 180 structures to 3 binding pockets 160

5.3.1 Preparing hGALK1 crystals for fragment soaking and data collection 160

5.3.2 Processing and overall analysis of the fragment data sets 162

5.3.3 Fragments **1** and **2** bind at the entrance to the ATP binding site 167

5.3.4 Fragments **3-5** bind close to where the protein termini converge 168

5.3.5 Fragments **6-28** bind in a binding hotspot distal to the active site 170

6. Characterisation and elaboration of hGALK1 fragment hits 180

6.1 Validation of two activity assays for the detection of hGALK1 inhibition 180

6.1.1 Principle of the Kinase Glo activity assay 181

6.1.2 Principle of the Amplex Red activity assay 182

6.1.3 Kinetic properties of hGALK1 determined by the Kinase Glo assay 183

6.1.4 Kinetic properties of hGALK1 determined by the Amplex Red assay 184

6.1.5 Measurement of inhibition of hGALK1 by tool compounds 186

6.1.6 Determination of the mode of hGALK1 inhibition for tool compound **T1** 189

6.2 Testing effects of fragments on hGALK1 activity 192

6.2.1 Single concentration screen of fragments against hGALK1 by Kinase Glo.....192
6.2.2 Determination of inhibition mode for fragment hits **4** and **8**.....195
6.3 Merging of fragments from the binding hotspot.....198
6.3.1 Design and selection of merged compounds.....199
6.3.2 Single concentration screen of merged compounds in hGALK1 activity assays...200
6.3.3 Characterisation of merged compounds **M1-M3** that inhibit hGALK1.....201
6.3.4 Characterisation of merged compound **M4** in hGALK1 activity assays.....205
6.3.5 Characterisation of merged compound **M5** in hGALK1 activity assays.....209
7. Discussion and Conclusions.....213
7.1 Structural biology and small molecule effectors of hHAO1.....213
7.2 Structural biology and small molecule effectors of hGALK1.....219
7.3 Potential of substrate reduction therapy for other metabolic disorders.....230
8. References.....233
Appendix 1: Publications.....251
Classic galactosaemia/GALK1.....251
Primary hyperoxaluria/HAO1.....251
Other targets.....251
Appendix 2: Crystallographic refinement statistics.....253
Appendix 3: Characterisation of synthesised compounds.....263

List of Figures

Figure 1.1: Flow chart summarising the consequences of loss-of-function mutations that cause inborn errors of metabolism. 16

Figure 1.2: Chemical structures of the stereoisomers of galactose 18

Figure 1.3: Pathway diagram of hepatic galactose metabolism. 19

Figure 1.4: Pathway diagram and chemical structures of key metabolites of hepatic glyoxylate metabolism. 22

Figure 1.5: Flow chart illustrating therapeutic modalities for the treatment of an inborn error of metabolism in which loss-of-function mutations lead to a metabolic block 24

Figure 1.6: Simplified flow chart illustrating the drug discovery process. 31

Figure 1.7: Cartoon representation comparing library selection, hit discovery and hit-to-lead development in HTS and fragment screening 32

Figure 1.8: Representative structures of the three human sugar kinase fold groups 37

Figure 1.9: Comparison of human GHMP kinases and of galactokinases from different species. 39

Figure 1.10: Key structural features of hGALK1. 42

Figure 1.11: Evolution of spiro-benzoxazole series of hGALK1 inhibitors from an HTS campaign hit to low-μM lead compounds 44

Figure 1.12: Sequence alignment of representative members of the (S)-2-hydroxyacid oxidase family 46

Figure 1.13: Superimposed structures of representative members of the (S)-2-hydroxyacid oxidase family 47

Figure 1.14: Sequence alignment of hHAO1 and hHAO2 orthologues. 49

Figure 1.15: Surface and ribbon representations of the hHAO1 tetramer observed in crystal structures 50

Figure 1.16: Close-up views of the hHAO1 active site and gating loop 51

Figure 1.17: Chemical structures of HAO inhibitors for which co-crystal structures have been determined to elucidate their binding modes. 52

Figure 1.18: Close-up of published crystal structures showing stick and line representations of hydroxyacid oxidase active site residues with bound inhibitor TACA. 53

Figure 1.19: Close-up of published crystal structures showing stick and line representations of hydroxyacid oxidase active site residues with bound inhibitor CDST 54

Figure 1.20: Close-up of published crystal structures showing stick and line representations of hydroxyacid oxidase active site residues with bound inhibitor CCPST. 56

Figure 1.21: Close-up of published crystal structures showing stick and line representations of hydroxyacid oxidase active site residues with bound inhibitor TKP. 57

Figure 2.1: Nested constructs of hHAO1. 60

Figure 2.2: Chemical structures of tool compound **T1** – **T7**, known inhibitor series for hGALK1. 65

Figure 3.1: Nested constructs of hHAO1. 77

Figure 3.2: Purification of nine hHAO1 constructs by an adapted PREPX protocol. 78

Figure 3.3: Representative crystal images of different hHAO1 constructs. 80

Figure 3.4: Size exclusion chromatogram, SDS-PAGE and representative crystal images for large scale purification of hHAO1-c002. 81

Figure 3.5: Small angle X-ray scattering (SAXS) of purified hHAO1-c002. 83

Figure 3.6: Comparison of hHAO1-FMN-glyoxylate and hHAO1-FMN-CCPST crystal structures 84

Figure 3.7: Comparison of electron densities for hydroxyacid oxidase structures with bound ligand 86

Figure 3.8: Comparison of hHAO1-FMN-glycolate, hHAO1-FMN-glyoxylate and hHAO1-FMN- sulphate. 88

Figure 3.9: Overview of hHAO1 fragment screening campaign. 89

Figure 3.10: Overview of fragments binding to hHAO1 crystals 92

Figure 3.11: Fragments **1** and **2** bind in the glycolate binding pocket of the active site 94

Figure 3.12: Fragments **3** and **4** bind at the protein surface close to the gating loop. 97

Figure 3.13: Fragment **5** binds in a hydrophobic pocket partially formed by the gating loop…. 98

Figure 3.14: Sequence alignment of representative members of the (S)-2-hydroxyacid oxidase family highlighting residues involved in fragment binding 100

Figure 3.15: Fragments **6** and **7** bound at the interface between subunits of the hHAO1 tetramer 102

Figure 4.1: Principle of the assays used to characterise the effect of fragments on hHAO1 in solution 104

Figure 4.2: Determination of Amplex Red assay parameters for the kinetic characterisation of hHAO1 in vitro activity 105

Figure 4.3: Inhibition of hHAO1 by fragment hits using Amplex Red activity assay 107

Figure 4.4: Measurement of CCPST and fragment binding to immobilised hHAO1 in solution by SPR 108

Figure 4.5: Chemical structures of follow-up compounds derived from active site fragments 1 and 2..... 110

Figure 4.6: Single-point inhibition of hHAO1 by follow-up compounds derived from active site fragment **1**, measured using Amplex Red activity assay 113

Figure 4.7: Single-point inhibition of hHAO1 by follow-up compounds based on active site fragment **2**, measured using Amplex Red activity assay 116

Figure 4.8: Inhibition of hHAO1 by follow-up compounds based on active site fragments, measured using Amplex Red activity assay..... 119

Figure 4.9: Crystal structures of hHAO1 bound with follow-up compounds **19** and **47**, derived from active site fragments **1** and **2** respectively..... 122

Figure 4.10: Chemical structures of follow-up compounds based on fragment **5** at the gating loop. 124

Figure 4.11: Single-point inhibition of hHAO1 by follow-up compounds derived from fragment **5**, measured using Amplex Red activity assay..... 127

Figure 4.12: Inhibition of hHAO1 by follow-up compounds derived from fragment **5**, measured using Amplex Red activity assay..... 129

Figure 4.13: Crystal structures of hHAO1 bound with follow-up compound **79**, fragment **5** and glycolate..... 130

Figure 4.14: Synthesis of follow-up compounds based on fragment **6**..... 132

Figure 4.15: Single-point inhibition of hHAO1 by follow-up compounds based on fragment **6**, measured using Amplex Red activity assay..... 135

Figure 4.16: Inhibition of hHAO1 by follow-up compounds derived from fragment **6**, measured using Amplex Red activity assay..... 136

Figure 4.17: Native mass spectrometry analysis of hHAO1..... 137

Figure 4.18: Comparison of native mass spectra of hHAO1 in the presence of DMSO and follow-up compounds **123** – **135**..... 138

Figure 4.19: Comparison of native mass spectra of hHAO1 in the presence of DMSO, compound **125** and compound **126**..... 139

Figure 4.20: Crystal structures of hHAO1 bound with fragment **6**, follow-up compound **123**, and glycolate..... 141

Figure 5.1: SDS-PAGE, size exclusion chromatogram and intact mass spectrum of representative hGALK1 purification..... 145

Figure 5.2: Crystal images and maximal resolution for crystals of hGALK1 purified in HEPES buffer..... 146

Figure 5.3: Measurement of buffer and substrate effects on hGALK1 thermal stability by DSF 149

Figure 5.4: Optimisation of hGALK1 crystals in Morpheus-A9 condition 150

Figure 5.5: Co-crystallisation of hGALK1 with benzo-spiroxazole inhibitors produces robust crystals 153

Figure 5.6: Comparison of hGALK1-galactose-AMPPNP and hGALK1-galactose-ADP crystal structures. 156

Figure 5.7: hGALK1 structures in complex with spiro-benzoxazole tool compounds 158

Figure 5.8: Close-up views of the hGALK1 ATP binding site with bound tool compounds. 160

Figure 5.9: Overview of hGALK1 fragment screening campaign 161

Figure 5.10: Overlay of protein chains from various hGALK1 structures 163

Figure 5.11: Fragment hits excluded from further analysis 164

Figure 5.12: Fragment chemical structures and their binding sites from a screening campaign against hGALK1-galactose-**T3** crystals. 166

Figure 5.13: Two fragments bind close to compound **T3**, at the entrance to the ATP binding site 167

Figure 5.14: Three fragments bind in a pocket close to where the C and N termini of hGALK1 meet. 169

Figure 5.15: Twenty-two fragments bind at a binding hotspot distal to the active site 171

Figure 5.16: Fragments in the binding hotspot share common structural motifs and can be grouped to four binding poses. 173

Figure 5.17: Binding interactions of group 1 fragments at the binding hotspot. 174

Figure 5.18: Binding interactions of group 2 fragments at the binding hotspot. 176

Figure 5.19: Binding interactions of group 3 fragments at the binding hotspot. 177

Figure 5.20: Binding interactions of group 4 fragments at the binding hotspot 178

Figure 6.1: Principle of hGALK1 activity assays used in this work. 181

Figure 6.2: Confirmation of Kinase Glo assay parameters for the kinetic characterisation of hGALK1 *in vitro* activity 183

Figure 6.3: Optimisation of Amplex Red assay for the kinetic characterisation of hGALK1 *in vitro* activity 185

Figure 6.4: Inhibition of hGALK1 by spiro-benzoxazole tool compounds **T1**- **T7**. 187

Figure 6.5: Characterisation of *in vitro* inhibition of hGALK1 by tool compounds. 188

Figure 6.6: Kinetic characterisation of hGALK1 inhibition by compound **T1** 191

Figure 6.7: Change in hGALK1 activity in the presence of 5 mM fragment measured in Kinase Glo assay 193

Figure 6.8: Kinetic characterisation of hGALK1 inhibition by fragment **4** 196

Figure 6.9: Kinetic characterisation of hGALK1 inhibition by fragment **8** 197

Figure 6.10: Illustration of approaches to fragment optimisation. 198

Figure 6.11: Merging of fragments from the binding hotspot 199

Figure 6.12: Measurement of hGALK1 activity in the presence of 1 mM merged compounds **M1**-**M5** 200

Figure 6.13: Characterisation of in vitro inhibition of hGALK1 by merged compounds **M1**- **M3** 202

Figure 6.14: Kinetic characterisation of hGALK1 inhibition by merged compounds **M1**-**M3** 204

Figure 6.15: *In vitro* activation of hGALK1 by merged compound **M4** 205

Figure 6.16: Characterisation of the substrate-dependence of hGALK1 activation by **M4** 207

Figure 6.17: Kinetic characterisation of hGALK1 activation by **M4** 208

Figure 6.18: Characterisation of *in vitro* activation and inhibition of hGALK1 by merged compound **M5** 209

Figure 6.19: Kinetic characterisation of GALK1 inhibition by **M5** 210

Figure 7.1: Potential strategies to optimise active site inhibitors of hHAO1 213

Figure 7.2: Proposed synthetic strategy to optimise gating loop inhibitors of hHAO1. 215

Figure 7.3: Potential strategies to merge fragments at the entrance to the ATP binding site of hGALK1. 220

Figure 7.4: Potential strategies to optimise protein termini site inhibitors of hGALK1 222

Figure 7.5: Sequence alignment of GHMP kinase family members comparing secondary structures at the binding hotspot of hGALK1 224

Figure 7.6: Structural models of hGALK1 variants proposed to investigate the novel binding pockets described in this work 227

List of tables

Table 2.1: Condition details for follow-up crystallisation screens. 66

Table 3.1: Purified protein details for hHAO1 constructs 79

Table 4.1: Chemical properties for active site follow-up compounds 112

Table 4.2: Chemical properties for follow-up compounds derived from fragment **5** 126

Table 4.3: Chemical properties for follow-up compounds derived from fragment **6** 133

Table 5.1: Buffer screen for hGALK1 by DSF 148

Table 5.2: Tool compounds from spiro-benzoxazole series of hGALK1 inhibitors 151

1. Introduction

Inborn errors of metabolism (IEMs) are a group of more than 1000 inherited single gene disorders [1] that are individually rare but collectively occur in more than 1:2000 live births [2] with up to 4-fold greater occurrence in high risk populations [3], [4]. Drug development for these disorders is challenging, due to their individual rarity and genetic heterogeneity and the lack of precedence in developing therapies for loss-of-function disorders. The development of target-specific assays and recruitment of sufficient patient numbers for clinical trials, prerequisites of standard drug discovery campaigns, is both difficult and expensive in IEMs, given the rarity and diversity of these disorders.

1.1 Inborn errors of metabolism perturb function by blocking a specific metabolic pathway

The term 'inborn errors of metabolism' was coined by Archibald Garrod in 1908 to describe a group of four disorders – alkaptonuria, albinism, cystinuria and pentosuria - that were caused by "the congenital lack of some particular enzyme in the absence of which a step is missed and some normal metabolic change fails to be brought about" (figure 1.1) [5], [6].

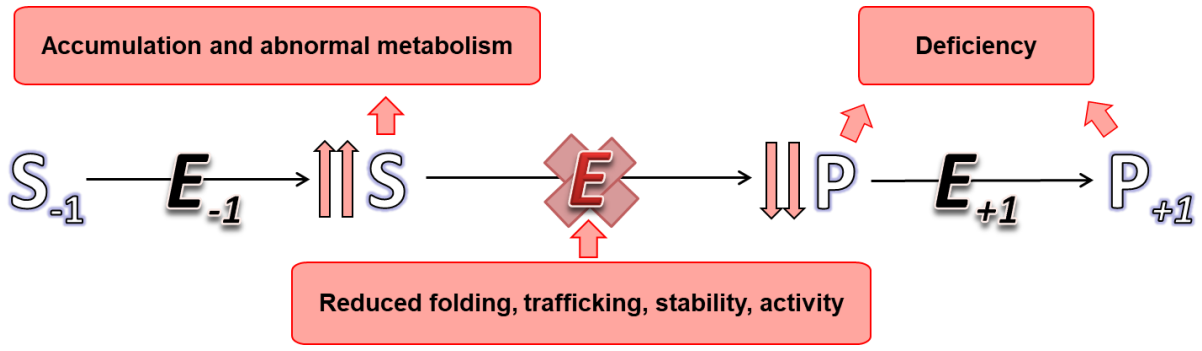


Figure 1.1: Flow chart summarising the consequences of loss-of-function mutations that cause inborn errors of metabolism. *Pathogenic variants reduce substrate turnover so that the substrate accumulates (left) and there is a shortage of the downstream metabolites (right).*

This definition served for more than 100 years during which time the number of reported disorders has grown exponentially. A recent formal classification of IEMs, expanded to include mutations in accessory proteins such as transporters, chaperones and epigenetic regulators, describes a total of 1239 disorders that disrupt normal metabolism [1], [7]. The majority are inherited in an autosomal recessive manner (80%), and almost all result in the loss of enzyme function (97%,

consequences illustrated in figure 1.1). A simplified classification of IEMs was later proposed to group disorders into 5 classes based on their pathophysiology [8]. These include toxic accumulation of small molecules (e.g. amino acids, sugars, nucleotides); impaired synthesis or transport of essential small molecules; toxic accumulation of complex macromolecules (e.g. glycogen, bile acids, triglycerides); impaired synthesis and recycling of complex molecules; and finally energy defects that impair the synthesis, transport or utilisation of energetic molecules (e.g. glucose, pyruvate, NADPH, coenzyme Q) [8].

The two IEMs studied in this work, classic galactosemia and primary hyperoxaluria, are small molecule accumulation disorders, where pathogenesis stems from the toxicity of accumulated substrate(s) preceding the defective step.

1.1.1 Galactose metabolism primarily occurs in the liver via the Leloir pathway

Galactose is a monosaccharide that occurs predominantly in its D-configuration as both α and β anomers [9] (figure 1.2). Galactose is an essential energy source, particularly in young animals with a milk-based diet and has a key structural role in glycolipids and glycoproteins, for example in galactocerebroside, the predominant lipid in myelin. The majority of exogenous galactose comes from the hydrolysis of dietary lactose, from milk and other dairy products, into β-galactose and α- or β-glucose by lactase (EC3.2.1.108) in the intestinal lumen [10], [11]. The primary endogenous sources of galactose are the hydrolysis of glycoproteins, proteoglycans and glycolipids in the lysosome [12], [13]. Endogenous galactose production is highest in infants and young children [12] and has been suggested to peak in the prenatal period [9], [14]. Free galactose, predominantly in the β-anomer, is transported into the liver where it is epimerised to α-galactose by galactose mutarotase (GALM, EC5.1.1.3) [15], for entry into the Leloir pathway, the main route of galactose metabolism [16].

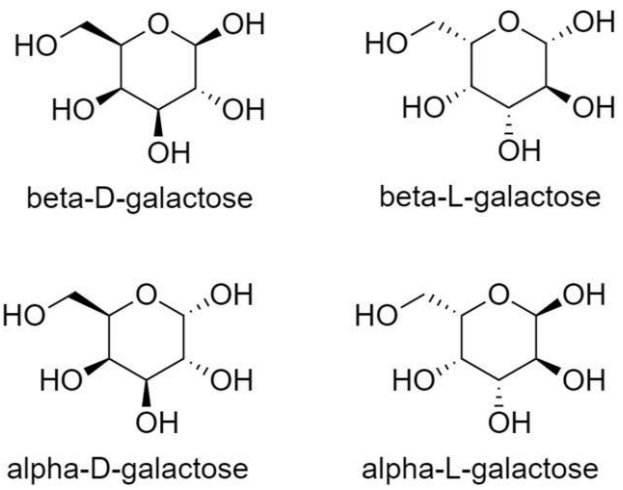


Figure 1.2: Chemical structures of the stereoisomers of galactose.

There are three pathways of galactose metabolism (figure 1.3). Firstly, galactose can be reduced to galactitol by aldose reductase (EC1.1.2.14), using NADPH, in the polyol pathway [17], [18]. Secondly, galactose can be oxidised to galactonate, by currently uncharacterised mechanism(s) [19]-[21]. Galactitol cannot be further metabolised, does not diffuse well and is prone to accumulate in cells, whereas galactonate can be excreted in urine or further metabolised to enter the pentose phosphate pathway via xylulose. Finally, galactose can be converted to glucose-1-phosphate and UDP-galactose via the Leloir pathway (figure 1.3) [22].

The Leloir pathway consists of three enzymes that convert α-galactose to glucose-1-phosphate for further metabolism by the glycolytic, gluconeogenic and pentose phosphate pathways, and to UDP-galactose which is used as the source of galactose for galactosylation reactions (figure 1.3, blue box) [16], [23]-[26]. The first step is the ATP-dependent phosphorylation of the 1 position hydroxyl of galactose by galactokinase 1 (GALK1; EC 2.7.1.6). The second step is an uridylyl transfer from UDP-glucose to galactose-1-phosphate, catalysed by galactose-1- phosphate uridylyltransferase (GALT; EC 2.7.7.12), to form glucose-1-phosphate and UDP- galactose. In the third step, UDP-glucose is regenerated from UDP-galactose by UDP- galactose-4-epimerase (GALE; EC 5.1.3.2) [27].

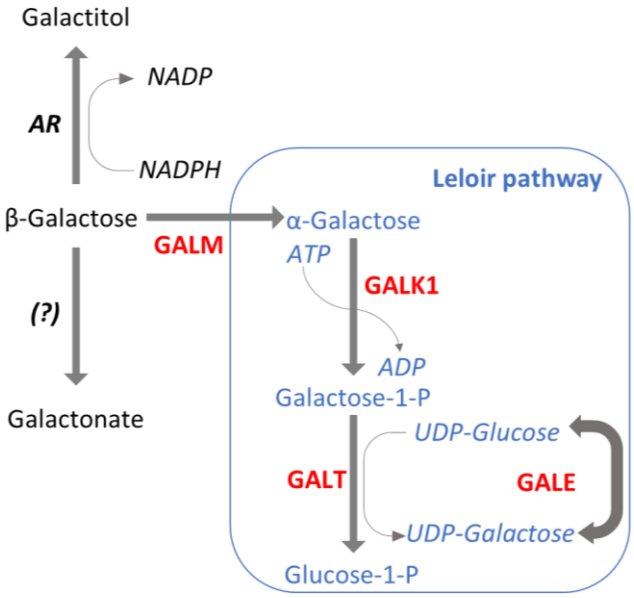


Figure 1.3: Pathway diagram of hepatic galactose metabolism. *Galactose can be oxidised to galactonate, reduced to galactitol or metabolised via the Leloir pathway. The Leloir pathway is outline in blue and the names of enzymes with reported disease-causing loss-of-function variants are shown in red.*

1.1.2 Classic galactosemia is a disorder of galactose metabolism

Classic galactosaemia is caused by inherited loss of function in any one of the three Leloir pathway enzymes (reviewed in [28]). Type 1 classic galactosaemia (OMIM 230400), caused by mutations in the *GALT* gene [29]-[31], is the most common form with an estimated global prevalence of 1:48'000 live births [32]. A 100-fold higher prevalence of type I classic galactosaemia is observed in the Irish Traveller population (prevalence of 1:480, in contrast to 1:30'000 in the non-Traveller Irish population) and the only GALT variant reported in this population is p.Q188R (cf. 89% of non-Traveller Irish patients), the most common variant to cause classic galactosaemia [33]. Patients with type 1 galactosaemia sicken in the neonatal period, with liver, kidney, intestinal and central nervous system toxicity caused by the high galactose content in human and formula milk [31], [32]. If lactose is not removed from the patient's diet, progressive liver and brain damage lead to death or severe disability.

Type 2 classic galactosaemia, caused by mutations in the *GALK1* gene [34], [35], has a population-dependent incidence rate ranging from 1: 1'000'000 to 1:52'000 live births [36]. An accurate estimate of its exact prevalence is difficult to obtain as many patients remain symptom

free until their third or fourth decade of life, and often the only clinical indication is presenile cataract formation, caused by the build-up of galactitol in the lens cells of the eye [37]. Other rarer forms of galactosemia are caused by mutations in the *GALE* gene (type 3 galactosaemia, also called generalised GALE deficiency to differentiate from the mostly asymptomatic peripheral form; OMIM 230350), reported in 7 patients [38]-[42], and by recently identified mutations in *GALM* gene, reported in 8 patients [43].

All four types of classic galactosaemia include cataract formation as one of the clinical outcomes. This is due to conversion of galactose to galactitol in the lens cells of the eye by the unusually high levels of aldose reductase in these cells [44]. Accumulated galactitol cannot be further metabolised or transported out of the lens cells and consequently it accumulates, causing cellular swelling and apoptosis [44].

Given the mild phenotype of GALK1 and GALM deficiencies relative to GALT and GALE deficiencies, there is good evidence for ascribing pathogenicity to the accumulation of galactose-1-phosphate, but no clear mechanism of action has been delineated. Various *in vitro* studies have shown that galactose-1-phosphate inhibits other key metabolic enzymes including UDP-glucose pyrophosphorylase [45], inositol monophosphatase [46] and galactosyltransferases [47]. The Leloir pathway is conserved from bacteria to humans and several disease models based on knockout or inactivation of GALT genes have been developed and these *in vivo* systems are essential in understanding the pathophysiology of galactosaemia. These models include knockout of hGALT orthologues in yeast (*S. cerevisiae Gal7* gene, [46], [48], [49]), fruit fly (*D. melanogaster dGALT* gene, [50], [51]), zebrafish (*D. rerio galt* gene [52]) and mouse (*M. musculus GalT* gene [21], [53], [54]) models.

1.1.3 Glyoxylate metabolism primarily occurs in the liver via peroxisomal and mitochondrial pathways

Glyoxylate is a highly reactive two-carbon keto-acid generated from mitochondrial or peroxisomal intermediary protein metabolism in the hepatocytes (chemical structures of glyoxylate and related intermediates are shown in figure 1.4A) [55]. In the mitochondria, hydroxyproline, resulting from the degradation of collagen and dietary protein (e.g. meat), is metabolised to glyoxylate (figure 1.4B, yellow shaded box) through sequential catalysis by 5 enzymes. Hydroxyproline dehydrogenase (PRODH2; EC 1.5.5.2) catalyses the FAD-dependent dehydrogenation of hydroxyproline to pyrroline-5-carboxylate [56], which hydrolyses spontaneously to 4-hydroxy-glutamate semialdehyde. Delta-1-pyrroline-5-carboxylate dehydrogenase (ALDH4A1; EC 1.5.1.12) catalyses the NAD-dependent dehydrogenation of 4-hydroxy-glutamate semialdehyde to 4-hydroxy-glutamate [57]. Aspartate aminotransferase (GOT2; EC 2.6.1.1) catalyses the PLP-dependent transamination of 4-hydroxy-glutamate to 4-hydroxy-2-oxoglutarate [58], which is cleaved to glyoxylate and pyruvate by 4-hydroxy-2-oxoglutarate aldolase (HOGA; EC 4.1.3.16) [59]. Glyoxylate is reduced to glycolate by the activity of glyoxylate reductase – hydroxypyruvate reductase (GRHPR; EC 1.1.1.79) [60], [61].

Glycolate, glyoxylate and glycine can all diffuse freely into and out of the peroxisome where enzymatic interconversion occurs (figure 1.4B, blue shaded box). Glycolate is oxidised to glyoxylate in the peroxisome by FMN-dependent 2-hydroxyacid oxidase 1 (HAO1; EC 1.1.3.15), also known as glycolate oxidase [62]. Glyoxylate is used in the PLP-dependent transamination reaction with alanine to produce glycine and pyruvate, catalysed by alanine- glycine aminotransferase (AGXT; EC 2.6.1.44) [63] or further oxidised by lactate dehydrogenase (LDH-A; EC 1.1.1.27) in the hepatocyte cytosol to form oxalate.

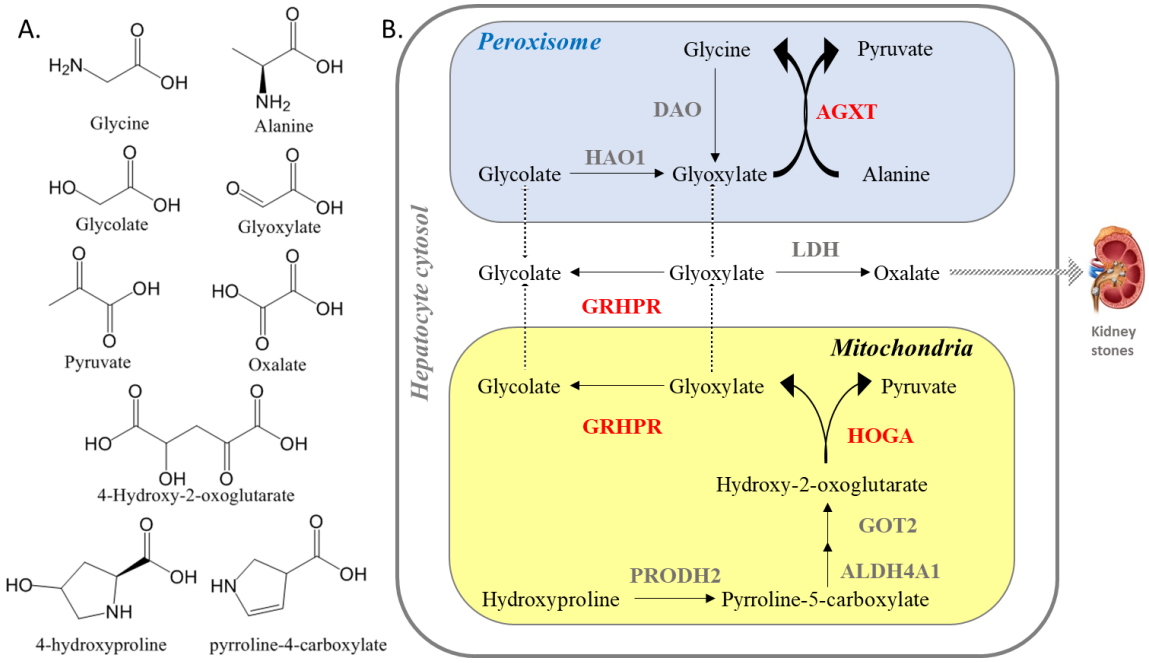


Figure 1.4: Pathway diagram and chemical structures of key metabolites of hepatic glyoxylate metabolism. *A. Chemical structures of metabolites involved in glyoxylate metabolism that are discussed in the text. B. Pathway diagram showing routes of glyoxylate synthesis and breakdown that take place within the hepatocyte. The names of enzymes for which loss-of-function variants are associated with primary hyperoxaluria are in red text and the remaining enzymes are in grey text. Reactions that take place in the peroxisome are shaded blue and reactions that take place in the mitochondria are shaded yellow. Dashed lines indicate passive transport and solid lines indicate enzyme-catalysed reactions.*

1.1.4 Primary hyperoxaluria – disorder of glyoxylate metabolism

In primary hyperoxaluria, inherited mutations in glyoxylate-metabolising enzymes cause glyoxylate to accumulate in the liver. Accumulated glyoxylate is readily converted to oxalate by lactate dehydrogenase subunit A (LDH-A), an abundant enzyme in the hepatocyte cytosol (figure 1.4B). Oxalate, a metabolic end-product with limited solubility, is then transported to the kidneys for excretion. When oxalate levels exceed the renal filtration rate, calcium oxalate crystals deposit in kidney tissues, compromising renal function [64]. When renal function is sufficiently impaired, systemic oxalate deposition (oxalosis) occurs, causing life-threatening damage to cardiovascular and skeletal tissues [65]. Primary hyperoxaluria is divided into three types, according to the mutated gene. Type 1 primary hyperoxaluria (PH1; OMIM 259900) is caused by mutations of peroxisomal *AGXT* [66]; type 2 (PH2; OMIM 260000) is caused by mutations of *GRHPR* [67], a ubiquitous protein present in both the cytosol and mitochondria of

hepatocytes; and type 3 (PH3; OMIM 613616) is caused by mutations of *HOGA1* [68], expressed in the mitochondria of both hepatic and renal cells.

PH1 accounts for approximately 80% of primary hyperoxaluria cases with an estimated occurrence of 1:100'000-150'000 live births [69]–[71]. The majority of identified *AGXT* mutations result in single nucleotide substitutions that abolish or severely reduce functional AGXT protein levels in the peroxisome (reviewed in [72]). Most of these variants show significant residual activity *in vitro* but minimal activity *in vivo* because they result in mistargeting of catalytically active AGXT protein to the mitochondria instead of the peroxisome [73]–[75] and/or cause aggregation of the variant protein [76], [77]. Clinical symptoms in PH1 patients can vary from end stage renal failure during infancy to onset of progressive renal inflammation and fibrosis in early adulthood, with a median age of onset of 5 years old [71]. Prognosis is generally poor, with end stage renal failure occurring within 3-4 decades of life.

PH2 accounts for a further 10% of primary hyperoxaluria patients [78] with an estimated occurrence of 1:310'000 live births [71]. The clinical phenotype of PH2 is similar to PH1 but is generally less severe with only around 20% of PH2 patients develop end stage renal failure [79], [80]. PH3 is the most recently identified, rarest form of primary hyperoxaluria and has a significantly milder phenotype, generally consisting of recurrent urinary tract infections and/or kidney stones but little impairment of kidney function [81].

The liver-specificity of AGXT means that, unlike many other metabolic disorders, patient fibroblasts are not suitable for studying primary hyperoxaluria. Consequently, several alternative disease models have been constructed to explore the pathogenesis of, and potential treatment approaches to, primary hyperoxaluria. Initial animal models of primary hyperoxaluria involved induction of high oxalate levels in rats by supplementing their drinking water with the glyoxylate precursor ethylene glycol [82]–[84]. An alternative model, inducing hyperoxaluria by feeding rats chow that contained high levels of hydroxyproline, has also been developed [85]. More

recent models have employed genetic knock out of AGXT orthologues in yeast (*S. cerevisiae Agx1* gene [76]), fruit fly (*D. melanogaster dAGXT* gene [86]) and mouse (*M. musculus Agxt1* gene [87]). Finally, recent studies have employed a Chinese hamster ovary (CHO) cell model that lacks an endogenous glyoxylate pathway but is transformed with hHAO1, hAGXT (wild type or disease variants) and/or hGRHPR [88]–[90].

1.2 Treatment of inborn errors of metabolism

The goal in treating an inborn error of metabolism is to normalise the levels of metabolites within the affected pathway. On the basis that a metabolic enzyme defect generally forms a metabolic block, accumulating preceding substrates and depleting succeeding products, there are multiple routes to reach this goal (figure 1.5).

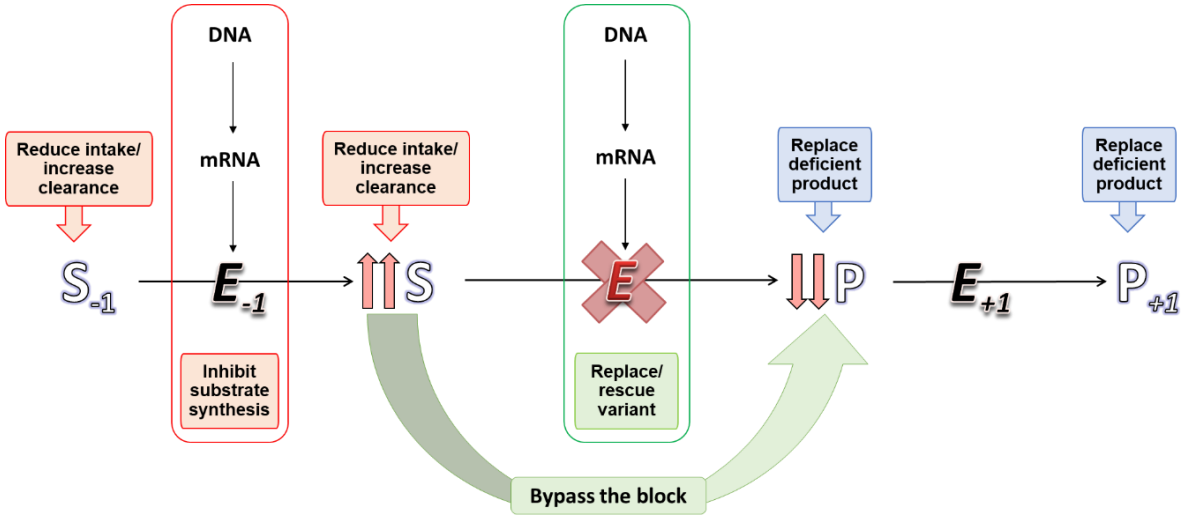


Figure 1.5: Flow chart illustrating therapeutic modalities for the treatment of an inborn error of metabolism in which loss-of-function mutations lead to a metabolic block. *Therapies may clear accumulating substrate/s (red), replace depleted product/s (blue) or mitigate the block itself (green).*

These routes include reducing the accumulation of toxic substrate(s) of the affected enzyme (substrate reduction, dietary restriction), metabolising the accumulating substrates through an alternative reaction (metabolic bypass) and/or supplementing deficient metabolites downstream of the metabolic block. Alternatively, therapies may aim to rescue the activity of the defective protein by stabilising residual variant enzyme (pharmacological chaperones, proteostasis and proteasome modulators) or by replacing it with functional enzyme (gene therapy, enzyme replacement and organ transplantation). As understanding of the causative gene mutations and

variant proteins involved in IEMs has increased, a wealth of alternative therapeutic approaches have been developed [91]–[93]. For galactosemia and primary hyperoxaluria, the following four approaches have been intensively explored.

1.2.1 Dietary therapy aims to restore steady state balance of the affected pathway

Dietary therapy, applicable in most IEMs, aims to reduce dietary sources of the accumulating substrate(s) and/or their precursors and to supplement essential downstream metabolites that are deficient. The classic example of this approach is the dietary restriction of phenylalanine as a transformative therapy for phenylketonuria (PKU) patients, in which loss- of-function mutations in phenylalanine hydroxylase (PAH) cause an accumulation of phenylalanine [94], [95]. Dietary therapy ameliorates the progressive neurological degeneration observed in the absence of such measures and has been applied for over six decades [96], [97]. In some cases, residual enzyme activity can be increased by supplementing vitamins or cofactors of the native enzyme [98]–[100].

For galactosemia, dietary restriction of galactose (and its precursor lactose) is the mainstay treatment [32], [101]. If a galactose-restricted diet is followed, the probability of neonatal survival is very high, although long-term complications occur even in cases where galactose intake is restricted from birth [102]. Specifically, despite dietary compliance, 85% of patients develop neurological impairments (e.g. developmental delay, speech disorders) and 80% of female patients develop primary ovarian insufficiency [31]. Additionally, the galactose- restricted diet is burdensome, negatively impacting the quality of life of both patients and their family and leading to non-compliance issues [102], [103].

For primary hyperoxaluria, dietary measures to reduce oxalate crystal formation include: high fluid intake to maintain dilute urine; supplementation of magnesium or potassium citrate and pyrophosphates to increase oxalate solubility and excretion; and prevention of calcium deficiency to avoid increased intestinal oxalate absorption [104]. Dietary restriction of oxalate

is generally considered unnecessary as the contribution of exogenous oxalate to overall oxalate levels is negligible relative to endogenously produced oxalate [105]. The precursors to oxalate (hydroxyproline, glycine and glycolate) are impractical targets for restriction since they come from metabolism of a wide range of essential nutrients.

1.2.2 Organ transplantation aims to replace the defective pathway

Organ transplantation can be curative in disorders where the defective protein and/or resultant phenotype is organ specific, e.g. liver transplantation corrects underlying defects of hepatocyte-specific enzymes [106]. However, after transplantation patients require life-long immunosuppression and are at risk of several serious complications and consequently the benefits of organ transplantation must be carefully weighed against the risks incurred.

For galactosemia, liver transplantation to replace the major source of defective galactose metabolism is generally considered too high risk, because acute manifestations are prevented by dietary restriction while chronic symptoms do not decrease life expectancy [107]. Additionally, despite primary localisation of Leloir metabolism in the liver, a case report describing liver transplant in a patient who had suffered liver failure as a result of undiagnosed galactosaemia found that transplantation lowered but did not normalise galactose-1-phosphate and galactitol levels, indicating appreciable galactose turnover outside of the liver [108].

For primary hyperoxaluria, a combined liver and kidney transplant is recommended to replace both the source of defective glyoxylate metabolism (liver) and the organ most damaged by oxalate accumulation (kidney). Isolated kidney transplantation is not recommended as the new organ would simply begin accumulating oxalate [109], [110]. The lack of genotype-phenotype correlation between AGXT mutations raises concerns about pre-emptive liver transplantation: this would be curative, but the high risk associated with lifelong immunosuppression and transplant rejection must be weighed against the likelihood of developing severe disease [111].

1.2.3 Alternative methods to restore defective protein function

A more conservative method, relative to organ transplantation, is to restore the missing metabolic function by provision of the wild type, or correction of the defective, gene or protein. Pre-clinical studies have shown successful correction of galactosaemia in GALT-deficient mouse by provision of wild type GALT mRNA [112] and successful correction of an aberrant splice site GALT mutation in patient cells by locked nucleic acid oligonucleotides that shield the new splice site, allowing translation of a functional GALT transcript (antisense therapy) [113]. For primary hyperoxaluria, phenotypic rescue of an AGXT deficient mouse model has been demonstrated upon transient expression of hAGXT [87] or long-term expression of hAGXT mediated by adeno-associated viral (AAV) vector transfer [114] or helper-dependent adenoviral vectors [115].

Small molecules have also been used in proof-of-concept and pre-clinical studies to restore defective enzyme functions. Based on their modes of action, these small molecules can be broadly classified as: (1) proteostasis regulators that modulate protein folding and protect cells against protein degradation, ER stress and cellular autophagy; (2) chemical chaperones that act non-specifically to stabilise aggregation-prone proteins through osmotic mechanisms; and (3) pharmacological chaperones that bind specifically to the defective protein to facilitate its expression, folding and trafficking.

For galactosemia, induction of endoplasmic reticulum stress upon galactose challenge of GALT-deficient organisms have been shown in yeast [116], Drosophila [117], mouse [118] and patient cells [119]. Treatment of GALT-knockout mice with Salubrinal, a proteostasis regulator that protects against ER stress through induction of eukaryotic translation initiation factor 2 subunit alpha (eIF2α) phosphorylation, slowed the development of cerebellar ataxia and subfertility [54], [118]. More than 80% of GALT variants are predicted to be less stable than wild type [120] and misfolding and aggregation has been shown experimentally for several of the more common GALT variants [121], [122]. Arginine, as a chemical chaperone, stabilises

the most common GALT mutation (p.Q188R) when expressed recombinantly in *E. coli* [123] but had no effect in patient fibroblasts or on patients ([124], NCT03580122).

For primary hyperoxaluria, four of the most common missense AGXT variants (p.F152T, p.G170R, p.I244T and p.P319L) demonstrate a mis-targeting phenotype i.e. catalytically active AGXT protein is transported to the mitochondria rather than the peroxisome. Mitochondrial mis-targeting has been corrected in CHO cell models by repurposing approved drugs known to inhibit mitochondrial import (dequalinium chloride [125]; monesin, salinomycin and nigericin [126]), thus biasing AGXT fate towards the peroxisome. Mis-targeted variants also show increased interactions with molecular chaperones and decreased kinetic stability [127]. Correct peroxisomal targeting of variant proteins expressed in CHO cells was partially restored by reduced growth temperature, supporting reduced variant stability as a contributory mechanism in protein mis-targeting [128]. Mis-targeted variants can also be partially rescued by supplementation with pyridoxine, the metabolically stable precursor to PLP, a cofactor required for AGXT activity [129]–[131]. This has now become an approved pharmacological chaperone therapy for primary hyperoxaluria, effective in approximately 30% of PH1 patients [129], [132].

1.2.4 Evolution from dietary supplementation to other substrate reduction therapies

Substrate reduction therapy (SRT) encompasses a broad range of therapeutic modalities that aim to reduce the bioavailability of the compound for which normal metabolism is impaired in the target disease to ameliorate clinical manifestations associated with toxicity of the accumulating compound or abnormal metabolites derived from it [90]. Dietary restriction of the accumulating substrate and/or its precursors, as discussed in section 1.2.1 above, is often described as the original application of substrate reduction. A more recent approach to substrate reduction is to reduce synthesis of the accumulating substrate and/or its precursor by inhibiting a protein upstream of the defective step. This neatly addresses the conceptual challenge of

developing therapies for loss-of-function disorders by instead drawing on the extensive expertise of the drug development industry in discovering small molecule inhibitors.

An SRT approach to classic galactosemia, namely inhibition of the phosphorylation of galactose to galactose-1-phosphate by hGALK1, was first proposed for classic galactosaemia in 2002 [37] based on the milder phenotype of GALK1 deficiency. This is supported by studies showing rescue of galactose toxicity in yeast [23], [116] and correction of movement disorder in Drosophila [117] disease models of classic galactosemia by genetic deletion of the respective GALK1 orthologue. Consequently, a number of high throughput screening campaigns have been performed against hGALK1 to identify small molecule inhibitors [107], [133]–[139]. Translation of these small molecules to therapeutic inhibitors is hampered by three factors: the limited potency and chemical diversity of the discovered small molecules; the lack of consensus regarding the pathogenic mechanism of classic galactosemia; and the lack of rescue observed in a *Drosophila* model upon deletion or knockdown of GALK1 orthologue [51]. The existing small molecule inhibitors of hGALK1 are described in more detail in section 1.4.4 and chapters 5 and 6 of this work focus on identifying novel starting points for the development of hGALK1 small molecule inhibitors.

For primary hyperoxaluria, three hepatic enzymes have been proposed and validated as SRT targets. These are: PRODH2 which catalyses the first step in the mitochondrial cascade to convert hydroxyproline to glyoxylate; LDH-A which catalyses the oxidation of glyoxylate to oxalate in the hepatocyte cytosol; and hHAO1 which catalyses the oxidation of glycolate to glyoxylate in the peroxisome. The safety of all three targets is supported by the mostly benign phenotypes of their loss of function mutations in humans: PRODH2 deficiency, hydroxyprolinaemia, is asymptomatic [140], [141] as is HAO1 deficiency [142], [143], while deficiency of LDH-A is characterised by muscle pain and stiffness during heavy exercise [144], [145].

RNA interference (RNAi) has emerged as a popular tool to validate these targets by gene silencing in the mouse model of primary hyperoxaluria. RNAi mediates sequence-specific gene silencing through short, double-stranded fragments of RNA (siRNA) that bind to the target mRNA and facilitate its degradation by the endogenous RNA-induced silencing complex (RISC) (reviewed in [146]). Treatment of AGXT knockout mice with siRNA targeting PRODH2, transported in lipid nanoparticles, reduced urinary oxalate excretion [147]. RNAi targeting LDH reduces oxalate excretion and protects against ethylene glycol mediated kidney damage in both AGXT knockout and GRHPR knockout mice [148].

HAO1 is the most extensively studied SRT target for primary hyperoxaluria, and HAO1 knockdown has been confirmed to reduce oxalate levels via siRNA targeting endogenous HAO1 in mouse and rat hyperoxaluria models and healthy nonhuman primates [149], [150], and via CRISPR/Cas9 knockdown of HAO1 in AGXT knockout mice [151]. Phase 1 and 2 clinical trials of siRNA targeting hHAO1 (ALN-GO1, Lumasiran; NCT02706886) showed 43- 92% oxalate reduction in PH1 patients, supporting the commencement of phase 3 trials (NCT03681184; NCT03905694). However, current siRNA therapies are associated with immunoinflammatory adverse reactions [152], [153], high cost of treatment [154] and toxicity of excipients necessary for siRNA delivery or stability (e.g. nanoparticles, peptides) [155]– [157], supporting the development of small molecule inhibitors as adjunct or alternate therapies. The existing small molecule inhibitors of hHAO1 are described in more detail in section 1.5.4 and chapters 3 and 4 of this work focus on identifying novel starting points for the development of hHAO1 small molecule inhibitors.

1.3 Structure-guided hit discovery

The drug discovery pipeline can be broadly delineated into the various stages shown in figure 1.6. Given a validated therapeutic target, hit discovery is the crucial step to provide entry into the drug development pipeline. This involves the identification of compounds (or biomolecules) that

modulate the target's activity in the desired manner and requires a suitable library of compounds to screen against, a robust screening method to identify initial hits, and orthogonal assays to remove false positives and characterise true positives.

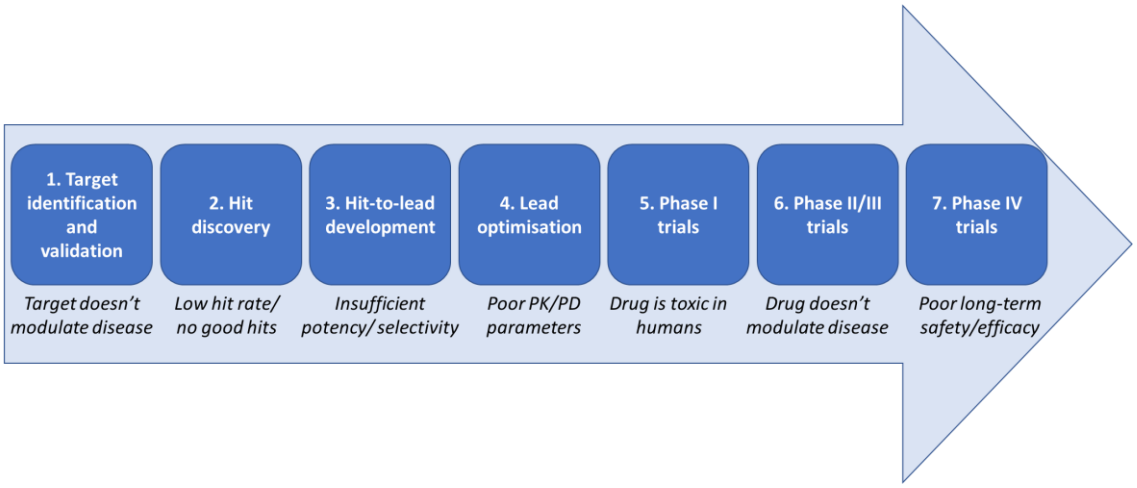


Figure 1.6: Simplified flow chart illustrating the drug discovery process. *1. Target identification and validation: identify biomolecule that can modify the course of the target disease. 2. Hit discovery: identify compounds that modulate the target's activity in the desired manner. 3. Hit-to-lead development: cycles of chemistry and in vitro testing to generate a molecule, or series of molecules, that are potent and selective for the identified target. 4. Lead optimisation: preclinical development to optimise the therapeutic, pharmacokinetic, pharmacodynamic and toxicological profiles of lead compounds using cellular and animal disease models. 5. Phase I clinical trials: determine safety profile in healthy volunteers. 6. Phase II/III clinical trials: determine therapeutic effect in patients. 7. Phase IV clinical trials: post-approval safety and efficacy testing. The most common reason for attrition of drug candidates at each stage is shown in italics.*

1.3.1 High Throughput Screening (HTS)

HTS form the backbone of pharmaceutical drug discovery and entails screening millions of compounds against a purified protein using a cheap, automated *in vitro* assay. However, these libraries are expensive to compile and screen, prone to assay artefacts and hits are often difficult to optimise. The latter limitation stems from the complexity of the drug-like compounds typically found in HTS libraries, which identifies hits where only a small portion of the whole compound is required for binding to and/or activity against the target. Optimisation of such compounds to achieve therapeutic potency generally either results in very large molecules, requiring considerable medicinal chemistry efforts to maintain acceptable solubility and cellular permeability, or requires considerable time and expense to deconvolute the active components of a hit compound (figure 1.7A, right). The high number of compounds required, both in

screening and optimisation, also limits the viability of this method outside industrial campaigns for common diseases.

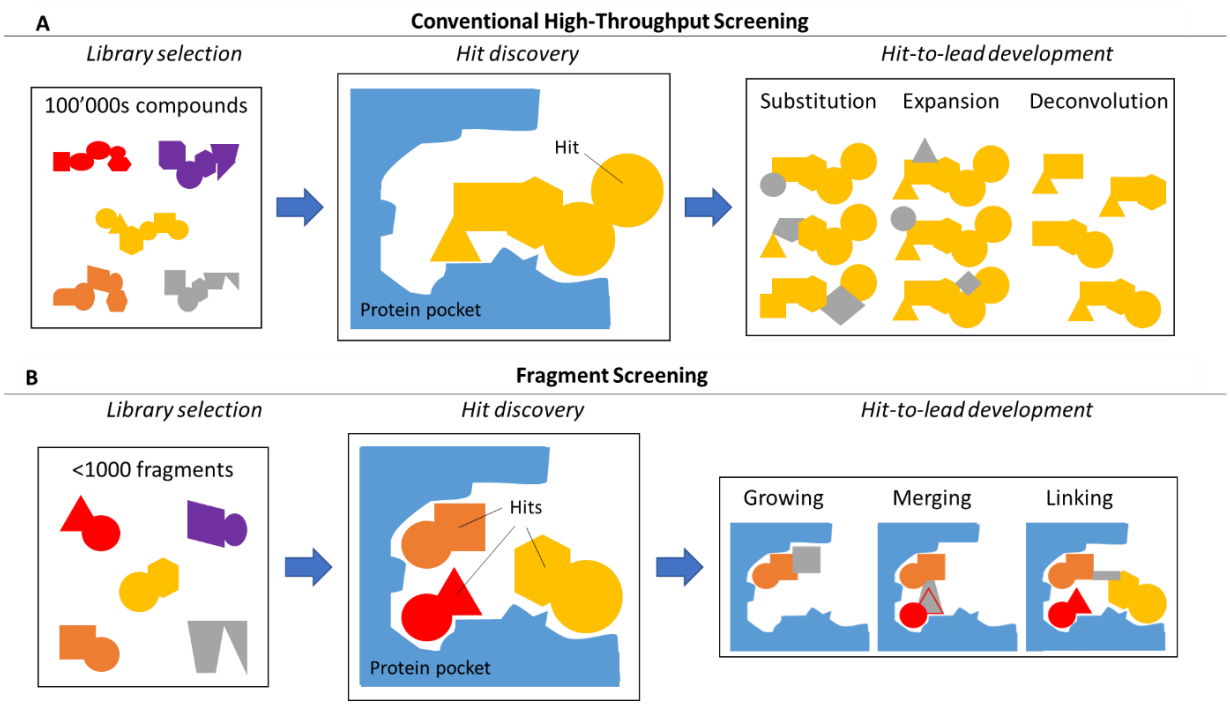


Figure 1.7: Cartoon representation comparing library selection, hit discovery and hit-to-lead development in HTS and fragment screening. *A. In an HTS campaign, 100'000s of compounds are screened, the hit rate is generally low, and optimisation of the discovered hits is often a long and complicated process requiring intensive medicinal chemistry efforts and screening of a large number of follow-up compounds. B. In a fragment screening campaign, less than 1000 compounds are needed to cover the same chemical space, providing better mapping of protein pockets and a higher hit rate. Hits can be optimised by growing single fragment hits or by merging or linking multiple fragment hits.*

1.3.2 Fragment screening

One alternative to HTS is to screen a library of a few hundred to a few thousand smaller compounds called fragments (figure 1.7B). Fragments contain an average of 15 non-hydrogen atoms, in contrast to up to 30 non-hydrogen atoms found in HTS compounds. Fragments are often characterised by 'rule-of-3' compliance in their physical properties, being < 300 Daltons in size and having < 3 hydrogen bond acceptors and donors and a cLogP ≤ 3 [158]. Additional suggested properties include < 3 rotatable bonds and a polar surface area ≤ 60 $Å^2$ [159].

As well as ensuring molecular simplicity and thus applicability to a broad range of targets, adherence to the above-mentioned physiochemical property guidelines promotes the use of

highly soluble, low lipophilicity fragments resulting in hits with ideal pharmacokinetic profiles for drug development. In contrast to the estimated 10^{63} drug-like compounds containing up to 30 non-hydrogen atoms [160], the estimated number of possibilities containing up to 17 non-hydrogen atoms is a much more manageable 10^{11} compounds [161]. This means that fragment libraries containing 2-3 orders of magnitude less compounds than a typical HTS library can still sample >10 orders of magnitude greater chemical space, allowing high quality drug discovery campaigns to be conducted in smaller industrial and academic labs.

In addition to improved sampling of chemical space, the higher hit rate observed in fragment screening campaigns (estimated to be 1 -10 % depending on the protein target) can be attributed to the higher probability of observing good complementarity between the target protein and a small fragment that need only make a few interactions to bind efficiently [162], [163]. This latter concept of low molecular complexity also contributes to ligand efficiency, which states that maximal efficiency is achieved when every atom of the ligand forms a favourable interaction with the target protein. This is much less likely in larger compounds where unfavourable interactions, such as electrostatic clashes, do not prevent binding of the complementary chemical moieties. For a fragment to bind, however, it must overcome a large entropic barrier and thus minimal unfavourable interactions are tolerated.

While fragment screening often yields higher quality starting points than HTS campaigns, optimisation of fragment hits still requires considerable medicinal chemistry effort, generally achieved by linking, merging or growing fragment hits to develop a single high affinity lead (figure 1.7B, right panel; see section 6.3 for details). One library that aim to reduce the chemistry effort required is the Diamond-SGC poised library (DSi-Poised), used in this work, consisting of 776 chemistry-ready fragments, at 100 – 500 mM concentrations, from which follow-up analogue libraries of each are either commercially available or readily synthesised by robust, well-established reactions [164].

These fragments are ‘poised’ for chemistry in that they are composed of one or more building blocks or ‘synthons’ that can be rapidly and simply disconnected and synthetically or commercially elaborated to generate follow-up compounds. In this approach, the synthon that demonstrates complementarity with the target is preserved and analogues are generated to map the binding pocket and/or explore structure-activity relationships (SAR) within the discovered series. This allows for rapid optimisation from a low-mM affinity fragment to a nM affinity lead compound, a process that is greatly simplified by the availability of fragment-bound structures to facilitate logical, structure-guided design [165], [166].

Although fragment binding is highly efficient, it is mediated by only a few interactions with the target protein and thus fragment hits often have weak affinities. Due to fragment affinities being in the high-µM to low-mM range, initial hits typically display limited functionality that is often below the detection limit of routine biochemical assays [167]. Additionally, development of high concentration bioassays is hampered by false positives and assay artefacts. These limitations can be largely avoided by employing biophysical methods of sufficient sensitivity to detect weak binding of fragments [168]. Several methods have been used to screen fragments, the most established of which are NMR spectroscopy, SPR and X-ray crystallography. Fragment screening by X-ray crystallography, the approach used in this work, is discussed in more detail in the next section.

NMR was the first method successfully used to screen fragments and involves structural mapping by comparison of the atomic spin of specific isotopes within a magnetic field when stimulated by radiation [169]. Protein-observed NMR is a highly robust and sensitive method, often considered the gold-standard for fragment screening, but requires large amounts of isotope labelled protein and is typically limited to proteins less than 30 kDa in size. Ligand- observed NMR does not require such high quantities of protein and no protein labelling or size limit is involved but this method is less robust and provides minimal information about where and how fragments bind.

Another method that is widely used for fragment screening is surface plasmon resonance (SPR),

an optical technique that quantifies changes to the reflection of polarised light induced by binding of fragments to immobilised protein [170]. In practice, quantification of fragment binding is rare as K_D determination requires ligand solubility significantly higher than the K_D. Additionally, deconvolution of specific and nonspecific interactions is often difficult for weak ligands tested at high concentrations. Overall, SPR is most suited to the comparison of analogues of fragment hits that are of sufficient potency to be detected by other methods.

1.3.3 Crystallography-based fragment screening

Most fragment screening campaigns require fragment-bound structures from X-ray crystallography to guide hit-to-lead development by providing direct information about where and how fragments bind. This method is superior both in terms of the range of ligand affinities that can be detected - from sub-nM to mM, limited only by the ligand's solubility – and the size of the target protein – from minimal protein scaffolds to megadalton complexes, limited only by crystallisation tractability. X-ray crystallography is also one of the most sensitive screening methods, as evidenced by identification of fragments by X-ray crystallography that were detectable by other methods only after hit elaboration [171].

However, determination of crystal structures is often a bottleneck in drug discovery efforts and fragment screening by X-ray crystallography requires a robust crystallisation system and, until recently, could not be performed with sufficient throughput to be a viable primary screening approach. Recent advances in beamline automation and detector technology at the Diamond Light Source synchrotron, coupled with high throughput crystal soaking, harvesting and data analysis methods developed in collaboration with the SGC, have increased the throughput of X-ray crystallography to allow screening of hundreds of fragment-soaked crystals within a week [46]. Using this pipeline, pre-formed protein crystals (typically of ligand-free, wild-type protein) are each soaked with single fragments at 10-100 mM and X-ray diffraction data is collected to identify binders [172] (figure 1.7B, central panel).

The pre-requisite for this method is the production of highly reproducible crystals for the target protein with consistent atomic level diffraction and high tolerance to an organic solvent (in which fragment compounds are dissolved). While fragment screening by X-ray crystallography yields atomic level information, the data is binary, with no affinity information to prioritise hits for optimisation. Thus, downstream validation methods, such as binding and functional assays, are required to quantify fragment binding in solution.

X-ray crystallography is particularly suited to the identification of small molecule modulators in inborn errors of metabolism because it identifies binders regardless of whether they inhibit or activate a target, and consequently can be used to screen for stabilisers of the defective enzyme as well as inhibitors of upstream enzymes. Additionally, structural genomics projects have provided structural data for many IEM-related proteins, providing tractability to this approach [173]. This is exemplified in the following section by the available structures for hGALK1 and hHAO1, the two therapeutic targets investigated in this work.

1.4 Structural biology of hGALK1

Section 1.2.4 has provided the rationale for the development of hGALK1 inhibitors to treat classic galactosemia via the substrate reduction approach. In this section, the published structures and small molecule inhibitors of hGALK1 developed to date are summarised.

1.4.1 GALK1 is a sugar kinase belonging to the GHMP kinase superfamily

The human genome encodes 663 kinases, a class of enzymes that catalyse the transfer of a phosphoryl group from a cation-nucleoside triphosphate, most commonly magnesium (Mg^{2+})-adenosine triphosphate (ATP), to a protein, lipid, nucleoside or sugar substrate [174], [175]. Kinases are separated into 12 groups (or ‘structural clans’) based on their overall fold [175]. Kinases that act on sugar substrates belong to three of these groups [174], [176] (figure 1.8).

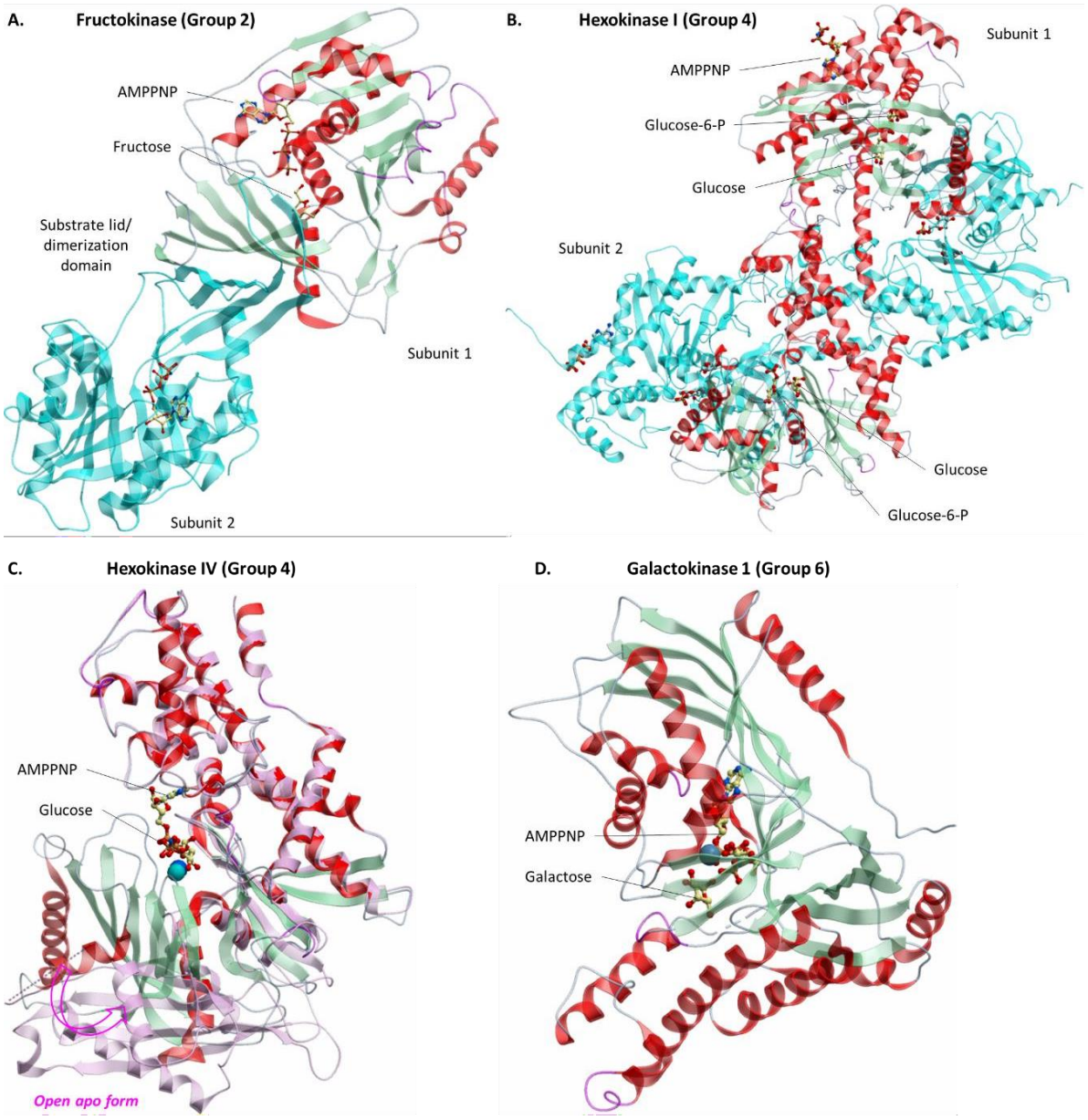


Figure 1.8: Representative structures of the three human sugar kinase fold groups. *A. Ribbon diagram of group 2 kinase member fructokinase (PDB 3nbv) dimer. One subunit of the dimer is coloured by secondary structure (α-helices in red, β-sheets in green and loops in purple) and the other subunit is blue. B. Ribbon diagram of group 4 kinase member containing two ASHKA domains, hexokinase I (PDB 1qha) dimer. One subunit of the dimer is coloured by secondary structure and the other subunit is blue. C. Ribbon diagram of group 4 kinase member containing one ASHKA domain, hexokinase IV (glucokinase) monomer in the closed, substrate-bound conformation (PDB 3fgu), coloured by secondary structure, superimposed with the open, apo conformation (PDB 1v4t) in pink. D. Ribbon diagram of group 6 kinase member galactokinase 1 (PDB 1wuu) monomer coloured by secondary structure.*

Group 2 kinases have a Rossmann fold with an ATP-binding core domain composed of three layers of repeats with the central β-sheet mostly parallel (figure 1.8A). The human sugar kinases in this group, fructokinase [177] and ribokinase [178], comprise a central 9-stranded β-sheet, 10 flanking α-helices and an additional four-stranded β-sheet, which acts as a substrate lid and

dimerization interface (figure 1.8A) [179], [180].

Group 4 kinases have a ribonuclease-H like fold (βββαβαβα) with three layers (α/β/α) [174]. The human sugar kinases in this group are the hexokinase isoforms within the acetate/sugar kinase/hsc70/actin (ASKHA) superfamily with either one (figure 1.8C) or two (figure 1.8B) substrate binding domains [181]. The single-domain hexokinase isoform (glucokinase) is a monomer with significantly different open and closed conformations whereas the two-domain hexokinase isoforms (hexokinase I-III) are dimers for which no major conformational changes have been reported.

Group 6 kinases, also collectively known as the GHMP kinases, are monomers that fold into two domains with the ATP binding site at the cleft between domains (figure 1.8D). The overall architecture of this group is an N-terminal domain containing 2 β-sheets and 4 α- helices, including an unusual left-handed βαβα component, and a C-terminal domain consisting of a ferredoxin-like core (βαββαβ) with 4 additional α-helices [174], [182].

1.4.2 GHMP family of kinases

The group 6 kinase structural clan contains only the GHMP kinase superfamily, named after its four original members - galactokinase, homoserine kinase, mevalonate kinase and phosphomevalonate kinase. Other human members include N-acetylgalactosamine kinase (hGALK2) and mevalonate 5-diphosphate decarboxylase (hMVDD). It is important to note that hGALK2 can phosphorylate galactose as well as its preferred substrate N-acetylgalactosamine, however the low affinity of hGALK2 for galactose likely indicates a lack of biological relevance within normal metabolism [183]. Crystal structures of 4 human GHMP kinases (hGALK1– PDB 1wuu, [184] and structures from this work; hGALK2– PDB 2a2c, 2a2d, [185]; hMVK– PDB 2r3v, [186]; and hMVDD – PDB 3d4j, [187]) reveal high structural similarity between family members (figure 1.9A), despite their low sequence identity (26.2, 27.2 and 32.5 % between hGALK1 and hMVK, hGALK1 and hMVDD, and hGALK1 and hGALK2 respectively; figure 1.9C).

All four human GHMP kinases, share two conserved glycine rich loops (shown as black ribbons in figure 1.9B and red boxes in figure 1.9C-D): Motif II (PxxxGLxSSA; ^{134}PLGGGLSSSA^{143} in hGALK1), a novel phosphate-binding loop that wraps around the α- and β-phosphate groups of ATP; and Motif III (xS/TGxGxx; ^{343}MTGGGFGG^{350} in hGALK1), which is close to the sugar substrate hydroxyl being modified [188], [189], [190]. An additional conserved motif, Motif I (e.g. VNL/IIGEH in galactokinases (Val38-His44 in hGALK1) as illustrated in figure 1.9D), is shared between orthologues and has been shown to mediate sugar substrate recognition [177], [188].

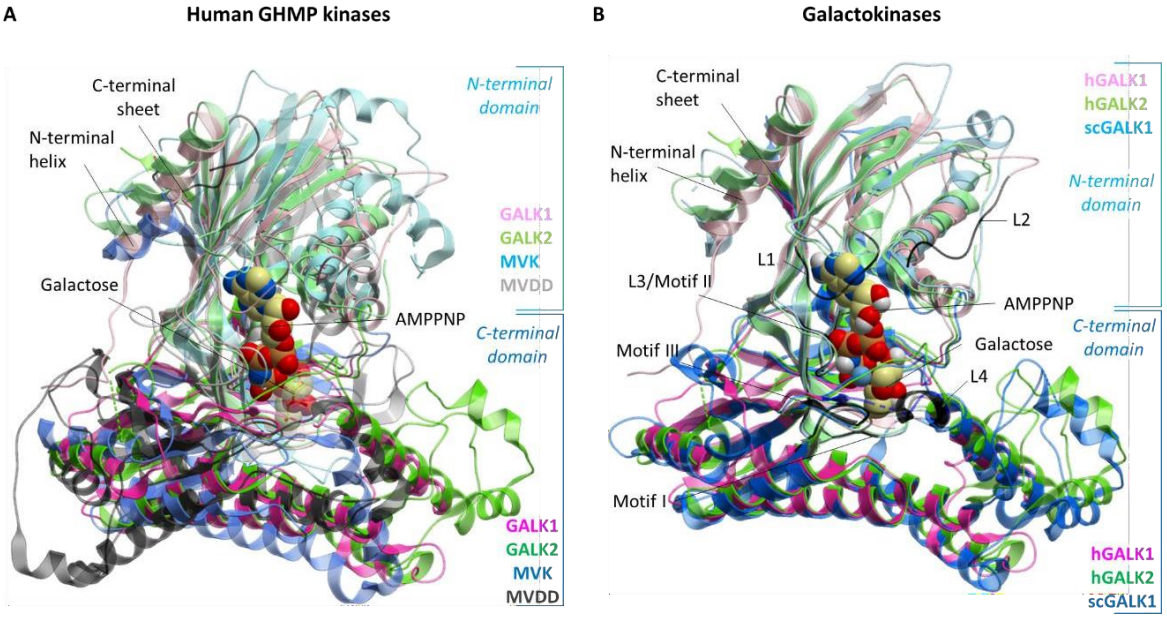


Figure 1.9: Comparison of human GHMP kinases and of galactokinases from different species. *A. Superimposed structures of hGALK1 (pink; PDB 1wuu), hGALK2 (green; PDB 2a2c), hMVK (blue; PDB 2r3v) and hMVDD (grey; PDB 3d4j), the four human GHMP kinases for which crystal structures have been solved. Galactose and AMPPNP are shown in spacefill. N-terminal and C-terminal domains are differentiated by lighter and darker shades, respectively. AMPPNP from hGALK1 structure (PDB 1wuu) is the only nucleoside shown but a similar orientation is seen in all structures. B. Superimposed structures of hGALK1 (pink; PDB 1wuu), hGALK2 (green; PDB 2a2c) and scGALK1 (Saccharomyces cerevisiae; blue; PDB 2aj4). Galactose and AMPPNP are shown as spacefill. N-terminal and C-terminal domains are differentiated by lighter and darker shades, respectively. Structural motifs and substrate binding loops of the hGALK1 structure are in black.* *Continued on the next page.*

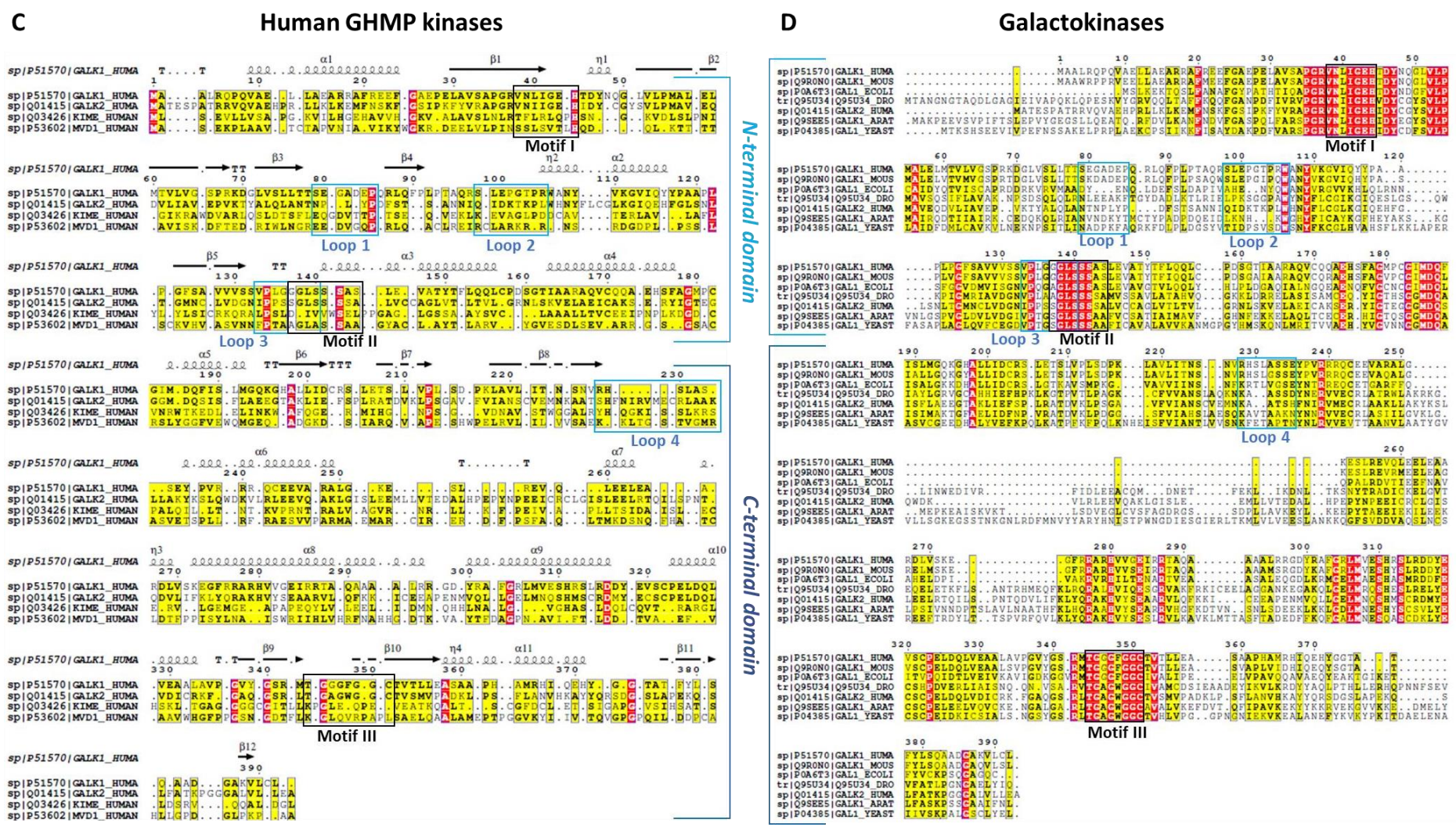


Figure 1.9 continued. *Sequence alignments of human GHMP kinases (C) and of galactokinases from different organisms (D) prepared using the MultiAlin tool* [192], [193]. *Invariant residues are shown as white characters on a red background, residues that are similar in >70% of sequences are shown as bold characters on a yellow background and residues that are similar in >50% of sequences also have a yellow background. The GHMP kinase motifs are shown in black boxes and the hGALK1 ATP binding loops are shown in blue boxes.*

Of the four loops involved in binding of ATP to hGALK1 (L1 – L4; shown as black ribbons in figure 1.9B and as blue boxes in figure1.9C-D), only L3 (Val133-Ser140) includes a motif conserved across GHMP kinases (Motif II). The other three loops are not conserved between GHMP kinases (figure 1.9C, blue boxes) or even between galactokinase orthologues (figure 1.9D, blue boxes). These ATP binding loops are discussed in more detail in the next section. Comparison of galactokinases from different organisms yields surprising results: GALK1 from *Escherichia coli* (ecGALK1) and from human (hGALK1) are fairly similar (57% sequence similarity) whereas GALK1 from *Saccharomyces cerevisiae* (scGALK1) is more similar to human GALK2 (hGALK2) (51% sequence similarity; RMSD between crystal structures of 4.7 Å over 446 Cα) than it is to hGALK1 (43% sequence similarity; RMSD between crystal structures of 7.2 Å over 390 Cα).

1.4.3 Structure of human hGALK1

The published structure of hGALK1 bound to the non-hydrolysable ATP analogue, AMPPNP, and galactose reveals a monomer with the characteristic GHMP kinase V-shaped overall fold, divided into N- and C-terminal domains that form a central cleft where the galactose and ATP binding sites are situated [185] (PDB 1wuu; figure 1.10A).

The N-terminal domain (Gln8-Met180) begins with a long α-helix of 15 residues (α1), which flanks one side of the six-stranded mixed β-sheet. A four helical bundle surrounds the other side of this β-sheet. The C-terminal domain (Pro181-Leu392) contains six α-helices and two layers of anti- parallel β-sheet, each containing four strands, with the last sheet (β12) folding back towards, and hydrogen bonding with, the first N-terminal sheet (β1, figure 1.10A) [184]. Previous studies showed that in hGALK1 catalysis, ATP binds first, stabilising the active site loops that were mentioned in the last section (L1, Ser79-Pro85; L2, Ser98-Arg105; L3, Val133-Ser140; and L4, Arg228-Glu235) to facilitate galactose binding [1]. Figure 1.10B shows the position of these loops in the published structure relative to AMPPNP and galactose.

Figure 1.10C shows the sidechain orientations of active site residues observed in the published structure, elucidating the binding modes of both galactose and ATP. The nucleotide adenine base is stabilised by a stacking interaction with the indole ring of the Trp106 sidechain and the nucleotide ribose is coordinated by hydrogen bonding with the sidechain hydroxyl of Tyr109. The three serine residues of Motif II (Ser140, 141 and 142) aid in anchoring the phosphoryl groups of the nucleotide within the active site cleft. Galactose is anchored by hydrogen bonds with the side chains of Glu43, Asp46, Asp186 and Tyr236 and the backbone of His44 and Gly183 (figure 1.10C).

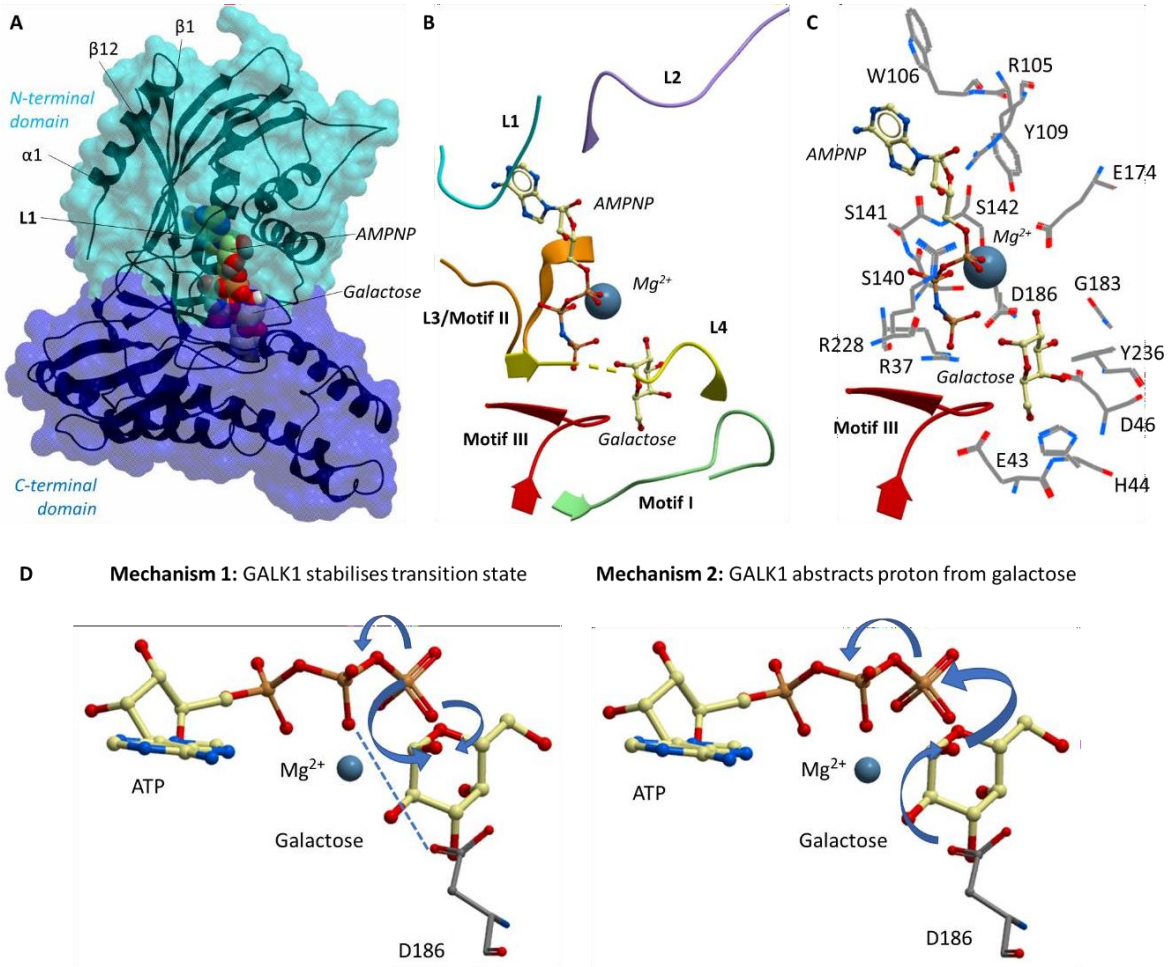


Figure 1.10: Key structural features of hGALK1. *A. Surface representation of the published hGALK1 structure (PDB 1wuu). The N-terminal domain is in light blue and the C-terminal domain is in dark blue. AMPPNP and galactose are shown as spacefill and protein secondary structures are shown as ribbons. B. Close-up view of secondary structure features of hGALK1 active site including Motif I (light green), Motif II/L3 (orange), Motif III (red), L1 (blue), L2 (lilac) and L4 (yellow). AMPPNP and galactose are shown as sticks. C. Close-up view of residue sidechains in the hGALK1 active site. Protein sidechains are shown as lines and a ribbon representation of Motif III is included for comparison with panel B. AMPPNP and galactose are shown as sticks. D. Diagrams of the two proposed mechanisms for hGALK1 catalysis: abstraction of the proton by an oxygen of the ATP phosphoryl, stabilised by hGALK1 (mechanism 1) and abstraction of the proton by Asp186 of hGALK1 (mechanism 2).*

The crystal structure of hGALK1 also illuminates the positions of other key catalytic residues: Arg37, Arg105, Glu174, Asp186 and Arg228 (figure 1.10C). Arg105 and Glu174 are positioned to form hydrogen bonds that restrict the mobility of ATP bound to the active site [194]. The side chain of Arg37 is positioned between the galactose C1 hydroxyl group, which is phosphorylated by hGALK1, and the ATP γ-phosphoryl oxygen [24] where it may facilitate phosphoryl transfer. The carboxyl group of Asp186 is positioned to either act as a catalytic base, abstracting a proton from the C1 hydroxyl of galactose to promote nucleophilic attack of the ATP γ-phosphorous (mechanism 2 in figure 1.10D), or to stabilise the transition state allowing an oxygen from the ATP γ-phosphoryl to abstract the proton from galactose (mechanism 1 in figure 1.10D) [194], [195]. It is worth noting that the two mechanisms are not necessarily mutually exclusive.

Finally, Arg228 has been hypothesised to facilitate phosphoryl transfer by reducing the pKa of either the galactose C1 hydroxyl (mechanism 1 in figure 1.10D) or the Asp186 carboxyl group (mechanism 2 in figure 1.10D) [195]. In the published structure of hGALK1, the η-amine groups of Arg228 are 7.2 and 7.3 Å away from the γ-phosphoryl of AMP-PNP and 12.5-13.8 Å away from the sidechain carboxyl oxygen atoms of the Asp186 sidechain. This could either indicate that Arg228 does not act in the hypothesised role or that the crystal structure reflects an inactive conformation. Determination of additional structures of hGALK1 that reveal different conformations (e.g. in the presence of galactose-1-phosphate or ADP) would help elucidate the role of this residue in catalysis.

1.4.4 Small molecule inhibitors of hGALK1

The availability of structural data to guide drug development has prompted three HTS campaigns to date aimed at identifying inhibitors of hGALK1 to treat classic galactosaemia. All of them have used a well-established, luminescence-based activity assay, Kinase Glo [196], miniaturised to screen a total of more than 500’000 compounds. In this assay, the activity of recombinant purified hGALK1 is quantified in the presence of varying compound concentrations by measuring the amount of ATP remaining after a set period of hGALK1 activity. ATP is

quantified through luminescence released upon ATP-dependent oxidation of luciferin by the luciferase enzyme.

One such campaign involved screening 274'000 compounds from the NIH Molecular Repository and identified 65 hGALK1 inhibitors in six chemotypes [133], typically containing a heterocyclic aromatic group able to act as an ATP mimetic [138]. The spiro-benzoxazole series, predicted to mimic the binding mode of ATP by molecular dynamic simulations and docking studies, was prioritised for initial hit-to-lead development as the most 'drug-like' scaffold [133], [135], [138]. An IC_{50} value of 6 µM was calculated for the initial hit (compound **1** in [135]; CID1286615). Removal of the two methyl groups from the cyclopentane of this compound slightly improved potency (IC_{50} 1 µM) and the resulting compound was released as a hGALK1-specific probe but demonstrated poor aqueous solubility (compound **3** in [135], CID664331; Probe ID ML152) (figure 1.11) [133], [135].

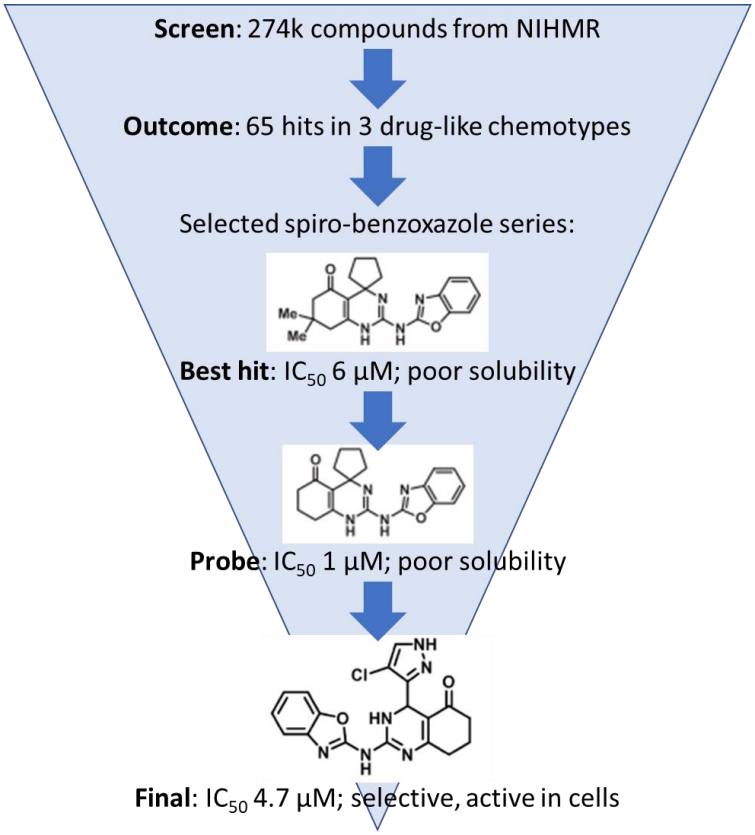


Figure 1.11: Evolution of spiro-benzoxazole series of hGALK1 inhibitors from an HTS campaign hit to low-µM lead compounds. *Flowchart summarising the HTS campaign and subsequent medicinal chemistry outcomes for the spiro-benzoxazole series of hGALK1 inhibitors.*

Replacement of the spiro ring with a chloro-pyrazole (compound **36** in [135], CID46926062) improved solubility and this compound showed fair stability in mouse liver microsomes and plasma (half-life of 34 minutes and >60 minutes respectively) [135]. An IC_{50} of 4.7 μM was calculated for compound **36**, and the Kinase Glo assay demonstrated competition of this inhibitor with ATP for binding to hGALK1. Compound **36** showed no activity against hGALK2, and minimal inhibition of other human kinases when tested in a broad selectivity panel. Finally, compound **36** elicited a dose-dependent reduction in galactose-1-phopshate levels of galactose-challenged fibroblasts from galactosaemia patients, without compromising cell survival. To progress to therapeutic lead molecules, considerably higher potency would be required for this ATP-competitive compound series to effectively inhibit hGALK1 *in vivo* due to the millimolar concentrations of ATP within the cell.

Optimisation efforts to date have failed to improve the potency of this series, primarily due to the high molecular weight and complexity of the original hit and the poor solubility and stability profiles of most of the tested analogues. To address these limitations, this work describes structural characterisation of inhibitors from the spiro-benzoxazole series to elucidate protein- ligand interactions and provide guidance on how they should be optimised (section 5.2.3) and performance of a fragment screening campaign (section 5.3) to identify novel scaffolds and binding sites free from the solubility, complexity and kinetic limitations of the existing compounds.

1.5 Structural biology of hHAO1

Section 1.2.4 has provided the rationale for the development of hHAO1 inhibitors to treat primary hyperoxaluria via the substrate reduction approach. In this section, the published structures and small molecule inhibitors of hHAO1 published to date are summarised.

1.5.1 HAO1 belongs to the (*S*)-2-hydroxyacid oxidase family

(*S*)-2-hydroxyacid oxidases (EC1.1.3.15) catalyse the oxidation of L-2-hydroxyacids to

ketoacids, using molecular oxygen as an electron acceptor, and preferentially oxidise either short-chain aliphatic, long-chain or aromatic 2-hydroxyacids. There are two (*S*)-2-hydroxyacid oxidases in the human genome, hHAO1 and hHAO2. hHAO2 is highly expressed in the kidney and can metabolise medium to long chain 2-hydroxyacids (2-hydroxyoctanoate and 2-hydroxypalmitate), whereas hHAO1 (also known as glycolate oxidase) is highly expressed in the liver and can metabolise a broad range of 2-hydroxyacids from the 2-carbon glycolate, its preferred substrate, to the 16-carbon 2-hydroxypalmitate [62].

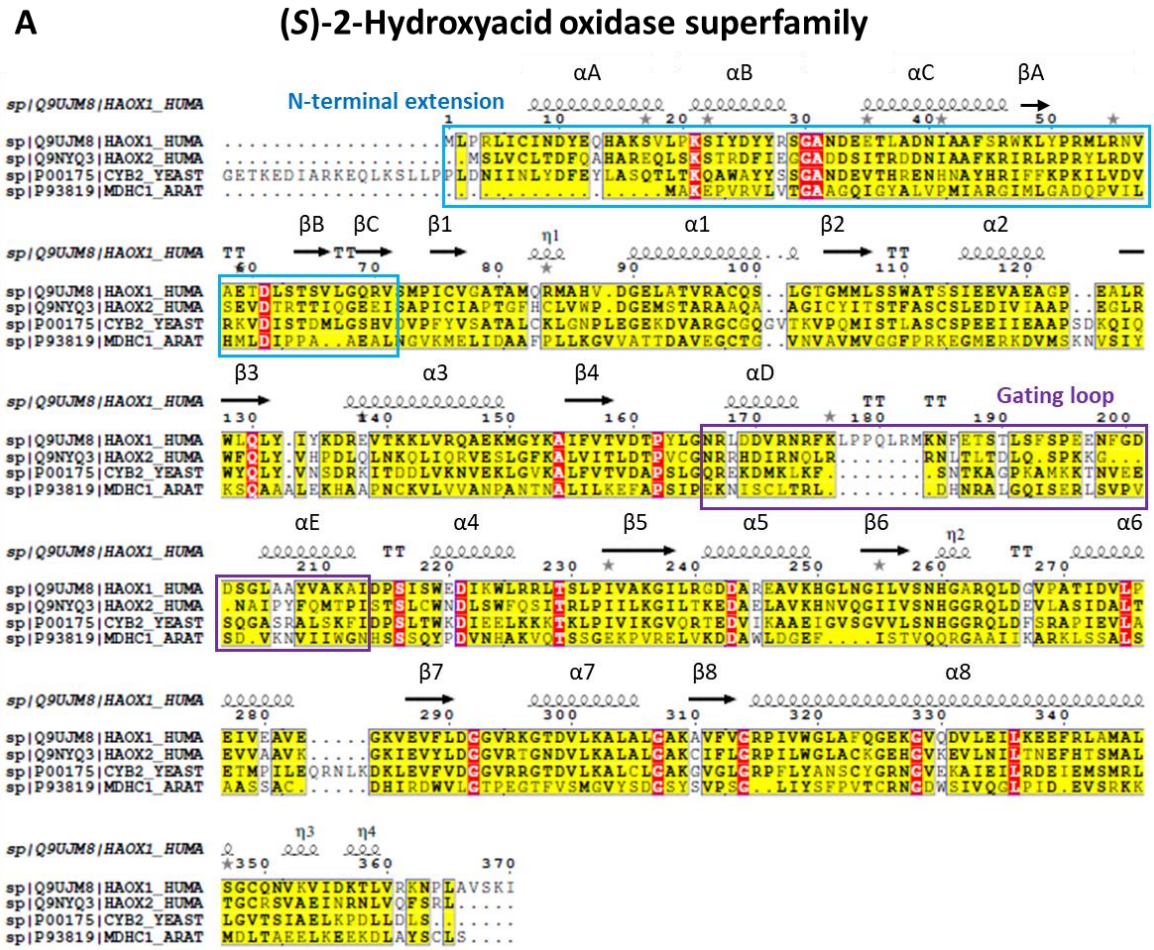
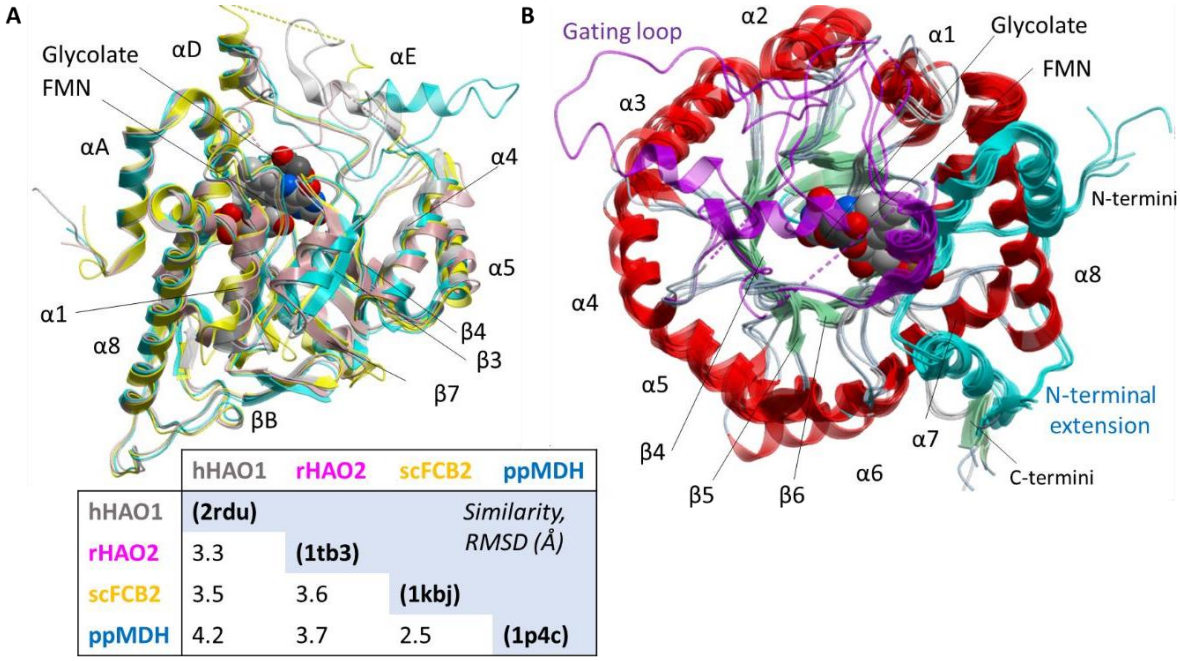


Figure 1.12: Sequence alignment of representative members of the (S)-2-hydroxyacid oxidase family. *Sequence alignment of human hHAO1, human HAO2, lactate oxidase domain of yeast CYB2 and mandelate dehydrogenase (MDHC1) of A. thaliana. The alignment was prepared using the MultiAlin server* [192], [193]. *Invariant residues are shown as white characters on a red background, residues that are similar in >70% of sequences are shown as bold characters on a yellow background and residues that are similar in >50% of sequences also have a yellow background. The N-terminal extension is boxed in blue and the gating loop in purple.*

Other members of this family catalyse the oxidation of substrates of intermediate size such as the

aromatic 2-hydroxyacid mandelate (mandelate dehydrogenase, MDH, EC1.1.99.31, [197]), the 3-carbon lactate (lactate oxidase, EC1.13.12.4, [198]) and the 5-carbon leucate (rat long chain 2-hydroxyacid oxidase, rHAO2, EC 1.1.3.15, [199]). All family members display fairly high sequence similarity (figure 1.12) and structural similarity (figure 1.13A).



	hHAO1	rHAO2	scFCB2	ppMDH
hHAO1	(2rdu)			*Similarity, RMSD (Å)*
rHAO2	3.3	(1tb3)		
scFCB2	3.5	3.6	(1kbj)	
ppMDH	4.2	3.7	2.5	(1p4c)

Figure 1.13: Superimposed structures of representative members of the (S)-2-hydroxyacid oxidase family. *A. Ribbon diagram showing the superimposed structures of hHAO1 (grey, PDB 2rdu), rHAO2 (pink, PDB 1tb3), lactate oxidase domain of yeast CYB2 (yellow, PDB 1kbj) and P. putida MDH (blue, PDB 1p4c). FMN and glycolate from the hHAO1 structure are shown as spacefill models. Inset: Comparison table showing the RMSD values between the superimposed structures. B. Top-down ribbon diagram of the superimposed structures of representative members of the (S)-2-hydroxyacid oxidase family, coloured according to secondary structure – α-helices are red, β-sheets are light green, the N-terminal extension is blue, and the gating loop is purple. The structures shown are the same as in A.*

Monomers of (*S*)-2-hydroxyacid oxidase enzymes form a typical TIM barrel fold composed of eight βα repeats that organise into a central barrel made up of a single eight stranded parallel sheet surrounded by the helices (figure 1.13B) [200]. Appended to the TIM barrel are an additional 70 residues at the N-terminal (α/β A-C; blue in figure 1.13B) and an insertion (equivalent to Leu168-Ile213 in hHAO1) between strand β4 and helix α4 of the barrel (purple in figure 1.13B). This insertion contains two short helices (αD and αE) separated by a flexible loop and is referred to as either loop 4, due to its position between strand β4 and helix α4, or the gating loop (used in this work), as it forms a lid-like structure outside the TIM barrel, guarding the

entrance to the active site [201]–[203]. The gating loop is highly flexible, as demonstrated by the variable orientation and degree of disorder observed in the published crystal structures, and is poorly conserved in both sequence and length within the hydroxyacid oxidase protein family (purple box, figure 1.12A) [203].

1.5.2 The human 2-hydroxyacid oxidases, hHAO1 and hHAO2, are FMN-dependent

hHAO1 and hHAO2 are among the 17 flavin-dependent oxidases in the human proteome which require flavin mononucleotide (FMN) as the enzyme cofactor [204]. FMN contains a tricyclic heteroaromatic isoalloxazine ring that can reversibly accept and donate one or two electrons, making it ideal for the catalysis of redox reactions [205]. There is a high degree of sequence similarity within and between HAO1 orthologues (e.g. hHAO1, spinach GOX, rat HAO1) and HAO2 orthologues (e.g. hHAO2, rHAO2, dHAO2) (figure 1.14).

Crystal structures of HAO1 from human (hHAO1, e.g. PDB 2nzl, 2rdu), spinach (*Spinacia oleracea*; sGOX e.g. PDB 1gox, 1al7) and tobacco plant (*Nicotiana benthamiana*; nbGOX, e.g. PDB 5zbm, 5zbn) are available in the protein databank. The crystal structure of hHAO2 has not been solved, however it is 74% identical in sequence to the rat orthologue rHAO2 (also called long chain 2-hydroxyacid oxidase, LCHAO), for which a 2.3 Å resolution crystal structure (PDB 1tb3) has been solved [206]. The structural (RMSD of 2.2 ± 1 Å over 332 Cα) and sequence (similarity > 60%) similarity between hHAO1 and rHAO2 enzymes is similar to that observed across the (*S*)-2-hydroxyacid oxidase family as a whole. Similarly, moderate variation is observed in the length and sequence of the gating loop (figure 1.14, purple box).

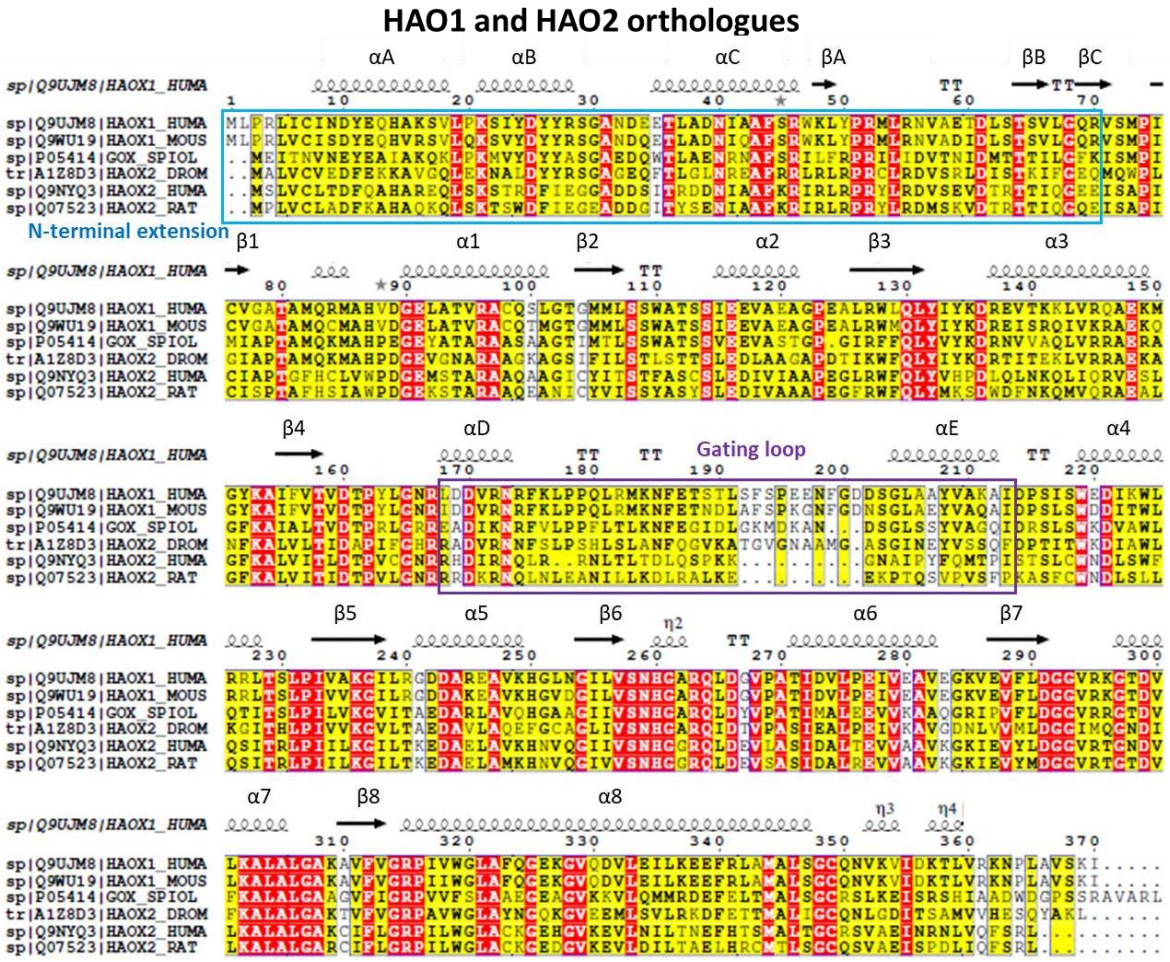
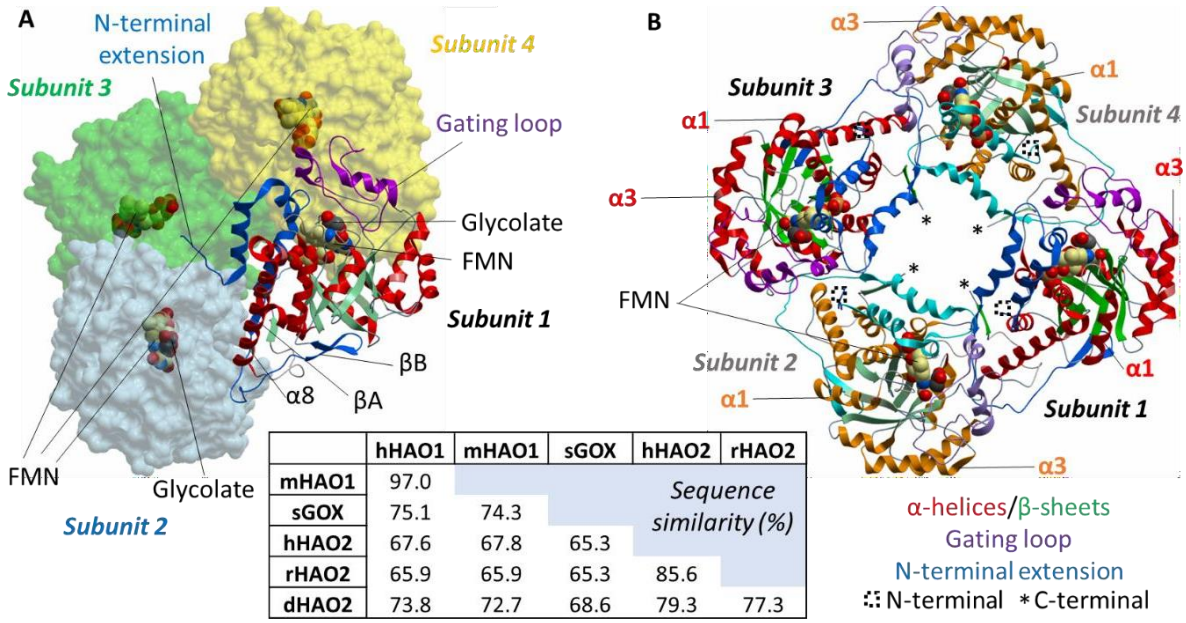


Figure 1.14: Sequence alignment of hHAO1 and hHAO2 orthologues. *Sequence alignment of 3 short-chain 2-hydroxyacid oxidases (human, mouse and spinach HAO1) and 3 medium/long-chain 2-hydroxyacid oxidases (human, rat and fruit fly HAO2) prepared using the MultiAlin server* [192], [193]. *Invariant residues are shown as white characters on a red background, residues that are similar in >70% of sequences are shown as bold characters on a yellow background and residues that are similar in >50% of sequences also have a yellow background.. The N-terminal extension is boxed in blue and the gating loop in purple.*

hHAO1 forms a homotetramer both in solution [203] and when crystallised (e.g. PDB 2nzl, 2rdu, 6gmb) (figure 1.15A). The homotetramer contains four identical subunit interfaces, formed by the N-terminal (loop βA-βB) and C-terminal (helix α8) ends of one subunit, and the α-helices of the gating loop (αD and αE) from the adjacent subunit (figure 1.15A). The helix αB of each monomer forms the inner face of the tetramer (blue helices in figure 1.15B) and the C-termini also point inwards to the centre of the tetramer (* in figure 1.15B).



	hHAO1	mHAO1	sGOX	hHAO2	rHAO2
mHAO1	97.0				
sGOX	75.1	74.3		*Sequence similarity (%)*	
hHAO2	67.6	67.8	65.3		
rHAO2	65.9	65.9	65.3	85.6	
dHAO2	73.8	72.7	68.6	79.3	77.3

Figure 1.15: Surface and ribbon representations of the hHAO1 homotetramer as observed in crystal structures. *A. Surface representation of the hHAO1 homotetramer observed in the deposited structure PDB 2rdu. Three subunits of the tetramer are shown as surface representations and the fourth subunit is shown as ribbon coloured according to secondary structure (α-helices are red, β-sheets are light green, the N-terminal extension is blue, and the gating loop is purple). FMN and glycolate are shown as spacefill. B. Face view of the hHAO1 homotetramer observed in the deposited structure PDB 2rdu, shown as ribbon representation. Subunits are coloured by secondary structure and lighter shades are used in alternate subunits for clarity. Two helices (α1 and α3) are labelled to aid orientation. The C- termini are marked by asterisks and the N-termini are marked by black boxes. Inset: Comparison table showing the sequence similarity (%) between human, mouse and spinach HAO1 and human, rat and fruit fly HAO2 orthologues.*

1.5.3 Catalytic mechanism of hHAO1 revealed by crystal structures

Closer inspection of the published HAO1 structures reveals that binding of the glycolate substrate is coordinated in a similar manner between sGOX [201] and hHAO1 [203], with the carboxylate group stabilised by hydrogen bonds with conserved active site residues Tyr26, Arg167 and Arg263 (hHAO1 numbering; figure 1.16A). The α-carbon of glycolate sits above the N5 nitrogen of the FMN, which is in turn the putative binding site for the electron acceptor, molecular oxygen [203]. To oxidise glycolate, His260, activated by Asp160, helps transfer protons from glycolate to FMN, releasing the product glyoxylate, and then the FMN is re- oxidised by molecular oxygen, producing hydrogen peroxide [207]. Tyr132 interacts with the hydroxyl group of glycolate or the carbonyl group of glyoxylate and has been shown to stabilise the transition state [208] and Lys236 has been proposed to lower the pKa of FMN's proton- accepting nitrogen (N5) to facilitate the electron transfer [202], [209] (figure 1.16A).

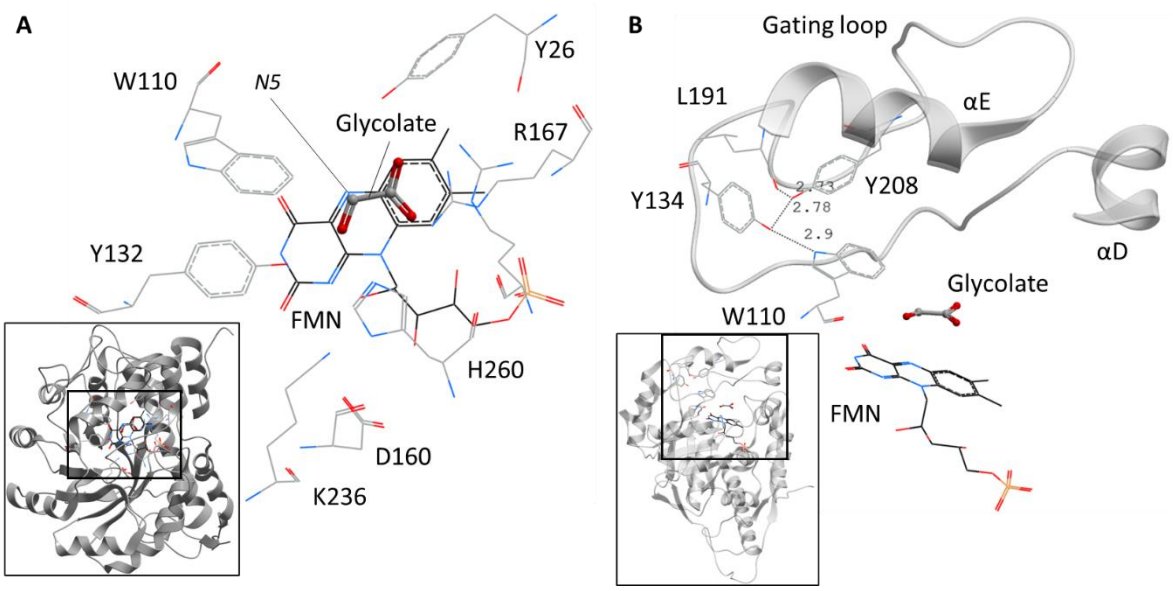


Figure 1.16: Close-up views of the hHAO1 active site and gating loop. *A. Close-up view of the active site of hHAO1 bound to glycolate (PDB 2rdu). FMN and protein residues are shown as lines and glycolate is shown as sticks. B. Illustration of the hydrogen bonding network that locks the gating loop in place when substrate is bound (PDB 2rdu). The gating loop (Arg167-Ile213) is shown as ribbon, protein residues and FMN are shown as lines and glycolate is shown as sticks. Dotted lines indicate the hydrogen bonds between Trp110, Tyr134, Tyr208 and Leu191. Inset: Overall structures of hHAO1 shown as ribbon with the viewpoint of each panel highlighted.*

Trp110 is also an important active site residue, with the sidechain indole ring pointing in towards the bound glycolate to stabilise binding. In contrast to the majority of active site residues, Trp110 is not conserved across the (*S*)-2-hydroxyacid oxidase family and is hypothesised to be essential in allowing hHAO1 and its orthologues to bind the 2-carbon glycolate. This is supported by the observation that scCYB2 (Saccharomyces cerevisiae; mitochondrial cytochrome b2 or lactate dehydrogenase [Cytochrome]) has a smaller residue, leucine, at this position and preferentially oxidises lactate while some plant MDH homologues have an alanine to preferentially oxidise the aromatic mandelate [203].

The relationship between the size of the residue sidechain at this position and the size of substrate accommodated is less clear in mammalian orthologues. For example, rHAO2 has a tyrosine residue at this position and can oxidise substrates covering a broad size range with a preference for medium-chain 2- hydroxyacids [210], whereas hHAO2 has a phenylalanine at this position and is most efficient in oxidising the 16-carbon 2-hydroxypalmitate [62] and neither HAO2 orthologues can metabolise glycolate.

In hHAO1, the Trp110 sidechain is very flexible and can either point in towards the active site to metabolise the very small substrate glycolate or rotate outwards, presumably to bind larger substrates (or inhibitors, discussed in the next section) [203], [211], [212]. When Trp110 is rotated outwards to accommodate larger molecules, this disrupts the network of hydrogen bonds that locks the gating loop into place, occluding access to the active site, when glycolate or glyoxylate is bound (figure 1.16B). Specifically, the sidechain hydroxyl of Tyr134 hydrogen bonds with the sidechain nitrogen of Trp110 and the sidechain hydroxyl of Tyr208 hydrogen bonds with the backbone carbonyl of Leu191 [203].

1.5.4 Small molecule inhibitors of hHAO1

Two general scaffolds have been explored for the inhibition of hHAO1 and related enzymes. Both scaffolds consist of a polar head group containing a 5- or 6-membered heterocycle substituted with a carboxylic acid (scaffold 1) or two carbonyl groups (scaffold 2) and a hydrophobic group containing one or more aromatic rings or long aliphatic chains (figure 1.17).

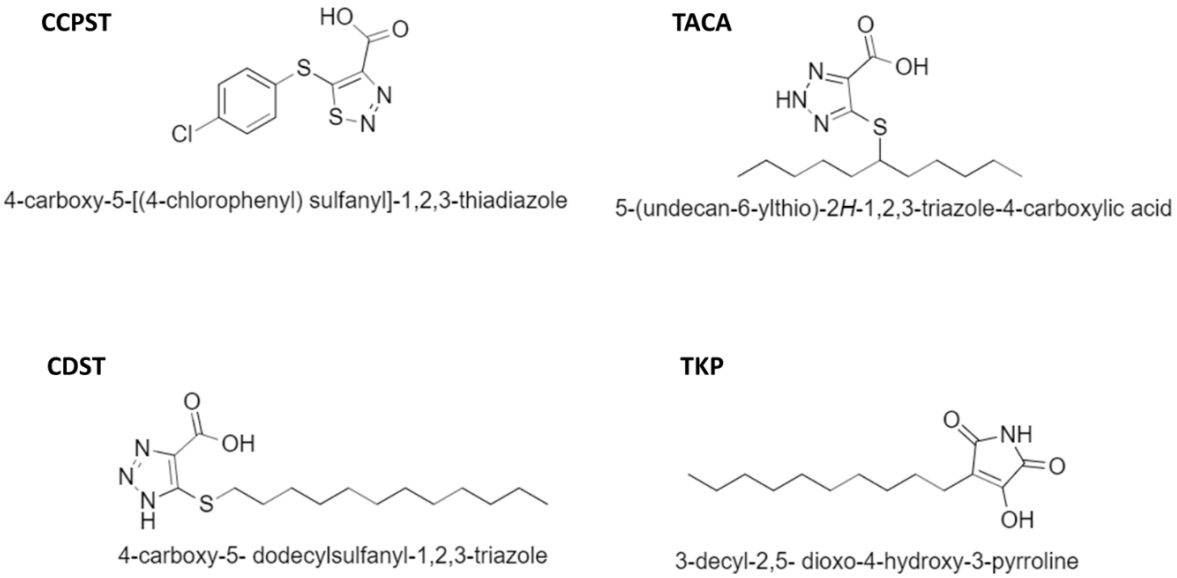


Figure 1.17: Chemical structures of HAO inhibitors for which co-crystal structures have been determined to elucidate their binding modes.

The first scaffold includes related series of carboxylic acid-substituted naphthoic acid [213], pyrazole [211] and triazole [203], [212] compounds. The binding modes for 3 compounds from this scaffold (TACA, CDST, CCPST; chemical structures shown in figure 1.17) have been

determined from co-crystal structures with sGOX (TACA) [212] or hHAO1(CDST [202], CCPST [211]). The second scaffold incudes related series of pyrrole [212], [214] compounds substituted with two carbonyl groups and one compound from this scaffold (TKP) has been characterised by co-crystallisation with sGOX.

The first HAO1-inhibiting triazole described was 5-(undecan-6-ylthio)-*2H*-1,2,3-triazole-4-carboxylic acid (TACA), a herbicide that inhibits sGOX and consists of a carboxylic acid-substituted triazole head group and two 5-carbon chains as the hydrophobic group [212] (figure 1.18). The crystal structure of sGOX-FMN-TACA (green in figure 1.18) shows that the carboxylic acid group of the inhibitor superimposes with the carboxylic acid moiety of glycolate likewise forming hydrogen bonds with Arg263, Tyr26 and Arg167 (hHAO1 numbering).

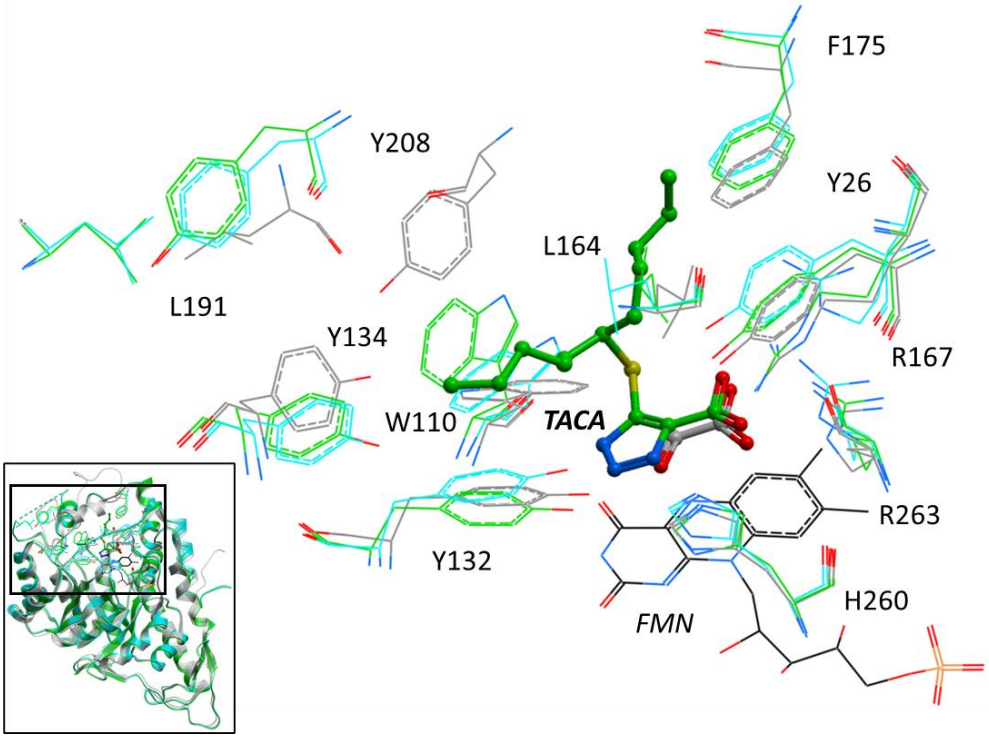


Figure 1.18: Close-up of published crystal structures showing stick and line representations of hydroxyacid oxidase active site residues with bound inhibitor TACA. *Superimposed active site residues of hHAO1(PDB 2rdu, grey), sGOX in the presence of glycolate (PDB 1gox, blue) and sGOX bound to TACA (PDB 1al7, green). Protein residues and FMN are shown as lines. Glycolate and TACA are shown as sticks. Residue numbering is that of hHAO1.*

The N3 nitrogen of TACA mimics the hydroxyl moiety of glycolate, hydrogen bonding with the catalytic histidine, His260 (hHAO1 numbering). A small displacement of the sidechain of

Tyr132 (hHAO1 numbering) is observed in the sGOX-FMN-TACA structure, relative to the glycolate bound structure of sGOX (PDB 1gox, blue in figure 1.18), resulting in a hydrogen bond between the sidechain hydroxyl of Tyr132 and the N2 nitrogen of TACA [212]. Both branches of the TACA hydrocarbon tail are well resolved in the crystal structure: one branch binds in a pocket created by displacement of Trp110 (hHAO1 numbering), where it is coordinated by hydrophobic interactions with the sidechains of Tyr132 and Tyr134, while the other branch forms hydrophobic interactions with the sidechains of Phe175 and Leu164 [212].

The scaffold represented in TACA, namely a 1,2,3-triazole-4-carboxylic acid with a 5-thioether group followed by a hydrocarbon tail, is recapitulated in the most potent hHAO1 inhibitor reported to date: CDST (4-carboxy-5-dodecylsulfanyl-1,2,3-triazole; Ki for hHAO1 of ~15 nM [203]). CDST varies from TACA only in that rather than a branched tail of two 5-carbon chains, CDST has an unbranched 12-carbon chain [203] (figure 1.17). The carboxylic acid and triazole moieties of CDST interact with hHAO1 in a similar manner to that described above for TACA and sGOX, despite somewhat different ring orientations (figure 1.19).

One exception is the ligand-induced changes to the sidechain orientation of Trp110. In the hHAO1-FMN-CDST structure (green in figure 1.19), ligand binding causes a 180° displacement of the Trp110 sidechain out from the active site, whereas in sGOX this sidechain already points outward in the holoenzyme structure (blue in figure 1.18) and is instead displaced upwards upon ligand binding (green in figure 1.18) [203], [212]. In both cases, changes to the position of the Trp110 sidechain, combined with displacement of Tyr208 and Leu191 sidechains to avoid clashing with the displaced Trp110 indole ring, disrupts the hydrogen bonding network between Trp110, Tyr134, Tyr208 and Leu191 described in section 1.5.3, so that the gating loop becomes more mobile. Leu205 of hHAO1, for which there is no structural equivalent in sGOX, is also displaced upon binding of CDST, to accommodate the beginning of the aliphatic chain (figure 1.19) [203].

In contrast to the well-resolved carbon chains of TACA, only the sulphur and first two carbon atoms of the CDST aliphatic chain can be unambiguously resolved in the crystal structure (see section 3.2.3 for further details), so that the presence of the rest of the tail can only be inferred from the disordering of the gating loop [203].

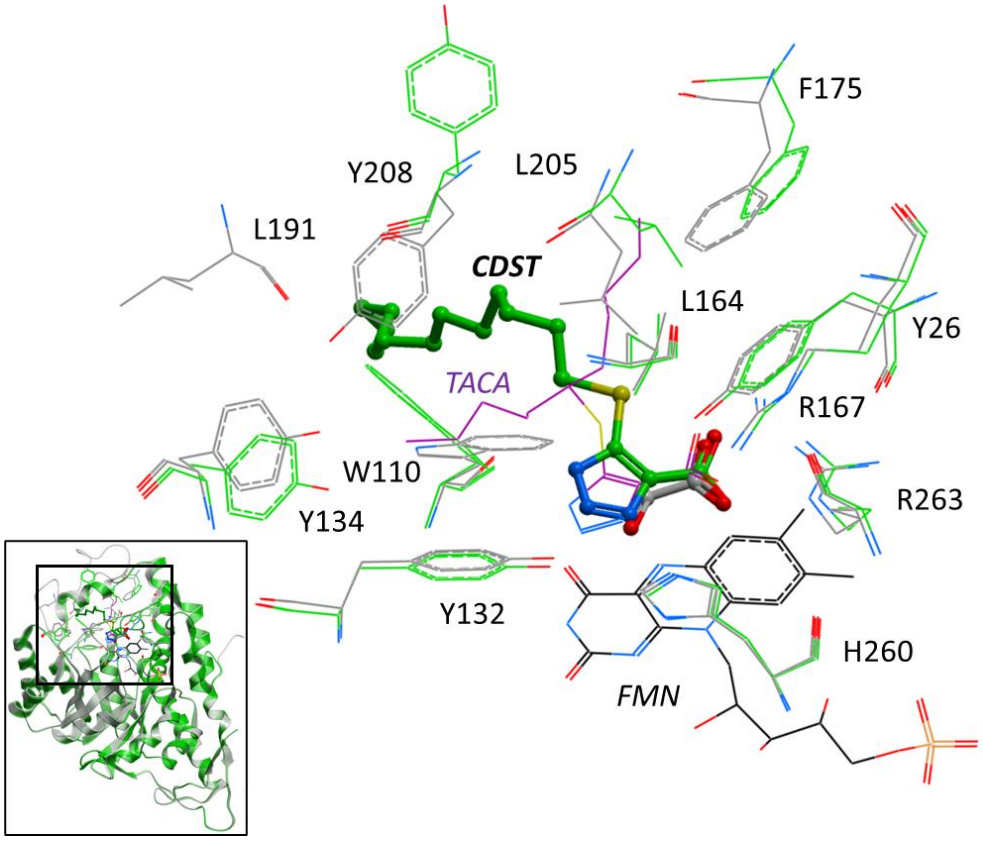


Figure 1.19: Close-up of published crystal structures showing stick and line representations of hydroxyacid oxidase active site residues with bound inhibitor CDST. *Superimposed active site residues of hHAO1 bound to glycolate (PDB 2rdu, grey) and bound to CDST (PDB 2rdt, green). Protein residues, TACA (purple) and FMN (grey) are shown as lines. Glycolate and CDST are shown as sticks. Residue Leu191 is not shown for the hHAO1-FMN-CDST structure (PDB 2rdt) as Lys176-Gly204 is disordered in this structure.*

A similar hHAO1 inhibitor, identified from a chemical screen of CDST-related compounds, is 4-carboxy-5-[(4-chlorophenyl)sulfanyl]-1,2,3-thiadiazole (CCPST, IC_{50} 1 μM), in which the aliphatic chain of CDST is replaced by a 4-chlorophenyl ring and the N1 nitrogen of CDST is replaced with a sulphur atom [211]. The hHAO1-FMN-CCPST crystal structure (green in figure 1.20) shows that the five-membered ring, carboxylic acid and sulfanyl groups of CCPST superimpose well with the equivalent moieties in CDST (purple in figure 1.20), stabilised by similar interactions with the invariant active site residues Tyr26, Tyr132, Arg167, His260 and

Arg263 [211].

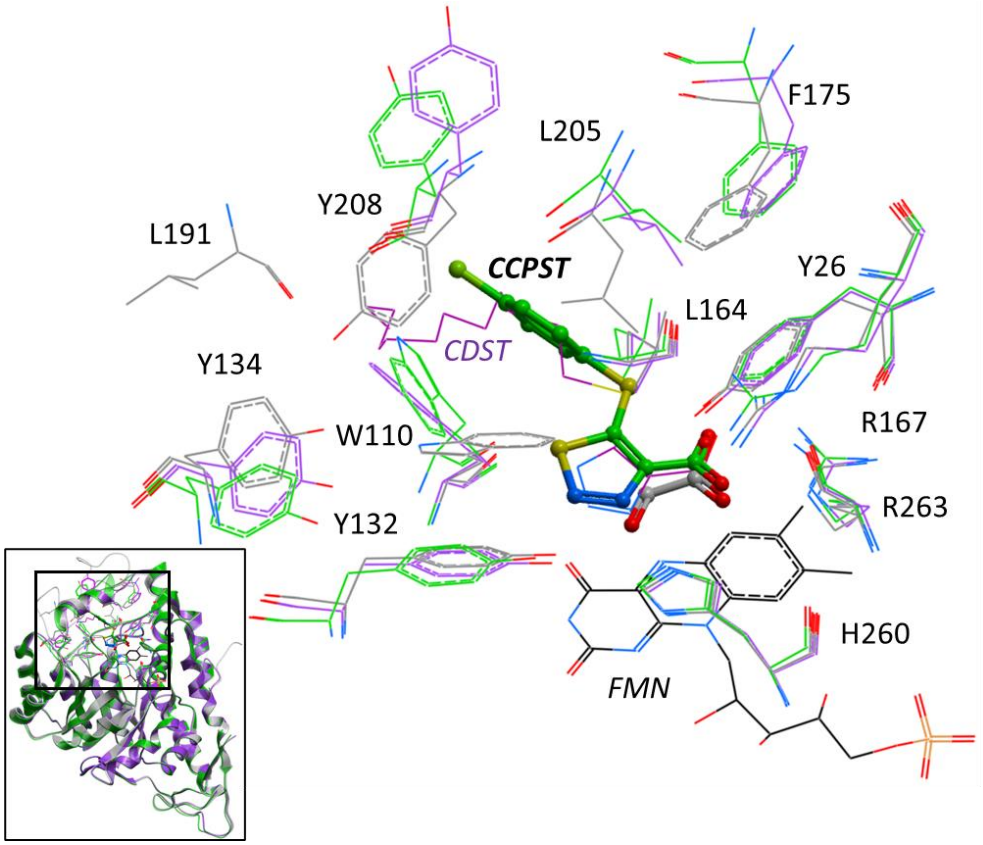


Figure 1.20: Close-up of published crystal structures showing stick and line representations of hydroxyacid oxidase active site residues with bound inhibitor CCPST. *Superimposed active site residues of hHAO1 bound to glycolate (PDB 2rdu, grey), CCPST (PDB 2w0u, green) and CDST (PDB 2rdt, purple). Protein residues, CDST (purple) and FMN (grey) are shown as lines. Glycolate and CCPST are shown as sticks. Residue Leu191 shown only for the hHAO1-FMN-glycolate structure because Lys176-Gly204 and Pro179-Ser203 are disordered in the hHAO1-FMN-CDST or hHAO1-FMN-CCPST structures respectively.*

The 4-chlorophenyl substituent of CCPST binds in the same pocket as the CDST aliphatic chain, causing similar side chain movements for Tyr134, Leu205 and Tyr208 as well as a lesser displacement (~90° rather than 180°) of the Trp110 indole sidechain [211]. Testing of CCPST in a mouse model of primary hyperoxaluria showed effective reduction of oxalate production but only when administered at high concentrations, limiting its suitability as a clinical candidate [214].

The second published scaffold of HAO inhibitors contains a 3-hydroxy-1H-pyrrole-2,5-dione polar head group with a hydrophobic substituent at position 4. This series has been tested against porcine HAO1 *in vitro* and a hyperoxaluric rat model *in vivo* [215].

The only crystal structure of a HAO1 orthologue bound to a compound from this series is the sGOX-FMN-TKP structure (10-carbon aliphatic chain, PDB 1al8, green in figure 1.21) [212]. This structure shows the hydroxyl and one of the carbonyl groups of TKP forming equivalent interactions to the carboxylic acid group of glycolate, as described for inhibitors belonging to the first scaffold, including hydrogen bonds to Tyr26, Arg167 and Arg263. The single nitrogen of the pyrrole ring of TKP hydrogen bonds with His260 and the remaining carbonyl hydrogen bonds with the sidechain hydroxyl of Tyr132. The sidechain indole of Trp110 is displaced by ~45° upon binding of TKP [212].

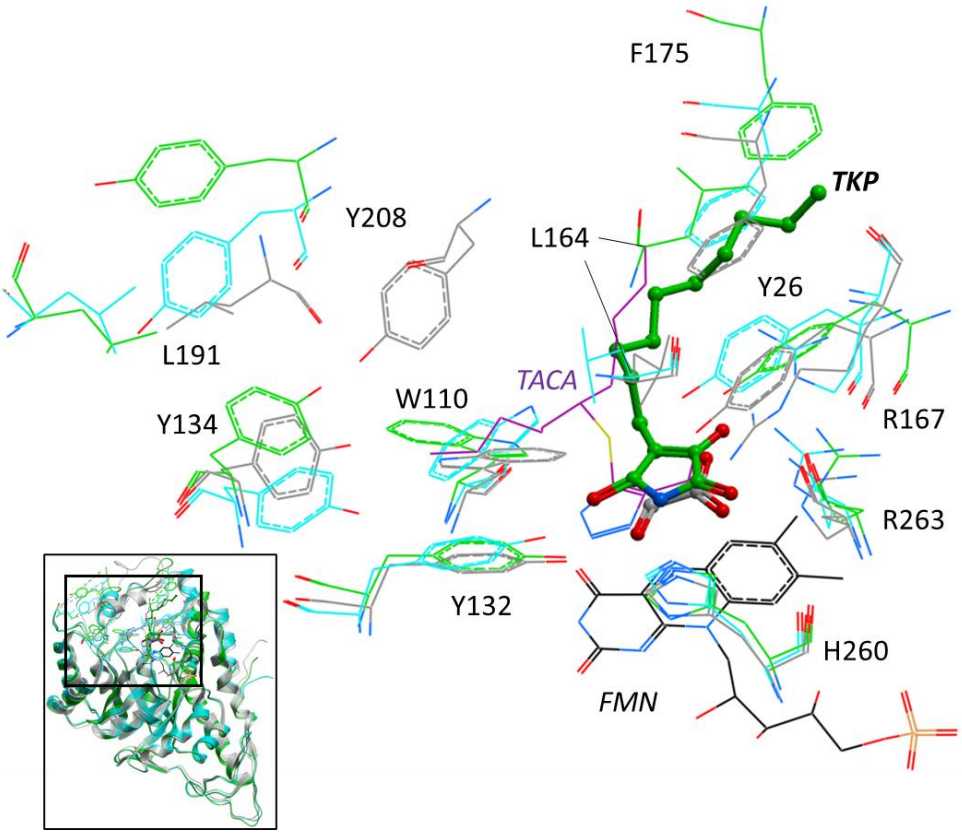


Figure 1.21: Close-up of published crystal structures showing stick and line representations of hydroxyacid oxidase active site residues with bound inhibitor TKP. *Superimposed active site residues of hHAO1-FMN-glycolate (2rdu, grey), sGOX-FMN-glycolate (1gox, blue) and sGOX-FMN- TKP (1al8, green). Protein residues, TACA (purple) and FMN (grey) are shown as lines. Glycolate and TKP are shown as sticks. Residue numbering is that of hHAO1.*

Finally, in the absence of electron density to guide ligand placement (figure 3.7D), the carbon chain of TKP was modelled to form hydrophobic interactions with the sidechains of Phe175 and Leu164, as described for one of the carbon chains of TACA (shown as purple lines in figure 1.21 for comparison). This model is supported by the significant displacement of these residues,

potentially to accommodate the carbon chain. However, the sidechains of Tyr134 and Tyr208 are also considerably displaced, supporting binding of the carbon chain in the other hydrophobic pocket identified in the sGOX-FMN-TACA structure.

To summarise, several series of potent inhibitors of hHAO1 and its homologues have been developed, each with a heterocycle-containing polar head group that interacts with the glycolate-coordinating residues of the active site and a hydrophobic moiety that primarily interacts with nonpolar residues of the gating loop. While some of these inhibitors exhibit nM potency, they are not suited for therapeutic use due to poor cell permeability, cellular toxicity and/or inhibitory activity against hHAO2. The lack of chemical diversity between the identified inhibitor series also suggests that screening a chemically distinct and diverse library not biased towards this chemotype would be of considerable benefit.

1.6 Thesis Overview

This thesis describes the discovery, characterisation and initial optimisation of compounds that provide novel starting points for developing small molecule substrate reduction therapies in classic galactosaemia and primary hyperoxaluria by inhibiting hGALK1 and hHAO1 respectively. Chapter 2 describes the methods used to produce, crystallise, screen and characterise both targets and their discovered modulators. Chapters 3 and 5 describe the optimisation of protein crystallisation and screening of a small fragment library by X-ray crystallography for hHAO1 and hGALK1, respectively. Chapters 4 and 6 describe the development and validation of orthogonal binding and activity assays to characterise the identified fragments in solution, and the selection, testing and characterisation of follow-up compounds, for hHAO1 and hGALK1, respectively. Chapter 7 summarises the outcomes and future directions from this work and discusses the relevance of the described findings to primary hyperoxaluria, classic galactosaemia and the wider field of inborn errors of metabolism.

2. Methods

2.1 Recombinant production of hHAO1 and hGALK1 protein

Plasmids for all hHAO1 constructs used in this work were available in house before the start of this project. All constructs had been cloned into the pNIC28-Bsa4 vector (Genbank EF198106), conferring an N-terminal 6-histidine tag with tobacco etch virus (TEV) protease cleavage site and kanamycin resistance, by the SGC Biotechnology group. The hGALK1 plasmid was generously provided by Prof. Kent Lai from the University of Utah. This plasmid contained full length hGALK1 (Ala2-Leu392) preceded directly by an N-terminal 6-histidine tag, cloned in the Pet21d vector (GenBank NM_153280) conferring ampicillin resistance. The construct contained two mutations designed to improve crystallisation, Glu252Ala and Lys253Ala.

All centrifugation steps were performed at 4000 rpm for 30 minutes at 4 °C unless otherwise stated. All purification steps were performed at 4 °C. Samples were concentrated by centrifugation (4000 rpm, 4 °C, Eppendorf 5810 benchtop centrifuge) in Amicon® concentrators with a 30 kDa molecular weight cut off (Sigma Aldrich). Protein concentration was determined from the A_{280nm} absorbance, measured with a NanoDrop™ spectrophotometer, and each construct's extinction coefficient, as calculated by the ProtParam tool available online from the ExPASy server [216]. Protein containing fractions were identified from inspecting A_{280nm} chromatograms and confirmed by polyacrylamide gel electrophoresis with denaturing and reducing sodium dodecyl sulfate and a discontinuous buffer system (SDS-PAGE) analysis. Protein identity was confirmed by tryptic digest coupled to tandem mass spectrometry analysis of excised SDS- PAGE gel bands and intact, denaturing mass spectrometry analysis of each final purified protein.

2.1.1 HAO1 PREPX protein production

For the nine hHAO1 constructs tested in this work (c100, c002, c003, c010, c011, c012, c020, c021 and c022; figure 2.1), 1 µL of each plasmid at 50 – 150 ng/µL was added to 30 µL of *E. coli* BL21(DE3) chloramphenicol-resistant competent cells containing the pRARE2 plasmid (carrying genes for seven rare-codon tRNAs) and incubated on ice for 30 minutes. The

transformation mixture was heat shocked at 42 °C for 45 seconds and chilled on ice for 2 minutes before addition of 250 µL lysogeny broth (LB) and further incubation at 37 °C for 60 minutes at 200 rpm. The mixture was centrifuged at 2000 rpm for 5 minutes (Eppendorf 5418 benchtop centrifuge) and 200 µL of the supernatant was discarded. The cell pellet was gently resuspended in the remaining solution by pipetting and plated onto LB-agar plates (1.5% agar, 25 g/L LB) containing 50 µg/mL kanamycin and 34 µg/mL chloramphenicol. The plates were incubated at 37 °C overnight to allow colonies to grow.

Several colonies were scraped from each plate and used to inoculate two 250 mL baffle-bottomed flasks, each containing 100 mL of auto-induction terrific broth (AI-TB) media (ForMedium, AI-TB with trace minerals prepared according to the manufacturer's instructions plus 2% glycerol) containing 50 µg/mL kanamycin and 34 µg/mL chloramphenicol. The cells were cultured for 4 hours at 37 °C in a Multitron Standard incubator (InforsHT) with shaking at 180 rpm and then the incubator temperature was reduced to 18 °C for a further 48 hours incubation with shaking at 180 rpm.

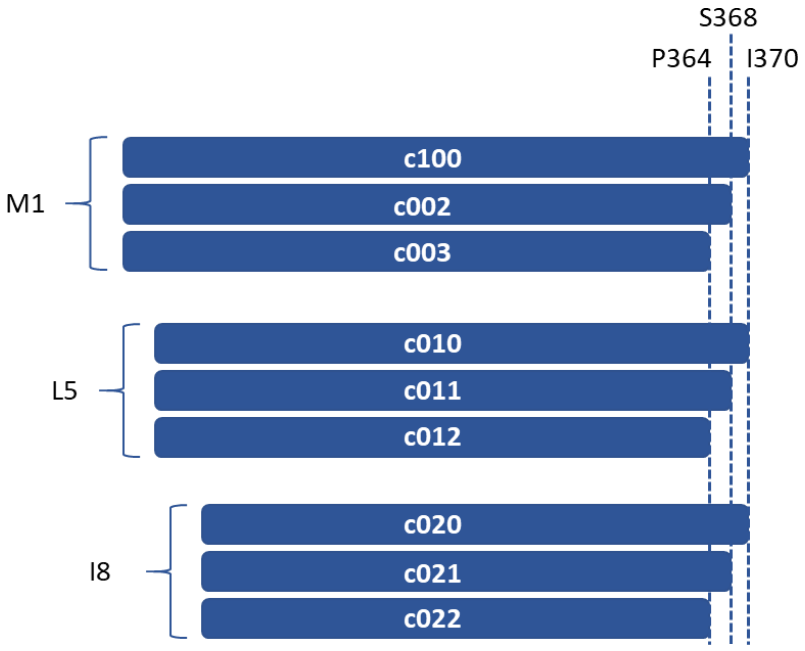


Figure 2.1: Nested constructs of hHAO1. *Diagram showing amino acid boundaries of hHAO1 constructs for E. coli expression and their respective SGC construct IDs.*

The following buffers were used for purification: *Lysis buffer*: 50 mM 4-(2-hydroxylethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.5, 500 mM sodium chloride, 5% glycerol, 0.1 mM tris(2-carboxyethyl)phosphine (TCEP), 0.5 mg/mL lysozyme, 0.1 mg/mL benzonase; *Base buffer*: 50 mM HEPES pH 7.5, 500 mM sodium chloride, 5% glycerol, 0.1 mM TCEP; *Binding buffer*: 50 mM HEPES pH 7.5, 500 mM sodium chloride, 5% glycerol, 0.1 mM TCEP, 10 mM imidazole; *Elution buffer*: 50 mM HEPES pH 7.5, 500 mM sodium chloride, 5% glycerol, 0.1 mM TCEP, 250 mM imidazole; *Anion exchange no salt buffer*: 50 mM HEPES pH 8.2, 5% glycerol, 0.1 mM TCEP; *Anion exchange high salt buffer*: 50 mM HEPES pH 8.2, 2 M sodium chloride, 5% glycerol and 0.1 mM TCEP.

Cells were harvested by centrifugation (Beckman Coulter Avanti JXN-26) and cell pellets were re-suspended in 30 mL of lysis buffer. After 30 minutes incubation at room temperature, cells were vortexed and 1% Triton X-100 (Sigma Aldrich) was added before freezing at -80 °C for 30 minutes to lyse cells. 10 mM imidazole was added to cell lysates after thawing at room temperature for approximately 30 minutes, and cell debris was pelleted by centrifugation (Eppendorf 5810 benchtop centrifuge). Clarified lysates were applied to pre-packed 1 mL His GraviTrap columns (Ni-NTA resin; Bio-Rad), pre-equilibrated with binding buffer, and washed twice with 10 mL binding buffer before elution with 3.5 mL elution buffer.

Arginine and glutamate, dissolved at 1 M in ddH_2O and adjusted to pH 7.5, were added to each elution sample to a final concentration of 100 mM. All samples were concentrated to reach 1 mL in volume and transferred to auto-sampler vials (Bio-Rad) for serial size exclusion on the Bio-Rad NGC (Next Generation Chromatography) Purifier using a 24 mL SEC 650 column (Bio-Rad) with base buffer as the mobile phase. Fraction collection was triggered when the A_{280} signal exceeded 50 mAU and 2 mL was collected per fraction in 5 mL glass test tubes.

Protein-containing fractions were pooled into first peak, second peak and tailing fractions for each construct. Pooled samples were concentrated or diluted as appropriate so that each had a

volume of 3.5 mL. Each sample was loaded onto a pre-packed 2.5 mL bed volume PD-10 desalting column (Sephadex G-25 resin, Bio-Rad) and eluted with 4 mL of anion exchange no salt buffer. Each sample was split into 2 x 2mL samples for anion exchange chromatography, performed on the Bio-Rad NGC Purifier system using a 5 mL Macro-Prep High Q column. Fractions were collected as described above and protein-containing fractions were pooled and concentrated to either 12 mg/mL or 75 μL, whichever was reached first. Purified protein was immediately applied to crystallisation trials (section 2.2).

2.1.2 HAO1 large-scale protein production

Transformation of hHAO1-c002 was performed as described above. After overnight incubation of the LB-agar plate at 37 °C, a few colonies were scraped and added to a 250 mL baffle-bottomed flask containing 50 mL of LB media with 50 μg/mL kanamycin and 34 μg/mL chloramphenicol. This starter culture was incubated overnight at 37 °C with shaking at 200 rpm. The next morning, 8 mL of starter culture was added to each of six 3 L baffle-bottomed flasks containing 1 L AI-TB media with 50 μg/mL kanamycin and 34 μg/mL chloramphenicol.

Cultures were grown for 6 hours at 37 °C in a Multitron Standard incubator with shaking at 180 rpm and then incubator temperature was reduced to 18 °C for a further 48 hours incubation with shaking at 180 rpm. Cell pellets were harvested by centrifugation (Beckman Coulter Avanti JXN-26) and stored at -80 °C until purification. The purification buffers used were the same as those described for PREPX purification (section 2.1.1) with the addition of *Wash buffer*: 50 mM HEPES pH 7.5, 500 mM sodium chloride, 5% glycerol, 0.1 mM TCEP, 40 mM imidazole.

After thawing at room temperature for approximately 1 hour, cell pellets were resuspended in lysis buffer (5 mL buffer was added per 1 g of wet cell weight) and lysed on ice by cycles of sonication (5 seconds on and 10 seconds off) for 15 minutes at 35% amplitude (Sonics Materials VCX 750 Ultrasonic Sonicator from Thermo Fisher Scientific). Cell lysates were clarified by centrifugation (Eppendorf 5810 benchtop centrifuge) and 2.5 mL of Ni-NTA affinity resin (Ni-NTA resin; GE Healthcare), pre-equilibrated in binding buffer, was added to the separated

supernatant prior to incubation at 4 °C for 1 hour with rotation at 12 rpm. The lysate with Ni-NTA resin was centrifuged (Eppendorf 5810 benchtop centrifuge) with reduced deceleration. The Ni-NTA pellet was transferred to a 1.5x10 cm column (BioRad) under gravity flow and washed with 2 x 25 mL binding buffer and 2 x 25 mL wash buffer. The bound protein was eluted with 5 x 5 mL of elution buffer.

Each elution of 5 mL was applied to a 124 mL size exclusion chromatography column (Sephadex 200 HiLoad 16/60 column; GE Healthcare) connected to an AKTA-Express purification system (GE Healthcare) with base buffer as the mobile phase. The protein was loaded and eluted at a constant rate of 1 mL/minute in base buffer and 1.8 mL fractions were collected throughout the run. Protein-containing fractions were pooled and concentrated to 13.7 mg/mL. Some protein was applied immediately to crystallisation trials (section 2.2) and the remainder was flash-frozen in liquid nitrogen and stored at -80 °C.

2.1.3 GALK1 standard protein production

Transformation, large-scale growth and purification of hGALK1 was performed as described for hHAO1-c002 in section 2.1.2 above, except that 50 μg/mL ampicillin was added to culture media instead of kanamycin and a 1:1000 dilution of Merck EDTA-free protease inhibitors was included in the lysis buffer. Some protein was applied immediately to crystallisation trials (section 2.3.1) and the remainder was flash-frozen in liquid nitrogen and stored at -80 °C.

2.1.4 GALK1 optimised protein production

Transformation and large-scale growth were performed as described in section 2.1.3. Purification was performed as above except that the following buffers were used (changed components are underlined): *Lysis buffer*: 50 mM sodium phosphate pH 7, 500 mM sodium chloride, 5% glycerol, 0.1 mM TCEP, 50 mM galactose, 0.5 mg/mL lysozyme, 0.1 mg/mL benzonase, 1:1000 EDTA-free protease inhibitors; *Binding buffer*: 50 mM sodium phosphate pH 7, 500 mM sodium chloride, 5% glycerol, 0.1 mM TCEP, 50 mM galactose, 10 mM imidazole;

Wash buffer: 50 mM sodium phosphate pH 7, 500 mM sodium chloride, 5% glycerol, 0.1 mM TCEP, 50 mM galactose, 40 mM imidazole; *Elution buffer*: 50 mM sodium phosphate pH 7, 500 mM sodium chloride, 5% glycerol, 0.1 mM TCEP, 50 mM galactose, 250 mM imidazole; *Base buffer*: 50 mM sodium phosphate pH 7, 500 mM sodium chloride, 5% glycerol, 0.1 mM TCEP, 50 mM galactose. After purification, some protein was applied immediately to crystallisation trials (section 2.2) and the remainder was flash-frozen in liquid nitrogen and stored at -80 °C.

2.2 X-ray crystallography studies of hHAO1 and hGALK1

Crystallisation experiments were performed using the sitting drop vapour-diffusion method in 96-well, three-sub-well SWISSCI plates (SWISSCI AG) with coarse screens covering 96 precipitant solutions. 20 µL of the precipitant solution was pipetted into each plate well to be equilibrated against 3 x 150 nL drops in the three sub-wells. The sub-well drops were set up by dispensing protein-to-precipitant ratios of 2:1, 1:1 and 1:2 using the TTP Labtech's Mosquito Crystal liquid handling robot. Crystallisation plates were incubated at 4 °C or 20 °C and regularly imaged at set intervals (12 hours, 1 day, 4 days, 7 days, 14 days, 28 days and 56 days) over a period of two months using an automatic inspection system (Rigaku). The following random matrix screens, commercially available or produced in-house at the SGC, were set up at both 4 °C or 20 °C with either hHAO1 or hGALK1 protein: Hampton Research Index (HIN; specifications available at https://hamptonresearch.com/product-Index-Index-HT-5); Joint Center for Structural Genomics (JCSG [217]); Ligand friendly (LFS, equivalent to the PACT screen [217]); Basic ChemSpace (BCS [218]); and Morpheus [219] screens.

2.2.1 Crystallisation trials

For hHAO1 constructs purified using the PREPX approach, each construct was tested for crystallisation at 4 °C in the LFS6 and BCS6 coarse screens and in the LFS-B1-B3 follow-up screen described in the next section at 4.5 – 13.5 mg/mL protein concentration, depending on the final protein yield. After the large-scale purification of hHAO1-c002, crystallisation trials were

set up with 13.7 mg/mL hHAO1-c002 in the LFS6-B1-B3 follow-up screen at 4 °C.

For hGALK1 purified under standard conditions (section 2.1.3), 50 mM galactose and/or 5 mM AMPPNP or ADP was added to 10 – 15 mg/mL hGALK1 before setting up crystallisation trials in JCSG7, BCS6, LFS6, HIN3 and Morpheus coarse screens at both 4 and 20 °C. For GALK1 purified using the optimised conditions (section 2.1.4), 5 mM AMPPNP or 0.5 – 5 mM tool compound (**T1**–**T7**; figure 2.2), dissolved in either DMSO or NMP, was added to 15 – 25 mg/mL hGALK1 before setting up crystallisation trials in the same coarse screens at 20 °C.

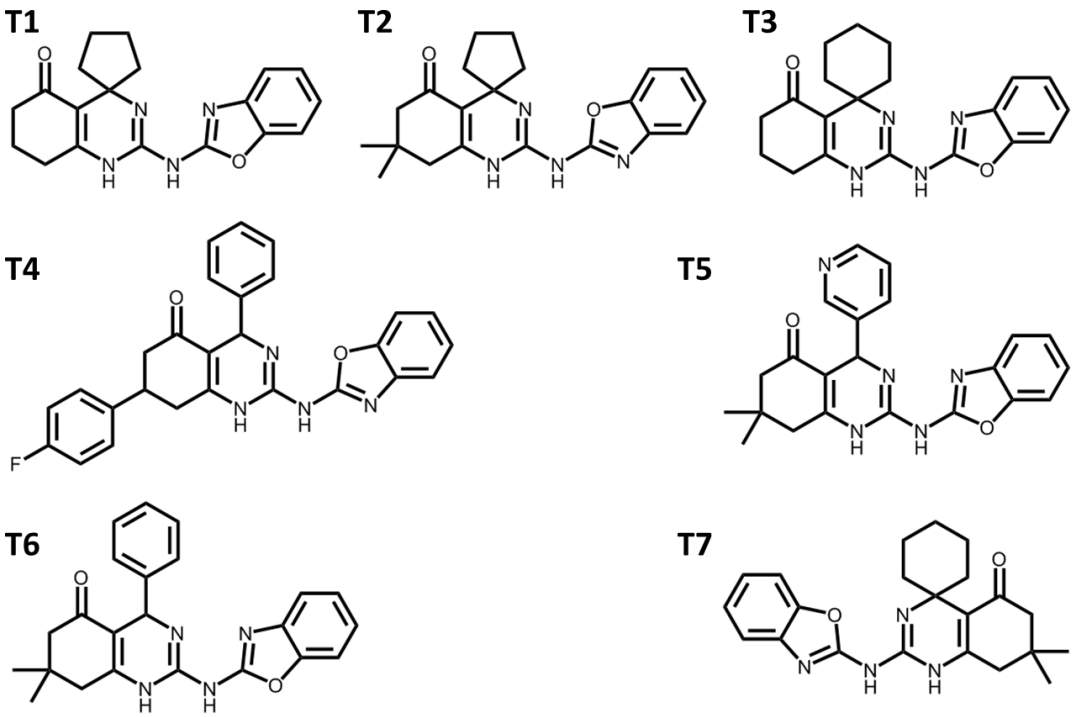


Figure 2.2: Chemical structures of tool compound T1 – T7, from the spiro-benzoxazole series of known hGALK1 inhibitors.

2.2.2 Crystallisation follow-up screen

96-well follow-up crystallisation screens were designed based around conditions in which crystals or crystalline precipitation was observed. These screens were designed as an array of twelve precipitant concentrations, differing in 2.5% increments either side of the coarse screen condition concentration, versus eight buffer pH values, differing in 0.2 pH unit increments either side of the coarse screen condition. The screens were made from ready-made stock solutions (Hampton Research/Molecular Dimensions) dispensed using the Multiprobe II plus liquid

handler (PerkinElmer) to make up 0.8 mL solution per well. The follow-up screen for hHAO1 was designed around condition LFS6-B1-B3 and various follow-up screens were tested for hGALK1, designed around conditions HCS3-A9, HCS3-B6, Morpheus-A9 and Morpheus-G8. The precipitant and pH ranges of these screens are shown in table 2.1. All follow-up screens were set up with the same drop volumes and ratios described for coarse screens (section 2.2.1).

Table 2.1: Formulation details of follow-up crystallisation screens. *The LFS-B1-B3 screen was used to grow hHAO1 crystals whereas the remaining screens were used to grow hGALK1 crystals.*

Screen	Buffer	Precipitant	Salt/ Additive
LFS6-B1-B3	0.1 M MIB buffer, pH 5.5 – 8.25	20 – 37.5 % PEG1000	None
HCS3-A9	0.1M citrate buffer, pH 4.5 – 6.2	20 – 37.5 % PEG4000	0.2 M ammonium acetate
HCS3-B6	0.1 M sodium cacodylate buffer pH 5.5 – 7.4	20 – 37.5 % PEG8000	0.2 M magnesium acetate
Morpheus-A9	0.1 M buffer system 3 (bicine-tris), pH 7.5 – 9.5	25 – 50 % precipitant mix 1	0.06 M Divalents mix
Morpheus-G8	0.1 M buffer system 2 (sodium HEPES- MOPS), pH 6.5 – 8.5	25 – 50 % precipitant mix 4	0.1 M Carboxylic acids mix

For co-crystallisation experiments of hHAO1, 13.7 mg/mL hHAO1-c002 was incubated on ice for 30 minutes with either 2 mM CCPST (known hHAO1 inhibitor, described in section 1.5.4) or 5 mM 2-hydroxypalmitate (low affinity 16-carbon substrate of hHAO1) before setting up crystallisation trials in the LFS6-B1-B3 follow-up screen at 4 °C. For fragment screening against hHAO1, 13.7 mg/mL hHAO1-c002 was used to set up 6 crystallisation plates of the LFS6-B1-B3 follow-up screen at 4 °C.

A co-crystal structure of hGALK1, galactose and ADP was obtained from crystals grown in the Morpheus-A9 follow-up screen at 20 °C after addition of 5 mM ADP to 24 mg/mL hGALK1

prior to crystallisation. Co-crystal structures for hGALK1 with **T1**, **T2**, **T3** and **T7** were obtained from crystals grown in the Morpheus screen at 20 °C after addition of tool compounds to 24 mg/mL hGALK1 prior to crystallisation. The crystallisation conditions were: Morpheus-H01: 0.1 M Amino acids, 0.1 M Buffer system 1 pH 6.5, 50% v/v Precipitant Mix 1 (5 mM **T1**, 10% NMP); Morpheus-F08: 0.12 M Monosaccharides, 0.1 M Buffer system 2 pH 7.5, 50% v/v Precipitant Mix 4 (2.5 mM **T2**, 5% NMP); Morpheus-G08: 0.1 M Carboxylic acids, 0.1 M Buffer system 2 pH 7.5, 50% v/v Precipitant Mix 4 (2.5 mM **T3**, 5% NMP); and Morpheus G04: 0.1 M Carboxylic acids, 0.1 M Buffer System 1, pH 6.5, 50 % v/v Precipitant Mix 4 (2.5 mM **T7**, 5% NMP). For fragment screening against hGALK1, 24 mg/mL hGALK1 was incubated on ice for 30 minutes with 2.5 mM **T3**, dissolved in NMP (final NMP concentration of 5%) before set up of 10 pates of the Morpheus-G8 follow-up screen at 20 °C.

2.2.3 Crystal soaking and mounting

For co-crystallisation experiments of hHAO1, 1.5 µL of the reservoir solution was mixed with 1 µL of 10mM stock of the desired ligand (CCPST or 2-hydroxypalmitate) and 1 µL of this solution was added to the crystal drop. After a few minutes' soaking time, single crystals from these drops were mounted and flash cooled in liquid nitrogen for storage before screening. Similarly, for-co-crystallisation experiments of hGALK1, 1.5 µL of the reservoir solution was mixed with 1 µL of 5 – 50 mM stock of the desired ligand (ADP, **T1** – **T7**) and 1 µL of this solution was added to the crystal drop for a few minutes before crystal mounting. For fragment screening, the TexRank crystal imaging system [220] was used to select crystallisation drops containing single, cuboid crystals of hGALK1 or hHAO1 for fragment soaking. The ECHO acoustic liquid dispensing system (Labcyte) was used to add 37.5 nL of one fragment (from supersaturated stock at 100 - 500 mM in DMSO) per selected crystallisation drop for 269 selected hGALK1 drops and 508 selected hHAO1 drops. After 1-hour incubation at room temperature, crystals were mounted using the Shifter microscope stage [221] and flash cooled in liquid nitrogen for storage.

2.2.4 Data collection, processing and refinement

Diffraction data from co-crystallisation experiments were collected at the Diamond Light Source on beamlines I03, I04, I04-1 or I24. Data were indexed, integrated and scaled automatically via the Diamond Light Source synchrotron automated Xia2 pipeline [222]. The number of molecules in the asymmetric unit was determined in the CCP4 suite of programs from the calculated Matthews's coefficient (Vm, crystal volume per unit of protein molecular weight). Phases were determined by molecular replacement performed in PHASER [223] using the hGALK1-galactose-AMPPNP structure (PDB 1wuu) or the hHAO1-FMN-glyoxylate structure (PDB 2nzl) as appropriate. Solved structures were refined through alternating rounds of automated refinement by REFMAC5 [224] and manual model building in COOT [225].

Diffraction data from fragment soaking experiments were collected in unattended mode at the Diamond Light Source synchrotron on beamline I04-1. Data were indexed, integrated and scaled automatically via the Diamond Light Source synchrotron automated Xia2 pipeline [222]. Electron density maps were imported into the XChemExplorer software suite [226]. Molecular replacement and initial refinement were performed in parallel by DIMPLE [227] using the hGALK1-galactose-AMPPNP structure (PDB 1wuu) or the hHAO1-FMN-glycolate structure (PDB 6gmb) as appropriate. After exclusion of datasets diverging in space group or with resolution >2.5 Å, DIMPLE maps were analysed together by the PanDDA software [228] to identify potential ligand binding events. The PanDDA interface was used to inspect identified hits with fragment structures and changes in protein sidechains modelled manually before refinement by the REFMAC program integrated in the XChemExplorer pipeline. All refined structures were validated with MolProbity [229] and the atomic coordinate files deposited in the Protein Data Bank. Molecular graphics figures were generated using Molsoft ICM-Pro software [230], [231].

2.3 Amplex Red fluorescence-based activity assays

Initial assays were performed using the assay buffer provided within commercially available Amplex Red assay kits (ThermoFisher). Subsequent experiments were performed in 50 mM

sodium phosphate pH 7.5, 200 mM potassium chloride, 0.01% Triton-X and either 2 mM magnesium chloride (for hHAO1 assays) or 10 mM magnesium chloride (for hGALK1 assays). The Amplex Red reagent consisted of 100 μM Amplex Red dye (from a 10 mM stock prepared in DMSO) and 0.2 U/mL horseradish peroxidase (from a 10 U/mL stock prepared in water) in assay buffer, prepared immediately before addition to assay microplates. The Amplex Red reagent used to assay hGALK1 activity assay also contained 4 U/mL galactose oxidase (from a 100 U/mL stock prepared in assay buffer). In both assays, control samples containing the appropriate amount of protein, with and without compound, and buffer instead of glycolate (HAO1 assay) or ATP (GALK1 assay), were included and DMSO concentration was kept constant across all samples. All measurements were made in technical triplicates. All microplates were covered with foil seals during incubation to protect them from light. All incubation steps were performed at room temperature.

Each plate included an array of hydrogen peroxide concentrations (0 – 75 μM) that were used to generate a standard curve describing the relationship between hydrogen peroxide concentration and fluorescence signal. Fluorescence signal was measured on a Pherastar FSX microplate reader using the FI 540 590 optics module (BMG Labtech) set for excitation and emission at 570 nm and 585 nm respectively. Fluorescence data were converted to reaction rate (μM H_2O_2 generated) or % inhibition as appropriate in Microsoft Excel before plotting and analysis of concentration-response curves, Michaelis-Menten plots and Cornish-Bowden plots using GraphPad Prism.

2.3.1 HAO1 Amplex Red fluorescence-based activity assay

Experiments to determine suitable assay parameters, validate the assay set up and characterise fragments were performed in flat-bottom, black 96-well non-binding surface plates (Corning®, Sigma Aldrich). The reaction volume was 50 μL and the reactant concentrations were 100 nM hHAO1 and 720 μM glycolate unless otherwise stated. Single concentration and concentration-response experiments with follow-up compounds were performed in flat-bottom, black 384- well Small Volume™ HiBase microplates (Greiner Bio-One). The reaction volume was 10 μL and

the reactant concentrations were 10 nM hHAO1 and 36 μM glycolate unless otherwise stated. Reaction mixtures were incubated for 10 minutes before addition of an equal volume of Amplex Red reagent, followed by a further 10 minutes incubation, measurement of the fluorescence signal and data analysis as described above.

To determine the assay's linear range, Amplex Red reagent was incubated with an array of hydrogen peroxide concentrations from 0 – 150 μM and measured as described. To determine substrate kinetics and check for interference of substrates with the Amplex Red reagent, an array of hHAO1 substrate concentrations (0-10 mM of glycolate or 2-hydroxypalmitate) was added to assay buffer with or without hHAO1 prior to addition of the Amplex Red reagent. Fragment concentration-response experiments were performed by incubation of 1 μM, 10 μM, 100 μM, 250 μM, 1 mM, 2.5 mM and 10 mM of fragments **1**, **2**, **5** and **6** with hHAO1 and either glycolate or an equivalent volume of buffer prior to addition of the Amplex Red reagent.

For the single concentration screen of follow-up compounds, 1 mM compound (**8**-**135**) was incubated with hHAO1 and either glycolate or an equivalent volume of buffer before addition of the Amplex Red reagent. For the concentration-response assay, hHAO1 was incubated with an array of 12 compound concentrations (0 – 1 mM) and either glycolate or an equivalent volume of buffer before addition of the Amplex Red reagent.

2.3.2 GALK1 Amplex Red fluorescence-based activity assay

All experiments were performed in black 384-well Small Volume™ HiBase microplates (Greiner Bio-One). The reaction volume was 20 μL and the reactant concentrations were 250 nM hGALK1, 100 μM ATP and 50 μM galactose unless otherwise stated. Reaction mixtures were incubated for 1 hour before addition of an equal volume of Amplex Red reagent (including galactose oxidase), followed by a further 40 minutes incubation before fluorescence measurement. To determine the assay's linear range and optimal incubation time, Amplex Red reagent was incubated with an array of galactose concentrations from 0 – 200 μM and the

fluorescence signal was measured after incubation for 10, 20, 30, 40 or 50 minutes. To determine the optimal hGALK1 concentration, measure ATP kinetics and check for interference of ATP with the Amplex Red reagent, an array of ATP concentrations (0-0.5 mM) was added to assay buffer with 0, 10, 25, 50, 100, 250 or 500 nM of hGALK1 prior to addition of the Amplex Red reagent.

For the single concentration assays, 1 mM tool compound (**T1**-**T7**) or 2 mM merged compound (**M1**-**M5**) was incubated with hGALK1, galactose and either ATP or an equivalent volume of buffer before addition of the Amplex Red reagent. For concentration-response and inhibition mode determination, arrays of 6 fragment concentrations (0 – 15 mM) or 12 compound concentrations (0 – 2 mM) versus seven ATP concentrations (0-400 μM) were prepared and incubated with hGALK1 and galactose before addition of the Amplex Red reagent.

GraphPad Prism was used to generate Michaelis-Menten plots for each inhibitor concentration and to interpret changes in $Vmax^{apparent}$ and $Km^{apparent}$ (*P* value for the null hypothesis that the parameter is the same in all datasets). Cornish-Bowden plots were also generated using GraphPad Prism and linear regression was used to fit the crossing point of lines. Compounds were tested against two different protein batches (hGALK1-1 and hGALK1-2), each measured in technical triplicates. Inhibition mode was determined by comparison of $Vmax^{apparent}$ and $Km^{apparent}$ values calculated in Michaelis-Menten plots and interpretation of the intersect of lines in Cornish-Bowden plots.

2.4 Additional solution assays used to characterise hGALK1 and hHAO1

2.4.1 Kinase Glo luminescence-based assay to measure hGALK1 activity

The Kinase Glo reagent, containing the enzyme luciferase and its substrate luciferin, was prepared according to the manufacturer's instructions. Control samples containing the appropriate amount of protein, with and without compound, and buffer instead of galactose, were included and DMSO concentration was kept constant. All measurements were made in technical

triplicates. All experiments were performed in white 384-well non-binding surface plates (Corning®, Sigma Aldrich) and microplates were covered with foil seals during incubation to protect them from light. All incubation steps were performed at room temperature. An array of ATP concentrations (0 – 60 μM) was included in each plate to allow standard curve generation, describing the relationship between ATP consumption and luminescence signal. Luminescence was measured by the Pherastar FSX microplate reader using a luminescence optics module. Luminescence data were converted to reaction rate or % inhibition as appropriate in Microsoft Excel before analysis with GraphPad Prism.

The reaction volume was 20 μL and the reactant concentrations were 10 nM hGALK1, 35 μM ATP and 100 μM galactose unless otherwise stated. Reaction mixtures were incubated for 1 hour before addition of an equal volume of Kinase Glo reagent, followed by a further 20 minutes incubation and measurement of the luminescence signal.

To determine the assay's linear range, Kinase Glo reagent was incubated with an array of ATP concentrations from 0 – 100 μM and measured as described. To determine galactose kinetics and check for interference of galactose with the Kinase Glo reagent, an array of galactose concentrations (0-2.5 mM) was added to assay buffer with or without hGALK1 prior to addition of the Kinase Glo reagent. For the single concentration assays, 1 mM tool compound (**T1**-**T7**), 5 mM fragment (**1** – **28**) or 2 mM merged compound (**M1**-**M5**) was incubated with hGALK1, ATP and either galactose or an equivalent volume of buffer before addition of the Kinase Glo reagent. For concentration-response and inhibition mode determination, an array of 6 fragment concentrations (0 – 15 mM) or 12 compound concentrations (0 – 1 mM) versus seven galactose concentrations (0-250 μM) was prepared and incubated with hGALK1 and ATP before addition of the Kinase Glo reagent. Concentration-response curves were generated, and inhibition mode determined as described in section 2.3.2.

2.4.2 Small-angle X-ray scattering (SAXS) analysis of hHAO1

HAO1-c002 at 15 mg/mL was subjected to High Pressure Liquid Chromatography Small Angle X-ray Scattering (HPLC-SAXS) on a KW403-4F column (Shodex) in buffer consisting of 25 mM HEPES, pH 7.5; 150 mM sodium chloride, 0.5 mM TCEP, 2% glycerol and 1% sucrose, by flowing sample through an in-vacuum quartz capillary of 1.6mm diameter at 0.16 mL/min. Scattering data were collected using a Pilatus2M detector (Dectris, CH) with a sample- detector distance of 3914 mm and a wavelength of 1 Å. The data covering the peak were radially averaged and the scattering of the buffer was subtracted before analysis using Scatter [232] and the ATSAS suite of programs [233].

2.4.3 Surface plasmon resonance to measure fragment binding to hHAO1

The binding of fragments to hHAO1 was measured by surface plasmon resonance. Purified hHAO1-c002 (30 µg/mL) was attached via the N-terminal His-tag to a Ni-NTA chip to a density of 5000 RU. The assay buffer was 20 mM HEPES pH 7.5, 0.05 % TWEEN20, 200 mM sodium chloride, 0.5 mM TCEP, 5% DMSO. The system was validated using the known inhibitor CCPST as a tool compound before measurement of fragment binding. A serial dilution (11 concentrations) was prepared in the above buffer for each analyte (CCPST, fragments) by two-fold dilution from 100 µM to 0.05 µM and the subsequent solution were passed over the chip at a flow rate of 30 µL/min.

2.4.4 Native mass spectrometry to measure compound binding to hHAO1

Purified hHAO1-c002 was exchanged from the purification buffer to the analysis buffer, 50 mM ammonium acetate pH 6.5, by overnight dialysis and concentrated to 2 mg/mL. Samples were prepared by 30 minutes incubation at room temperature with follow-up compounds based on fragment **6** (compounds **123**-**135**) at a concentration of 200 µM (final DMSO of 0.1%). Native mass spectrometry analysis was generously performed by Dr. Rod Chalk at the SGC. Briefly, samples were directly infused into a Quantitative Time-of-Flight (Q-TOF) 6530 mass spectrometer attached to a standard electrospray ionization source (Agilent Technologies), with

a constant flow rate of 6 μL/min, using a 1.0-ml gas-tight positive displacement syringe (Hamilton) and PEEK capillary tubing (inner diameter, 0.005 inches). The mass spectrometer was operated in positive ion mode, using the 1-GHz detector mode with a scan range of 100–20,000 m/z and a fragmentor voltage of 430 V. Mass spectra were then analysed using the MassHunter Qualitative Analysis software (Agilent Technologies).

2.5 Synthesis of hHAO1 follow-up compounds derived from fragment **6**

A commercially available dichloropyrimidine building block and 13 diverse amines, available in-house and greater than 150 Daltons in size, were selected for coupling through nucleophilic aromatic substitution (S_NAr). The reaction mixture for each amine contained 50 mg of the dichloropyrimidine in 2 mL acetonitrile, 2.2 molar equivalents of *N, N*-diisopropylethylamine (DIPEA) and 1.05 molar equivalents of the amine building block to give a reaction molarity of 0.168 M. Reactions were first allowed to proceed overnight at room temperature with stirring at 660 rpm. If the amine building block was still detected by liquid chromatography- mass spectrometry (LC-MS, in analytical mode) after 18 hours, further incubation periods at 50 °C (overnight) or 80 °C (overnight or for 48 hours) were tested until complete conversion was observed for all 13 compounds.

Acetonitrile was removed from completed reactions by evaporation (GeneVac® EZ-2 centrifugal vacuum evaporator, 1 hour at the "low BP", "lamp OFF" setting). After resuspension in methanol, compounds were purified by LC-MS in preparative ES+ mode at pH 6 with a flow-rate of 20 mL/min. The collected fractions were lyophilised using a GeneVac® EZ-2 centrifugal vacuum evaporator (12 hours, "HPLC-Lyo" setting). Purity of selected fractions was confirmed by analytical LC-MS and 1H-NMR. Compound stocks were then prepared by dissolving to 200 mM in DMSO. Compound characterisation details are provided in appendix 2. Compounds were synthesised and validated with help from Dr. Tamas Zsommer of the Target Discovery Institute, University of Oxford.

3. Structural biology of hHAO1 inhibition

HAO1 is a promising target for substrate reduction therapy of primary hyperoxaluria, as discussed in section 1.2.4. *In vitro* characterisation of the structure and kinetics of hHAO1 [60], [90], [203], [234], has provided a sound basis for the development of potent, selective inhibitors. To date, most published inhibitors of hHAO1 (section 1.5.4) share a similar chemical scaffold with a carboxylic acid or di-carbonyl substituted heterocycle that forms a polar interacting group, and a hydrophobic interacting group that consists of one or more aromatic rings or long hydrocarbon chains.

Crystal structures of hHAO1 or its homologues bound to some of these inhibitors confirm their binding to the active site with their polar moiety coordinated by the substrate binding residues (Tyr26, Tyr132, Arg167, His260 and Arg263 of hHAO1; illustrated in figures 1.18-1.21), and their hydrophobic moiety coordinated by hydrophobic residues proximal to the active site (Tyr26, Trp110, Tyr132, Tyr134, Leu164, Phe175 and/or Tyr208 of hHAO1). An additional structural feature of note is that, in the two crystal structures of hHAO1 bound to inhibitors (CDST, PDB 2rdt, [203] and CCPST, PDB 2w0u, [211]), the poorly conserved, highly flexible gating loop is disordered.

A promising approach to overcome the limited diversity of inhibitor scaffolds identified to date is to screen a library of small, diverse compounds, using X-ray crystallography to generate structural data in parallel, providing high quality chemical starting points for lead development. Additionally, screening by X-ray crystallography can help overcome bias towards orthosteric inhibitors, revealing binding pockets often overlooked by activity-based screening methods.

This chapter describes crystallography-based fragment screening for hHAO1. Section 3.1 describes the optimisation of hHAO1 recombinant protein production and crystallisation to facilitate a fragment screening campaign. Section 3.2 expands the structural data available for

hHAO1 to include high resolution structures with the inhibitor CCPST and substrate glycolate. Finally, section 3.3 reports the screening of a fragment library against hHAO1 crystals, discovering 7 fragment hits clustering in 3 binding pockets.

3.1 Optimisation of hHAO1 protein production and crystallisation for drug discovery

Purification and crystallisation of recombinant hHAO1 produced in *E. coli* has been reported previously [60], [234]. This section describes the scale-up and optimisation of these procedures to generate >500 hHAO1 crystals, of which most (>80%) were of high resolution (better than 2.5 Å), to facilitate fragment screening.

3.1.1 Comparison of crystallisability of hHAO1 constructs

Nine hHAO1 constructs, designed and cloned for *E. coli* expression during the original project to solve the structure of hHAO1 at the SGC, were screened to identify a construct yielding crystals in sufficient quantity and quality for fragment screening. These include a full-length hHAO1 construct (Met1-Ile370, hHAO1-c100) that was used to solve the original structure (PDB 2nzl), and a nested set of constructs with varying N-terminal (Met1, Leu5, Ile8) and C-terminal (Pro364, Ser368, Ile370) boundaries (figure 3.1). All nine constructs were subcloned in the pNIC28-Bsa4 vector which includes an N-terminal 6-histidine tag preceding a TEV protease cleavage site.

A high throughput protein production pipeline developed at the SGC, PREPX (Parallel Rapid Expression & Purification for X-ray crystallography) [162], [235], was adapted to express and purify all nine hHAO1 constructs in parallel for crystallisation trials (figure 3.2A). This approach employs auto-induction media to allow high levels of protein production from unattended growth of small volume cultures (100s of mL rather than 10s of L) and combines pre-packed affinity columns with automated serial chromatography to rapidly purify these cultures in parallel. Using this pipeline, a few milligrams of purified protein can be obtained for up to 48 constructs in less than two weeks. Typically, three to four crystal plates can be set up per milligram of purified

protein, so that in a further 4-14 days initial diffraction screening of crystals can commence.

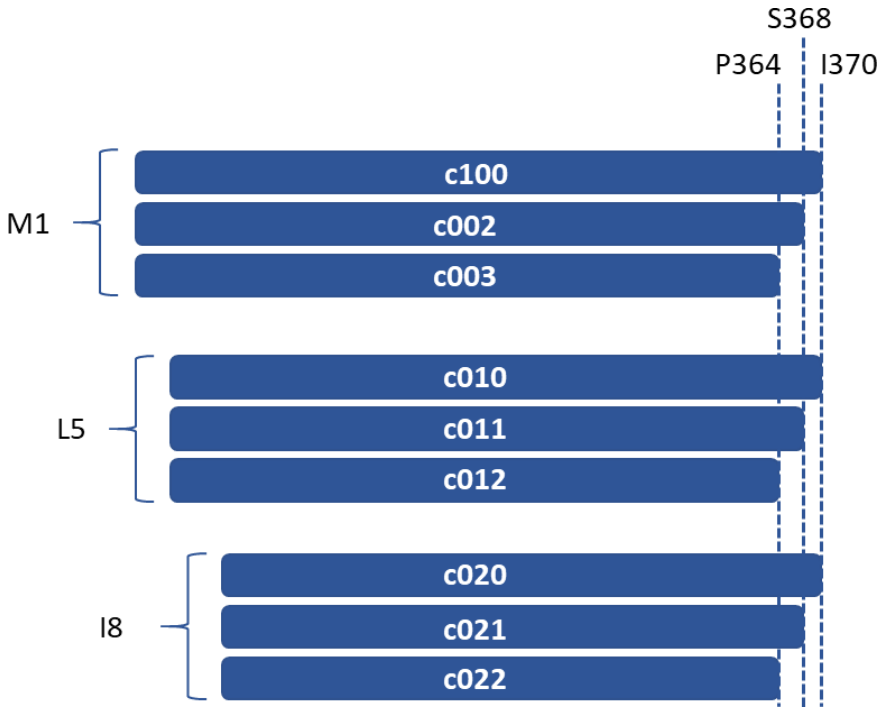


Figure 3.1: Nested constructs of hHAO1. *Diagram showing amino acid boundaries of hHAO1 constructs for E. coli expression and their respective SGC construct IDs.*

The nine His-tagged hHAO1 constructs were grown in 200 mL autoinduction media and subjected to freeze-thaw lysis before parallel purification of the clarified lysates. The first purification step was affinity purification using pre-packed Ni-NTA columns with a 1 mL bed volume (figure 3.2B). High protein yield was observed for all constructs except hHAO1-c020 (Ile8-Ile370), which was excluded from further purification. Each Ni-NTA elution was concentrated to 1 mL and further purified by size exclusion chromatography (SEC) using the BioRad serial chromatography system and an ENrich SEC 650 column with a 24 mL bed volume (figure 3.2C, D).

All chromatograms showed twin peaks that varied in co-factor loading – the first peak was darker in colour and had a higher absorbance at 360 and 450 nm, when tested at equal concentration, than the second peak indicating a higher concentration of FMN in the first peak.

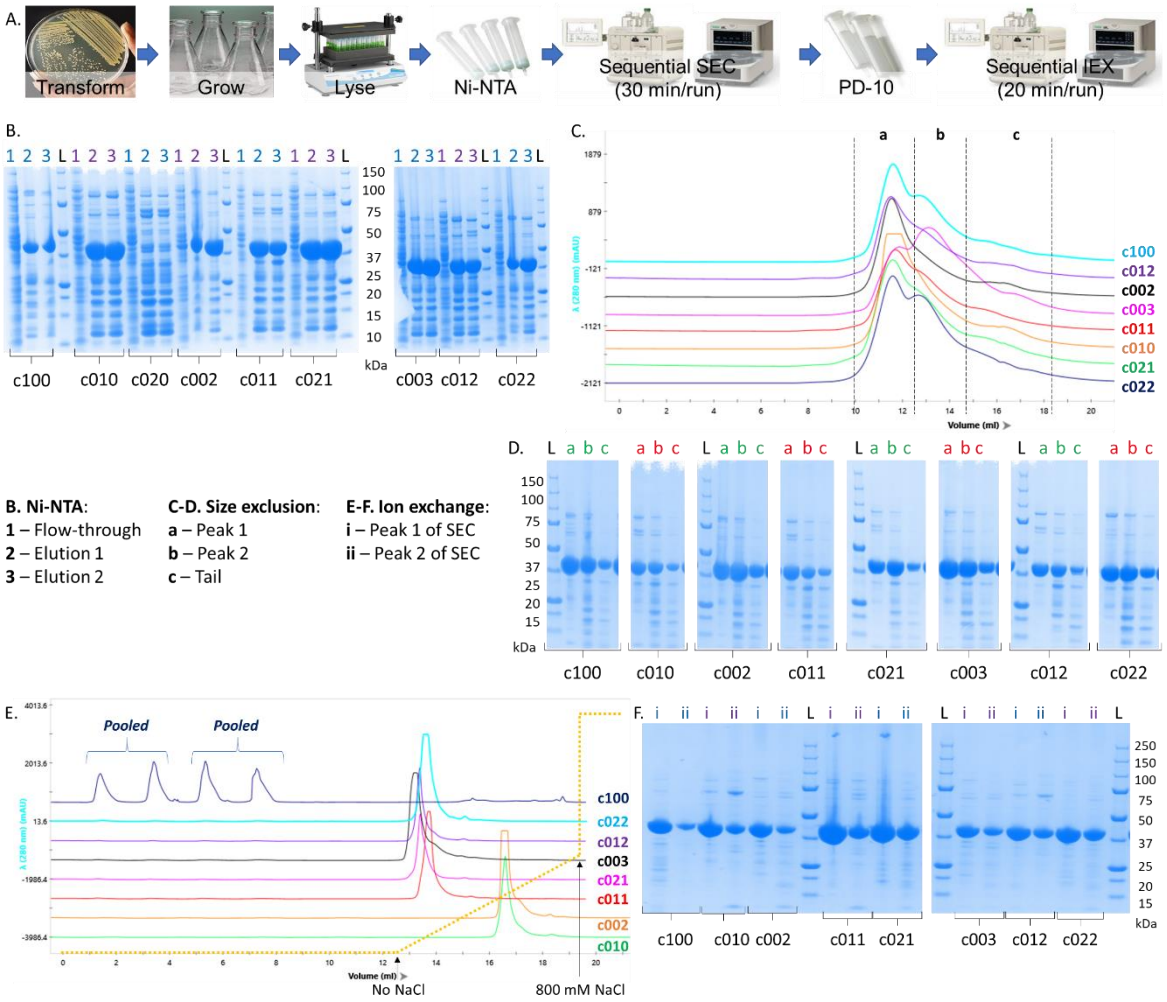


Figure 3.2: Purification of nine hHAO1 constructs by an adapted PREPX protocol. *A. Flow diagram illustrating the adapted PREPX protocol used in this work. B. SDS-PAGE analysis of Ni-NTA purification of hHAO1 constructs. Lanes labelled 1 are flow through samples from the Ni-NTA column and lanes labelled 2 and 3 represent the first and second 4 mL elution fraction. C. Superimposed and offset size exclusion chromatograms for the first elution sample of the eight hHAO1 constructs yielding enough protein after Ni-NTA for further purification, coloured according to construct. Size exclusion column was an ENrich SEC 650 column with a 24 mL bed volume. Fractions labelled 'a' represent the first purification peak, fractions labelled 'b' represent the second purification peak and fractions labelled 'c' represent the tail fraction. D. SDS-PAGE analysis of the three pooled fractions from size exclusion chromatography (a, b and c) of the eight hHAO1 constructs purified by size exclusion chromatography. E. Superimposed and offset anion exchange chromatograms for the peak 1 size exclusion samples of eight hHAO1 constructs. Anion exchange column was a Macro-Prep High Q column with a 5 mL bed volume. The salt gradient is indicated as a dashed yellow line. F. SDS-PAGE analysis of the two pooled fractions from anion exchange chromatography of the peak 1 sample of eight hHAO1 constructs. Lanes labelled 'i' represent the main ion exchange purification peak and lanes labelled 'ii' represent the tail fraction.*

Next, each of the two SEC peaks was exchanged into high-pH, no-salt anion exchange buffer (pH 7.5 to pH 8.2, 500 mM to no sodium chloride) by desalting over a pre-packed PD-10 column (Sephadex G-25 resin, 2.5 mL bed volume) and a final purification step was performed by anion exchange chromatography (AEX) using the BioRad serial chromatography system and a Macro-

Prep High Q column with a 5mL bed volume (figure 3.2E, F). SEC peak 1 from construct hHAO1-c100 (Met1-Ile370) did not bind to the AEX column at low salt and showed considerable precipitation after overnight incubation of the eluant at 4 °C. No signal was observed on the AEX run of the second SEC peak 2 sample from this construct. Collectively this indicated instability of the full length hHAO1 in the no salt condition chosen for anion exchange.

Two constructs (HAO1-c002, Met1-Ser368 and hHAO1-c010, Leu5-Ile370) eluted at a higher salt concentration (approximately 450 mM sodium chloride) relative to the other constructs, which eluted between 100 and 200 mM sodium chloride (figure 3.2E). The enhanced binding of c010 and c002 to the High Q resin indicates that the surface of these two proteins was either more negatively charged or more hydrophobic than the others, since High Q resin is positively charged and slightly hydrophobic. The final protein yields (table 3.1) from 200 mL auto-induction cultures varied between 0.9 and 5.3 mg, and the two size exclusion peaks were concentrated separately to 75 μL (sufficient volume for 3 crystal plates) or ~12 mg/mL, whichever was reached first.

Table 3.1: Purified protein details for hHAO1 constructs. *For each construct, there were two final protein samples, pooled from the first and second peaks from the size exclusion chromatography purification step.*

Construct ID	**Peak 1** Yield, final concentration	**Peak 2** Yield, final concentration
HAO1-c100	1.2 mg, 12 mg/mL	Ion exchange unsuccessful
HAO1-c002	2.9 mg, 13.1 mg/mL	2.4 mg, 15.8 mg/mL
HAO1-c003	2.0 mg, 13.5 mg/mL	0.5 mg, 7.0 mg/mL
HAO1-c010	1.0 mg, 12.7 mg/mL	0.6 mg, 7.6 mg/mL
HAO1-c011	0.9 mg, 12.3 mg/mL	0.5 mg, 6.7 mg/mL
HAO1-c012	0.6 mg, 8.5 mg/mL	0.3 mg, 4.5 mg/mL
HAO1-c021	0.8 mg, 10.5 mg/mL	0.8 mg, 11.1 mg/mL
HAO1-c022	1.8 mg, 11.8 mg/mL	0.5 mg, 6.5 mg/mL

Crystallisation trials were set up for the two final protein samples of each construct, prioritising the crystallisation conditions previously identified for hHAO1. Two coarse screens, BCS6 and LFS6, and one follow-up screen, based on three conditions from LFS6 (B1-B3: 0.1 M MIB pH 5.5 – 8.25 and 20 – 37.5% PEG1000), were set up at 4 °C for each protein sample. In all

constructs except c010 (Leu5-Ile370), both AEX peaks yielded crystals in the follow-up screen after 7 days (figure 3.3).

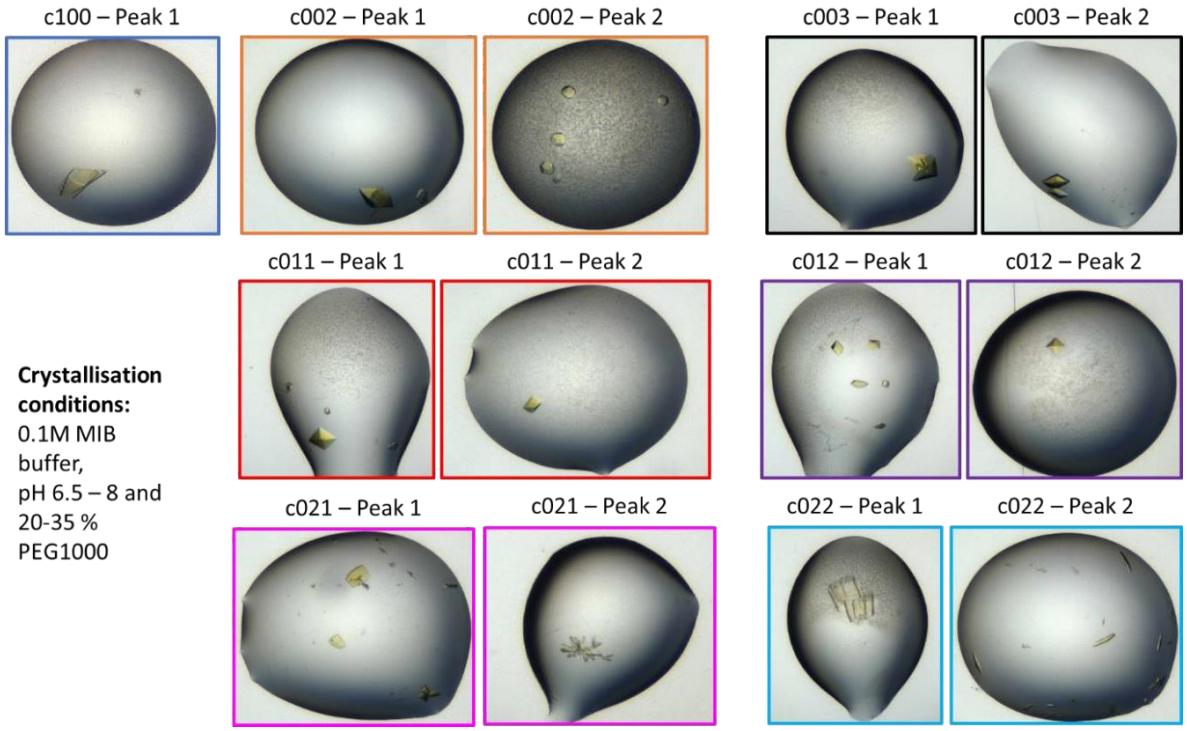


Figure 3.3: Representative crystal images of different hHAO1 constructs.

Initial diffraction screening focused on the two longest hHAO1 constructs, c100 and c002. Diffraction screening at the Diamond Light Source showed good diffraction (2.5 Å) for only one of five hHAO1-c100 crystals tested, whereas all 6 hHAO1-c002 crystals diffracted better than 2 Å. This construct yielded more than double the final protein amount relative to the others and is nearest full-length after c100 so for these reasons, c002 was selected to be scaled-up for fragment screening.

3.1.2 Scale-up of protein production and crystallisation for fragment screening

After identification of hHAO1-c002 as the most promising construct in terms of protein yield and crystal quality, protein production and crystallisation were scaled up to facilitate fragment screening. Briefly, two litres of auto-induction media expressing recombinant hHAO1-c002 were purified by Ni-NTA and SEC (figure 3.4). SDS-PAGE analysis showed a single band of the expected molecular weight, 43.2 kDa. Additional AEX purification was considered unnecessary due to the high protein purity after SEC as observed by SDS-PAGE and intact mass spectrometry

analysis. Purified hHAO1 was concentrated to 13.7 mg/mL prior to crystallisation trials. Crystallisation trials were set up in the follow-up screen described in section 2.2.2. Screening of 15 crystals (example image is shown in the inset of figure 3.4) after 7 days of growth showed diffraction of 2.5 Å for one crystal with the remaining 14 diffracting between 1.4 and 1.8 Å, confirming suitability of this crystal system for fragment screening.

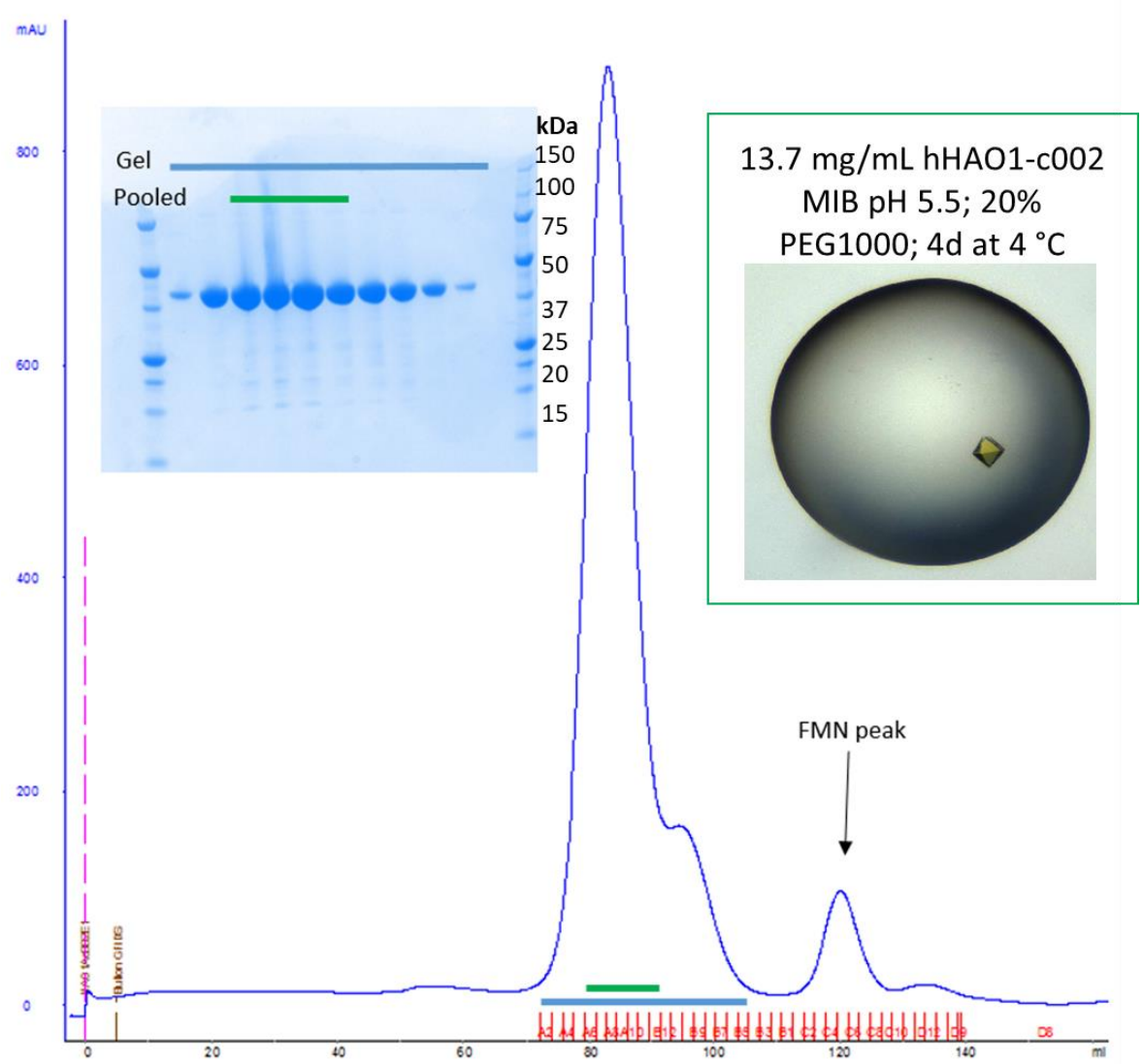


Figure 3.4: Size exclusion chromatogram, SDS-PAGE and representative crystal image for large scale purification of hHAO1-c002. *SDS-PAGE analysis and chromatogram of size exclusion purification after Ni-NTA elution of hHAO1-c002 by HiLoad 16/600 Superdex 200 column. Fractions analysed by SDS-PAGE are indicated in blue. Fractions that were pooled and concentrated as the final purification product are indicated in green. Inset: A representative crystal image and condition for hHAO1-c002.*

3.2 Expanding the structural repertoire of hHAO1

Before performance of a fragment screen (section 3.3), this crystal system was first used to extend the known structural repertoire for hHAO1 by co-crystallisation with the inhibitor CCPST

(section 3.2.2) and with the low-affinity substrate 2-hydroxypalmitate (section 3.2.3).

Additionally, small angle X-ray scattering (SAXS) was used to build a low-resolution model of the structure of hHAO1 in solution, providing information about the oligomeric state of hHAO1 (3.2.1). These data are an essential precursor to explore whether homo-oligomer disruption is a viable approach to inhibit hHAO1.

3.2.1 Small angle X-ray scattering (SAXS) data of purified hHAO1

In the size exclusion chromatography (SEC)-SAXS technique applied in this work, a monodispersed protein sample is isolated by high performance liquid chromatography (HPLC, 'SEC') and briefly exposed to high-flux X-rays (SAXS), typically under flow to minimise radiation damage [236]–[238]. The measured intensities (I) over a range of scattering angles are radially averaged to obtain a one-dimensional intensity profile as a function of the momentum transfer vector *q*. This profile can then be used to derive parameters such as the particle's molecular weight, radius of gyration (Rg, a measure of the protein radius relative to its centre-of-mass) and volume, to build *ab initio* shape envelopes of the sample, and for comparison with theoretical scattering profiles of atomic resolution coordinates or models.

In this work, scattering data were collected from purified hHAO1-c002 at the Diamond Light Source beamline b21 (figure 3.5A). The processed scattering data resulted in a linear Guinier fit (ln I(q) vs. q^2; figure 3.5B), indicating a monodispersed sample. The calculated Rg was 37.9 ± 0.4 Å. The Kratky plot (q^2I(q) vs q; figure 3.5C) indicated a folded, multi-domain (or multimeric) species in solution. The molecular weight of the protein, derived directly from the scattering curve, was 175955 Daltons, consistent with a tetramer of 43.2 kDa hHAO1 subunits (theoretical tetramer 172800 Da).

The *ab initio* envelope calculated from these data, with a resolution of 20 ± 2 Å (figure 3.5D) and the resultant fitting of the crystallographic tetramer (PDB 2nzl) to the scattering data and to the *ab initio* model is shown in figures 3.5E and 3.5F respectively (Chi^2 = 2.687). These data are

consistent with hHAO1 forming the same tetrameric tertiary structure in solution as that observed in the crystal structures and is consistent with previous size exclusion and cross- linking experiments [202]. The SAXS data therefore lend support to the development of ligands targeting the hHAO1 oligomeric interface from hits identified by X-ray crystallography (section 3.3.6).

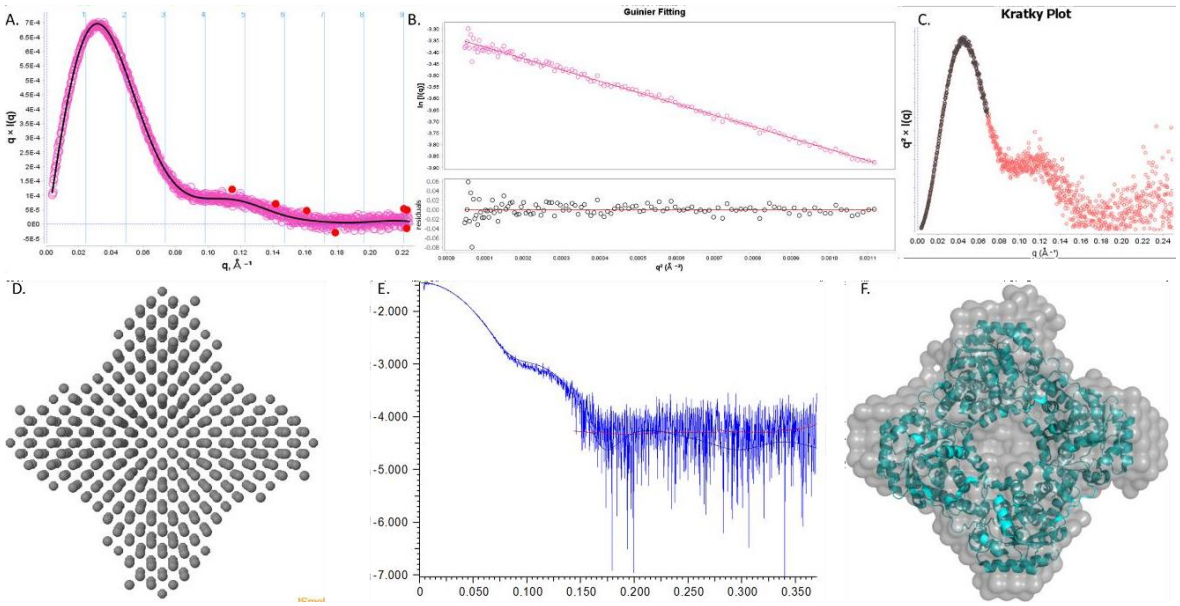


Figure 3.5: Small angle X-ray scattering (SAXS) of purified hHAO1-c002. *A. Scattering data obtained for hHAO1-c002 at the b21 beamline of Diamond Light Source synchrotron. The plot shows q x I(q) against q. Outliers are indicated as red points. B. Guinier plot of ln[I(q)] against q^2 showing linear fitting. C. Kratky plot of I(q) against q^2 showing a folded, multi-domain (or multimeric) species. D. A 3-dimensional bead model calculated from hHAO1-c002 SAXS intensity data. E. Plot showing the fit of the theoretical scattering profile, back-calculated from the tetramer from the published crystal structure (PDB 2nzl), to the experimental scattering profile collected in this work. F. A 3-dimensional surface model calculate from hHAO1-c002 SAXS data showing the fit of the hHAO1 tetramer from the published crystal structure (PDB 2nzl).*

3.2.2 Structure of hHAO1 in complex with CCPST

The structure of hHAO1-FMN-CCPST has been solved previously at 2.84 Å resolution (blue in figure 3.6A-B; PDB 2w0u). In this work, crystals of hHAO1 were grown in the presence of 2 mM CCPST to facilitate the determination of a higher resolution structure to provide further insight into the ligand-protein interactions. These crystals belonged to the tetragonal spacegroup I4 (cell dimensions a = b = 97.348, c = 80.363, α = β = γ = 90.00) with one molecule in the asymmetric unit. The structure, at 1.2 Å resolution, was solved by molecular replacement using the published structure of hHAO1-FMN-glyoxylate (grey in figure 3.6A-B; PDB 2nzl). The final refined structure converged to Rfree of 0.155 and R of 0.134. Based on difference Fourier

density, one molecule of FMN and one partially occupied molecule of CCPST (occupancy ≈ 0.50; electron density shown in figure 3.6C) were built into the asymmetric unit. Data processing and structure refinement statistics are provided in table S2.1.

Overall, the structure of hHAO1-FMN-CCPST solved in this work (yellow ribbon in figure 3.6A) is very similar to the previous low resolution structure containing the same ligands (blue ribbon in figure 3.6A, PDB 2w0u; RMSD of 0.47 Å over 334 Cα atoms) and the structure of hHAO1-FMN-glyoxylate (grey ribbon in figure 3.6A, PDB 2nzl; RMSD of 0.17 Å over 339 Cα atoms). CCPST binding to the active site is coordinated by hydrogen bonds with sidechains of Tyr26, Tyr132, Arg167 and Arg263. To accommodate the inhibitor, the sidechains of Trp110, Tyr132 and Tyr208 were displaced outward from the glycolate binding pocket relative to their observed locations in hHAO1-FMN-glyoxylate (figure 3.6B), as described previously (PDB 2w0u, section 1.5.4, figure 1.20).

Despite co-crystallisation with 2 mM CCPST, the inhibitor is not fully occupied in the structure, (figure 3.6C) as reflected by the observed sidechain orientations of Trp110 and Tyr132. Two orientations are seen for Trp110 in this structure: the inward facing orientation observed in the hHAO1-FMN-glyoxylate structure and the outward facing orientation observed in the published hHAO1-FMN-CCPST structure (figure 3.6B). Similarly, the Tyr132 sidechain in the structure from this work occupies an intermediate position relative to that observed in the published hHAO1- FMN-glyoxylate and hHAO1-FMN-CPPST structures (figure 3.6B).

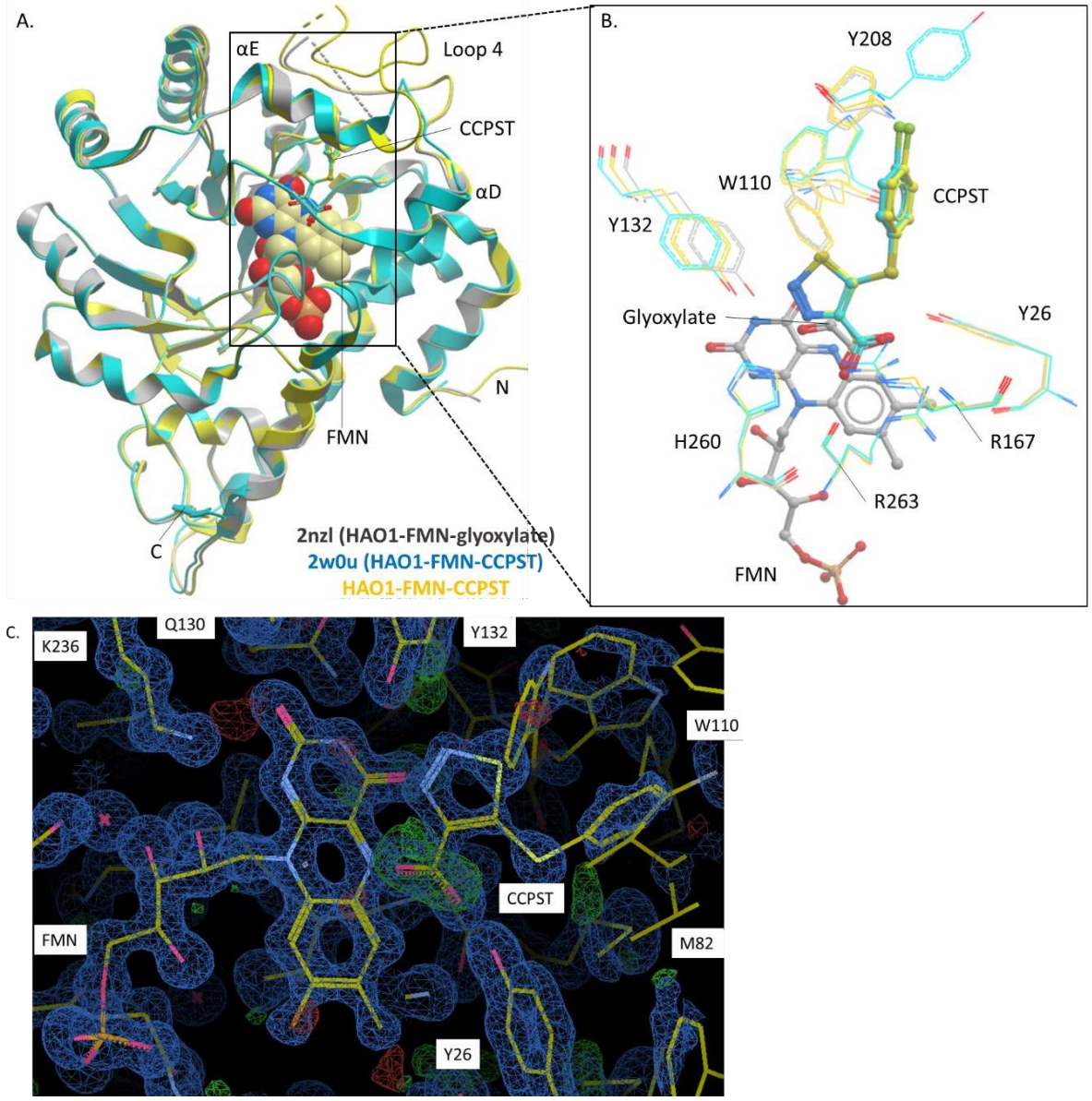


Figure 3.6: Comparison of hHAO1-FMN-glyoxylate and hHAO1-FMN-CCPST crystal structures. *A. Ribbon diagram showing superimposition of published hHAO1-FMN-glyoxylate structure (PDB 2nzl, grey), published hHAO1-FMN-CCPST structure (PDB 2w0u, blue) and the hHAO1-FMN-CCPST structure solved in this work (yellow). FMN is shown as spacefill model and CCPST is shown as sticks. B. Close-up view of active site residues in the three superimposed structures. Key active site residues are shown for the published hHAO1-FMN-glyoxylate structure (PDB 2nzl, grey lines), the published hHAO1-FMN-CCPST structure (PDB 2w0u, blue lines) and the hHAO1-FMN-CCPST structure solved in this work (yellow lines). FMN, glyoxylate and CCPST are shown as sticks. CCPST from the published structure is shown in blue and CCPST from the structure solved in this work is shown in yellow. C. Electron density map of the hHAO1 active site from the hHAO1-FMN-CCPST structure solved in this work, contoured at 1 σ. Positive and negative difference densities are displayed in green and red respectively and are contoured at 3 σ.*

The hHAO1-FMN-CCPST structure solved in this work confirms the protein-inhibitor interactions suggested by the lower resolution structure (PDB 2w0u) and validates CCPST as a suitable tool compound for use as a positive control in the activity and binding assays described

in Chapter 4.

3.2.3 Structure of hHAO1 in complex with substrate

Unlike hHAO2 and other long chain hydroxyacid oxidases, hHAO1 can oxidise both short hydroxyacids like glycolate and long hydroxyacids like 2-hydroxypalmitate. To gain structural insight into the capacity of hHAO1 to oxidise 2-hydroxypalmitate, hHAO1 was crystallised in the presence of 2-hydroxypalmitate.

Crystals of hHAO1 grown in the presence of 5 mM 2-hydroxypalmitate belonged to the same spacegroup as the hHAO1-FMN-CCPST crystals and likewise had one molecule in the asymmetric unit. The structure, at 1.35 Å resolution, was solved by molecular replacement using the same structure as above (PDB 2nzl). The final refined structure converged to Rfree of 0.130 and R of 0.110. Based on difference Fourier density, glycolate and FMN were built into the asymmetric unit. The added substrate, 2-hydroxypalmitate, could not be built into the structure, despite the high concentrations present during crystallisation, because no density was observed for C3 – C16 of the carbon chain (figure 3.7A). Data processing and structure refinement statistics are provided in table S2.1.

Glycolate was modelled rather than glyoxylate based on the density observed around the C2 atom: the C2 carbon of glycolate is bonded to the C1 carbon, a hydroxyl group and two hydrogen atoms forming a pyramidal shape, as observed in this structure (figure 3.7A), whereas the C2 carbon of glyoxylate is bonded to the C1 carbon, an oxygen and one hydrogen atom, allowing a planar conformation (figure 3.7B).

An absence of electron density for the hydrocarbon chain has been observed for other ligands bound to hydroxyacid oxidases including in the structure of hHAO1 bound to CDST (figure 3.7C) and the structure of sGOX bound to TKP (figure 3.7D). Consequently, it is not clear from this structure whether 2-hydroxypalmitate is present, but no density is visible for the majority of the carbon chain, or if glycolate was trapped in the active site due to the high concentration of 2-

hydroxypalmitate present in the solution during crystallisation.

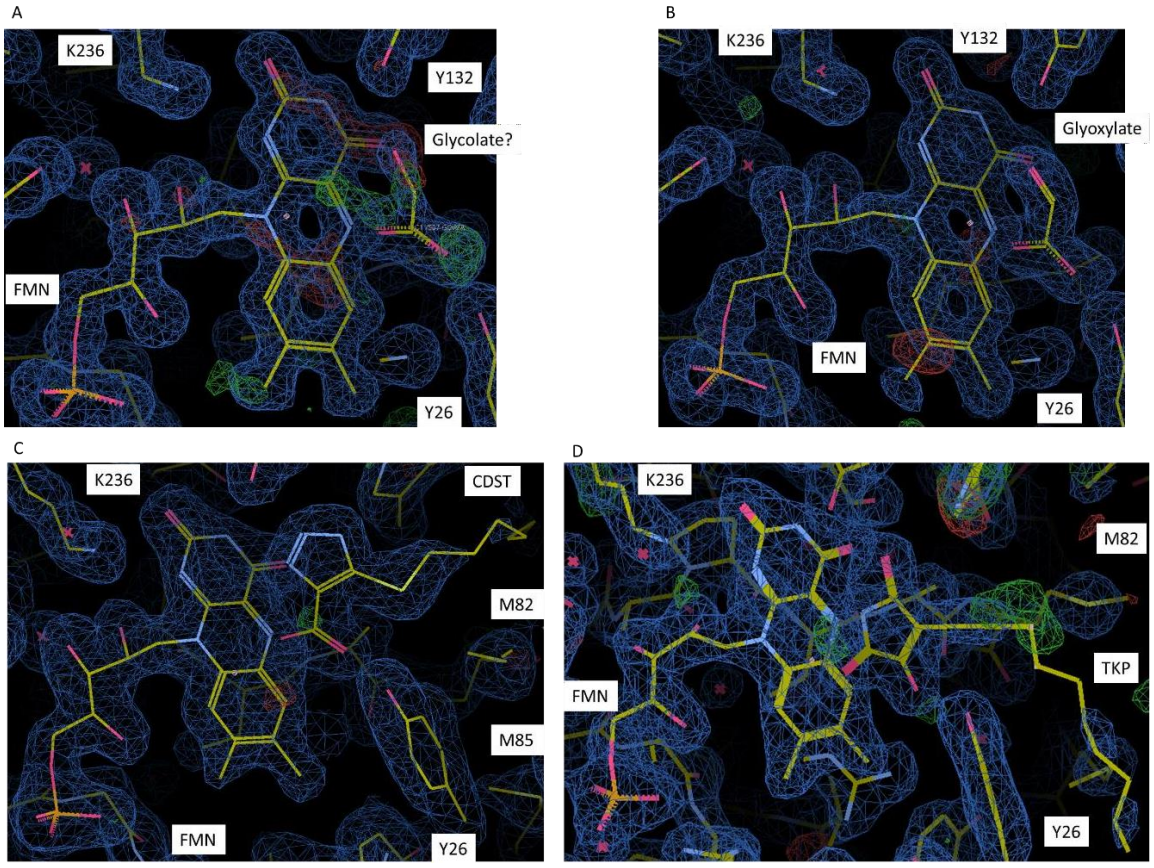


Figure 3.7: Comparison of electron densities for hydroxy-acid oxidase structures with bound ligands. *A. Close up view of the electron density map showing the active site of the hHAO1 substrate- bound structure solved in this work. The fitted model is shown as yellow sticks. The ligands FMN and glycolate have been placed in the active site. B. Electron density map showing the active site of the published hHAO1-FMN-glyoxylate structure (PDB 2nzl). The fitted model is shown as yellow sticks. C. Electron density map showing the active site of the published hHAO1-FMN-CDST structure (PDB 2rdt). The fitted model is shown as yellow sticks. D. Electron density map showing the active site of the published structure of sGOX bound to the inhibitor TKP (PDB 1al8). The fitted model is shown as yellow sticks. Residues are labelled using hHAO1 numbering. All electron density maps are contoured at 1 σ with positive and negative difference densities contoured at 3 σ in green and red respectively.*

Overall, the substrate-bound structure (lilac ribbon in figure 3.8) solved in this work is very similar to the previous high resolution structures of hHAO1-FMN-glyoxylate with a disordered gating loop (grey ribbon, PDB 2nzl; RMSD of 0.11 Å over 339 Cα atoms); hHAO1-FMN-glyoxylate with an ordered gating loop (green ribbon, PDB 2rdu; RMSD of 0.13 Å over 360 Cα atoms); and hHAO1-FMN-sulphate with an ordered gating loop (blue ribbon, PDB 2rdw; RMSD of 0.44 Å over 359 Cα atoms) (figure 3.8A) and no change was observed in the orientation of active site residues (figure 3.8B).

This structure was determined to be a more suitable molecular replacement model for the solution of fragment screening data than the preceding SGC structure (PDB 2nzl) because of its high resolution and the ordered state of the gating loop. Consequently, it was used as the molecular replacement model to process the fragment data, as described in the following section.

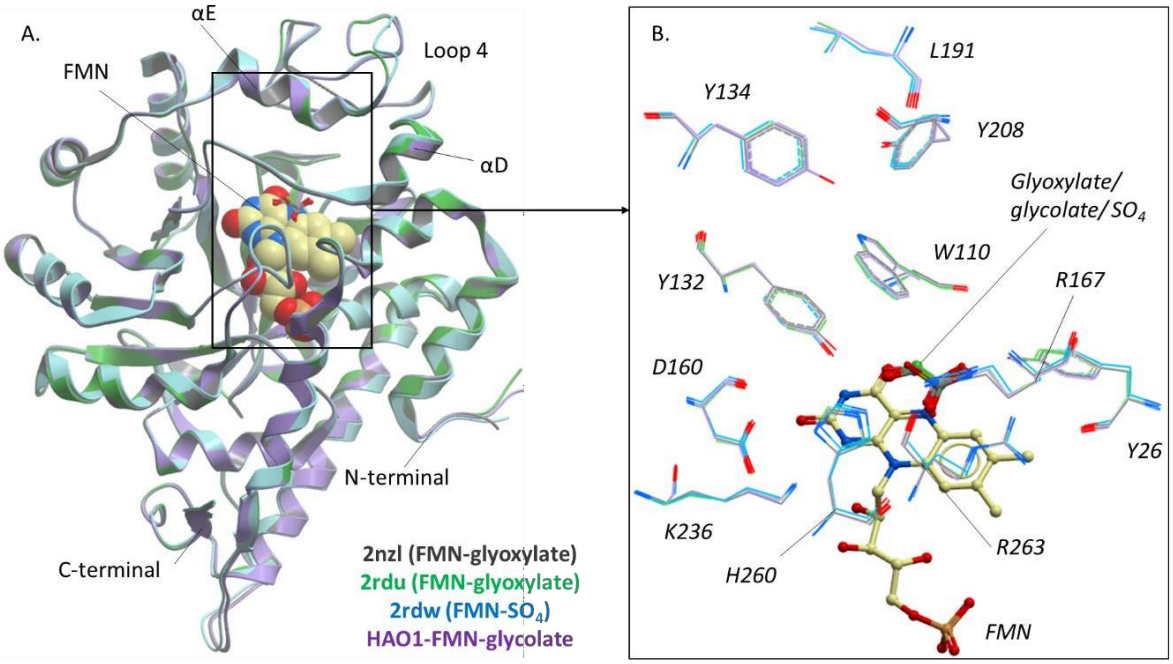


Figure 3.8: Comparison of hHAO1-FMN-glycolate, hHAO1-FMN-glyoxylate and hHAO1-FMN-sulphate structures. *A. Ribbon diagram showing superimposition of the published hHAO1-FMN-glyoxylate structure with a disordered gating loop (PDB 2nzl, grey), the published hHAO1-FMN-glyoxylate structure with an ordered loop (2rdu, green), published hHAO1-FMN-sulphate structure with an ordered loop (PDB 2rdu, blue) and the hHAO1 substrate-bound structure solved in this work (PDB 6gmb, lilac). FMN is shown as spacefill and the additional ligand is shown as sticks. B. Close-up view of active site residues in the four superimposed structures. Key active site residues are shown for the published hHAO1-FMN-glyoxylate structure with a disordered loop (2nzl, grey), published hHAO1-FMN-glyoxylate structure with an ordered loop (2rdu, green), published hHAO1-FMN-sulphate structure with an ordered loop (2rdw, blue) and the hHAO1 substrate-bound structure solved in this work (lilac). FMN, glyoxylate, sulphate and glycolate are shown as sticks.*

3.3 Fragment screening campaign: from 409 structures to 3 binding pockets

As discussed in section 1.3.3, the key benefits of using X-ray crystallography to screen for fragment binders are the increased sensitivity and hit-rate relative to more traditional solution-based approaches and the availability of structures to provide valuable direction in optimising hit compounds. hHAO1 is an ideal target for fragment screening by X-ray crystallography as it is a well-behaved protein that crystallises readily, the crystals show consistent atomic-level diffraction and high quantities of protein could be purified from construct hHAO1-c002.

3.3.1 Preparing hHAO1 crystals for fragment soaking and data collection

The optimised purification and crystallisation parameters described earlier in this chapter (section 3.1.2) were used to prepare 5 crystal plates, collectively containing >500 crystals, for fragment screening (figure 3.9). Approximately 100 cuboid crystals of 200 – 250 μm in size were selected from each crystallisation plate for fragment soaking. For the selected crystals, an ECHO acoustic liquid dispenser was used to add one fragment to each crystallisation drop, to a final concentration of 20-100 mM fragment and 20% DMSO. After incubation at room temperature for 1 hour, crystals were mounted using the Shifter, a microscope x-y stage that greatly improves crystal handling, and flash-frozen in liquid nitrogen [162], [221], [226].

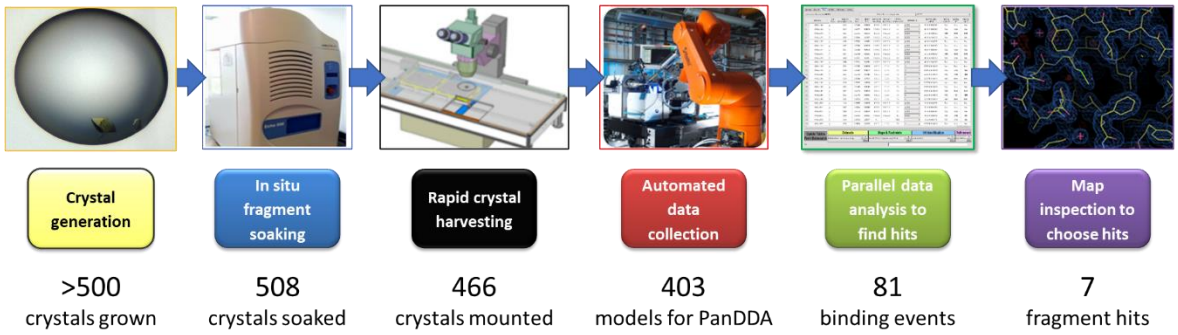


Figure 3.9: Overview of the hHAO1 fragment screening campaign. *Approximately 500 crystals of hHAO1-c002 were grown in the previously described crystallisation condition at 4 °C. An example crystal image is shown (yellow). 508 crystals of suitable size and cuboid appearance were selected, and one fragment was added to each crystallisation drop using the ECHO acoustic liquid dispensing system (blue). 42 of these crystals melted or were not correctly mounted after soaking so that a total of 466 fragment- soaked crystals were mounted for data collection (black). After exclusion of low-resolution datasets and datasets differing in space group from the majority, 403 models were submitted to the PanDDA software for difference density identification (red). PanDDA software identified 81 binding events, characterised by difference densities not present in the ground state (green). Visual inspection confirmed fragment binding in 17 structures and further analysis determined that 7 were non- artefactual, fragment-binding events (purple).*

In total, 466 out of 508 fragment-soaked hHAO1 crystals were successfully mounted. Of the 42 crystals that were soaked but not mounted, 18 appeared melted after incubation with the fragment, whereas the remaining failed crystals (n=24) were down to user error (loops broken or empty after mounting). Data collection was performed using the automated, unattended mode of the Diamond Light Source i04-1 beamline, with automated X-ray centring and the BART crystal

loading robot to facilitate rapid data collection. For 19 of the mounted crystals, automated X-ray centring failed and consequently no dataset was collected. Thus, a total of 447 datasets were collected for downstream processing.

3.3.2 Processing and overall analysis of fragment datasets

Diffraction data were integrated automatically through the DLS processing infrastructure in the xia2 pipeline [222], and the XChemExplorer interface [226] was employed to run molecular replacement and initial refinement on all suitable datasets in parallel (DIMPLE software [227]). The search model used by DIMPLE for molecular replacement was the hHAO1 substrate-bound structure described in the previous section. Next, the PanDDA software [228] was used to generate difference maps and search for potential fragment-binding events. PanDDA performs local alignment of electron density maps from all of the phased structures and generates spatial Z-score maps, quantifying the relative variability at each position in every dataset, which serves two purposes.

First, the low Z-score positions are used to generate an average map, called a ground state map, consisting of the best estimate of electron density shared by the majority of structures and thus representative of crystals in the absence of fragment binding. The ground state map is then used for background correction of all datasets to reduce experimental noise and minimise the confounding influence of stochastic variations between datasets. Second, clusters of high Z-score locations indicate electron densities that are absent from the majority of datasets, providing a statistically meaningful, objective measure of potentially interesting densities, referred to as binding events, such as fragment binding.

For datasets to be comparable by the PanDDA method, they must share the same space group and have similar unit cell dimensions and resolution [228]. Using these parameters, 16 of the collected hHAO1 datasets were unsuitable due to their space group (10 datasets in I 4 2 2, 4 in P 1 and one each in C 1 2 1 and F 2 2 2) and 36 were excluded due to low resolution (> 2.5 Å). PanDDA analysis was performed on 403 datasets, diffracting to 1.4 – 2.2 Å, belonging to the

tetragonal space group I 4 (unit cell dimensions a = b = 96 - 98 Å, c = 79 - 81 Å, α = β = γ = 90°) with 1 copy of hHAO1, containing one molecule of FMN, per asymmetric unit. Data processing and structure refinement statistics are provided in table S2.1.

PanDDA analysis of the 403 comparable hHAO1 models identified 81 binding events - weak electron densities of high Z-score clusters diverging from the ground state map – across 58 datasets. After visual inspection of the positive datasets, fragments could be confidently built into difference densities in 17 datasets. The high false positive rate observed in this fragment screening campaign is attributable to changes in the flexible gating loop region, where displaced loop residue densities were identified as hits by the PanDDA software, but no causative fragment binding could be detected. Consequently, manual building of the loop and additional refinement was required before the false positives could be excluded. After ligand placement and building of the gating loop, structures were refined in parallel using the REFMAC programme integrated in the XChemExplorer pipeline, followed by manual fitting of fragment binding residues and a final REFMAC run.

Of the 17 fragments that bound to hHAO1, 10 fragments were excluded as artefactual binders that made few interactions with hHAO1 instead binding due to the packing of hHAO1 molecules within the crystal lattice (figure 3.10A). Of the remaining 7 hits, two are bound in the glycolate binding pocket of the active site (fragments **1** and **2**, orange and yellow respectively in figure 3.10B-D), three are bound above the active site, close to the gating loop (fragments **3** – **5**, light green, dark green and blue respectively in figure 3.10B-D) and two are bound at the interface between two subunits of the hHAO1 tetramer (fragments **6** and **7**, purple and lilac respectively in figure 3.10B-D). Closer inspection of one hHAO1 chain of the superimposed structures shows that there is minimal conformational change in the presence of any of the fragments relative to the hHAO1-FMN-glycolate structure, described in the previous section, except that variable conformations are seen for the gating loop (figure 3.10C).

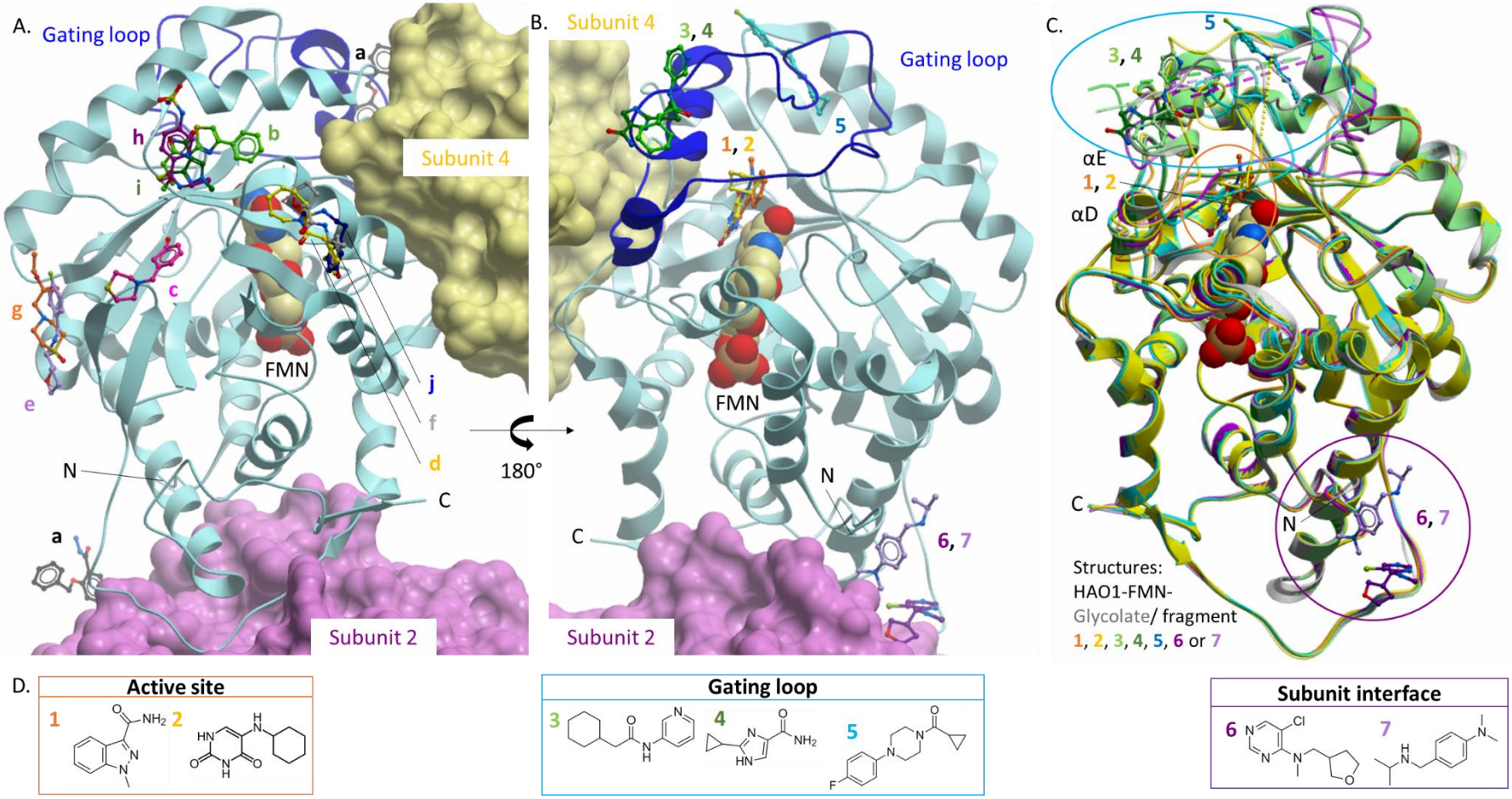


Figure 3.10: Overview of fragments binding to hHAO1 crystals. *A. Ribbon and surface diagram of the hHAO1 tetramer showing the positions of fragments* ***a*** *to* ***j*** *that formed few interactions with the protein and were excluded from further analysis as artefacts. One subunit of the tetramer is shown as light blue ribbon, one as purple surface representation and a third as yellow surface representation. The gating loop region (Tyr163-Ile213) is dark blue in the subunit shown in ribbon representation. FMN is shown as spacefill and fragments as sticks. B. Ribbon and surface diagram of the hHAO1 tetramer illustrating the position of fragments* ***1*** *–* ***7*** *that were selected for further analysis. C. Ribbon diagram of one hHAO1 subunit showing the superimposition of the seven selected fragment-bound structures and the hHAO1-FMN-glycolate structure described earlier in this work. Fragments are bound in 3 areas: in the glycolate binding site (****1****,* ***2****); close to the gating loop (****3*** *–* ***5****) and at subunit interface (****6****,* ***7****). FMN is shown as spacefill and fragments are shown as sticks. C. Chemical structures of the non-artefactual binders identified by fragment screening against hHAO1 crystals.*

The following sections provide a more detailed analysis of fragment binding events at the active site (section 3.3.3), at the gating loop (section 3.3.4) and at the subunit interface (section 3.3.5).

3.3.3 Fragments **1** and **2** bound in the glycolate binding pocket

Two fragments (**1**, **2**), sharing similar chemical scaffolds to the published inhibitors of hydroxyacid oxidases (discussed in section 1.5.4; chemical structures are shown figure 3.11A), were identified in the glycolate binding pocket of the active site (figure 3.10B-C). As mentioned in the previous section, the overall structures of hHAO1-FMN-fragment **1**, hHAO1-FMN-fragment **2** and hHAO1- FMN-glycolate are similar (RMSD 1.1 Å over 327 Cα or 2.8 Å over 346 Cα between hHAO1- FMN-glycolate and hHAO1-FMN-fragment **1** or hHAO1-FMN-fragment **2** respectively), except for the degree of ordering and the observed conformation of the gating loop (figure 3.10C). Most of the gating loop (Phe175-Ile213) is disordered in the hHAO1-FMN-fragment **1** structure (red ribbon) and part of the gating loop (Lys176-Ser203) is also disordered in the hHAO1-FMN- fragment **2** structure (orange ribbon) (figure 3.11B). The αE helix of the gating loop is displaced outwards from the active site in the hHAO1-FMN-fragment **2** structure by up to 5.8 Å at the furthest point, the carboxyl group of Gly204 (orange ribbon, figure 3.11B).

Fragment **1** binding is coordinated through interactions with invariant active site residues (Tyr132, Arg167, His260 and Arg263; invariant in sequence alignment shown in figures 1.14 and 3.14). Specifically, the carbonyl group of **1** overlaps with the hydroxyl moiety of glycolate and is positioned to form similar hydrogen bonds with either the sidechain hydroxyl of Tyr132 or the τ-nitrogen of His260 while the amine group of **1** is positioned to form hydrogen bonds with either the ε-nitrogen of Arg263, the η_1-nitrogen of Arg263 or the sidechain hydroxyl of Tyr26 (figure 3.11C). The methyl-indazole group of **1** displaces Trp110 and the N2 atom of the pyrazole can form a hydrogen bond with the side chain hydroxyl of Tyr132. Finally, the phenyl ring of the indazole displaces residues Met82, Leu205 and Tyr208 all of which are disordered in the hHAO1-FMN-fragment **1** structure.

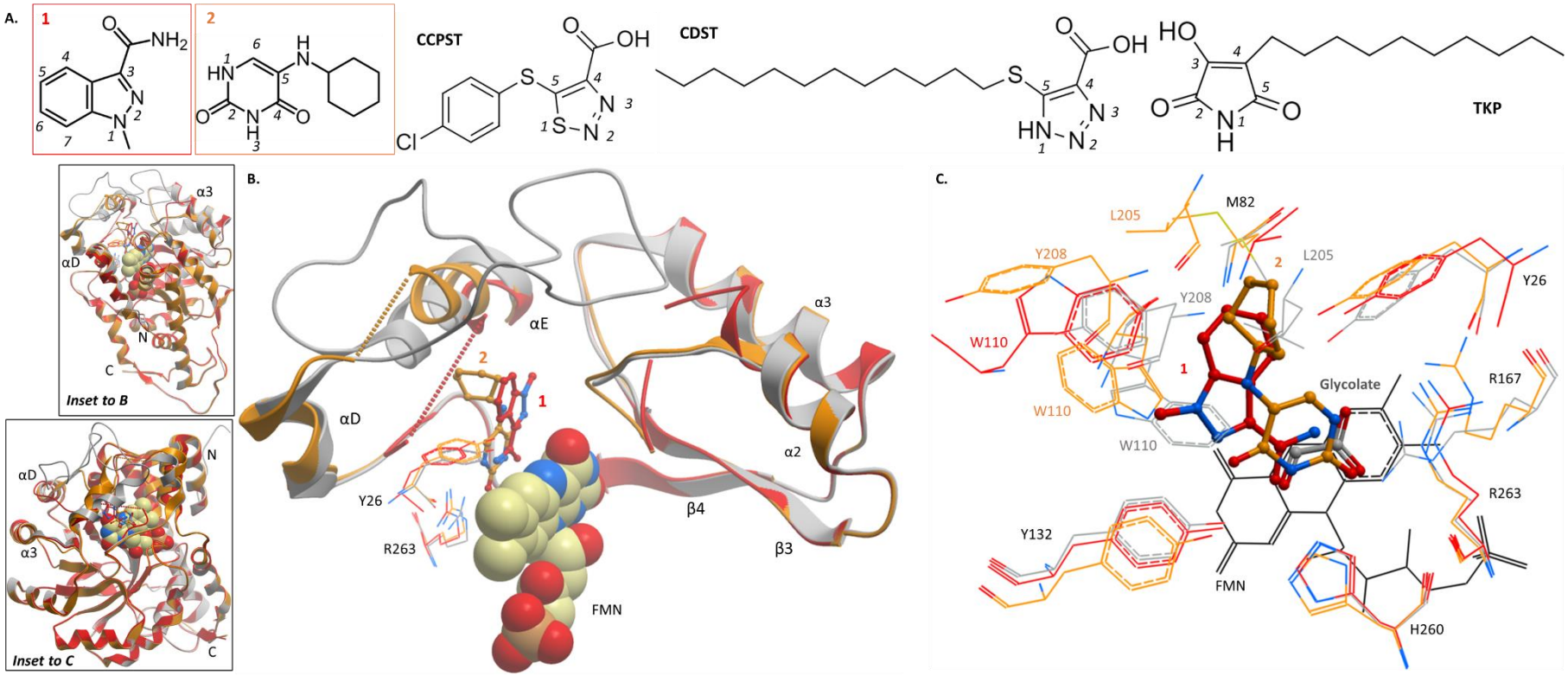


Figure 3.11: Fragments 1 and 2 bind in the glycolate binding pocket of the active site. *A. Chemical structures of fragment **1**, fragment **2** and published inhibitors of hHAO1. Ring constituents are numbered according to the IUPAC system. Resiudes Tyr26 and Arg263 are shown as lines to aid orientation. Glycolate is omitted for clarity. B. Close-up view of superimposed hHAO1-FMN-glycolate (grey), hHAO1- FMN-fragment **1** (red) and hHAO1-FMN-fragment **2** (orange) structures. Residues Ser108 to Ile213 are shown as ribbon representation. FMN is shown as spacefill and fragments are shown as sticks. C. Close-up view of the active site of superimposed hHAO1-FMN-glycolate (grey), hHAO1-FMN-fragment **1** (red), and hHAO1-FMN-fragment **2** (orange) structures. Key active site residues and FMN are shown as lines. Fragments and glycolate are shown as sticks. No electron density was observed for residues Leu205 or Tyr208 in the hHAO1-FMN-fragment **1** structure. Inset: ribbon diagrams showing overall orientation of superimposed hHAO1-FMN-glycolate (grey), hHAO1-FMN-fragment **1** (red) and hHAO1-FMN-fragment **2** (orange) structures in panels B and C. FMN is shown as spacefill and fragments **1** and **2** are shown as sticks.*

Fragment **2** binding is coordinated by the same active site residues. One of the carbonyl groups of **2** overlaps with the carboxylic acid hydroxyl moiety of glycolate and can form similar hydrogen bonds with the τ-nitrogen of His260, the ε-nitrogen of Arg263 or the η_1-nitrogen of Arg263 (figure 3.11C). This carbonyl group also displaces the guanidino group of the Arg167 sidechain by 3.7 Å, at a 45° angle away from the glycolate binding site, so that rather than hydrogen bonding with the carboxylic acid of glycolate, this sidechain now bonds with the sidechain of Asp170, part of the αD helix. The N3-atom of **2** can form hydrogen bonds with the hydroxyl sidechain of Tyr132, the π-nitrogen of the His260 sidechain or the sidechain guanidino of Arg263. The cyclohexylamine moiety of **2** displaces the sidechains of Met82, Trp110 and Leu205, so that both sidechains rotate 180° to point away from the glycolate binding site, and also displaces the αE helix as described above.

Fragment **1** is superficially similar to known inhibitors of hHAO1 such as CDST and CCPST (figure 3.11A, section 1.5.4). Specifically, they share the scaffold of a polar head group containing a five-membered heterocycle ring (carboxamide substituted pyrazole in **1**) and a hydrophobic group such as an aliphatic chain or aromatic ring (a phenyl ring fused to the pyrazole in **1**). Fragment **1** differed from published inhibitors in that it contained a carboxamide group rather than a carboxylic acid and the hydrophobic moiety was a phenyl ring fused to the pyrazole of the polar head group rather than connected via a thioether linker as in the known inhibitor CCPST. The presence of a methyl group on the N2 nitrogen of the fragment **1** pyrazole causes the polar moiety to be less embedded in the active site so that the carbonyl group of fragment **1** superimposes with the glycolate C2-hydroxyl moiety rather than the substrate carboxylic acid. The scaffold of fragment **2** is somewhat similar to a different published inhibitor, TKP, except that the fragment polar head group is a pyrimidine-2,4-dione rather than a pyrrole-2,5-dione, and the hydrophobic group of fragment **2** is a cyclohexylamine rather than an aliphatic chain as in TKP (figure 3.11A).

3.3.4 Fragments **3 - 5** bound near to the gating loop

Novel binding poses near the gating loop were observed for three fragments (**3** – **5**). Chemical starting points binding near the gating loop are more likely to be selective for hHAO1 (over hHAO2) because the gating loop is the least conserved secondary structure feature across the 2-hydroxyacid oxidase family, varying in both length and sequence (section 1.5.1, sequence alignments in figures 1.12 and 1.14). Rationale for targeting the gating loop to inhibit hHAO1 is also provided by published work indicating that the gating loop has a role in substrate recognition and specificity [203].

The overall structures of hHAO1-FMN-fragment **3** and hHAO1-FMN-fragment **4** are very similar to the hHAO1-FMN-glycolate structure (RMSD < 1 Å over 362 Cα; figure 3.12A). The gating loop is partially disordered in both fragment-bound structures and, unlike for the active site fragments described in the previous section, no change in conformation was observed in the visible portions of the gating loop for either fragment-bound structure. Taken together with the position of these two fragments at a crystal packing interface (figure 3.12B), these two fragments were determined to be poor starting points for initial hit development. Consequently, the main utility of these fragments would be for merging or linking to more buried compounds binding in nearby pockets.

In contrast to the weakly coordinated, surface-binding nature of fragments **3** and **4**, fragment **5** is mostly buried within the hHAO1 protein chain, in a hydrophobic pocket partially formed by the gating loop (figure 3.13A). The overall structures of hHAO1-FMN-fragment **5** is similar to that of hHAO1-FMN-glycolate (RMSD 2.7 Å over 335 Cα) except for the disordering and displacement of the gating loop (figure 3.13B). This gating loop pocket is formed by the α3 helix (Leu131-Met150) on one side, the αE helix (Val171-Ile213) opposite it and the αD- αE loop region of the ordered structure wraps around the remaining two sides (figure 3.13C). Fragment **5** consists of a hydrophobic fluorophenyl moiety, which is buried in the described pocket, and a polar cyclopropyl(piperazin-1-yl)methanone group that is more surface exposed.

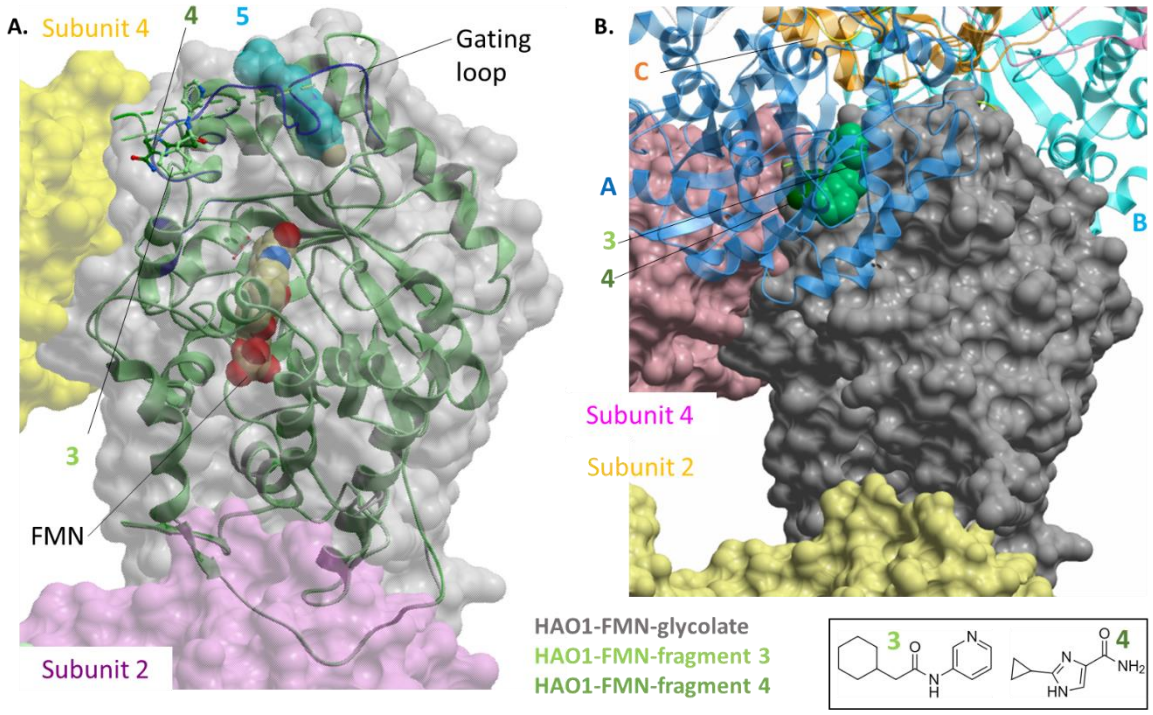


Figure 3.12: Fragments 3 and 4 bind at the protein surface close to the gating loop. *A. Ribbon diagram showing superimposition of hHAO1-FMN-glycolate (grey), hHAO1-FMN-fragment **3** (light green) and hHAO1-FMN-fragment **4** (dark green) structures. FMN and fragment **5** are shown as spacefill and fragments **3** and **4** are shown as sticks. The gating loop (Tyr163-Ile213) is shown in dark blue ribbon. Surface representations of adjacent subunits of the tetramer are shown in yellow, grey and purple. B. Surface representation of the hHAO1 tetramer with nearby protein chains within the crystal (labelled A-C) shown as ribbons. Fragments **3** and **4** are shown as light green and dark green spacefill respectively.*

Binding of the fluorophenyl group of fragment **5** is coordinated by hydrophobic interactions with the sidechains of Trp110, Tyr134 and Ile213. This group displaces the sidechain of Tyr134 to avoid a steric clash, disrupting the hydrogen bonding network between Trp110, Tyr134 and Tyr208 (detailed in section 1.5.3, figure 1.16). Additional minor displacement of the β2-α2 loop (loop 2) away from the glycolate binding site, including a 1.6 Å displacement of the sidechain of key active site residue Trp110, could also be attributed to the loss of this hydrogen bonding network (figure 3.13C-D). Binding of the cyclopropyl(piperazin-1-yl)methanone moiety of fragment **5** is stabilised by hydrogen bonding to the peptidic backbone of residues Tyr134-Asp136. This group occupies the space where the ordered gating loop is observed in the hHAO1-FMN-glycolate structure and displaces the αE helix by approximately 4 Å and 30° away from the αD helix. Consequently, the displaced residue Tyr208 is now mostly surface exposed and able to form Van der Waals interactions with the cyclopropyl and carbonyl groups of

fragment **5** (figure 3.13C-D).

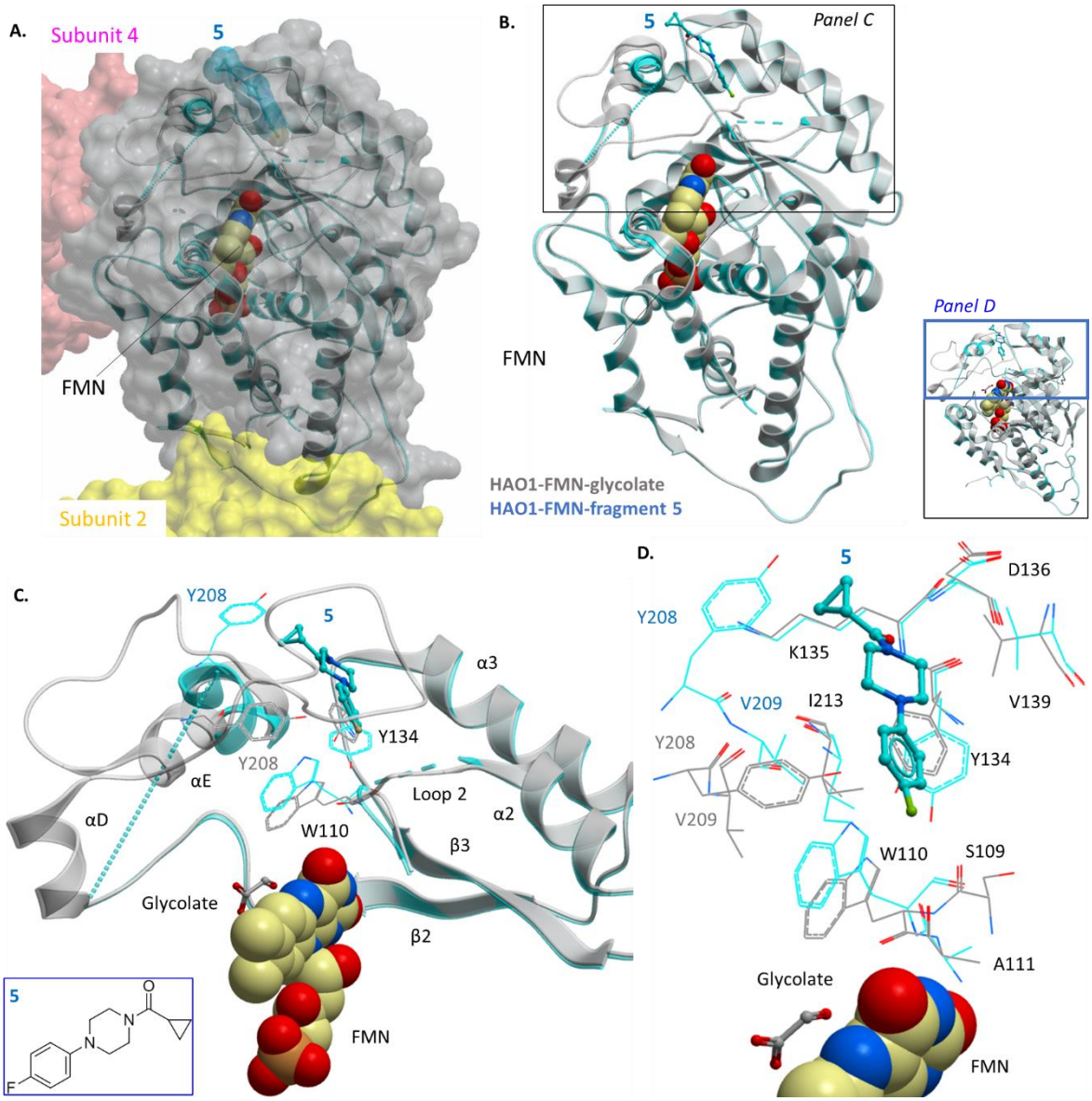


Figure 3.13: Fragment 5 binds in a hydrophobic pocket partially formed by the gating loop. *Surface representation of adjacent subunits showing the binding site of fragment **5** relative to hHAO1 tetramer subunits. Superimposed hHAO1-FMN-glycolate (grey) and hHAO1-FMN-fragment **5** (blue) structures are shown as ribbons. FMN is shown as spacefill and fragment 5 as both sticks and spacefill. Surface representations of adjacent subunits of the tetramer are shown in yellow, grey and pink. B. Ribbon diagram showing superimposition of hHAO1-FMN-glycolate (grey) and hHAO1-FMN- fragment **5** (blue) structures. FMN is shown as spacefill and fragment **5** as sticks. Black box indicates view in panel C. C. Ribbon diagram of hHAO1 secondary structures that form the fragment **5** binding site. hHAO1-FMN-glycolate (grey) and hHAO1-FMN-fragment **5** (blue) structures are shown as ribbons. Inset: Chemical structure of fragment **5**. D. Close-up view of the fragment **5** binding site with superimposed hHAO1-FMN-glycolate (grey) and hHAO1-FMN-fragment **5** (blue) structures. hHAO1 residues that coordinate fragment **5** binding are shown as lines. FMN is shown as spacefill and fragment **5** and glycolate are shown as sticks. Inset: ribbon diagrams showing the overall orientation of structures in panel D.*

The published hHAO1 inhibitors to date either bind to the active site (e.g. CCPST, CDST, TKP, TACA), coordinated by the highly conserved active site residues (red asterisks in figure 3.14),

or are predicted to do so.

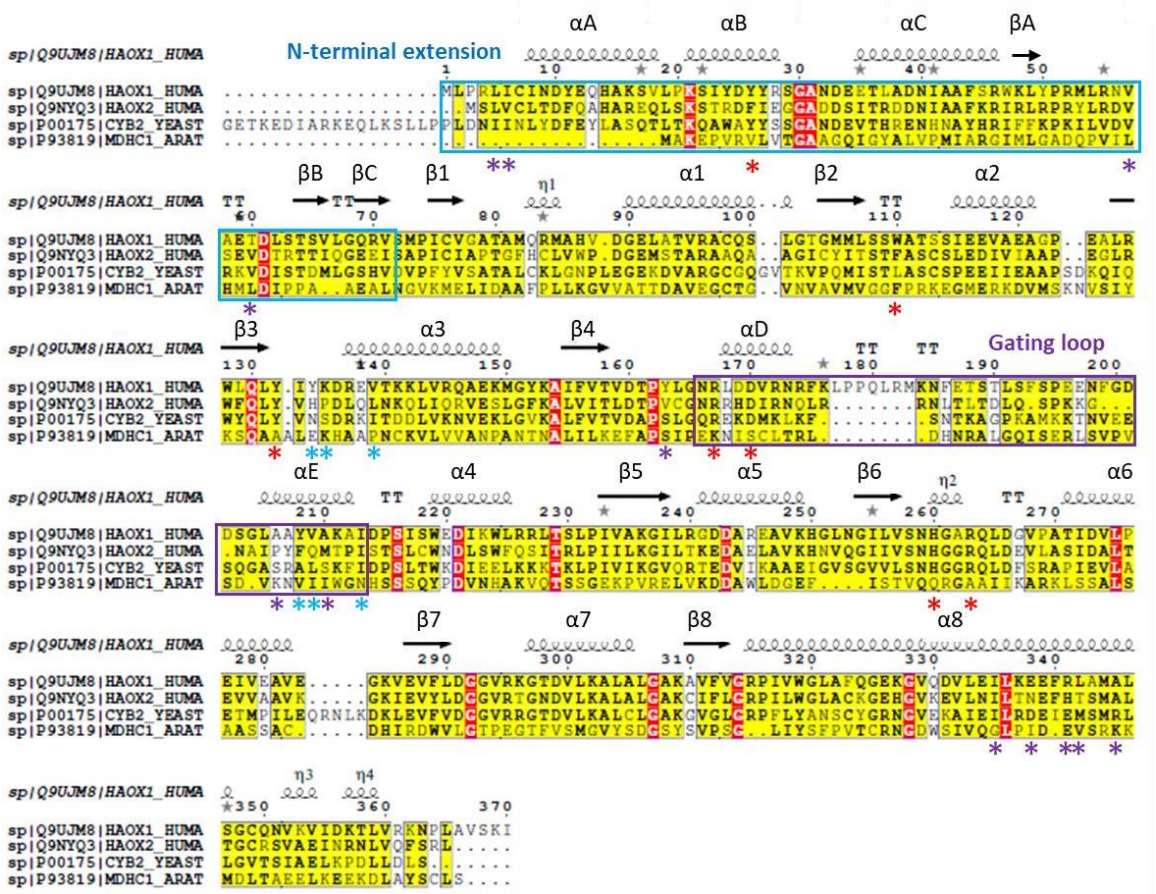


Figure 3.14: Sequence alignment of representative members of the (S)-2-hydroxyacid oxidase family highlighting residues involved in fragment binding. *Sequence alignment of human hHAO1, human HAO2, lactate oxidase domain of yeast CYB2 and A. thaliana mandelate dehydrogenase (MDHC1) – prepared using the MultiAlin software [192], [193]. Invariant residues are shown as white characters on a red background, residues that are similar in >70% of sequences are shown as bold characters on a yellow background and residues that are similar in >50% sequences also have a yellow background. The N-terminal extension is boxed in blue and the gating loop in purple. Protein residues interacting with active site fragments **1** and **2** – Tyr26, Trp110, Tyr132, Arg167, Asp170, His260 and Arg263 – are indicated by red asterisks. Protein residues interacting with gating loop fragment **5** – Tyr134, Lys135, Asp136, Val139, Tyr208, Val209 and Ile213 – are indicated by blue asterisks. hHAO1 residues interacting with subunit-subunit interface fragments **6** and **7** – Leu5, Ile6, Val57, Thr60, Tyr163, Ala206, Ala210, Ile335, Glu338, Arg341, Leu342 and Ala345 – are indicated by lilac asterisks.*

The high degree of homology observed between hHAO1 and its homologues (sections 1.5.1, 1.5.2) often translates to low selectivity of hHAO1 inhibitors over other 2-hydroxyacid oxidases with similar efficacy observed against both short and long chain isozymes e.g. characterisation of inhibition of hHAO1 [203] and rHAO2 [206] indicated similar binding interactions and potency for the active site inhibitor CCPST. This raises concerns about the safety profiles of drugs developed from the active site scaffolds reported to date that may inhibit hHAO2, which

plays a role in onco-suppression and correlates with increased metastasis and decreased survival in hepatocellular carcinoma when downregulated [239].

In contrast, the fragment **5** binding site is poorly conserved (blue asterisks in figure 3.14) and the αD-αE loop of the gating loop is 10 amino acids shorter in hHAO2 than in hHAO1 (figure 1.14; section 1.5.2). hHAO2 residues equivalent to those binding fragment **5** in hHAO1 are also different e.g. hHAO1 residues Tyr134, Val139, Lys135 and Val209 are equivalent to His132, Leu137, Pro133 and Gln195 of hHAO2. Thus, fragment **5** represents an exciting avenue for the development of highly specific hHAO1 inhibitors.

3.3.5 Fragments **6** and **7** bound at the subunit-subunit interface

Fragments **6** and **7** bound at the interface between subunits of the hHAO1 tetramer (figure 3.15A-B) and did not cause any significant changes to the overall structure of hHAO1 (RMSD < 1 Å over 359 Cα between both fragment-bound structures and the hHAO1-FMN-glycolate structure described earlier in this work). The interface binding pocket occupied by these fragments is formed by the N-terminal βA-βB loop and sections of the αA, α7 and α8 helices of one subunit and part of the gating loop, the β6-α6 loop and part of the αB and αC helices of the other subunit (figure 3.15C). Similar to the gating loop pocket, this pocket at the subunit-subunit interface may hold promise for the development of hHAO1-selective inhibitors as the residues that coordinate binding of fragments **6** and **7** are poorly conserved (lilac asterisks, figure 3.14).

Fragment **6** consists of a chloropyrimidine ring and a tetrahydrofuran ring connected by an amine linker (figure 3.15, inset). Binding of the chloropyrimidine group at the hHAO1 subunit interface is coordinated through π - π aromatic stacking with the side-chain aromatic ring of Tyr163 of one subunit (figure 3.15D, purple), through hydrogen bonding of the pyrimidine N1 nitrogen with the hydroxyl group of Thr60 and hydrophobic interactions of the chlorine group with Val57, Leu342 and Ala345 of the other subunit (figure 3.15D, blue).

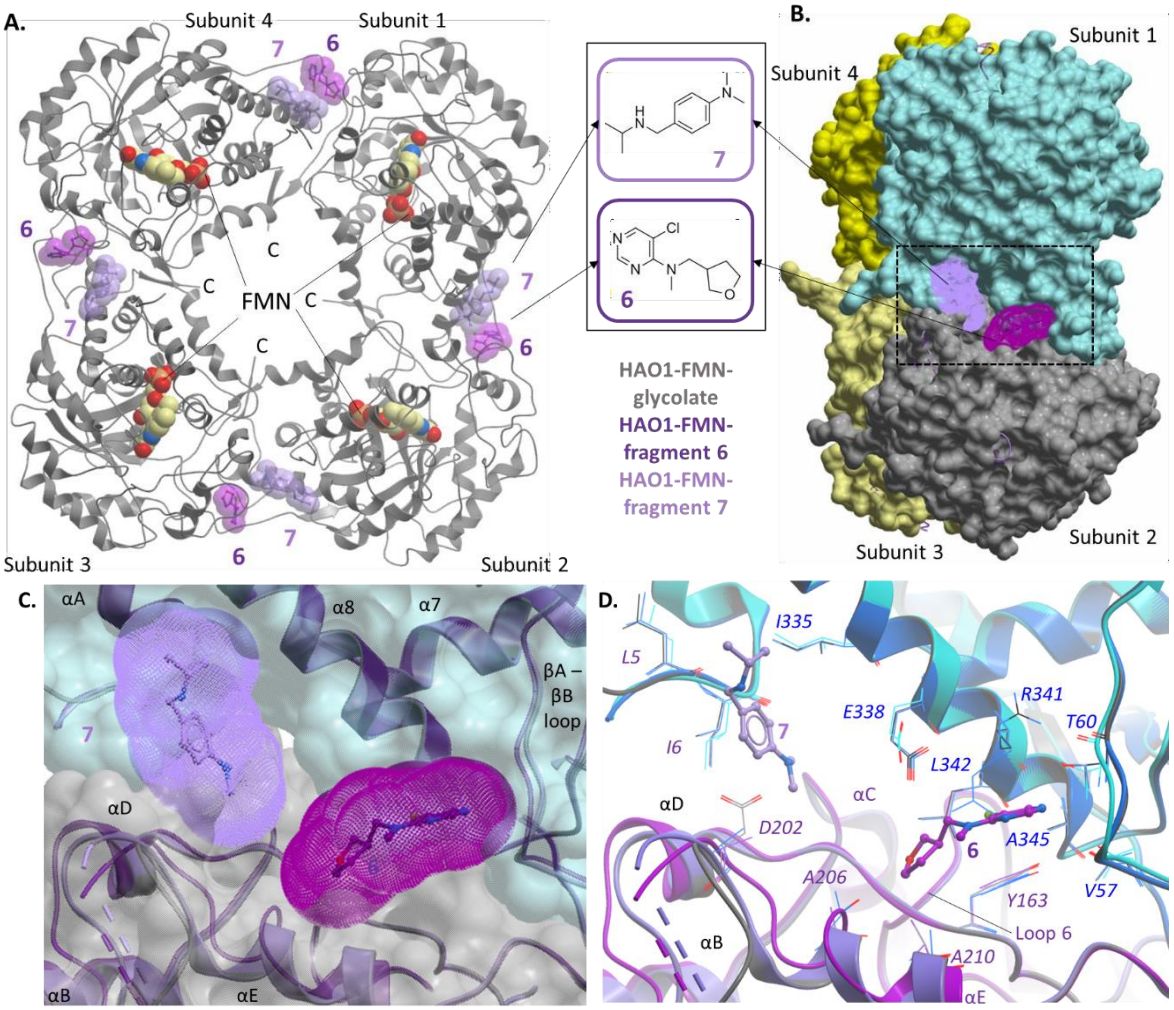


Figure 3.15: Fragments 6 and 7 bind at the interface between subunits of the hHAO1 tetramer. *A. Ribbon diagram showing where fragments **6** and **7** (purple and lilac spacefill and sticks) bind with respect to the hHAO1 tetramer. FMN is shown as spacefill. B. Side-view of hHAO1 tetramer with subunits shown as different coloured surface representations. Fragments **6** and **7** are shown as sticks with purple and lilac mesh surfaces respectively. Inset: Chemical structures of fragments **6** and **7**. C. Close-up surface representation of the binding environment of fragments **6** and **7**. Subunits surfaces are coloured blue or grey. Superimposed ribbons are of hHAO1-FMN-glycolate (grey), hHAO1-FMN-fragment **6** (purple) and hHAO1-FMN-fragment **7** (lilac) structures. Fragments **6** and **7** are shown as sticks with purple and lilac mesh surfaces respectively. D. Close up ribbon and stick diagram of the hHAO1 subunit interface where fragments **6** and **7** bind. Residues that interact with bound fragments are shown as lines and fragments are shown as sticks. Subunits of the hHAO1-FMN-glycolate are in grey and black; subunits of the hHAO1-FMN-fragment **6** are in purple and dark blue; and subunits of the hHAO1-FMN-fragment **7** structure are in lilac and light blue.*

The tetrahydrofuran ring of fragment **6** is stabilised by weak Van der Waals interactions with Ala206 and Ala210 of one subunit (figure 3.15D, purple) and Leu342 and Glu338 of the other subunit (figure 3.15D, blue). Due to the weak electron density in the crystal structure and the small size of tetrahydrofuran ring, the stereochemistry of this group could not be resolved. Finally, the nitrogen of the amine linker hydrogen bonds with the guanidino side chain of Arg341

of the first subunit (figure 3.15D, purple).

Fragment **7** consists of a dimethylaniline ring with a propan-2-amine chain and forms fewer interactions with hHAO1 than fragment **6**. The dimethylaniline ring is coordinated by hydrophobic interactions with one subunit only (residues Leu5, Ile6 and Ile335; figure 3.14, blue) whereas the propan-2-amine chain points out towards the solvent space (figure 3.15D). Consequently, only fragment **6** from the subunit-subunit interface was selected for hit characterisation and development studies, described in Chapter 4.

This chapter describes the expression, purification and crystallisation of recombinant hHAO1, extends the known structural data available for hHAO1 and describes the identification of 7 fragment binders of hHAO1 by X-ray crystallography. The four promising fragments described in this chapter were next characterised for their ability to bind to and inhibit hHAO1 in solution (chapter 4). Subsequent design and testing of follow-up compounds derived from these fragment hits is also described in Chapter 4.

4. Characterisation and optimisation of fragment hits for hHAO1

Chapter 3 described a fragment screen against hHAO1 by X-ray crystallography that identified four promising fragment hits in three different binding sites. This chapter describes the characterisation of these fragments in solution (section 4.1), and the subsequent design and testing of follow-up compounds based on fragments **1** and **2** at the active site (section 4.2), fragment **5** at the gating loop (section 4.3) and fragment **6** at the subunit interface (section 4.4).

4.1 Characterisation of hHAO1 fragment hits in solution

An activity assay and a binding assay were set up to characterise the effect of fragments **1**, **2**, **5** and **6** in solution, with the known active site hHAO1 inhibitor, CCPST, used as positive control.

The first assay was an Amplex Red fluorescence-based activity assay, previously published in a high throughput format for hHAO1 (figure 4.1A) [240]. The assay reagent contains horseradish peroxidase (EC 1.11.1.7) and its substrate Amplex Red (10-acetyl-3,7-dihydroxyphenoxazine). Hydrogen peroxide generated from the oxidation of glycolate by hHAO1 is used by horseradish peroxidase to reduce Amplex Red to its highly fluorescent product, resorufin (7-hydroxy-3-phenoxazin-3-one) which can be detected by fluorescence emission at 585 nm after excitation at 570 nm.

The resultant fluorescence signal is proportional to the amount of hydrogen peroxide produced and thus to hHAO1 oxidative activity, within the measurement window of the assay. At high concentrations of hydrogen peroxide, horseradish peroxidase further reduces resorufin to generate resazurin which does not fluoresce, demarcating the upper limit of the measurement window. The characterisation of fragment hits using the Amplex Red activity assay is described in section 4.1.1.

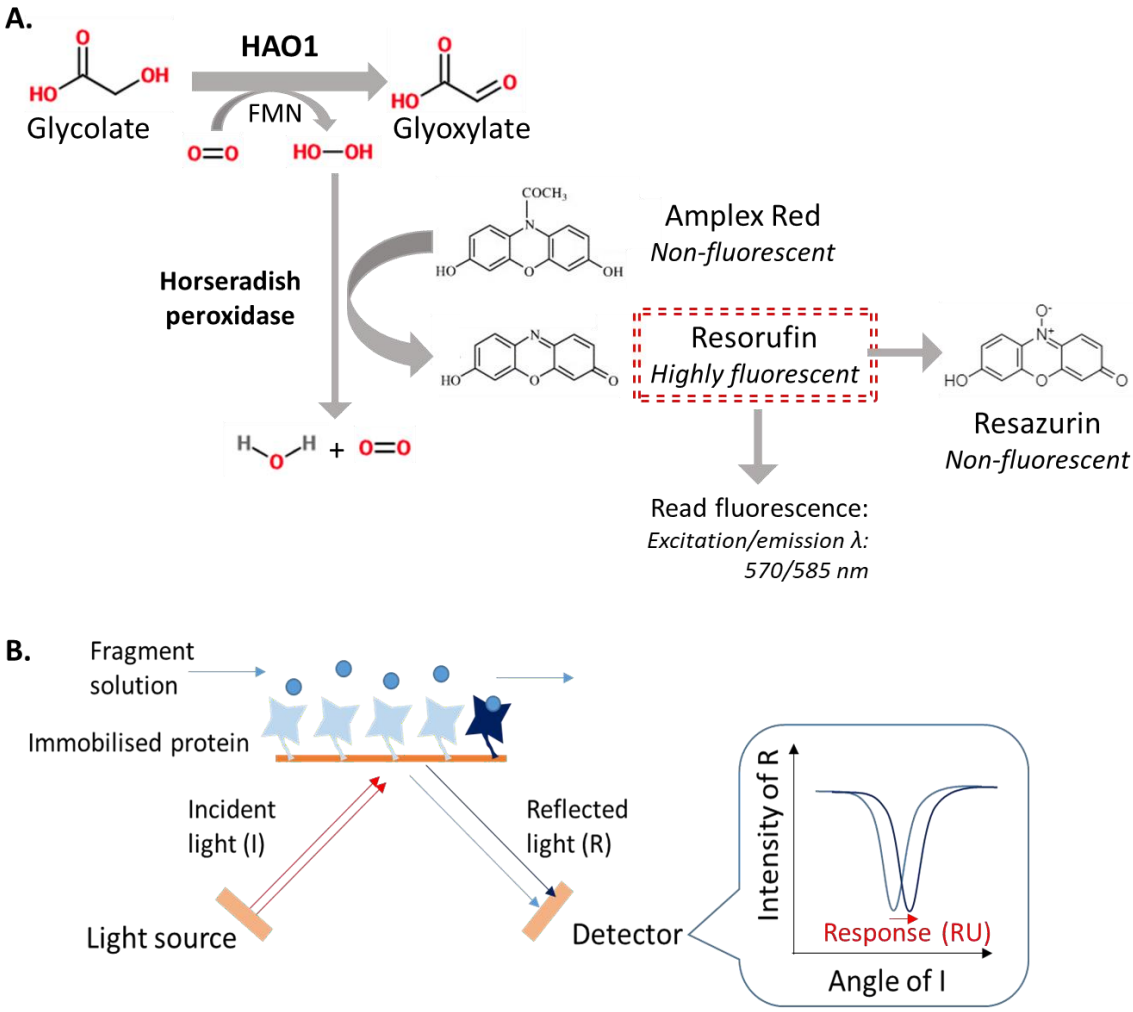


Figure 4.1: Principle of the assays used to characterise the effect of fragments on hHAO1 in solution. *A. Reaction scheme for the Amplex Red activity assay. Hydrogen peroxide produced by the oxidation of glycolate by hHAO1 is used by horseradish peroxidase to reduce Amplex Red, generating the highly fluorescent product, resorufin. Horseradish peroxidase can also further reduce resorufin to the non-fluorescent resazurin. B. Principle of surface plasmon resonance binding assay. Binding of fragments (analyte) to the immobilised protein (ligand) is detected by changes in the reflection of polarised light.*

The second assay employed in this work was surface plasmon resonance (SPR), a binding assay in which the protein of interest is immobilised onto a planar metal surface and a beam of polarised light is bounced off the surface (figure 4.1B). A solution containing the molecule of interest (termed the analyte, in this case varying concentrations of each fragment) is passed over the immobilised protein (referred to as the ligand, in this case His-tagged hHAO1-c002). When a fragment binds, this changes the intensity of light reflected at different angles and the degree of change (the 'response') is directly proportional to the strength of binding of the analyte. The characterisation of fragment binding using SPR is described in section 4.1.2.

4.1.1 Application of Amplex Red assay to hHAO1 fragment characterisation

Figures 4.2A and 4.2B show that the relationship between fluorescence signal and hydrogen peroxide concentration was linear up to 75 µM hydrogen peroxide (R^2 = 0.995). When the hydrogen peroxide concentration exceeds 75 µM, the hydrogen peroxidase in the assay reagent begins to reduce the fluorescent resorufin to the non-fluorescent resazurin (figure 4.2A).

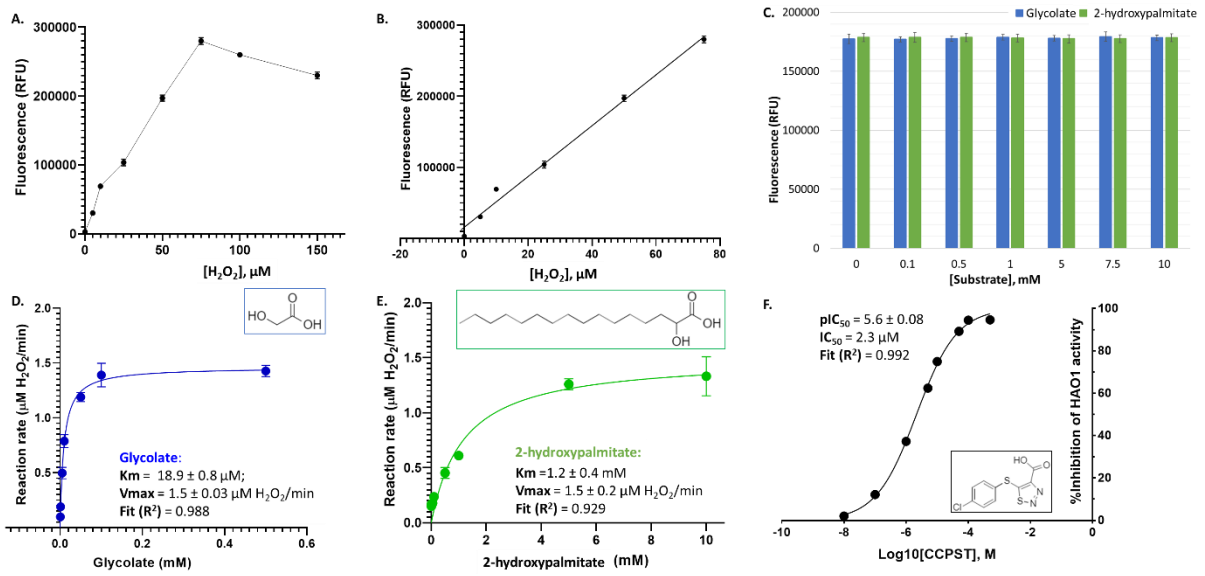


Figure 4.2: Determination of Amplex Red assay parameters for the kinetic characterisation of hHAO1 activity in vitro. *A. Plot of relative fluorescence generated from the Amplex Red reagent in the presence of increasing concentrations of hydrogen peroxide, showing non-linearity at higher hydrogen peroxide concentrations. B. Plot of relative fluorescence versus hydrogen peroxide concentration showing the standard curve fitted within the linear range of the Amplex Red reagent. C. Bar chart showing the relative fluorescence generated by the Amplex Red reagent in the presence of increasing concentrations of the hHAO1 substrates glycolate and 2-hydroxypalmitate. D. Michaelis-Menten plot of glycolate concentration versus reaction rate for 100 nM hHAO1. Inset: chemical structure of glycolate. E. Michaelis-Menten plot of 2-hydroxypalmitate concentration versus reaction rate for 100 nM hHAO1. Inset: chemical structure of 2-hydroxypalmitate. F. Concentration-response curve showing % inhibition of hHAO1 activity by increasing concentrations of the published inhibitor, CCPST. Inset: chemical structure of CCPST. Throughout this figure, error bars are standard deviation of three technical replicates.*

Figure 4.2C shows that neither of the two tested hHAO1 substrates (glycolate and 2-hydroxypalmitate) interfere with the Amplex Red reagent up to 10 mM concentration. To confirm purified recombinant hHAO1 in this work was active, the reaction rate of 100 nM hHAO1 was measured across a range of glycolate (figure 4.2D) and 2-hydroxypalmitate (figure 4.2E) concentrations. The same maximal velocity ($Vmax^{apparent}$ = 1.5 µM H_2O_2/minute) was reached at high concentrations of both substrates, albeit requiring an order of magnitude

greater concentration of 2-hydroxypalmitate than of glycolate. The calculated $Km^{apparent}$ glycolate was 18.9 ± 0.8 μM, which is fairly similar to the published value of 74.5 μM [240]. In agreement with published work, hHAO1 was much less active against 2-hydroxypalmitate with a calculated $Km^{apparent}$ of 1.2 ± 0.4 mM.

As an additional validation step, the inhibition of hHAO1 (100 nM) by the published inhibitor CCPST was measured over a range of inhibitor concentrations (10 nM to 500 μM) and a concentration-response curve was fitted to these data (figure 4.2F). As expected, concentration-dependent inhibition of hHAO1 by CCPST was observed, with a calculated IC_{50} value of 2.3 μM, which is in close agreement with the published value of 4.5 μM under the same conditions [206].

To test the effect of the four promising fragment hits (**1**, **2**, **5** and **6**) on hHAO1 activity, 100 nM hHAO1 was pre-incubated with varying concentrations of each fragment (1 μM - 10 mM) for 10 minutes before addition of substrate (720 μM glycolate) and hHAO1 activity was measured by adding the Amplex Red reagent. The two active site fragments (fragment **1** in red and fragment **2** in orange) and the fragment from the gating loop site (fragment **5** in blue) weakly inhibits hHAO1 in a dose-dependent manner, whereas no inhibition is observed by up to 10 mM of the fragment from the subunit interface (fragment **6** in purple) (figure 4.3).

The maximal inhibition observed for fragment **1** was 85 ± 0.8 % and a sigmoidal curve, consistent with hHAO1 inhibition proportional to fragment concentration, was fitted with high confidence (red curve in figure 4.3; $R^2 = 0.989$). The calculated IC_{50} of fragment **1** was 420 μM ($LogIC_{50} = -2.5 \pm 0.1$). The maximal inhibition observed for fragment **2** was 37 ± 1.5 % (orange curve in figure 4.3) and the maximal inhibition observed for fragment **5** was 19 ± 2.7 % (blue curve in figure 4.3). The weak potency of these two fragments precluded IC_{50} calculation, since 50% inhibition of hHAO1 activity was not achieved at the maximal tested fragment concentration of 10 mM.

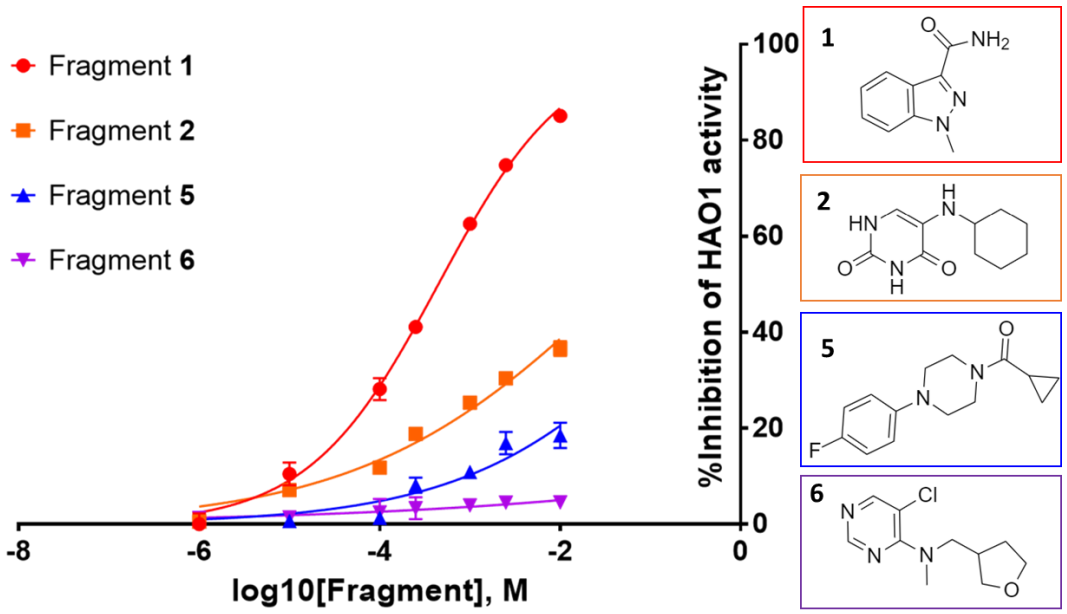


Figure 4.3: Inhibition of hHAO1 by fragment hits detected using Amplex Red activity assay. *Left: Plot of log_{10} concentration of fragments **1**, **2**, **5** and **6** versus %inhibition of hHAO1 activity. Right: Chemical structures of the tested fragments.*

4.1.2 Application of SPR assay to hHAO1 fragment characterisation

The use of SPR to measure small molecule binding to hHAO1 has not previously been described, therefore, this assay was first validated by measuring binding of the published inhibitor CCPST to immobilised hHAO1. To do so, a series of SPR titrations were performed, flowing increasing concentrations of CCPST (the analyte) over His_6-tagged hHAO1 immobilised on a Ni-NTA chip (the ligand). The relative responses at each CCPST concentration (0.05 to 100 µM) were plotted to calculate the affinity of CCPST for hHAO1 (figure 4.4A).

A maximal specific binding response (B_{max}) of 45.3 RU was observed when titrating up to 100 µM CCPST, consistent with saturation of four binding sites per molecule of tetrameric hHAO1. The calculated equilibrium dissociation constant (K_D) was 16.9 ± 2.3 µM. When data were used to fit specific binding only, the calculated Hill slope was 0.75, indicating negative cooperativity between binding of CCPST to different hHAO1 monomers i.e. binding of CCPST to one (or

more) subunit(s) of the tetramer reduces CCPST affinity for the remaining subunit(s) of the tetramer.

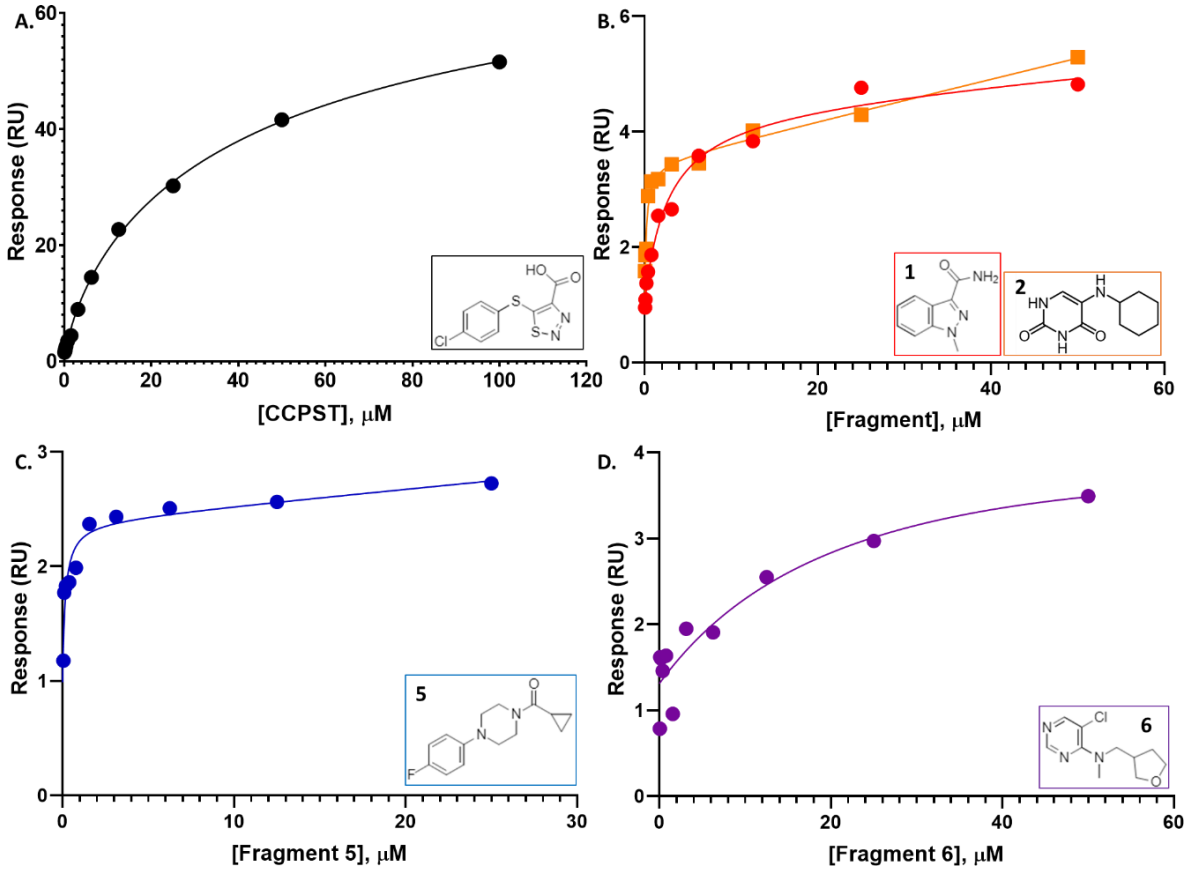


Figure 4.4: Measurement of CCPST and fragment binding to immobilised hHAO1 in solution by SPR. *A. Plot of relative binding response measured by SPR against CCPST concentration. B. Plot of relative binding response measured by SPR against concentration of fragments **1** (red) and **2** (orange). C. Plot of relative binding response measured by SPR against fragment **5** concentration. D. Plot of relative binding response measured by SPR against fragment **6** concentration. Higher fragment concentrations that result in predominantly non-specific binding are omitted from the curve fitting. The chemical structure of the analyte is shown inset for each binding curve. RU, response unit.*

Next, serial dilution of fragments were prepared at the same concentrations as in the CCPST binding experiment, and SPR titrations were performed to analyse fragment binding to immobilised hHAO1. Significant non-specific binding of fragments was observed at 100 µM for fragments **1**, **2** and **6** and at 50 µM and 100 µM for fragment **5**, and consequently binding response measurements at these concentrations were excluded from curve fitting. The active site fragments **1** and **2** demonstrated weak but specific binding to immobilised hHAO1 up to 50 µM concentrations (figure 4.4B). K_D could not be calculated due to the low response observed. Fragments **5** (figure 4.4C) and **6** (figure 4.4D) demonstrated more non-specific binding than the

active site fragments and concomitant weaker responses.

Shephard and co-workers [170] describe the following equation to calculate the minimum K_D of fragments that can be detected by SPR, based on the maximal response observed with a positive control and the concentration of fragment screened:

$$R_{eq} = (R_{MAX} \text{ x } c)/ (K_D + c)$$

where R_{eq} is the minimal response for fragments to be detected at equilibrium; R_{MAX} is the maximal response observed in the presence of the positive control; c is the concentration of fragment; and K_D is the minimum equilibrium dissociation constant required for a fragment to be detected. Using this equation, the K_D of fragments **1** and **2** can be estimated to be less than 550 µM ($K_D = (R_{MAX} \text{ x } c/ R_{eq}) – c = (45 \text{ x } 50 / 4) – 50 = 512.5$ µM). Application of this equation to fragments **5** and **6** suggest K_D values up to 1 mM and 700 µM respectively.

4.2 Follow-up compounds from active site fragments **1** and **2**

Fragment **1** demonstrated dose-dependent inhibition and specific binding to hHAO1 in the solution assays but has weak potency attributable to its small size (figure 4.3, 4.4B). To discover more potent inhibitors, 19 commercially available compounds containing the indazole-3-carboxamide scaffold were selected by performing a substructure search against the Enamine 'screening compounds' library (compounds **8** – **26**; figure 4.5). As well as sharing an indazole-3-carboxamide core, the selected follow-up compounds contain either small additions to the indazole ring (**8** – **16**), an extension of the carboxamide group (**17**, **20**, **21**, **23** and **26**), or both modifications (**18**, **19**, **22**, **24**, **25**).

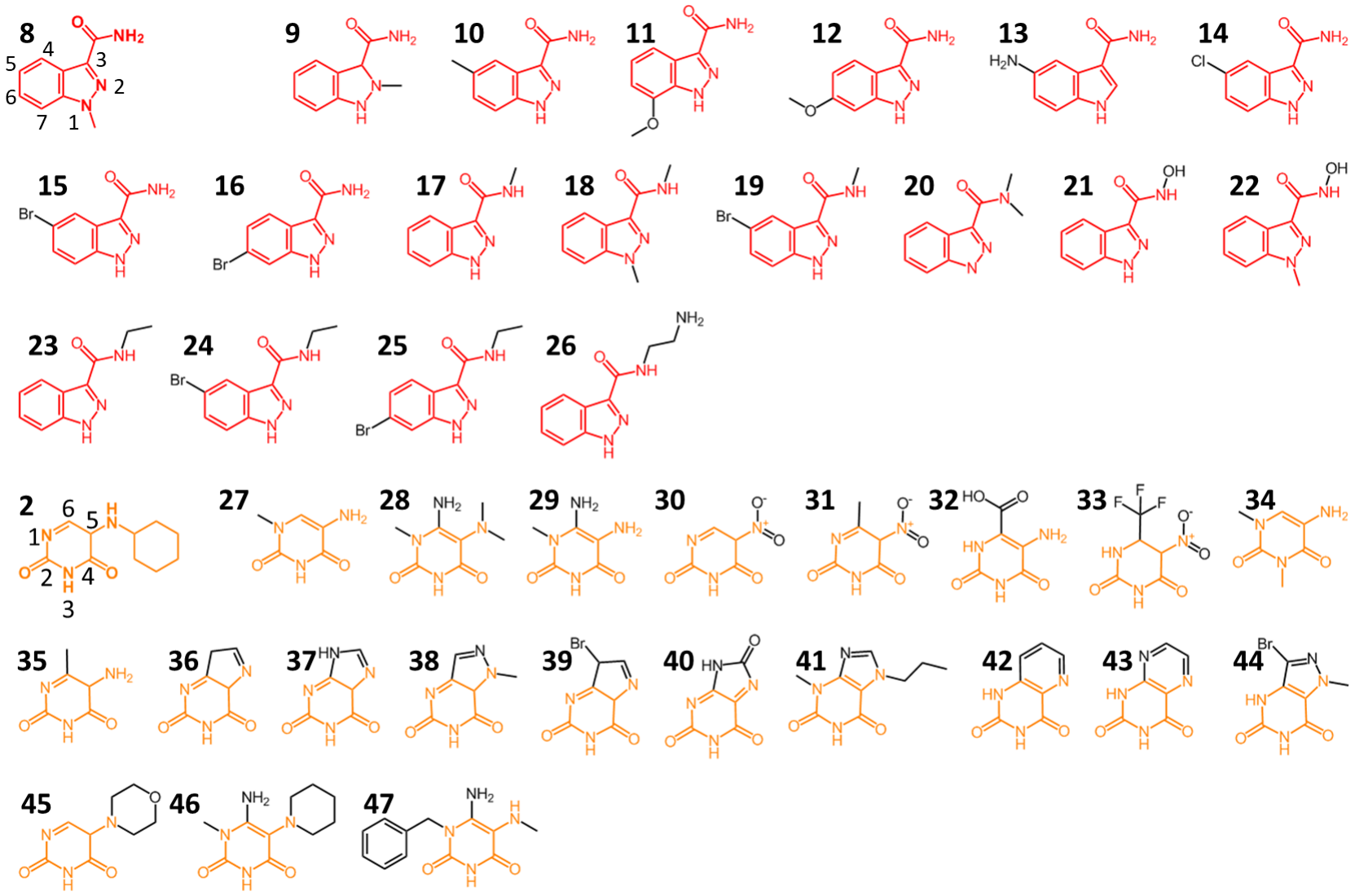


Figure 4.5: Chemical structures of follow-up compounds derived from active site fragments 1 and 2. *The fragment **1** scaffold is coloured red and the fragment **2** scaffold is coloured orange.*

Fragment **2** demonstrated dose-dependent inhibition and specific binding to hHAO1 in the solution assays, but with less potency and more non-specific binding than fragment **1** (figure 4.3, 4.4B). Analysis of the hHAO1-FMN-fragment **2** structure (figure 3.11C) showed that the pyrimidine-2,4-dione moiety of **2** forms specific interactions with hHAO1 active site residues while the 5-aminocyclohexyl moiety displaces nearby residues, which may indicate only the former group is essential for binding to hHAO1. A substructure search of 5-aminopyrimidine-2,4-dione against the Enamine screening compounds library yielded 21 commercially available compounds containing this scaffold (compounds **27** – **47**; figure 4.5). The selected follow-up compounds consist of the pyrimidine-2,4-dione scaffold, embellished with either small acyclic substituents (**27** – **35**), a fused 5-membered heterocycle (**36** – **41, 44**), a fused 6-membered heterocycle (**42** – **43**) or another 6-membered ring substituent (**45** – **47**).

Table 4.1 shows the range and average values for several chemical properties of the active site follow-up compounds. For comparison, the guideline properties of a good lead compound for drug development, known as Lipinski's rule-of-5 (Ro5), and similar guidelines adjusted for selecting good fragment starting points, known as the rule-of-3 (Ro3), are provided. It is clear that the active site follow-up compounds are more fragment-like than lead-like. The key benefit of screening many similar fragments is the potential to identify the most potent minimal scaffold for development into a lead compound.

Table 4.1: Chemical properties for active site follow-up compounds. *Chemical properties were obtained from the PubChem database* [241]. *Guidelines for chemical properties of fragments and lead (or drug-like) compounds were collated from references* [158], [242]–[246]. *The number of non- hydrogen atoms is an average value (*).*

Chemical property	**Fragment hits**		**FUP compounds with scaffold of 1**		**FUP compounds with scaffold of 2**		**Fragment (Ro3)/ Lead (Ro5) guidelines**
	1	**2**	Average	*Range*	Average	*Range*	
Molecular weight (Da)	175	209	203	*166 – 268*	182	*141 – 246*	<300/≤500
Polar surface area (Å^2)	61	70	68	*47– 98*	85	*67 – 122*	≤60/≤140
Molecular complexity score	219	306	228	*207 – 249*	276	*216 – 386*	--
Non-hydrogen atoms	13	15	14.6	*12 – 15*	12.6	*10 – 18*	~15*/38*
Hydrogen bond donors	1	3	1.9	*1 – 3*	2.4	*1 – 4*	≤3/≤5
Hydrogen bond acceptors	2	3	2.4	*2 – 3*	3.5	*2 – 7*	≤3/≤10
Rotatable bonds	1	2	1.3	*0 – 3*	0.5	*0 – 3*	≤3/<12

4.2.1 Single point assay of follow-up compounds from fragment **1**

The 40 active site follow-up compounds were screened for inhibition of 10 nM hHAO1 at 1 mM compound concentration in the Amplex Red activity assay to select the most potent compounds for concentration-response experiments. To assay all compounds in triplicate and in parallel, the Amplex Red assay was miniaturised to a 384-well format with the hHAO1 reaction performed at 10 nM hHAO1 and 36 μM glycolate in 10 μL reaction volume.

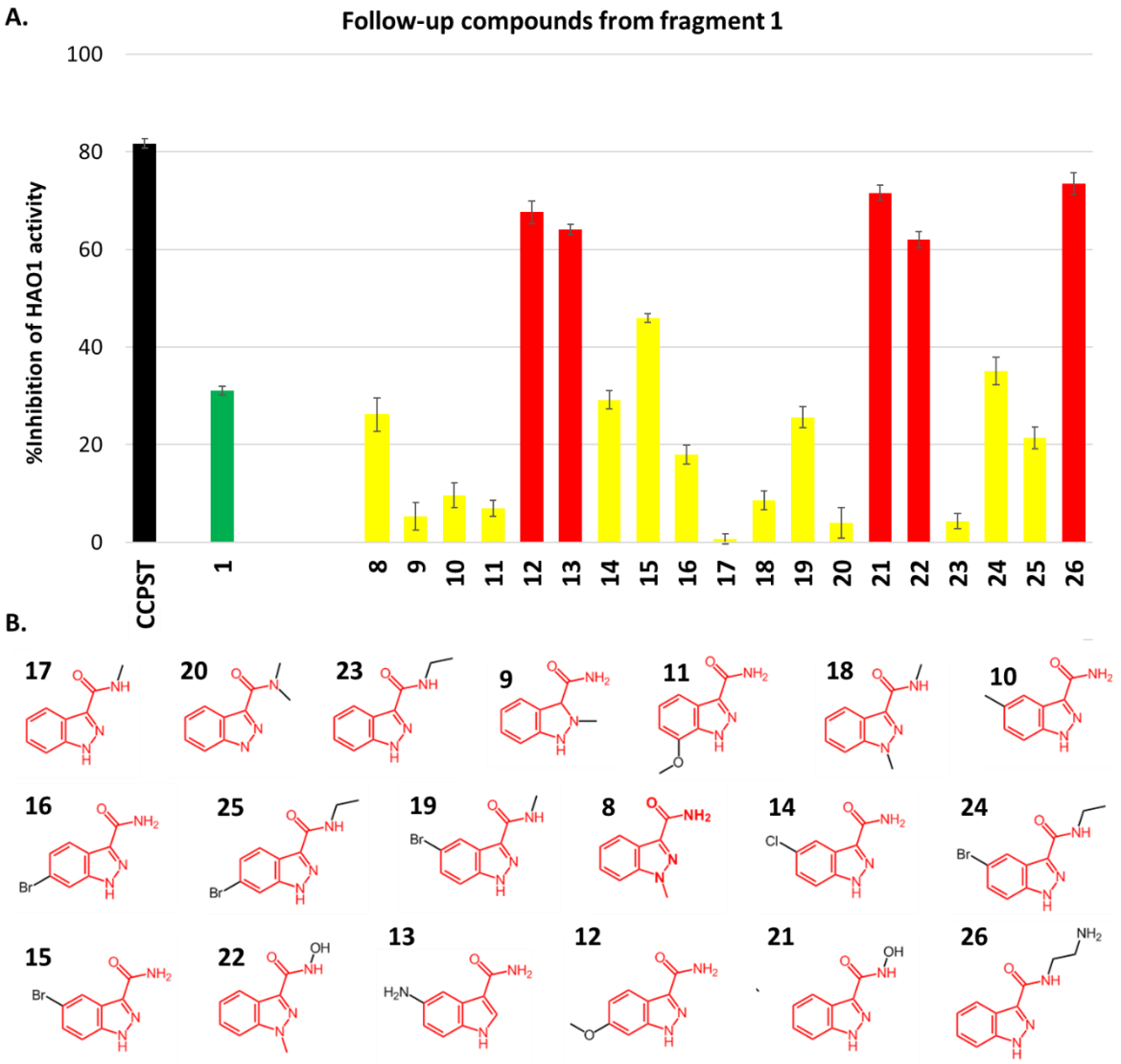


Figure 4.6: Single-point inhibition of hHAO1 by follow-up compounds derived from active site fragment 1, measured using Amplex Red activity assay. *A. Bar chart showing %inhibition of hHAO1 activity by follow-up compounds from fragment* ***1****. Compounds demonstrating >50% inhibition are shown as red bars whereas compounds demonstrating <50% inhibition are shown as yellow bars. CCPST (black bar) and fragment* ***1*** *(green bar) were included as positive controls. All compounds were tested at 1 mM. Error bars show standard deviation of technical triplicates. B. Chemical structures of follow-up compounds* ***8-26****, shown in increasing order of potency. The shared scaffold derived from fragment* ***1*** *is shown in red and the variable groups are in black.*

Most follow-up compounds from fragment **1** (figure 4.5) contained one or more substituents to the indazole ring, and comparison of these related compounds facilitates description of their initial structure-activity relationships. Movement of the methyl group from position 1 (**8**, 26 ± 3.4% inhibition) to position 2 (**9**, 5 ± 2.8% inhibition) or position 5 (**10**, 10 ± 2.5% inhibition) reduces compound potency (figure 4.6A).

Four substituents at position 5 of the indazole ring were tested: a methyl group demonstrating

minimal inhibition (**10**, 10± 2.5% inhibition), halide groups demonstrating fair inhibition (chloro- group in **14**, 29 ± 1.8% inhibition and bromo- group in **15**, 46 ± 0.9% inhibition), and an amine group demonstrating good inhibition (**13**, 64 ± 1.0% inhibition) (figure 4.6A). Two substituents at position 6 were tested: addition of a methoxy group improves inhibition relative to the starting fragment (**12**, 68 ± 2.3% inhibition), whereas addition of a bromide group reduces potency (**16**, 18 ± 2.0% inhibition). The only substituent tested at position 7 was a methoxy group, which demonstrated minimal inhibition (**11**, 7 ± 1.6% inhibition).

Addition of a methyl, dimethyl, hydroxy, ethyl or ethylamine substituent to the amide group of the carboxamide moiety was also investigated. Addition of a methyl or ethyl group reduces inhibitor potency e.g. compound **8** demonstrates higher inhibition than **18**, while **15** demonstrates higher inhibition than **19** or **24**. Conversely, addition of a hydroxyl group at the carboxamide position increased HAO1 inhibition two-fold i.e. from 26 ± 3.4% inhibition by 8 to 62 ± 1.6% inhibition by **22**. Potency was further improved by removal of the 1-methyl group (**21**), moderately increasing inhibition to 72 ± 1.6%. Finally, the most potent compound from this series contained an amino-ethyl group at this position as well as removal of the 1-methyl group (**26**), demonstrating 74 ± 2.2% inhibition.

To summarise, the following substituents of the indazole-3-carboxamide scaffold are identified as most advantageous: no substituent at position 1; a 2-amino-ethyl substituent added to the amine of the carboxamide group; an amino group at position 5; a methoxy group at position 6; and no substituent at position 7. Concentration-response experiments were performed for the five compounds demonstrating between 60 and 75% inhibition of hHAO1 activity at 1 mM (**12**, **13**, **21**, **22**, **26**) and are described in section 4.2.3.

4.2.2 Single point assay of follow-up compounds from fragment **2**

Nearly half of the follow-up compounds from fragment **2** (figure 4.5) consisted of a single 5-aminopyrimidine-2,4-dione ring with small acyclic substituents (**27** – **35**). Appending a methyl group to this scaffold at position 1 (**27**, 78 ± 4.9% inhibition) or at position 6 (**35,** 86 ± 0.8%

inhibition) greatly improved potency but further additions at positions 3, 5 and/or 6 decreased potency i.e. compounds **28**, **29** and **34** demonstrated 45 ± 1.2%, 29 ± 2.2% and 67 ± 0.8% inhibition respectively (figure 4.7A).

Similarly, addition of a nitro group to position 5 (**30**, 55 ± 1.5% inhibition) demonstrates considerably improved potency over the original fragment **2**, but further additions of 6-methyl (**31**, 1 ± 1.4% inhibition) or 6-trifluoro (**33**, 21 ± 0.5% inhibition) groups reduce inhibition. A carboxylate group at position 6 also demonstrated good potency (**32**, 76 ± 0.6% inhibition) but no other compounds substituted only at this position were tested, which precludes further interpretation of this position.

Three follow-up compounds have a ring substituent added to the pyrimidine-2,4-dione scaffold. Compound **45**, adding a 5-morpholino group to this scaffold (59 ± 1.3% inhibition), was the most potent. Compound **46**, with a piperidine ring at this position and other small additions at positions 1 and 6 (6-amino, 1-methyl), demonstrated similar potency (56 ± 0.7% inhibition). The ring substituent of compound **47** is a benzyl group at position 1 of the pyrimidine scaffold, together with two amine containing substituents (6-amino,5-methylamino). This compound was the least potent of the three (37 ± 1.3% inhibition) but still represented an improvement in potency relative to the starting fragment **2**. Similar to the findings of the single ring compounds, an increase in complexity appears to decrease potency.

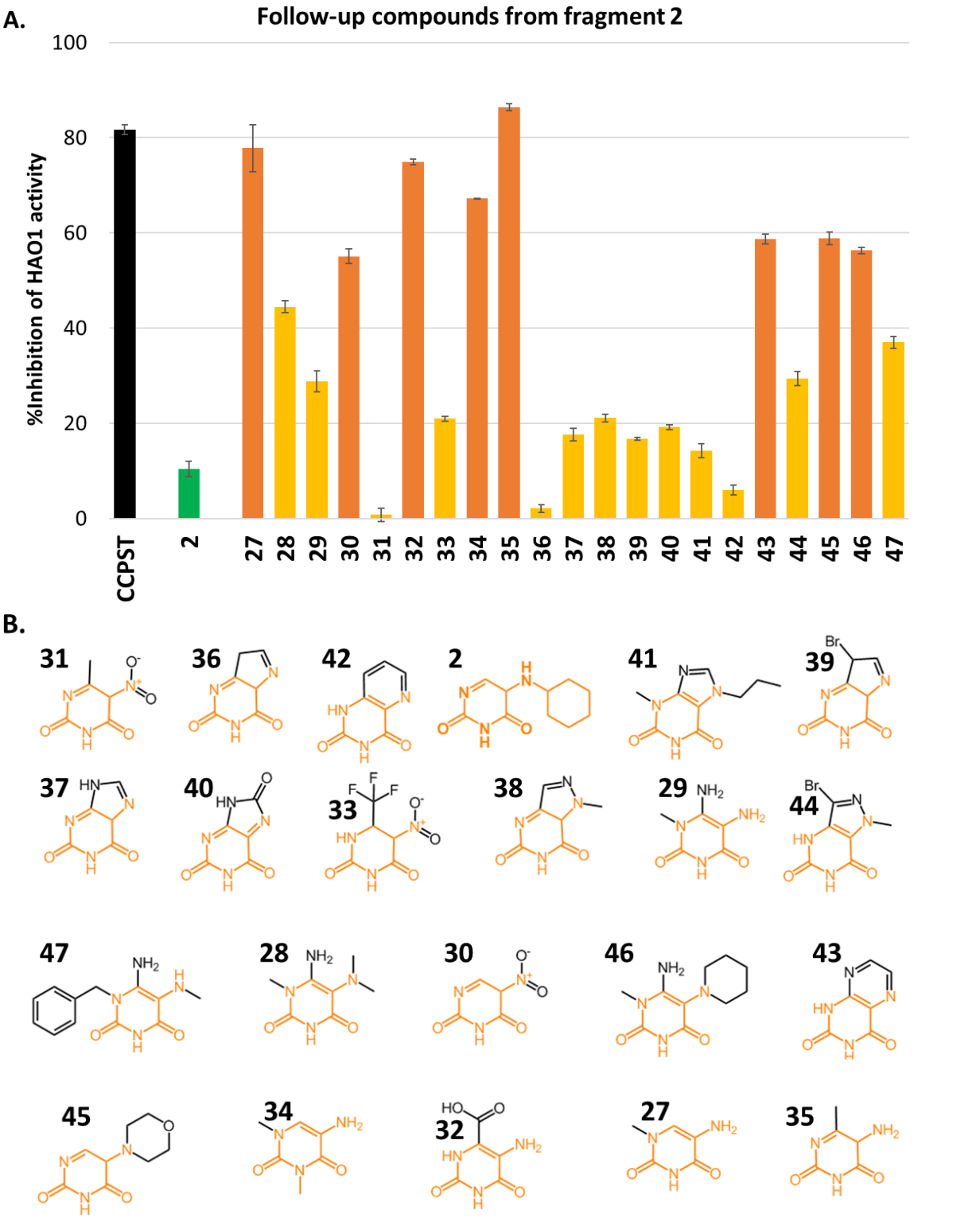


Figure 4.7: Single-point inhibition of hHAO1 by follow-up compounds based on active site fragment 2, measured using Amplex Red activity assay. *A. Bar chart showing %inhibition of hHAO1 activity by follow-up compounds from fragment* ***2****. Compounds demonstrating >50% inhibition are shown as orange bars whereas compounds demonstrating <50% inhibition are shown as yellow bars. CCPST (black bar) and fragment* ***2*** *(green bar) were included as positive controls. All compounds were tested at 1 mM. Error bars show standard deviation of technical triplicates. B. Chemical structures of follow-up compounds* ***27-47*** *in increasing order of potency. The shared scaffold derived from fragment* ***2*** *is shown in orange and the variable groups are in black.*

The nine remaining compounds fuse a 5- or 6-membered heterocycle to the pyrimidine-2,6-dione scaffold. The resultant fused rings are pyridopyrimidine (**42**), pteridine (**43**), pyrrolo-pyrimidine (**36**, **39**), pyrazolo-pyrimidine (**38**, **44**) and purine (**37**, **40**, **41**). Among these nine compounds, only compound **43** showed significant inhibition of hHAO1 (56 ± 0.7%).

To summarise, removal of the 5-aminocyclohexyl group from fragment **2** and addition of a 1-methyl (**27**), 6-carboxylic acid (**32**), 6-methyl (**35**) or 1,3-dimethyl (**34**) significantly improves compound potency. Concentration-response experiments were performed for the eight compounds demonstrating between 55 and 90% inhibition of hHAO1 activity at 1 mM (**27**, **30**, **32**, **34**, **35**, **43**, **45** and **46**) and are described in section 4.2.3.

4.2.3 Concentration-response curves for selected inhibitors

To test whether the active site follow-up compounds that demonstrated >50% inhibition of hHAO1 in the single concentration screen were able to inhibit hHAO1 in a dose-dependent manner, hHAO1 activity in the presence of 12 compound concentrations (0-1 mM) was measured. Figures 4.8A and 4.8C show the poor concentration-response curve fitting for the less potent compounds derived from fragment **1** (**13**, **15**, **22**) and fragment **2** (**30**, **43**, **45**, **46**) respectively.

An increase in molecular weight and lipophilicity can often result in increased inhibition of a lead compound, measured as improving IC_{50} values, without increasing interactions between the compound and target protein. To monitor protein-ligand complementarity independently, several ligand efficiency metrics have been developed to account for compound size, by correcting for the number of non-hydrogen atoms, and compound hydrophobicity, by inclusion of a lipophilicity measure such as the octanol-water partition coefficient (LogP) [159], [247].

One popular metric to compare ligand efficiency is the lipophilic ligand efficiency score developed at Astex Therapeutics (LLE_{AT}), which was used to complement IC_{50} value calculations in this work [189]. LLE_{AT} allows comparison of ligand potency adjusted for

lipophilicity and molecular weight and is calculated using the following formula:

$$LLE_{AT}\,(kcal/mol) = 0.111 + (1.37*(pIC_{50} - LogP)/HAC)$$

where pIC_{50} is equal to the negative log of the molar IC_{50}; LogP is the log of the octanol-water partition coefficient; and HAC is the number of non-hydrogen atoms. The target score using this metric is 0.3 with a guideline minimum value of 0.19 and the principle is that during efficient ligand optimisation the IC_{50} will improve in proportion to the increase in lipophilicity and number of atoms so that the LLE_{AT} score will remain the same or increase during hit-to- lead development [189]. It is important to note, however that a favourable LLE_{AT} score does not necessarily correlate to clinical suitability. For example, the HAO1 inhibitor used as a positive control in this chapter – CCPST – has a LLE_{AT} score of 0.31 kcal/mol but demonstrates very limited cell permeability with 30 times lower intracellular compound concentration observed when CCPST is added to media containing mouse primary hepatocytes and the dose required to achieve 50% reduction of oxalate levels in a mouse model of primary hyperoxaluria is too high to be safe in a clinical setting [214].

The concentration-response curves for the more potent follow-up compounds derived from fragment **1** (compounds **12**, **21** and **26**) are shown in figure 4.8B. Sigmoidal curves were fitted for these three compounds with high confidence ($R^2 > 0.98$). The inhibition curves are flat below 30 μM and do not reach a plateau at the maximum tested concentration (1 mM) but sufficiently high inhibition (>80%) was achieved to allow IC_{50} calculation. For comparison, fragment **1** had an IC_{50} of 420 μM and a LLE_{AT} score of 0.39 kcal/mol. The calculated IC_{50}s and LLE_{AT} scores for these compounds were 206 μM (pIC_{50} 1.9 ± 0.07; LLE_{AT} 0.40 kcal/mol) for compound **12**, 153 μM (pIC_{50} 2.2 ± 0.06; LLE_{AT} 0.44 kcal/mol) for compound **21** and 81 μM (pIC_{50} 2.3 ± 0.06; LLE_{AT} 0.47 kcal/mol) for compound **26**.

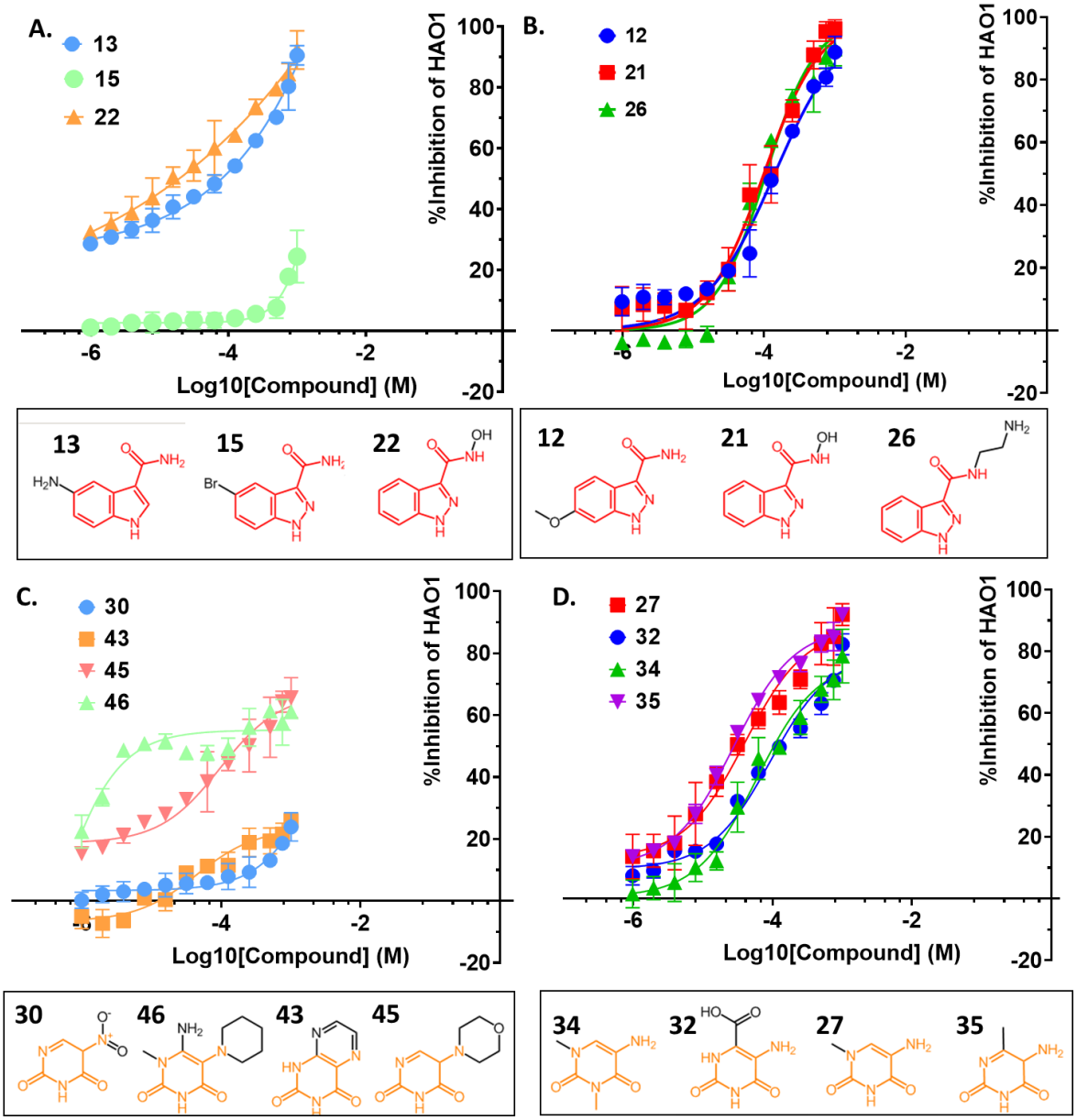


Figure 4.8: Inhibition of hHAO1 by follow-up compounds based on active site fragments, measured using Amplex Red activity assay. *A. Concentration-response curves (%inhibition vs log_{10} [compound]) for compounds **13** (light blue), **15** (light green) and **22** (orange), showing insufficient curve fitting for IC_{50} calculation. B. Concentration-response curves for compounds **12** (blue), **21** (red) and **26** (green), showing dose-dependent inhibition of hHAO1. Inset: Chemical structures of follow-up compounds. The shared scaffold derived from fragment **1** is shown in red and the variable groups in black. C. Concentration-response curves for compounds **30** (light blue), **43** (orange), **45** (pink) and **46** (light green), showing insufficient curve fitting for IC_{50} calculation. D. Concentration-response curves for compounds **27** (red), **32** (blue), **34** (green) and **35** (purple), showing dose-dependent inhibition of hHAO1. Inset: Chemical structures of follow-up compounds. The shared scaffold derived from fragment **2** is shown in orange and the variable groups in black.*

The concentration-response curves for the more potent follow-up compounds derived from fragment **2** (compounds **27**, **32**, **34** and **35**) are shown in figure 4.8D. Sigmoidal curves were

fitted for these four compounds with high confidence ($R^2 > 0.95$). Dose-dependent inhibition is observed for all four compounds from 4 μM compound concentration. Although no final plateau was reached at the maximal compound concentration, sufficiently high inhibition (>80%) was achieved to allow IC_{50} calculation. The calculated IC_{50}s and LLE_{AT} scores for these compounds were 40 μM (pIC_{50} 1.6 ± 0.09; LLE_{AT} 0.88 kcal/mol) for compound **27**, 93 μM (pIC_{50} 2.0 ± 0.1; LLE_{AT} 0.71 kcal/mol) for compound **32**, 60 μM (pIC_{50} 1.8 ± 0.1; LLE_{AT} 0.77 kcal/mol) for compound **34**, and 27 μM (pIC_{50} 1.4 ± 0.1; LLE_{AT} 0.89 kcal/mol) for compound **35**.

Concentration-response experiments confirmed the results of the single concentration screen (section 4.2.1) that 6-methoxy (**12**), *N*-hydroxy (**21**) or *N*-ethylamine (**26**) addition to, and removal of the 1-methyl group from, the fragment **1** scaffold significantly improved hHAO1 inhibition, and the tested compounds demonstrate good pharmacokinetic properties and high ligand efficiency. Substitution of the 6-aminopyrimidine-2,4-dione scaffold of fragment **2** with 1-methyl (**27**), 1,3-methyl (**34**), 6-carboxylic acid (**32**) or 6-amino (**35**) group also yielded promising starting points for hit-to-lead development, with very high LLE_{AT} scores.

4.2.4 Crystal structures of hHAO1 in complex with follow-up compounds

To rationalise the considerable changes in inhibitor potency observed with subtle changes to the fragment scaffolds, effort was made to obtain crystal structures of hHAO1 bound to active site follow-up compounds. However, crystal soaking and co-crystallisation experiments resulted in only two co-crystal structures with unambiguous difference density for the added follow-up compound. Specifically, structures of hHAO1 bound to compounds **19** and **47** were solved through soaking 20 mM compound into preformed hHAO1 crystals grown under the conditions described for fragment screening (chapter 3, section 3.3.1).

Both structures had the same space group and unit cell dimensions as the fragment-bound datasets with 1 copy of hHAO1 in the asymmetric unit. The structure of hHAO1 bound to compound **19** (at 1.5 Å resolution) converged to Rfree of 0.196 and R of 0.158. The

structure of hHAO1 bound to compound **47** (at 1.4 Å resolution) converged to Rfree of 0.190 and R of 0.149. Based on difference Fourier density, one fully occupied molecule of FMN and one partially occupied inhibitor molecule (**19** or **47**) were built into the hHAO1 protomer. Data processing and structure refinement statistics are provided in table S2.1.

The structures of hHAO1-FMN-**19** and hHAO1-FMN-**47** demonstrate similar overall conformations to hHAO1-FMN-glycolate, hHAO1-FMN-fragment **1** and hHAO1-FMN-fragment **2** structures (RMSD < 0.5 Å between hHAO1-FMN-**19** and hHAO1-FMN-**1** or hHAO1-FMN- glycolate and between hHAO1-FMN-**47** and hHAO1-FMN-**2** or hHAO1-FMN-glycolate) (figure 4.9A), except for the gating loop region (figure 4.9B). Significant disorder of the gating loop is observed upon binding of compound **47,** similar to that observed in the presence of fragments, while minor disorder of the gating loop is observed when compound **19** is bound. Unlike the fragment bound structures, however, there is no displacement or disorder observed for the helices αD and αE of the gating loop.

The binding mode of compound **19** is similar to that observed with the published pyrazole or triazole containing inhibitors (e.g. CDST; black lines in figure 4.9C). The carbonyl of compound **19** is positioned to hydrogen bond with the guanidino groups of either Arg167 or Arg263, while the pyrazole nitrogen atoms are hydrogen-bonded with the hydroxyl group of the Tyr132 sidechain and a sidechain nitrogen of His260. Additionally, the sidechain indole of Trp110 is displaced 180° out from the active site in the hHAO1-FMN-**19** structure with no additional movement of the associated loop.

The binding mode of compound **47** is very similar to that of fragment **2**, and also superimposes well with the published di-carbonyl inhibitor, TKP (black lines; figure 4.9D). Specifically, the carbonyl at position 2 of the pyrimidine ring hydrogen bonds with the nitrogen atom of the Trp110 sidechain indole; the amine at position 3 hydrogen bonds with the hydroxyl of the Tyr132 sidechain; the carbonyl at position 4 hydrogen bonds with the sidechain nitrogen of His260;

and the 6-benzyl of **47** binds in a similar position to the 1-cyclohexyl group of fragment **2**, displaces the sidechain of Trp110 by 180° from the active site.

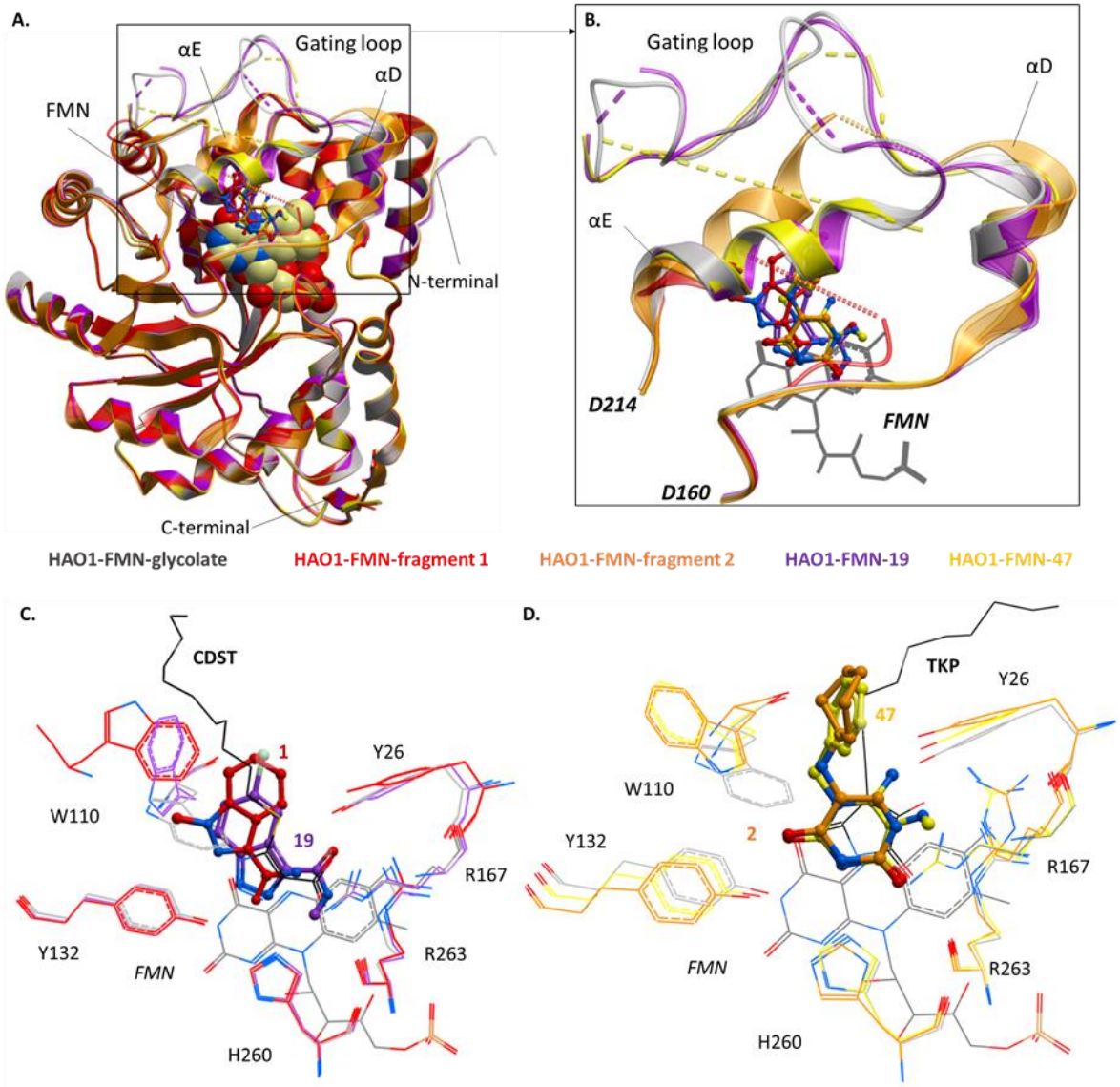


Figure 4.9: Crystal structures of hHAO1 bound with follow-up compounds 19 and 47, derived from active site fragments 1 and 2 respectively. *A. Ribbon diagram of superimposed hHAO1-FMN-glycolate (grey), hHAO1-FMN-fragment **1** (red), hHAO1-FMN-fragment **2** (orange), hHAO1-FMN-**19** (purple) and hHAO1-FMN-**47** (yellow) structures. FMN is shown as spacefill and compounds are shown as sticks. B. Close-up view of hHAO1 active site from the highlighted area in panel A. Residues Asp160-Asp214 from each structure are shown as ribbon, FMN is shown as grey lines and compounds are shown as sticks. C. Close-up view of hHAO1 active site in the superimposed hHAO1-FMN-glycolate (grey), hHAO1-FMN-fragment **1** (red), and hHAO1- FMN-**19** (purple) structures, showing the positions of key active site residues. Compounds are shown as sticks. hHAO1 residues, CDST and FMN are shown as lines. D. Close-up view of hHAO1 active site in the superimposed hHAO1-FMN-glycolate (grey), hHAO1-FMN-fragment **2** (orange), and hHAO1-FMN-**47** (yellow) structures. Compounds are shown as sticks. hHAO1 residues, TKP and FMN are shown as lines.*

Additionally, the 1-amino and 2-methylamino groups of **47** are positioned to hydrogen bond with the hydroxyl of the Tyr26 sidechain. Finally, two conformations were observed for the sidechain

of Arg167: that seen in the hHAO1-FMN-glycolate structure where it points in towards the glycolate binding site, and that observed in the fragment **2** bound structure where the sidechain is displaced 45° outward from the site.

4.3 Follow-up compounds from fragment **5** at the gating loop

Fragment **5** demonstrated concentration-dependent (though weak) inhibition of hHAO1 and specific binding to hHAO1 in the solution assays (figure 4.3, 4.4C). To improve potency, 75 commercially available follow-up compounds containing the 1-(4-phenylpiperzin-1-yl)ethanone scaffold were selected by performing a substructure search against the MolPort commercial library (compounds **48** – **122**) (figure 4.10). These compounds primarily replace the cyclopropyl group of fragment **5** with a larger group containing one or two rings and make small acyclic substitutions to the phenyl ring of the fragment scaffold.

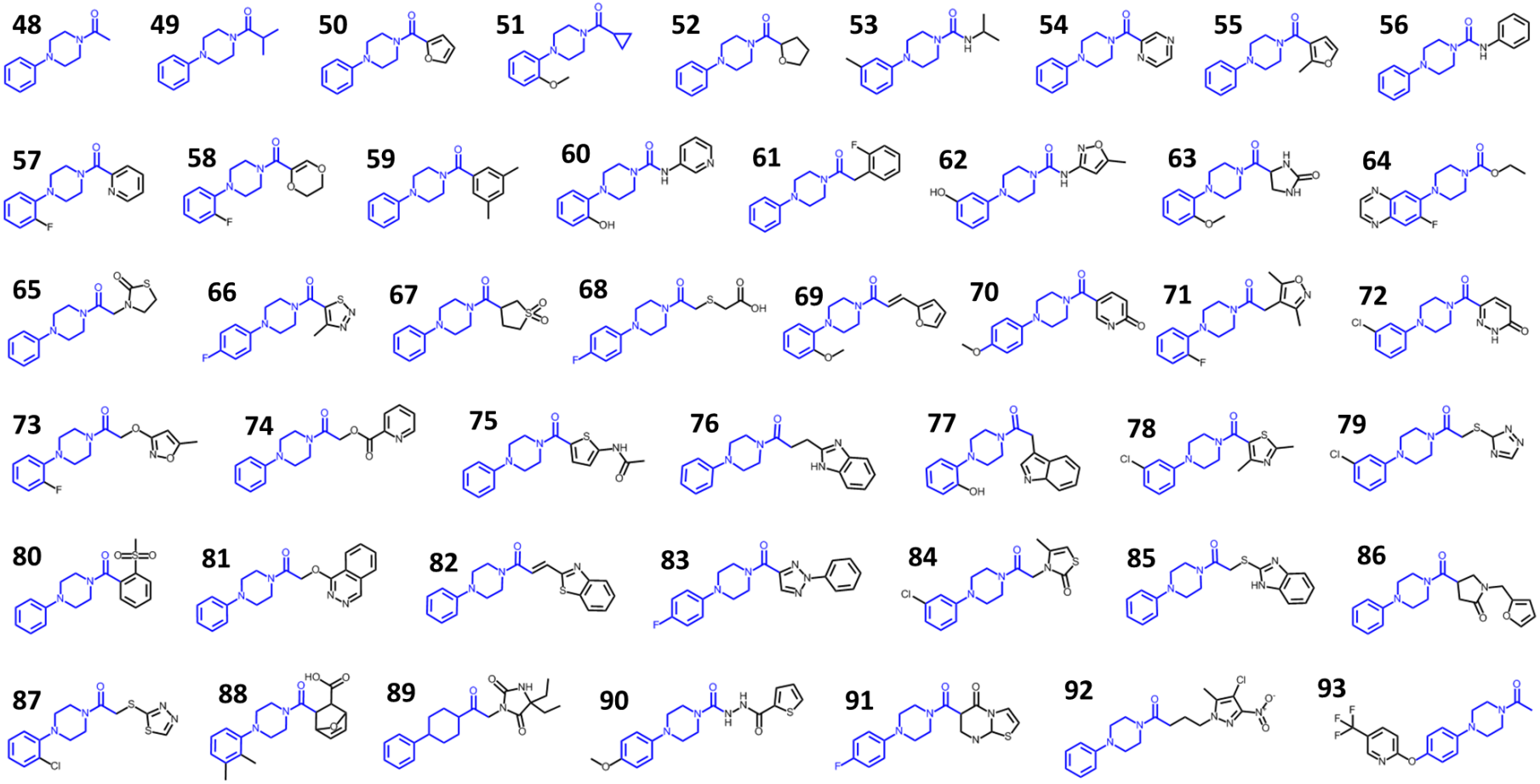


Figure 4.10: Chemical structures of follow-up compounds based on fragment 5 at the gating loop. *Compounds are arranged from the lowest molecular weight (**48**) to the highest molecular weight (**122**). The 1-(4-phenylpiperzin-1-yl)ethanone scaffold from fragment **5** is coloured in blue and the variable substituents in black. Structures are continued on the next page.*

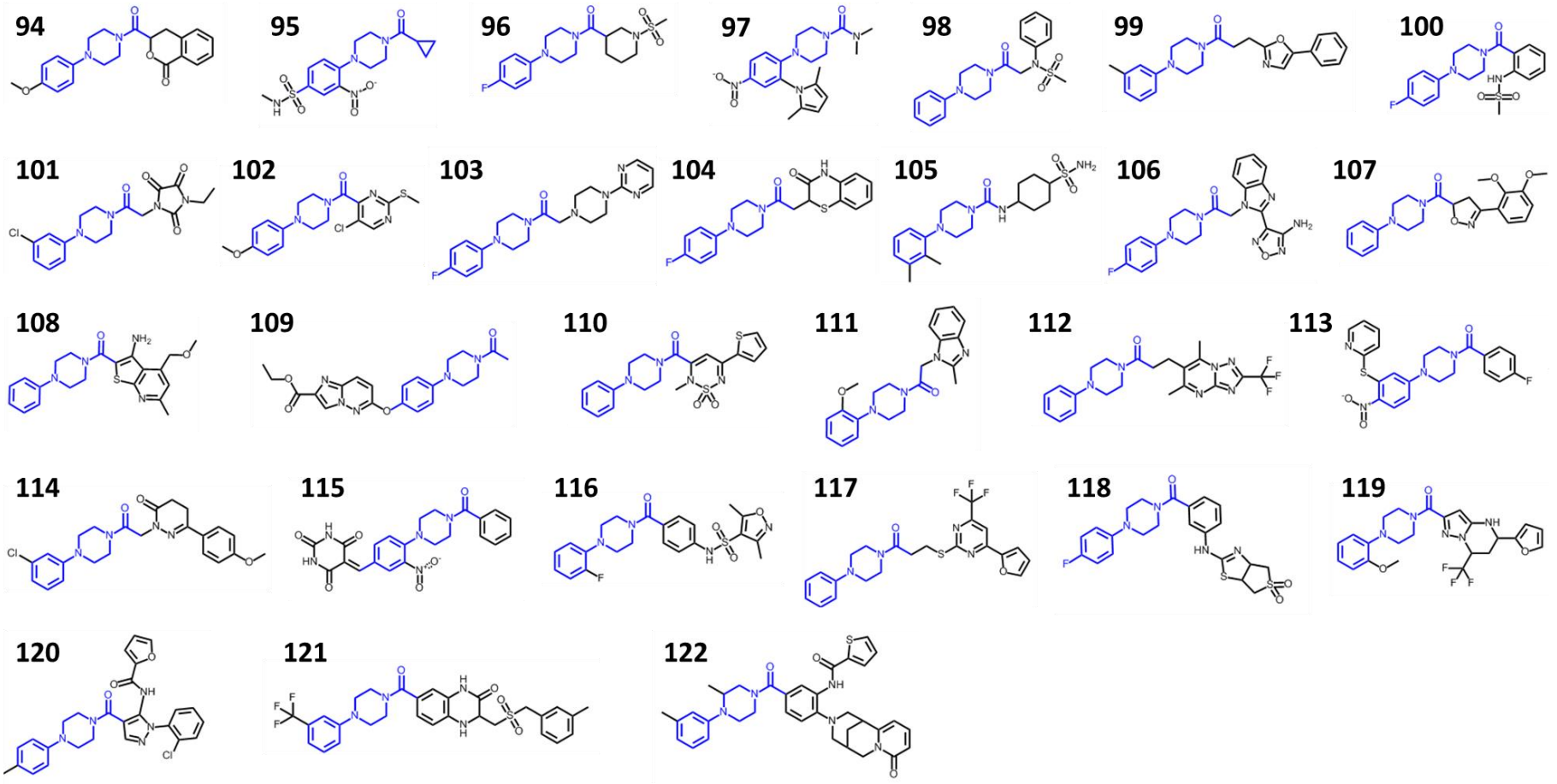


Figure 4.10 *continued.*

Table 4.2 shows the range and average values for chemical properties of follow-up compounds derived from fragment **5**. In contrast to the active site follow-up compounds (**8-35**), follow-up compounds derived from fragment **5** have properties that are more lead-like than fragment-like.

Table 4.2: Chemical properties for follow-up compounds derived from fragment 5 at the gating loop. *Chemical properties were obtained from the PubChem database* [241]. *Guidelines for chemical properties of fragments and lead (or drug-like) compounds were collated from references* [158], [242]–[246]. *The number of non-hydrogen atoms is an average value (*).*

Chemical property	**Fragment hit 5**	**FUP compounds with scaffold of 5**		**Fragment (Ro3)/ Lead (Ro5) guidelines**
		Average	*Range*	
Molecular weight (Daltons)	248	356	*204 – 608*	<300/≤500
Polar surface area (Å^2)	24	70	*24 – 145*	≤60/≤140
Molecular complexity	303	507	*216 – 1180*	--
Non-hydrogen atoms	18	25.3	*15 – 44*	~15*/38*
Hydrogen bond donors	0	0.5	*0 – 2*	≤3/≤5
Hydrogen bond acceptors	3	5	*2 – 9*	≤3/≤10
Rotatable bonds	2	3.4	*1 – 6*	≤3/<12

4.3.1 Single point assay of follow-up compounds from fragment **5**

The 75 loop site follow-up compounds were tested for inhibition of 10 nM hHAO1 in a single point assay (1 mM compound concentration), as described for the active site compounds (section 4.2.1 – 4.2.2).

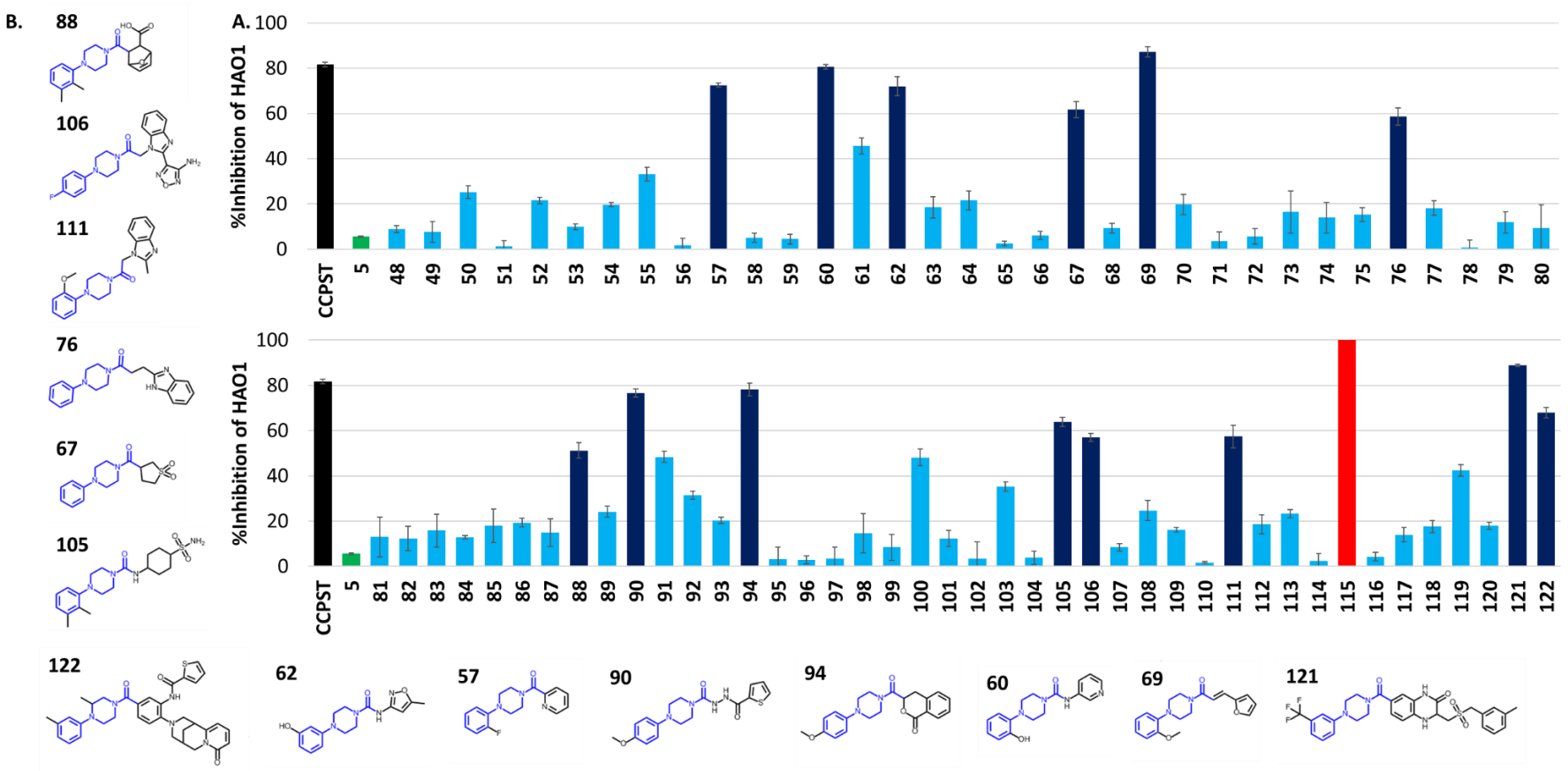


Figure 4.11: Single-point inhibition of hHAO1 by follow-up compounds derived from fragment 5, measured using Amplex Red activity assay. *A. Bar chart showing % inhibition of hHAO1 activity by follow-up compounds. Compounds demonstrating >50% inhibition are shown as dark blue bars, whereas compounds demonstrating <50% inhibition are shown as light blue bars. CCPST (black bar) and fragment* ***5*** *(green bar) were included as positive controls. All compounds were tested at 1 mM. Error bars show standard deviation of technical triplicates. Compound* ***115*** *(red bar) demonstrated interference with the assay reagent, suppressing the fluorescence signal independent of hydrogen peroxide generation. B. Chemical structures of follow-up compounds demonstrating >50% inhibition ordered from the least potent* ***(88)*** *to the most potent* ***(121)****. The shared scaffold derived from fragment* ***5*** *is shown in blue and the variable groups are in black.*

The single point assay revealed >50% inhibition of hHAO1 by 14 compounds (dark blue bars in figure 4.11A). One compound (**115**, red bar in figure 4.11A) demonstrated interference with the Amplex Red reagent, supressing the fluorescence signal independent of the hydrogen peroxide concentration, and so was excluded from further analysis. Two compounds were more potent than the positive control, CCPST, demonstrating 89 ± 0.5% inhibition (**121**) and 87 ± 2.3% inhibition (**69**) at 1 mM concentration. Of particular interest is the significant increase in potency upon replacement of the fragment cyclopropyl group with a single 5-membered (**62**, **67**, **69**) or 6-membered heterocycle (**57**, **60**).

4.3.2 Concentration-response curves for selected inhibitors

To test whether the follow-up compounds demonstrating >50% inhibition in the single concentration screen were able to inhibit hHAO1 in a dose-dependent manner, hHAO1 activity in the presence of 12 compound concentrations (0-1 mM) was measured for the 14 compounds identified from the single concentration screen. Concentration-response curves for less potent follow-up compounds that demonstrated minimal inhibition below 100 μM followed by an unusually steep curve without a plateau, often indicative of non-specific inhibition (**67**, **76**, **88**, **106** and **111**), are shown in figure 4.12A.

The plotted concentration-response curves for the remaining follow-up compounds are shown in figures 4.12B and 4.12C. Sigmoidal curves, consistent with hHAO1 inhibition proportional to compound concentration, were fitted for all nine compounds with high confidence ($R^2 > 0.95$). The inhibition curves do not reach a plateau at the maximum tested concentration (1 mM) but sufficiently high inhibition (>80%) was achieved to allow IC_{50} calculation. The calculated IC_{50}s and LLE_{AT} scores for the lower molecular weight compounds were 238 μM (pIC_{50} 3.6 ± 0.2; LLE_{AT} 0.22 kcal/mol) for compound **57**, 44 μM (pIC_{50} 4.4 ± 0.09; LLE_{AT} 0.31 kcal/mol) for compound **60**, 158 μM (pIC_{50} 3.8 ± 0.1; LLE_{AT} 0.25 kcal/mol) for compound **62**, 131 μM (pIC_{50} 3.9 ± 0.1; LLE_{AT} 0.18 kcal/mol) for compound **69**, and 144 μM (pIC_{50} 3.8 ± 0.08; LLE_{AT} 0.20 kcal/mol) for compound **90** (figure 4.12B)**.** These smaller compounds have favourable LLE_{AT}

scores except for compound **69** which falls below the guideline value due to its high lipophilicity.

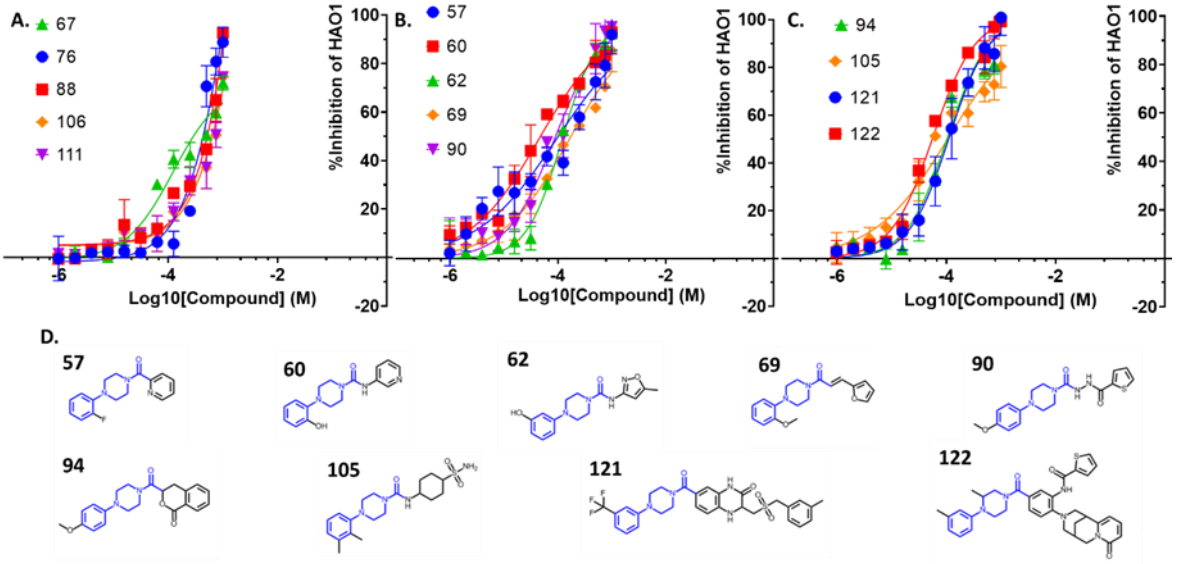


Figure 4.12: Inhibition of hHAO1 by follow-up compounds derived from fragment 5, measured using Amplex Red activity assay. *A. Concentration-response curves (% inhibition vs log_{10} [compound]) for compounds **67** (green), **76** (blue), **88** (red), **106** (orange) and **111** (purple), showing insufficient curve fitting for IC_{50} calculation. B. Concentration-response curves for compounds **57** (blue), **60** (red), **62** (green), **69** (orange) and **90** (purple), showing dose-dependent inhibition of hHAO1. C. Concentration-response curves for compounds **94** (green), **105** (orange), **121** (blue) and **122** (red), showing dose-dependent inhibition of hHAO1. D. Chemical structures of follow-up compounds for which concentration-response curves could be reliably fitted. The shared scaffold derived from fragment **5** is shown in blue and the variable groups in black.*

The calculated IC_{50}s and LLE_{AT} scores for the higher molecular weight compounds were 109 µM (pIC_{50} 4.0 ± 0.1; LLE_{AT} 0.17 kcal/mol) for compound **94**, 47 µM (pIC_{50} 4.3 ± 0.09; LLE_{AT} 0.23 kcal/mol) for compound **105**, 146 µM (pIC_{50} 3.8 ± 0.06; LLE_{AT} 0.10 kcal/mol) for compound **121** and 55 µM (pIC_{50} 4.3 ± 0.07; LLE_{AT} 0.08 kcal/mol) for compound **122** (figure 4.12C)**.** As a result of the larger size of these compounds, the observed inhibition does not translate to good ligand efficiency except for compound **105**, which has an improved LLE_{AT} score due to its high potency and low lipophilicity relative to the others.

Five of the nine compounds that demonstrated dose-dependent inhibition of hHAO1 contained only one additional ring moiety attached to the fragment **5** scaffold (**57, 60, 62, 90** and **105**) and had LLE_{AT} scores above the guideline value of 0.19, indicating that these compounds could serve as good starting points for further inhibitor development.

4.3.3 Crystal structure of hHAO1 after soaking follow-up compounds

As discussed for the active site follow-up compounds in section 4.2.4, every effort was made to obtain crystal structures of hHAO1 bound to follow-up compounds derived from loop site fragment **5**. Nevertheless, extensive crystal soaking and co-crystallisation experiments yielded only one crystal structure of hHAO1 bound to compound **79**, which replaces the 4-fluorophenyl of fragment **5** with a 3-chlorophenyl and replaces the cyclopropyl of fragment **5** with a methyl-thio-triazole.

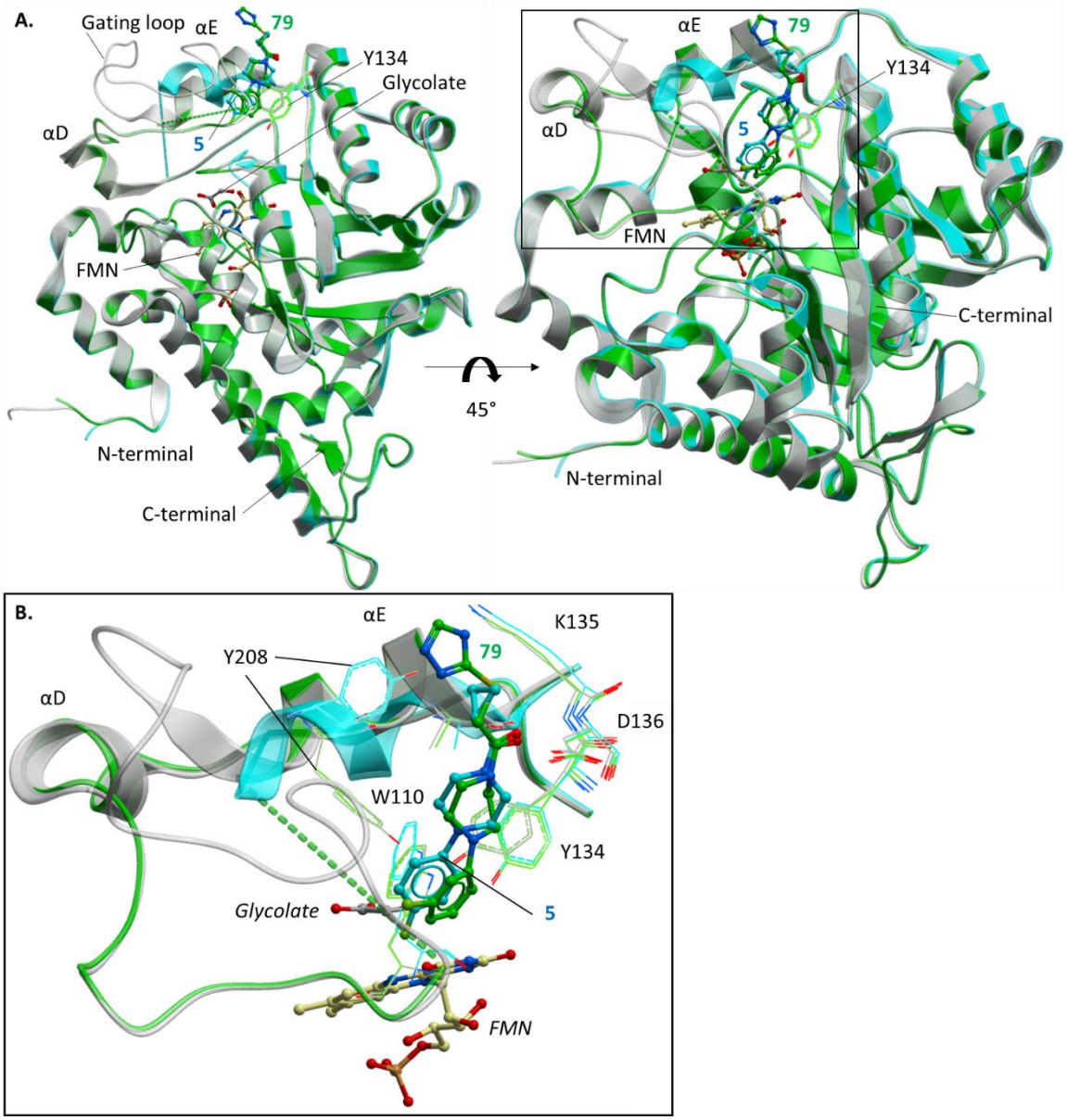


Figure 4.13: Crystal structures of hHAO1 bound with follow-up compound 79, fragment 5 and glycolate. *A. Ribbon diagram of superimposed hHAO1-FMN-glycolate (grey), hHAO1-FMN-fragment **5** (blue), and hHAO1-FMN-**79** (green) structures. FMN is shown as spacefill, compounds are shown as sticks, and key residues are shown as lines. Two orientations are shown. B. Close-up view of the gating loop region (black box in panel A) in hHAO1-FMN-glycolate (grey), hHAO1-FMN-fragment **5** (blue), and hHAO1-FMN-**79** (green) structures. Compounds, FMN and glycolate are shown as sticks.*

This structure had the same space group and unit cell dimensions as the fragment-bound datasets, with 1 copy of hHAO1 in the asymmetric unit. The final refined structure (at 1.6 Å resolution) converged to Rfree of 0.172 and R of 0.149. Based on difference Fourier density, one full occupancy molecule of FMN and one partial occupancy molecule of compound **79** were built into the asymmetric unit. Data processing and structure refinement statistics are provided in table S2.1.

The hHAO1-FMN-**79** structure had a similar overall conformation to the hHAO1-FMN-glycolate and hHAO1-FMN-fragment **5** structures described earlier in this work (RMSD < 0.3 Å over 330 Cα) (figure 4.13A). In contrast to the structure bound with fragment **5**, no displacement of the αD or αE helices is observed upon binding of compound **79** (figure 4.13B). The thio-triazole moiety added to **79** is oriented towards the solvent-exposed face of hHAO1 and does not appear to form additional interactions with hHAO1 residues, supporting the poor inhibition observed with this compound.

4.4 Follow-up compounds from fragment **6** at oligomeric interface

Fragment **6**, bound at the subunit-subunit interface, demonstrates very weak binding to hHAO1 in solution and no inhibition of hHAO1 was detected (figures 4.3, 4.4D). To explore whether compounds targeting the subunit-subunit interface could lead to inhibition of hHAO1, a small but diverse set of follow-up compounds were synthesised based on the fragment **6** scaffold and characterised in solution.

4.4.1 Synthesis of follow-up compounds by poised chemistry

As discussed in section 1.3.2, fragments in the DSi-Poised library, including fragment **6**, are 'poised' for easy deconvolution into their component synthons and rapid generation of analogues. The component synthons of fragment **6** are a chloropyrimidine and a tetrahydrofuran connected by an amide linker. In the hHAO1-FMN-fragment **6** structure, the chloropyrimidine ring is stabilised by π-π stacking with the Tyr163 sidechain of one subunit and hydrogen bonding with the Thr60 sidechain of the other subunit, while the tetrahydrofuran moiety forms few

interactions despite being more buried at the subunit interface (section 3.3.5; figure 3.15D).

A substructure search for compounds in the Enamine screening library yielded few commercially available compounds containing the chloropyrimidine moiety of interest. Therefore, chemical synthesis was performed to generate analogues from the chloropyrimidine synthon of fragment **6** (figure 4.14A).

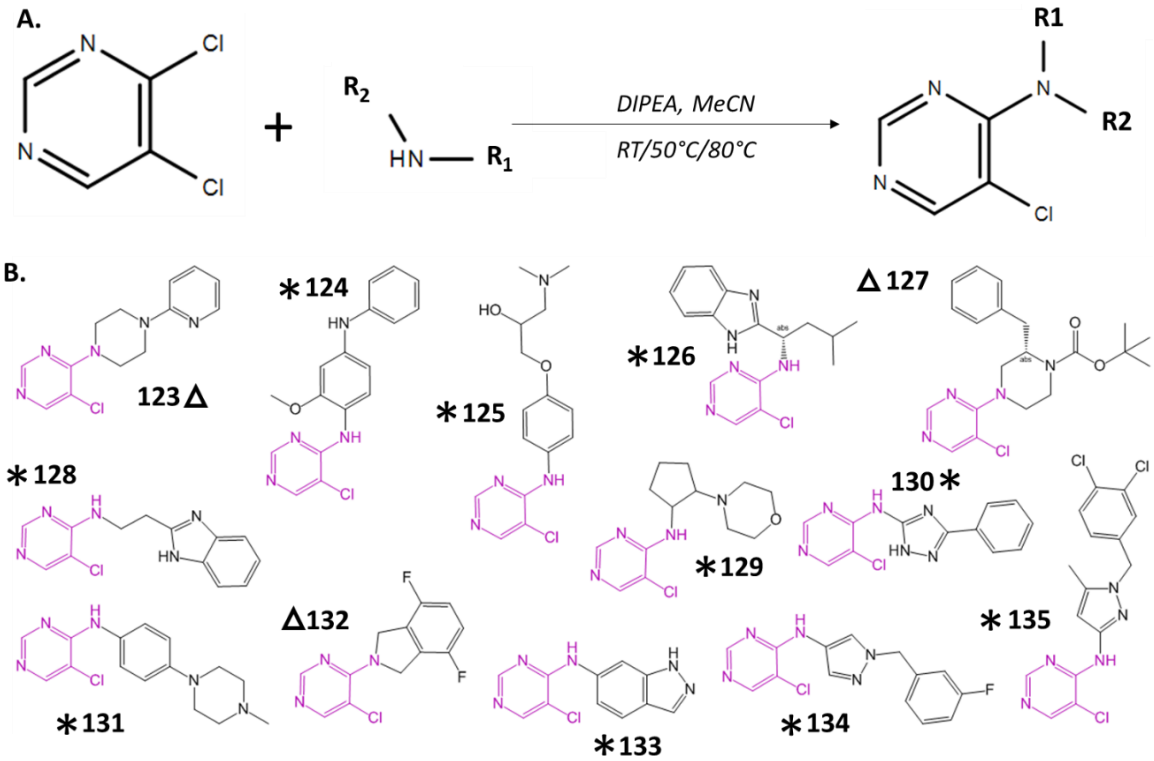


Figure 4.14: Synthesis of follow-up compounds based on fragment 6. *A. Reaction scheme for the nucleophilic aromatic substitutions performed to couple the dichloropyrimidine reagent with amine building blocks. Building blocks were either primary amines, where R_1 is a hydrogen atom and R_2 is a variable group containing one or more aromatic rings, or secondary amines, where R_1 and R_2 are part of a nitrogen-containing heterocycle of the variable group. The reaction was performed overnight in acetonitrile solvent, catalysed by the base DIPEA. Reactions were shaken at 200 rpm and performed at room temperature, 50 °C or 80 °C. B. Chemical structures of the 13 follow-up synthetic compounds based on the fragment **6** scaffold. The conserved scaffold from fragment **6** is in purple and the variable groups are in black. Primary amines are indicated by * and secondary amines are indicated by Δ.*

This work applied nucleophilic aromatic substitution (S_NAr), a robust, well-characterised medicinal chemistry reaction, to link the chloropyrimidine ring to amide building blocks larger than the tetrahydrofuran present in the fragment (figure 4.14A). A diverse set of building blocks were selected from available stocks to have two or more ring structures and a molecular weight >100 Daltons. This reaction was used to generate 13 compounds that maintained the chloropyrimidine moiety but added an additional synthon containing at least one aromatic ring,

selected to generate different compound shapes that help map the binding pocket at the oligomeric interface (figure 4.14B).

Briefly, the reagent (dichloropyrimidine) and an amide-containing building block were dissolved in acetonitrile, with *N*, *N*-diisopropylethylamine (DIPEA) acting as the base catalyst, and the reaction was allowed to proceed at room temperature overnight. Total substrate conversion was confirmed by HPLC-MS before purification of the product on the same system. Reactions showing incomplete substrate conversion were further incubated at either 50 or 80 °C until full conversion was observed. The identity of the 13 synthesised compounds was confirmed by 1H-NMR and stock solutions of 200 mM were prepared in DMSO.

Table 4.3: Chemical properties for follow-up compounds derived from fragment 6. *Chemical properties were calculated using the Molinsipration software* [248]. *Guidelines for chemical properties of fragments and lead (or drug-like) compounds were collated from references* [158], [242]–[246]. *The number of non-hydrogen atoms is an average value (indicated by *).*

Chemical property	**Fragment 6**	**FUP compounds with scaffold of 6**		**Fragment (Ro3)/ Lead (Ro5) guidelines**
		Average	*Range*	
Molecular weight (Da)	228	301	*246 – 389*	<300/≤500
Polar surface area (Å^2)	38	57	*29 – 79*	≤60/≤140
Non-hydrogen atoms	15	21	*17 – 27*	~15*/38*
Hydrogen bond donors	0	1.2	*0 – 2*	≤3/≤5
Hydrogen bond acceptors	4	5.1	*3 – 6*	≤3/≤10
Rotatable bonds	3	3.7	*1 – 7*	≤3/<12

Table 4.3 shows the range and average values for the chemical properties of these follow-up compounds alongside those of the original fragment (fragment **6**). The follow-up compounds

mostly adhere to the upper limits of guideline properties for fragments, with some overlap into lead-like compounds, particularly in the number of hydrogen bond acceptors.

4.4.2 Single point assay of follow-up compounds from fragment **6**

The synthesised follow-up compounds were assayed for inhibition of 10 nM hHAO1 at 1 mM compound concentration, as described for the previous follow-up compounds (sections 4.2.1, 4.2.2 and 4.3.1), to select the most potent compounds for concentration-response experiments.

The single concentration assay revealed >50% inhibition of hHAO1 by 6 compounds (dark purple bars in figure 4.15A). One compound demonstrated similar inhibition to the positive control, CCPST (black bar in figure 4.15A), with 85 ± 1.9% inhibition at 1 mM (**124**), while another compound demonstrated higher inhibition than CCPST, with 93 ± 5.7% inhibition (**132**). From this small dataset, no clear structure-activity relationships can be established for the synthesised compounds, though a few general observations can be made.

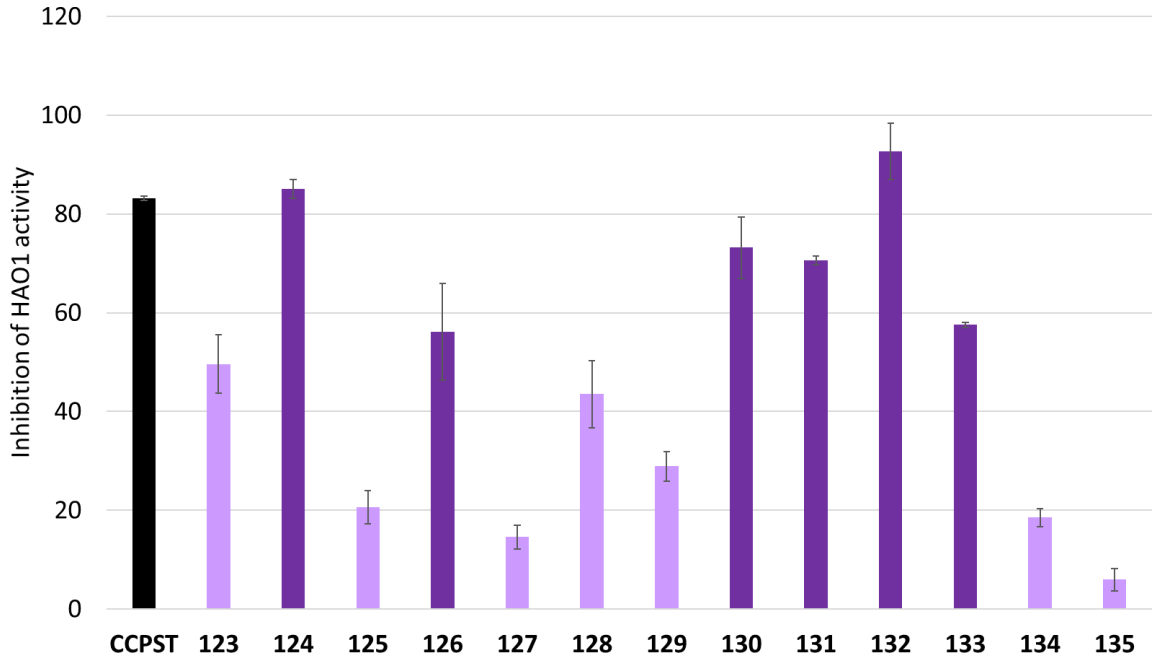


Figure 4.15: Single-point inhibition of hHAO1 by follow-up compounds based on fragment 6, measured using Amplex Red activity assay. *A. Bar chart showing the % inhibition of hHAO1 activity by follow-up compounds. Compounds demonstrating >50% inhibition are shown as dark purple bars whereas compounds demonstrating <50% inhibition are shown as light purple bars. CCPST (black bar) was included as a positive control. All compounds were tested at 1 mM. Error bars show standard deviation of technical triplicates.*

Specifically, compounds containing aromatic substituents at both ends of the structure showed lower potency when separated by a longer linking group (e.g. **134** and **135** were less potent than

132, and **128** was less potent than **133**), indicating a preference for compact, rigid ligands. Similarly, the addition of a long acyclic constituent (**125**) or an R-benzyl group (**127**) were not favoured. A more extensive screen, focusing on optimising the most potent compound, would be required to draw more precise conclusions in the future.

4.4.3 Concentration-response curves for selected inhibitors

To test whether the six follow-up compounds based on fragment **6** that demonstrated >50% inhibition in the single point assay were able to inhibit hHAO1 in a dose-dependent manner, hHAO1 activity in the presence of 12 compound concentrations (0-1 mM) was measured. Concentration-response curves for the less potent follow-up compounds, demonstrating minimal inhibition below 100 µM followed by an unusually steep curve without a plateau, which is often indicative of non-specific inhibition (**126**, **130** and **133**), are shown in figure 4.16A.

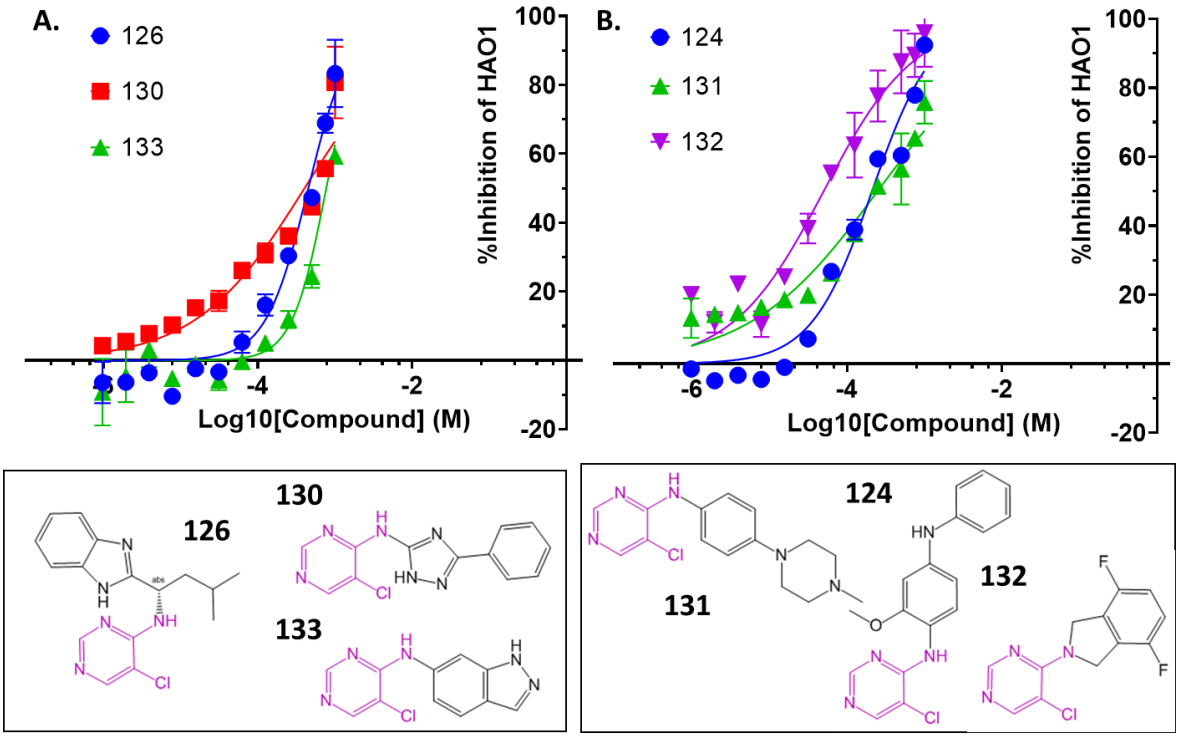


Figure 4.16: Inhibition of hHAO1 by follow-up compounds derived from fragment 6, measured using Amplex Red activity assay. *A. Concentration-response curves (% inhibition vs log_{10} [compound]) for compounds **126** (blue), **130** (red) and **133** (green), showing insufficient curve fitting for IC_{50} calculation. Inset: chemical structures of compounds **126**, **130** and **133**. The shared scaffold derived from fragment **6** is shown in purple and the variable groups in black. B. Concentration-response curves for compounds **124** (blue), **131** (green) and **132** (purple), showing dose-dependent inhibition of hHAO1. Error bars are standard deviation of technical triplicates. Inset: chemical structures of compounds **124**, **131** and **132**. The shared scaffold derived from fragment **6** is shown in purple and the variable groups in black.*

The concentration-response curves for the remaining three follow-up compounds are shown in

figure 4.16B. Sigmoidal curves, consistent with hHAO1 inhibition proportional to compound concentration, were fitted for all three compounds with high confidence (R^2 > 0.97). The inhibition curves do not reach a plateau at the maximum tested concentration (1 mM) but sufficiently high inhibition (>80%) was achieved to allow IC_{50} calculation. The calculated IC_{50}s for these compounds were 180 μM (pIC_{50} 3.7 ± 0.1; LLE_{AT} 0.22 kcal/mol) for compound **124**, 290 μM (pIC_{50} 3.5 ± 0.09; LLE_{AT} 0.11 kcal/mol) for compound **131**, and 82 μM (pIC_{50} 4.1 ± 0.1; LLE_{AT} 0.29 kcal/mol) for compound **132** (figure 4.16B)**.**

Compounds **124** and **132** represent exciting avenues for the development of hHAO1 inhibitors as they differ markedly from the known inhibitors of hHAO1 in both chemical structure and proposed binding site. To allow hit-to-lead development of these scaffolds, it would be helpful to elucidate the mechanism by which they inhibit hHAO1, as discussed in the next section.

4.4.4 Investigation of compound effect on hHAO1 tetramer by native mass spectrometry

Native mass spectrometry experiments, measuring the effect of follow-up compounds derived from fragment **6** on the size and shape of hHAO1 oligomers, were performed by Dr. Rod Chalk at the SGC to investigate whether their mechanism of action involved changes to the hHAO1 oligomeric structure and/or integrity. The MS spectra for purified hHAO1 (c002) at 2 mg/mL showed that hHAO1 was present in two oligomeric states, tetramer and octamer (figure 4.17A). The predominance of octamer can be reconciled with other experiments that find hHAO1 to be predominantly tetrameric by considering the octamer seen in the crystal structure of the spinach orthologue (sGOX, PDB 1gox). This octamer is a dimer of two tetramers that interact via only five residues, across four helices, that are within 4 Å of the other subunit (figure 4.17B).

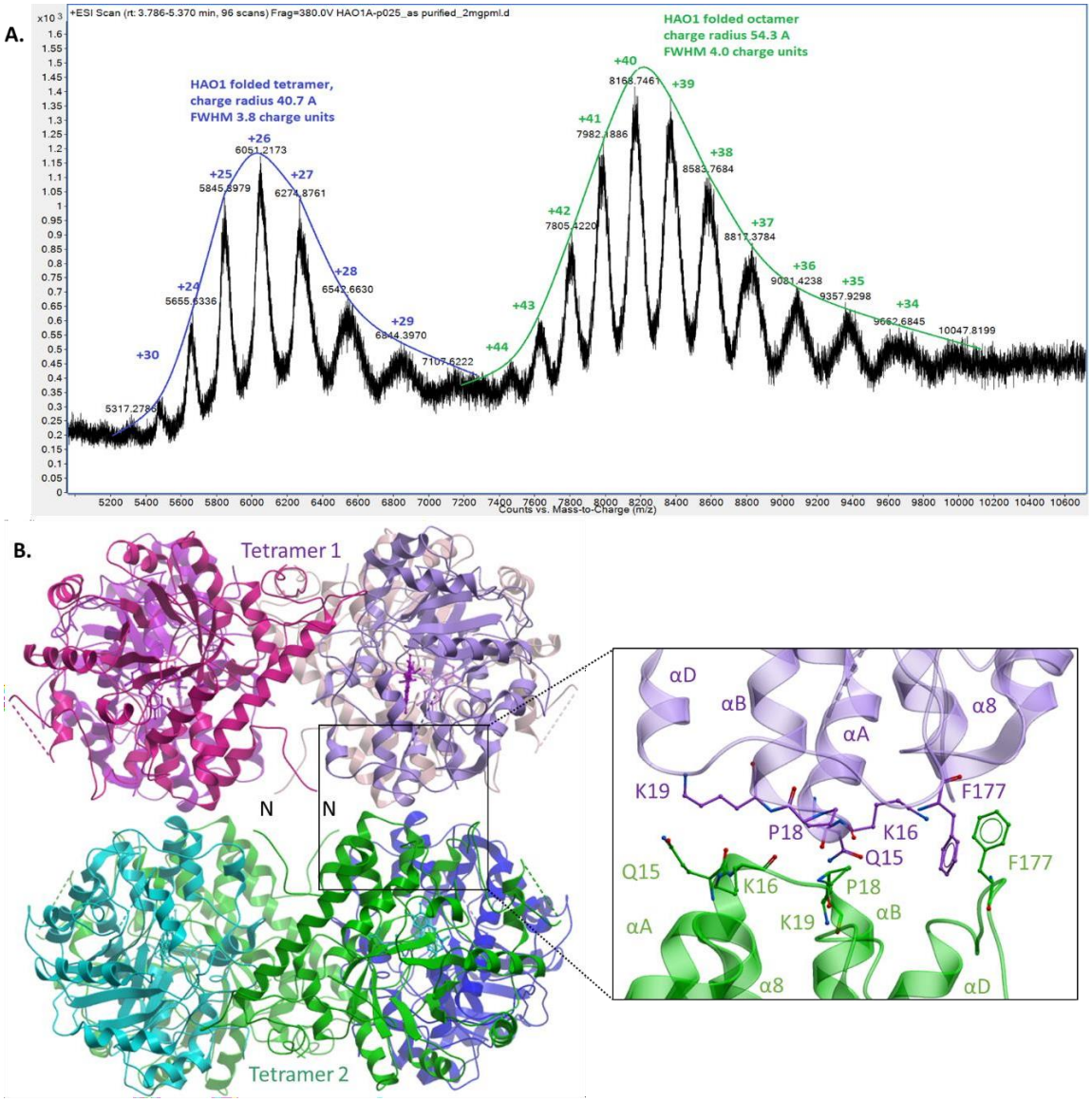


Figure 4.17: Native mass spectrometry analysis of hHAO1. *A. Native MS spectrum of hHAO1-c002 at 2 mg/mL showing two oligomeric states, tetramer and octamer. Each peak is labelled with the calculated oligomeric state (tetramer or octamer), surface area (charge radius) and resolution (full width at half maximum – FWHM – of the spectrum peak). B. Ribbon diagram showing the octamer observed in the crystal structure of spinach HAO1 homologue, sGOX (PDB 1gox). Subunit of one tetramer are shown in shades of pink and purple and subunits of the other tetramer are shown in shades of green and blue. Inset: Close-up of the octamer interface showing that few interactions occur at the interface between the two sGOX tetramers.*

This indicates formation of a weakly coordinated octamer, perhaps as a precursor to aggregation, that is not found under physiological conditions but rather is favoured by harsh *in vitro* conditions such as the tertiary-butanol used to crystallise sGOX [201], [212] or the no-salt and low pH buffer used for native MS experiments in this work. To characterise the effect of follow-up compounds derived from fragment **6** on the oligomeric state of hHAO1, each compound was added at 200 μM (0.1% final DMSO) to 2 mg/mL hHAO1 and incubated for 30 minutes on ice before measurement of native mass spectra. Comparison of the resultant m/z spectra in the presence of

each compound with the DMSO-only control demonstrated mass shifts of approximately 300 Daltons in the presence of compounds **125** and **126**, indicative of one compound molecule binding per tetramer of hHAO1 (red asterisks, figure 4.18) and mass shifts of approximately 500 Daltons in the presence of compounds **132**, **133** and **134**, indicative of two compound molecules binding per hHAO1 tetramer (black asterisks, figure 4.18).

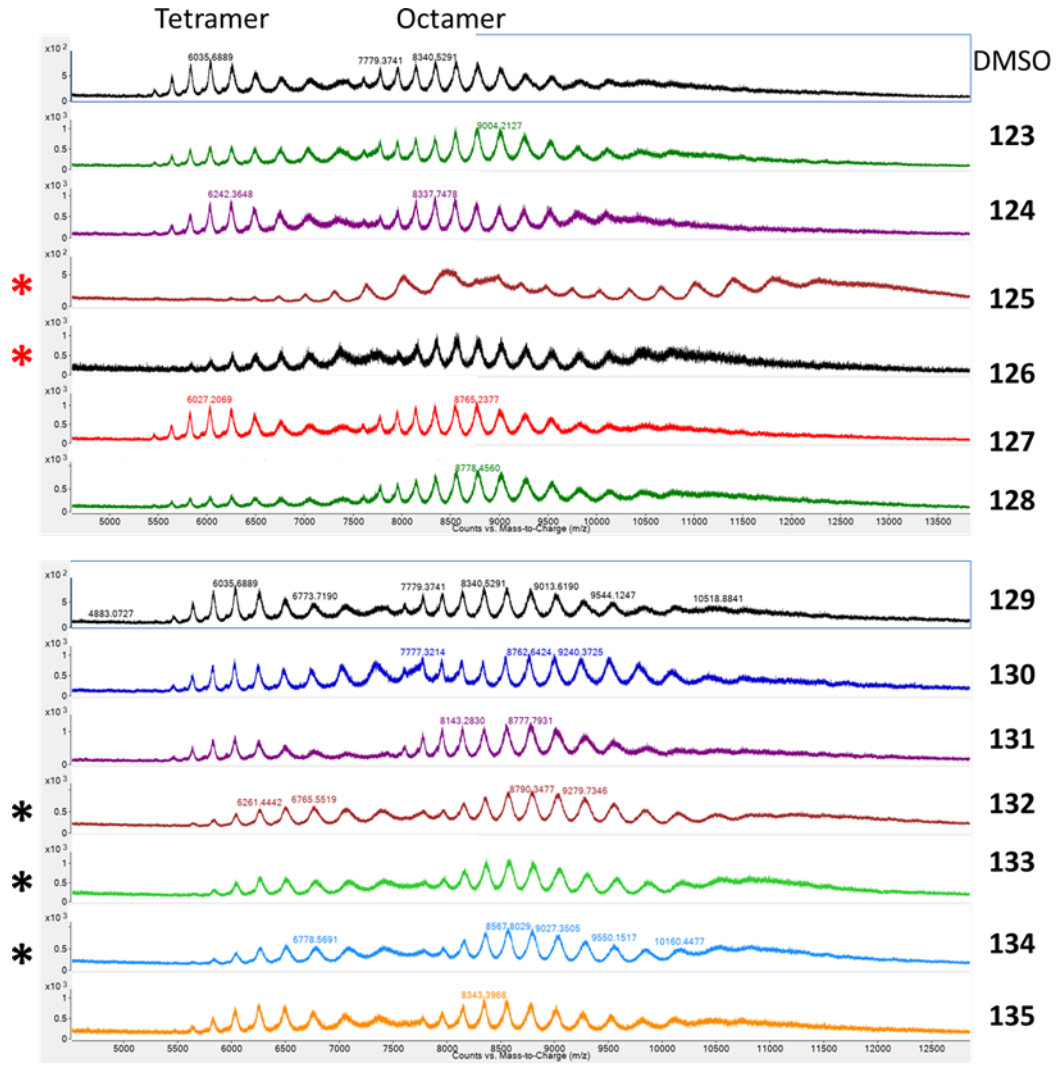


Figure 4.18: Comparison of native mass spectra of hHAO1 in the presence of DMSO and follow-up compounds 123 – 135. *m/z spectra of 2 mg/mL hHAO1 in the presence of 0.1% DMSO (top) or 200 μM of compounds* ***123*** *–* ***135*** *(final DMSO 0.1%). Red asterisks indicate spectra with a 300 Daltons mass shift and major conformational changes. Black asterisks indicate spectra with a 500 Dalton mass shift only.*

Compound **125** was the only compound that demonstrated the ability to disrupt the hHAO1 oligomer – in the m/z spectra, peaks are observed for heptameric (7 subunits) and monomeric species of hHAO1, with a concomitant reduction in the octamer peak (figure 4.19B). Addition of this compound also induces a reduction in charge radii of 8.5 Å for the tetramer and 2.7 Å for the octamer, corresponding to a major conformational change relative to the DMSO only control

(figure 4.19A). Similarly, addition of compound **126** induces a reduction in charge radii of 5.2 Å for the tetramer and 0.9 Å for the octamer, indicating a major conformational change in the tetramer and a minor conformational change in the octamer (figure 4.19C) relative to the DMSO only control (figure 4.19A).

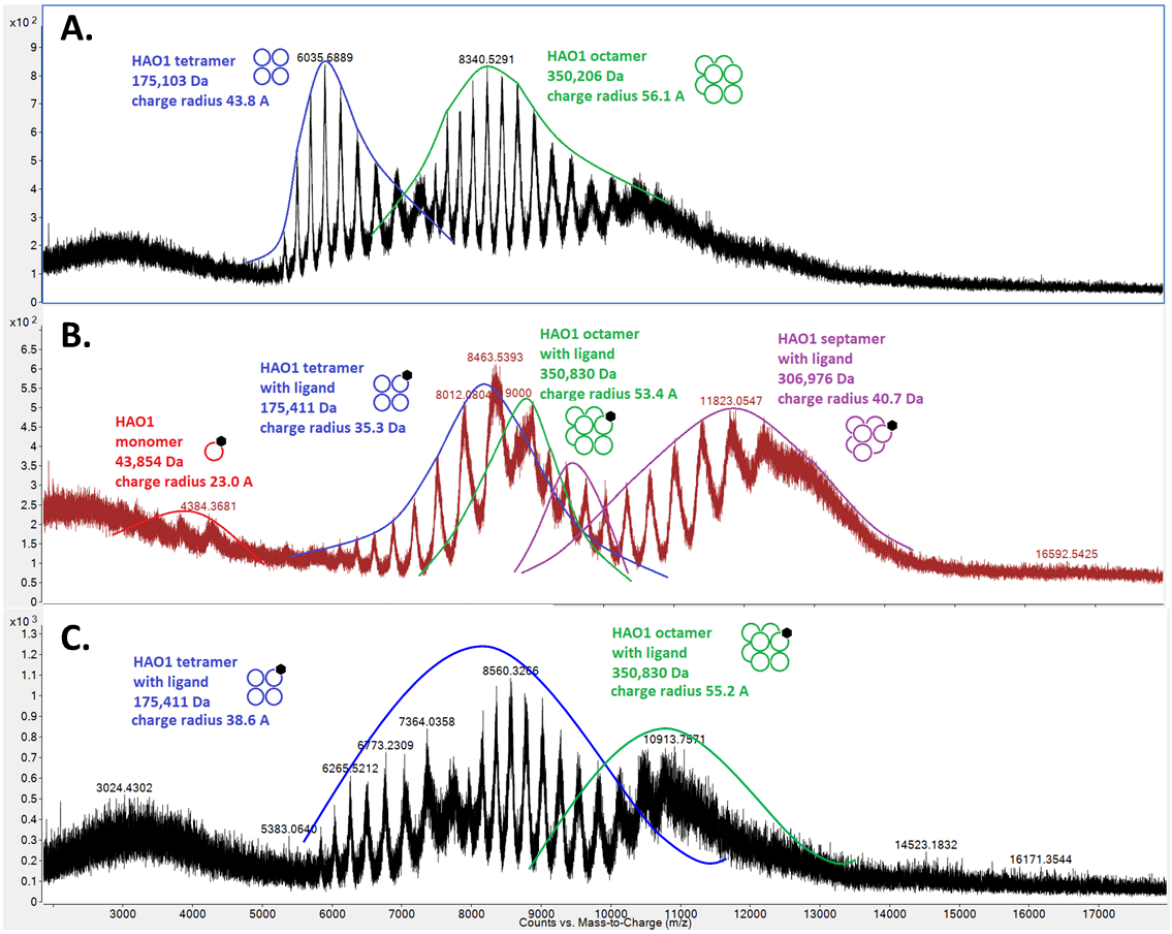


Figure 4.19: Comparison of native mass spectra of hHAO1 in the presence of DMSO, compound 125 and compound 126. *m/z spectra of 2 mg/mL hHAO1 in the presence of 0.1% DMSO (A), 200 µM compound* ***125*** *(B) or 200 µM compound* ***126*** *(C). Peaks are outlined according to the corresponding oligomeric state that they represent – monomer (red), tetramer (blue), heptamer (purple) or octamer (green).*

DMSO is not an ideal solvent for native MS analysis because it causes compaction of proteins, which can mask or exaggerate subtle conformational changes, and binding of a variable number of DMSO molecules to each protein molecule complicates the interpretation of mass shifts caused by ligand binding. Additionally, the resolution achieved in the experiments described in this work generated broad mass peaks that hamper mass assignment and lead to higher uncertainty in interpreting spectra. Initial attempts to improve resolution were not successful and consequently further development of this method, or investigation by orthogonal approaches, is

required to facilitate accurate analysis of these compounds.

Additional studies are required to optimise the native MS results describe here to gain a better understanding of how these compounds effect the hHAO1 tetramer. It is also important to confirm that these compounds inhibit hHAO1 activity by inducing conformational changes from binding at the subunit-subunit interface, rather than by simultaneous binding to the active site.

4.4.5 Crystal structure of hHAO1 after soaking follow-up compounds

As with the other follow-up compounds described in sections 4.2.4 and 4.3.3, crystal soaking and co-crystallisation experiments with follow-up compounds derived from fragment **6** were unsuccessful in generating multiple structures for comparative analysis. Only one co-crystal structure of hHAO1 was determined, bound with compound **123**, which replaces the tetrahydrofuran moiety of fragment **6** with 1-methyl-4-pyridinyl-piperazine.

This structure had the same space group and unit cell dimensions as the fragment-bound datasets with 1 copy of hHAO1 in the asymmetric unit. The structure of hHAO1 bound with compound **123** (at 1.4 Å resolution) converged to Rfree of 0.196 and R of 0.158. Based on difference Fourier density, one full occupancy molecule of FMN and one partially occupied molecule of compound **123** were built into the protomer. Data processing and structure refinement statistics are provided in table S2.1.

No significant changes were observed in either the overall structure of hHAO1 (RMSD < 1 Å across hHAO1-FMN-glycolate, hHAO1-FMN-fragment **6** and hHAO1-FMN-**123** structures), nor was significant movement of nearby sidechains observed when compound **123** was bound, although the activity assay shows fair inhibition for this compound (49.6 ± 5.9%, figure 4.15).

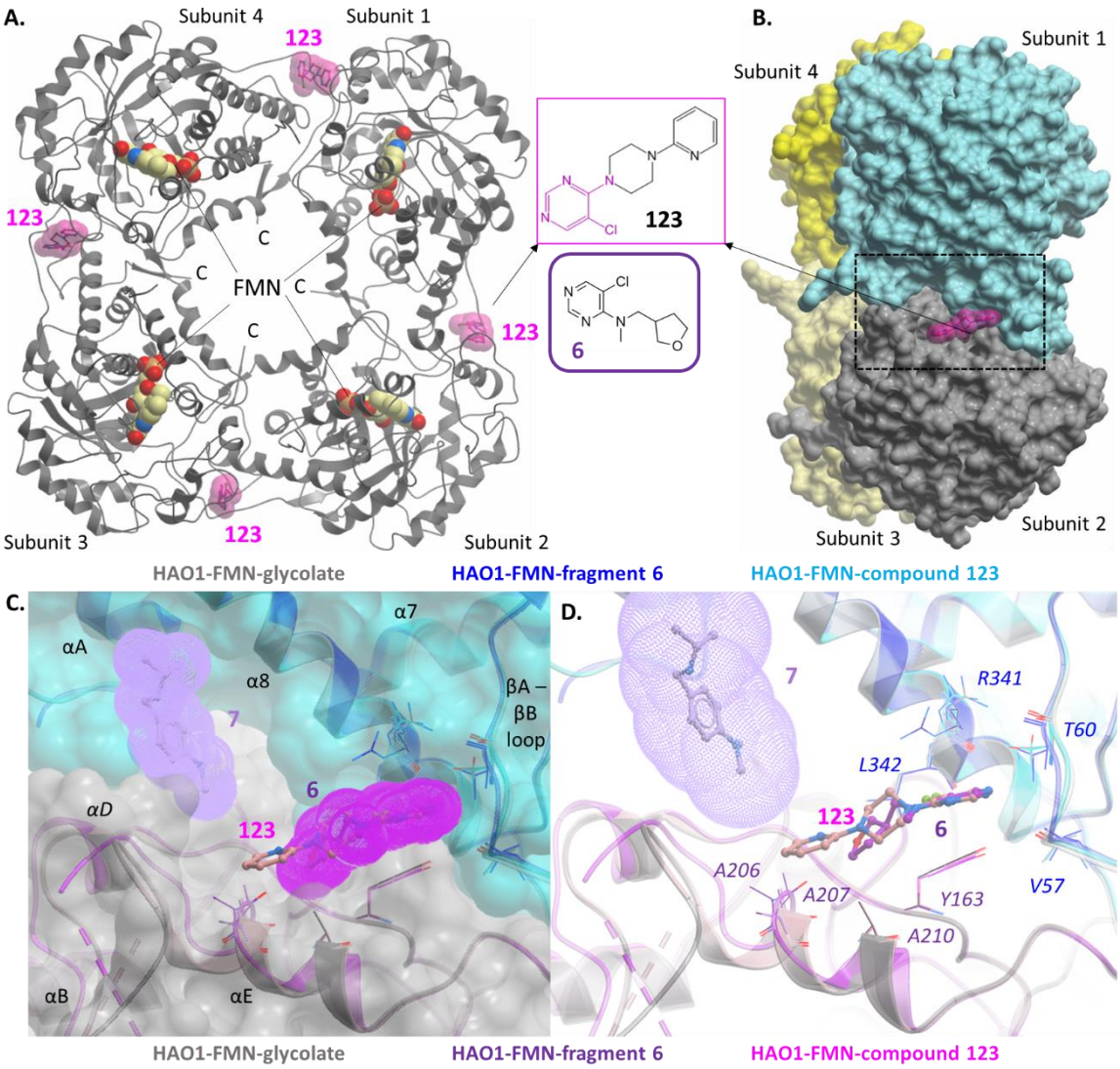


Figure 4.20: Crystal structures of hHAO1 bound with fragment 6, follow-up compound 123, and glycolate. *A. Ribbon diagram showing where compound **123** (purple spacefill and sticks) binds with respect to the hHAO1 tetramer. FMN is shown as spacefill. B. Side-view of hHAO1 tetramer with subunits shown as different coloured surface representations. Compound **123** is shown as sticks with mesh surfaces. Inset: Chemical structures of fragment **6** and compound **123**. C. Close up view of the subunit-subunit interface binding site as surface and ribbon (C) and ribbon and line (D)representations. Superimposed ribbons are of hHAO1-FMN-glycolate (grey), hHAO1-FMN-fragment **6** (dark blue/purple) and hHAO1-FMN-**123** (light blue/pink) structures. Fragments **6** and **7** are shown as sticks with mesh surfaces and compound **123** is shown as sticks.*

Compound **123** binds in the same binding pocket as fragment **6**, at the interface between each subunit pair of the hHAO1 tetramer (figure 4.20A, B). The binding mode of the chloropyrimidine ring of compound **123** is very similar to that observed for fragment **6**, while the pyridinyl-piperazine moiety reaches further into the hydrophobic pocket partially occupied by the tetrahydrofuran group in the hHAO1-FMN-fragment **6** structure (figure 4.20C, D). Specifically, the chloropyrimidine ring of compound **123** forms π- π stacking interactions with the Tyr163

sidechain of one subunit, and hydrogen bonds with the Thr60 sidechain of the other subunit. The 45° displacement of the guanidino group of Arg341 observed when binding fragment **6** (section 3.3.6) is not observed in the presence of compound **123**.

This chapter first describes the inhibition of, and binding to, hHAO1 in solution by the 4 promising fragment hits introduced in the previous chapter and then discusses the design, synthesis and *in vitro* characterisation of follow-up compounds from each fragment. Follow-up compounds demonstrating fair potency (IC_{50} < 100 μM) and good ligand efficiency (LLE_{AT} > 0.25) were obtained from each fragment starting point, representing promising compounds for hit-to-lead development. The significance of these findings and how they may be applied to the development of new therapeutic modalities for primary hyperoxaluria is discussed further in chapter 7.

5. Structural biology of hGALK1 inhibition

As discussed in section 1.2.4, hGALK1 is a promising target for substrate reduction therapy of classic galactosaemia. *In vitro* characterisation of wild type recombinant hGALK1, paired with mutagenesis studies, has elucidated the role of various conserved structural motifs and residues for hGALK1 substrate recognition, catalysis and stability, providing a sound basis for the development of potent, selective inhibitors [16], [24], [35], [182], [184], [195], [249]–[253]. Several hGALK1 inhibitors have been identified from HTS screening campaigns using a luminescence-based activity assay [107], [133]-[135], [139], [254].

Development of hit compounds into therapeutic molecules with suitable inhibitory, pharmacokinetic and pharmacodynamic properties has been hindered by two factors. Firstly, efficient hit-to-lead development for hGALK1 has been hampered by the composition of the screening libraries tested to date. In one campaign, the compounds were selected based on chemical diversity rather than clinical viability and included compounds known to be toxic or otherwise unsuitable for development as drug candidates [139]. In another two campaigns a subset of the NIH Molecular Libraries Small Molecule Repository (MLSMR) was screened, a collection of small molecules developed for the identification of chemical probes (i.e. for target validation, step 1 in figure 1.5) that are generally not suited for therapeutic use [190], [191].

The second confounding factor hindering hGALK1 hit-to-lead development is the absence of structural data to guide lead evolution. Characterisation of the published low-μM inhibitors of hGALK1 through activity assay kinetics and *in silico* modelling supports binding at the ATP binding site [133], [135], [254] but an accurate, detailed understanding of protein-inhibitor interactions is lacking. For example, interaction with Arg105 has been predicted to be an important factor in inhibiting hGALK1 but exploitation of this interaction would require more structural knowledge because molecular dynamics simulations suggest high flexibility of this residue [254].

A promising approach to overcome these limitations is to screen a library of smaller-sized compounds, using X-ray crystallography to generate parallel structural data, to identify the minimal scaffolds required for binding, leading to higher quality starting points for lead development. Screening by activity assay can bias results towards the discovery of orthosteric binders and may overlook high quality allosteric hits due to their lower initial inhibitory activity. Screening by X-ray crystallography can help overcome this bias, generating inhibitor scaffolds for hGALK1 that diverge from the ATP-mimetics identified to date.

The first section of this chapter (5.1) describes the optimisation of hGALK1 purification and crystallisation to facilitate fragment screening by X-ray crystallography. The second section of this chapter (5.2) expands the structural data available for hGALK1, including the first solved structures of hGALK1 bound to orthosteric inhibitors, the spiro-benzoxazole series reported in [133], [135]. The final section of this chapter (5.3) reports a fragment screening campaign against hGALK1 crystals, discovering 28 hits clustering in 3 binding pockets.

5.1 Optimisation of hGALK1 protein production and crystallisation for drug discovery

Purification and crystallisation of recombinant hGALK1 has been reported previously [22]. In traditional X-ray crystallography, many crystals, often of variable quality, are screened to solve a single structure but for fragment screening hundreds of highly reproducible crystals of consistent atomic level diffraction are required and considerable optimisation is often required to achieve this increase in scale.

5.1.1 Expression and purification of recombinant hGALK1 from *E. coli*

A DNA plasmid for *E. coli* expression of hGALK1 was generously supplied by Prof. Kent Lai, University of Utah (SGC ID: GALK1-c000). This construct, based on the published sequence of the known hGALK1 structure (PDB 1wuu, [184]), contains an N-terminal six histidine tag for nickel affinity purification and the full length (Ala2-Leu392) hGALK1 sequence within which two ‘surface entropy reduction (SER)’ substitutions were engineered: Glu252Ala and

Lys253Ala. The SER approach aims to improve crystallisation by substitution of surface exposed flexible side chains, particularly high entropy lysine and glutamate residues, that may interfere with the ability of a protein to form intermolecular contacts and thus to assemble into a crystalline lattice [255].

Recombinant hGALK1 was purified by affinity and size exclusion chromatography, in HEPES buffer, similar to published procedures [35], [184] (section 2.1.3; figure 5.1). Protein precipitation was observed within 15 minutes of elution from the nickel affinity column in high imidazole buffer. After centrifugation to remove precipitants, protein in the supernatant faction appeared in a non-void peak on size exclusion (i.e. it did not elute as high molecular weight or aggregated species; figure 5.1B) and no further precipitation was observed in buffer without imidazole.

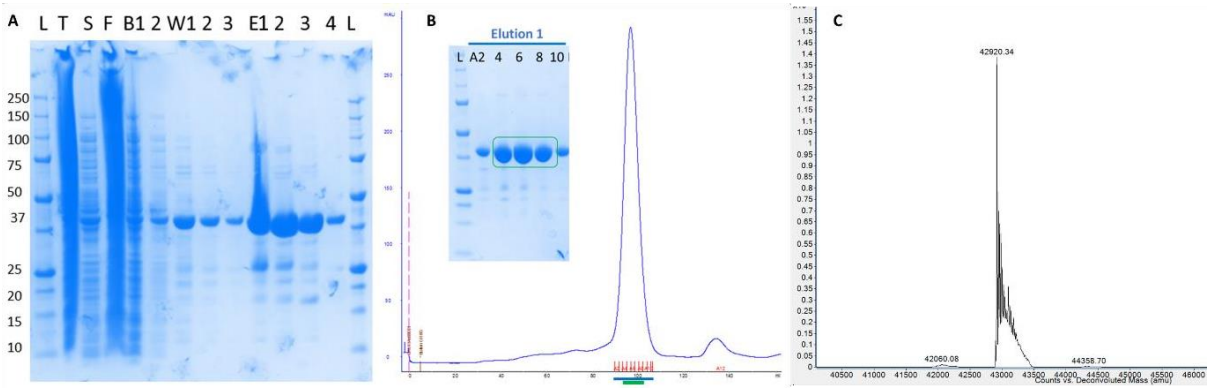


Figure 5.1: SDS-PAGE analysis, size exclusion chromatogram and intact mass spectrum of representative hGALK1 purification. *A. SDS-PAGE analysis of Ni-NTA purification. L – Molecular weight ladder (kDa); T –total cell lysate; S – soluble fraction; F – flow- through; B1-2 – fractions eluted with binding buffer; W1- 3 – fractions eluted with wash buffer; E1-4 – fractions eluted with elution buffer. B. SDS-PAGE analysis and chromatogram of size exclusion purification of E1 fraction by HiLoad 16/600 Superdex 200 column. Fractions that were analysed by SDS-PAGE are indicated in blue on the chromatogram. Fractions that were pooled and concentrated to generate the final purification product are indicated in green box. C. Intact mass spectrometry profile of the final purified hGALK1 product, showing a single peak of the expected molecular weight.*

5.1.2 Initial crystal trials for hGALK1 in HEPES buffer

Purified hGALK1 was concentrated to 11.7 mg/mL and 50 mM galactose and 10 mM AMPPNP were added prior to crystallisation trials in 150 nL sitting drops, containing mixtures of protein and precipitant with 20 µL precipitant solutions acting as reservoirs. Crystallisation trials were set up to explore different crystallisation conditions from commercial screens at both 4 °C and

20 °C and three ratios of protein: reservoir were tested for each condition. Two similar conditions from the Hampton Crystallisation Screen (HCS3 screen, conditions A09 and B06) yielded small plate (20 – 50 μm) or large needle-like (75 – 150 μm) crystals after 4-7 days at 20 °C (figure 5.2). These crystals generally diffracted to 3 – 5 Å resolution when tested at the Diamond Light Source. One further condition, A09 from the Morpheus screen, yielded crystals of similar diffraction.

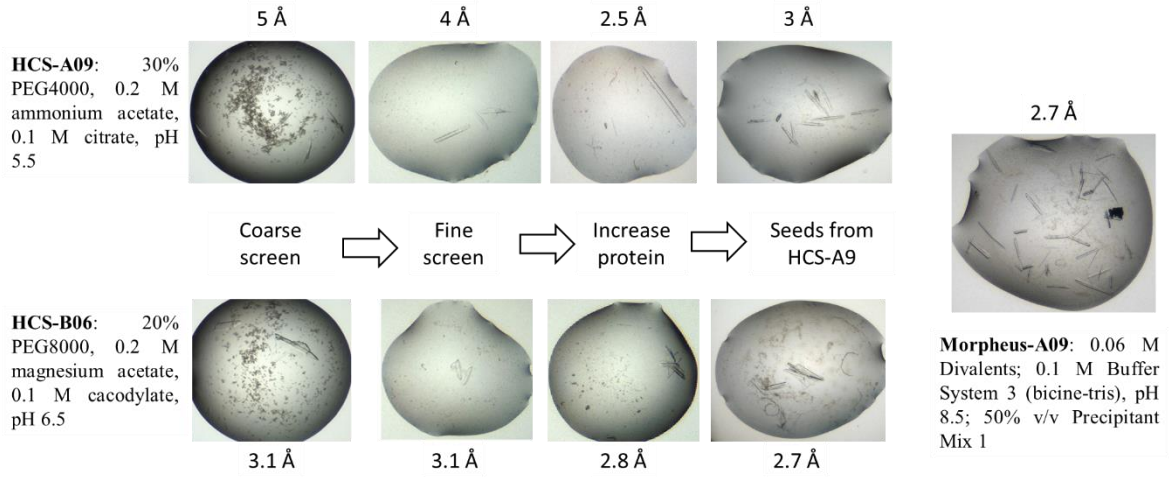


Figure 5.2: Crystal images and maximal resolution for crystals of hGALK1 purified in HEPES buffer. *Crystal images and maximal resolution achieved for hGALK1 crystals in HCS-A09 (top) and HCS-B06 (bottom) crystallisation conditions at 20 °C. Far right, the Morpheus-A09 condition, showing improved morphology but similar diffraction relative to coarse screen results from the HCS conditions. All crystals appeared between 4 and 14 days and were grown at 20 °C.*

To optimise crystals for better diffraction several standard approaches were tried. These included re-design of construct boundaries; fine-tuning of pH, salt and precipitant concentrations of the original crystallisation conditions in a grid plate; varying of protein concentration and crystallisation drop volume; micro-seeding to modulate crystal nucleation; and screening of generic crystallisation additives (figure 5.2). Collectively, these approaches improved the morphology and diffraction of hGALK1 crystals, but very few crystals had a resolution better than 2.5 Å and thus did not meet the threshold set in this work for fragment screening.

5.1.3 Optimisation of buffer conditions to improve diffraction of hGALK1 crystals

Based on the precipitation of hGALK1 seen during protein purification (section 5.1.1), and the limited improvement in diffraction consistency achieved by fine tuning crystallisation conditions

(section 5.1.2), protein stability was rationalised as an influential factor in crystallisation success. Protein stability encompasses two related characteristics: thermodynamic stability, relating to the equilibrium between the folded and unfolded states, and kinetic stability, relating to the rate of unfolding [256]. A protein with high thermodynamic stability favours an equilibrium with a low amount of unfolded and partially folded states whereas a protein with high kinetic stability has a low rate of irreversible denaturation so that protein remains in a folded, functional state over a given time period [257].

Thermodynamic stability can be readily measured *in vitro* as the amount of folded and/or functional protein remaining after application of an environmental stress such as temperature, chemical denaturants or proteolytic enzymes. Several studies have shown correlations between thermodynamic stability and protein crystal morphology and diffraction [258]–[262]. For example, a large-scale analysis of 657 proteins showed that a decreased thermodynamic stability, as measured by applying a temperature gradient using differential scanning fluorimetry (DSF), correlated to significantly less success in crystallisation trials [261].

DSF involves the determination of a protein's melting temperature 'Tm' (temperature at which half the protein is unfolded), by monitoring fluorescence released upon binding of an environmentallysensitive dye, such as Sypro Orange, typically using a real time PCR machine [263]. Sypro Orange, and similar probes, are hydrophobic molecules that preferentially bind to hydrophobic surfaces of a protein, unmasked by protein unfolding, relieving the fluorescence quenching of aqueous solution [263], [264]. Although initially developed as a rapid, cheap and high-throughput screening method for drug discovery [265], [266], DSF has since been validated for evaluating buffer components and additives that increase the thermodynamic stability of a target protein to increase protein yield and improve crystallisation [262], [263], [267], [268].

To determine whether an alternate buffer could improve the diffraction quality of hGALK1 crystals, a 32-condition buffer screen, varying both pH and buffer composition, was designed

based on a previous study [269]. The change in thermal stability of hGALK1 (ΔTm, change in melting temperature, °C) in the presence of 100 mM of each of these buffers, relative to the original purification buffer, was determined by DSF (table 5.1).

Buffer	pH:	4.5	5	5.5	6	6.5	7	7.5	8	8.5	9	10
Citrate		-14.8 ± 1.6	-9.4 ± 0.8									
K_2HPO_4			-3.4 ± 0.5		+0.7 ± 0.1	+1.6 ± 0.2	+1.7 ± 0.2	+1.9 ± 0.2				
Na_2HPO_4			-10.1 ± 1.6	-3.9 ± 0.9	+0.6 ± 0.1	+1.7 ± 0.2	+2.0 ± 0.2 **(49.1°C)**					
MES				-12.4 ± 0.6	-4.3 ± 0.4	-0.7 ± 0.4						
Bis-Tris					-0.7 ± 0.1	-1.1 ± 0.1	-3.6 ± 0.6		-0.9 ± 0.1			
HEPES							0.0 ± 0.1	0.0 ± 0.1 **(47.1°C)**	-0.8 ± 0.1	-1.3 ± 0.1		
Tris								-0.9 ± 0.2	-1.7 ± 0.2	-3.1 ± 0.2	-4.2 ± 0.1	
Bicine									-0.9 ± 0.1	-2.1 ± 0.2	-4.0 ± 0.1	
CHES											-9.4 ± 0.6	
Glycine												-5.1 ± 0.1

Table 5.1: Buffer screen for hGALK1 by DSF. *Numbers indicate the change in melting temperature (ΔTm) of purified hGALK1, relative to the original buffer (0.1 M HEPES pH 7.5), in buffers of different composition and pH. For the original buffer, 0.1 M HEPES pH 7.5, and the most optimal buffer, 0.1M sodium phosphate pH 7, the melting temperature is shown in brackets and highlighted. ΔTm values are the average of 3 technical replicates ± standard deviation. Conditions that lower the Tm by more than 1 °C are shaded gold and conditions that increase the Tm by more than 1 °C are shaded green.*

In the original purification buffer (HEPES pH 7.5) hGALK1 had a Tm of 47.1 °C whereas in the most stabilising buffer (sodium phosphate pH 7) hGALK1 had a Tm of 49.1 °C (table 5.1, highlighted; figure 5.3A), representing a modest but significant stabilisation of hGALK1 (ΔTm = 2 °C; P = 0.0001). All 7 buffer conditions that increased the Tm of hGALK1 were phosphate buffers (table 5.1, rows 2 and 3), consistent with the observation that nucleotide-binding proteins are preferentially stabilised by phosphate buffers, potentially due to weak binding of the phosphate anion to the nucleotide binding pocket [270].

Additional dose-dependent stabilisation of hGALK1 was observed with increasing

concentrations of its substrate, galactose, with a maximum increase in melting temperature of 5.9 °C in the presence of 100 mM galactose (figure 5.3). No Tm changes were observed upon addition of ATP, ADP or AMPPNP. From these results, sodium phosphate pH 7 buffer, supplemented with 50 mM galactose, was chosen to replace HEPES in subsequent purifications of hGALK1.

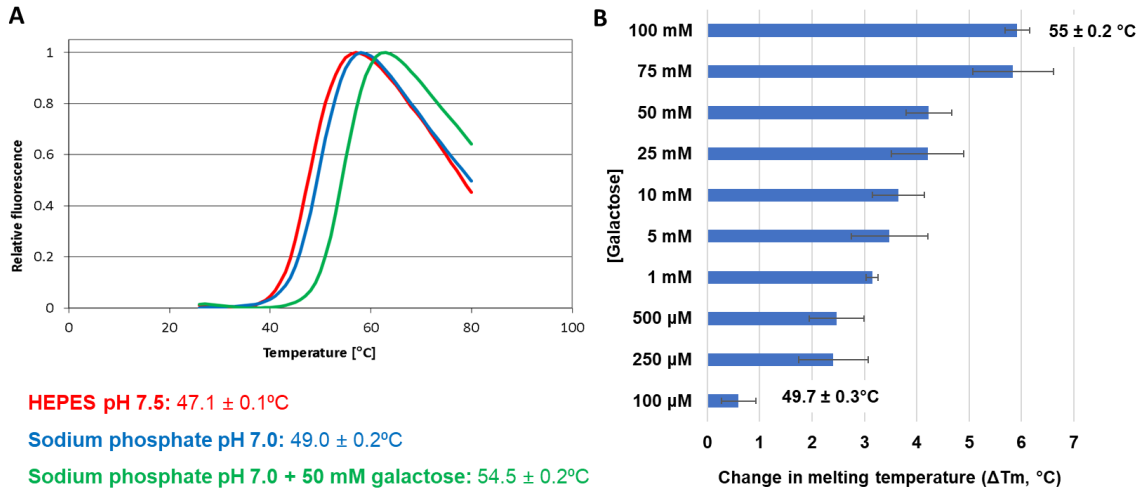


Figure 5.3: Measurement of buffer and substrate effects on hGALK1 thermodynamic stability as determined by DSF. *A. Example melting curves of hGALK1 protein (4.6 µM) under the original buffer condition (red, 0.1M HEPES pH 7.5), the optimised buffer condition (blue; 0.1M sodium phosphate pH7) and with the addition of substrate (green; 0.1M sodium phosphate pH 7, 50 mM galactose). B. Change in melting temperature of 4.6 µM hGALK1, in sodium phosphate pH 7 buffer, with increasing concentrations of galactose. In both panels, Tm values are averages of three technical replicates. Error bars in panel B are the standard deviation of technical triplicates.*

When purified in this buffer, hGALK1 protein did not precipitate after thawing or after extended incubation (>1 hour) on ice. An increase in protein stability was also observed during crystallisation trials – the protein concentration used to set up crystal plates had to be increased from 15 to 25 mg/mL as crystal drops were completely clear (indicating supersaturation was not achieved) at lower hGALK1 concentrations.

5.1.4 Crystal trials of hGALK1 in phosphate buffer

After increasing the protein concentration for crystallisation, numerous rod-like hGALK1 crystals (~75 µm in size), observed in the Morpheus and HCS coarse screens, were mounted for diffraction screening at the Diamond Light Source synchrotron. Of the ten crystals mounted from the Morpheus coarse screen, two did not diffract, but the remaining eight demonstrated fair

diffraction (between 2.7 and 5 Å) including two crystals that diffracted to 2.7 Å. Fine-tuning of the crystallisation condition that yielded the latter crystals (Morpheus-A09) and seeding from the same condition, led to larger, more straight-edged crystals (figure 5.4A). Despite apparent improvement in crystal morphology, the consistency in diffraction was similar with only a small proportion (~20%) of crystals diffracting to 2.7 Å.

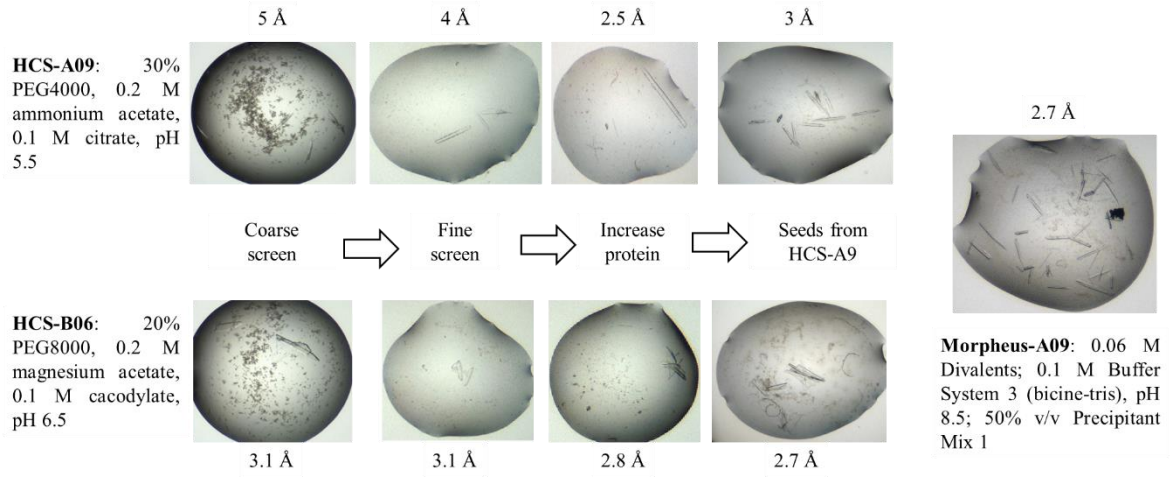
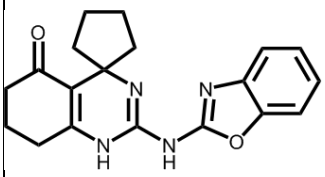
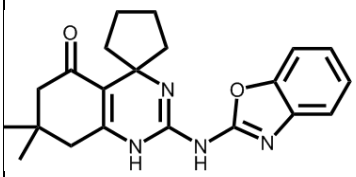
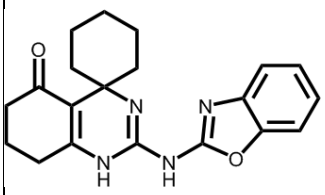
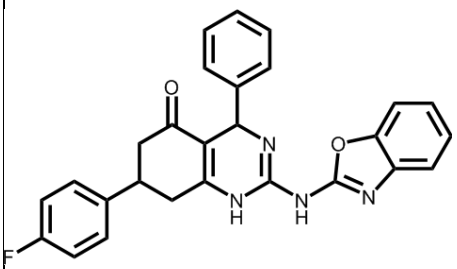
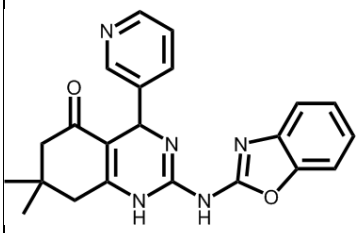
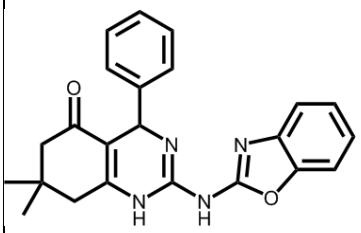
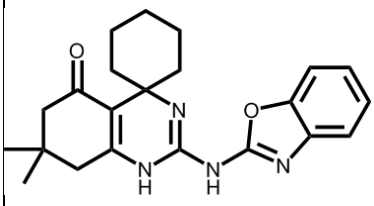


Figure 5.4: Optimisation of hGALK1 crystals in Morpheus-A09 condition. *A. Improved crystal size and shape with fine screen and seeding from the original condition. All crystallisation plates were set up as 300 nL drops of varying ratios of precipitant solution. All plates were set up with 25 mg/mL hGALK1 purified in sodium phosphate (pH 7) buffer supplemented with 50 mM galactose. B. Effect of DMSO and substrates on hGALK1 crystal morphology. All plates were set up with 25 mg/mL hGALK1 purified in sodium phosphate (pH 7) buffer without galactose.*

The next approach explored was the addition of small molecule ligands to purified protein prior to crystallisation trials, with the aim of improving diffraction by reducing protein conformational flexibility and/or heterogeneity [262], [270], [271]. Commercial crystallisation additives (from the Molecular Dimensions additive screen) did not significantly improve crystallisation. Screening of alternate ligand combinations (galactose alone, galactose + ADP and ADP alone) resulted in one 2.5 Å dataset from which the structure of hGALK1-galactose-ADP was determined (described in the next section). However, no further improvement in crystal reproducibility was observed (figure 5.4B). Finally, seven ATP- competitive inhibitors of hGALK1 (table 5.2; compounds **T1** – **T7**), selected from commercially available compounds belonging to the previously reported spiro-benzoxazole series of hGALK1 inhibitors (discussed in section 1.4.4; [133], [135]), were trialed as crystallisation additives.

Table 5.2: Tool compounds belonging to the previously reported spiro-benzoxazole series of hGALK1 inhibitors.

	Compound T1	**Compound T2**
Structure		
Reference	Compound **3** in [135] (IC_{50} 1 μM)	Compound **1** in [135] (IC_{50} 6 μM)
CID (Supplier)	CID664331 (STOCK4S-65953)	CID1286615 (STOCK4S-65744)
[Stock] DMSO/NMP	5/50 mM	7.5/50 mM
	Compound T3	**Compound T4**
Structure		
Reference	Compound **5** in [135] (IC_{50} 12 μM)	Not reported
CID (Supplier)	CID1294862 (STOCK4S-53098)	CID2956468 (#7810744)
[Stock] DMSO/NMP	10/50 mM	10/50 mM
	Compound T5	**Compound T6**
Structure		
Reference	Not reported	Not reported
CID (Supplier)	CID3161798 (STOCK5S-50757)	CID2953347 (STOCK4S-69599)
[Stock] DMSO/NMP	7.5/4.6 mM	7.5/25 mM
	Compound T7	
Structure		
Reference	Not reported	
CID (Supplier)	CID667019 (STOCK5S-50080)	
[Stock] DMSO/NMP	10/50 mM	

The seven spiro-benzoxazole tool compounds showed limited solubility in DMSO (maximal concentration of 10 mM), hence compound stocks were also prepared at higher concentrations in an alternate solvent, *N*-Methyl-2-pyrrolidone (NMP; 5 of 7 compounds were soluble up to

50 mM) and crystallisation trials were performed from both solvents in parallel (stock concentrations are listed in table 5.2).

These compounds, **T1** – **T7**, were selected for two reasons. Firstly, a single high-resolution structure from one or more of these inhibitors would illuminate compound-protein interactions, identifying bonded and non-bonded chemical constituents to guide hit-to-lead development. Secondly, if inclusion of an inhibitor improved diffraction quality and reproducibility to the standard needed for fragment screening, these inhibitor-bound crystals could be used to perform a fragment screen identifying primarily non-orthosteric binders.

The key benefits of allosteric, rather than orthosteric, inhibition are: increased selectivity against related proteins with similar functions; consistent target inhibition independent of substrate concentration; and potential to maintain high activity in the presence of high substrate concentrations [272]–[274]. In developing inhibitors of hGALK1, for example, ATP-analogous inhibitors may not be selective over other GHMP kinases (due to the shared ATP-binding motifs; section 1.4.2) and may be considerably less effective in the cell, where millimolar levels of ATP compete with the inhibitor for binding.

Crystallisation of hGALK1 was achieved in the presence of all 7 compounds, dissolved in both solvents, with 4 protein-compound mixtures yielding crystals diffracting to 2.5 Å or better when the compound was prepared in NMP (figure 5.5). It should be noted that galactose was present (50 mM from the purification buffer) in all protein samples used for crystal trials and was present in all solved structures. Crystals from the Morpheus coarse screen facilitated single structure solution for hGALK1 bound with compound **T1** (1/7 crystals diffracted better than 3 Å), compound **T2** (2/5 crystals diffracted better than 3 Å) and compound **T7** (2/9 diffracted better than 3 Å).

Importantly, crystals of hGALK1 bound with compound **T3** both facilitated single structure solution and identified suitable conditions for fragment screening (6/7 crystals diffracted better

than 2 Å). Crystal structures of hGALK1 bound with the tool compounds are discussed in section 5.2.

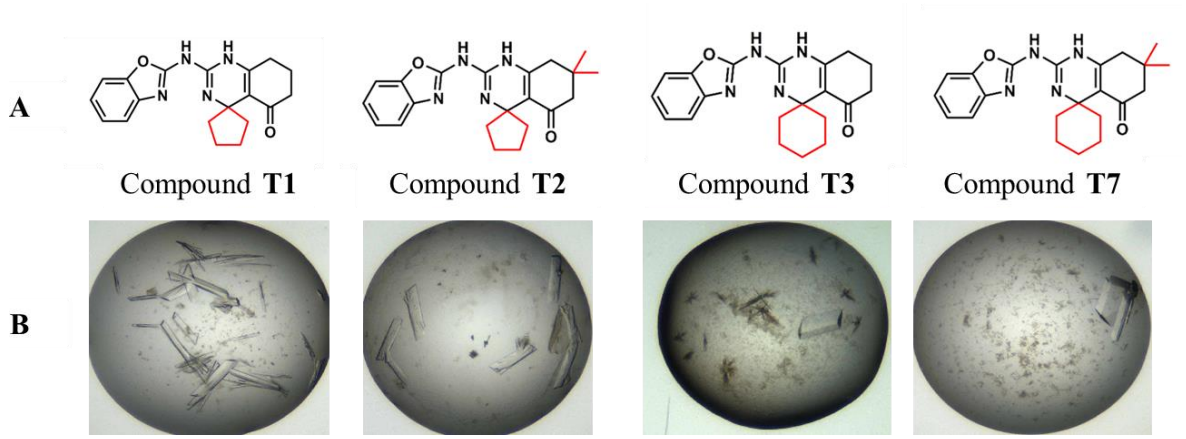


Figure 5.5: Co-crystallisation of hGALK1 with spiro-benzoxazole inhibitors produced robust crystals. *A. Chemical structures of tool compounds yielding high resolution co-crystals with hGALK1. The shared scaffold is coloured in black and the variable constituent is coloured in red. B. Representative crystal images of 25 mg/mL hGALK1, purified in sodium phosphate (pH 7) buffer supplemented with 50 mM galactose, and 2.25 mM compound (5% NMP), grown between 4 and 7 days at 20 °C. Crystallisation conditions are: Morpheus-H01, compound **T1**; Morpheus-F08, compound **T2**; Morpheus-G08, compound **T3**; and Morpheus-G04, compound **T7**.*

Specifically, crystallisation with 2.25 mM compound **T3** (final of 5% NMP) in Morpheus condition G08 (0.02 M sodium formate; 0.02 M Ammonium acetate; 0.02 M sodium citrate tribasic dihydrate; 0.02 M sodium potassium tartrate tetrahydrate; 0.02 M sodium oxamate; 0.1 M sodium HEPES:MOPS (acid) buffer pH 7.5; 12.5% v/v MPD; 12.5% PEG1000; 12.5% w/v PEG3350) yielded sufficient quantity and quality of crystals to perform fragment screening (section 5.3).

5.2 Crystal structures of hGALK1 liganded complexes

This section describes the structures of hGALK1 bound to galactose and either ADP (section 5.2.2) or one of the four tool compounds that yielded well diffracting crystals (section 5.2.3). Data processing and structure refinement statistics are provided in table S2.2.

5.2.1 Overall description of datasets collected with known ligands of hGALK1

Crystals of hGALK1 in the presence of galactose and ADP belonged to the triclinic space group P1 (cell dimensions were a = 78.22 Å, b = 110.32 Å, c = 118.23 Å, α = 113.85°, β = 102.26° and γ = 101.29°) with eight molecules in the asymmetric unit. The final refined structure, at 2.5 Å

resolution, converged to Rfree of 0.256 and R of 0.221. Based on difference Fourier density, galactose and ADP were built into all eight hGALK1 molecules within the asymmetric unit.

Crystals of hGALK1 in the presence of galactose and compound **T1** belong to the monoclinic space group P 2_1 (cell dimensions were a = 119.87 Å, b = 97.20 Å, c = 144.81 Å, α = 90.00°, β = 98.57° and γ = 90.00°) with eight molecules in the asymmetric unit. The final refined structure, at 2.4 Å resolution, converged to Rfree of 0.262 and R of 0.210. Based on difference Fourier density, galactose was built into eight hGALK1 molecules within the asymmetric unit and compound **T1** was built into seven molecules.

Crystals of hGALK1 in the presence of galactose and compound **T2** belong to the orthorhombic space group F2 2 2 (cell dimensions were a = 107.59 Å, b = 151.09 Å, c = 229.91 Å, α = β = γ = 90.00°) with two molecules in the asymmetric unit. The final refined structure, at 2.0 Å resolution, converged to Rfree of 0.240 and R of 0.195. Based on difference Fourier density, galactose was built into both hGALK1 molecules within the asymmetric unit and compound **T2** was built into one molecule. Crystals of hGALK1 in the presence of galactose and compound **T3** had the same space group and unit cell characteristics as crystals of hGALK1 in the presence of galactose and compound **T2**. The final refined structure, at 2.1 Å resolution, converged to Rfree of 0.231 and R of 0.191. Based on difference Fourier density, galactose and compound **T3** were built into both hGALK1 molecules within the asymmetric unit.

Crystals of hGALK1 in the presence of galactose and compound **T7** belong to the orthorhombic spacegroup C2 2 2_1 (cell dimensions were a = 115.51 Å, b = 219.18 Å, c = 151.24 Å, α = β = γ = 90.00°) with four molecules in the asymmetric unit. The final refined structure, at 1.8 Å resolution, converged to Rfree of 0.242 and R of 0.216. Based on difference Fourier density, galactose was built into all four molecules in the asymmetric unit and compound **T7** was built into one molecule.

5.2.2 Structure of hGALK1 in complex with galactose and ADP

Overall, the hGALK1-galactose-ADP structure (purple ribbon, figure 5.6A) determined in this

work is very similar to the previously published hGALK1-galactose-AMPPNP structure ([24]; grey ribbon in figure 5.6A; RMSD of 0.54 Å over 389 Cα atoms), with similar coordination of the nucleotide and galactose ligands (section 1.4.3, figure 5.6C). This overall similarity suggests that hGALK1 catalysis is not accompanied by significant conformational changes, unlike the monomeric glucokinase (hexokinase IV, the only monomeric group 4 sugar kinase; illustrated in figure 1.8C) that is known to undergo significant rotation of the N and C terminal domains relative to each other during catalysis [275].

GALK1 catalysis is shown to proceed by an ordered kinetic mechanism whereby ATP binds first, stabilising the active site loops (L1, Ser79-Pro85; L2, Ser98-Arg105; L3, Val133-Ser140; and L4, Arg228-Glu235) to facilitate galactose binding [35]. These loops are similar in conformation between the hGALK1-galactose-ADP structure described in this work (purple) and the published hGALK1-galactose-AMPPNP structure (grey), suggesting that the conformational state associated with galactose binding is maintained after removal of the γ-phosphoryl group from ATP (figure 5.6B).

Closer inspection of active site residues reveals displacement of two residue sidechains (Arg105 and Arg228) in the hGALK1-galactose-ADP structure, relative to the hGALK1-galactose-AMPPNP structure (figure 5.6B-C). Molecular dynamics simulations have implicated these two residues in providing the necessary positive charge (oxyanion hole) to facilitate phosphoryl transfer from ATP to galactose (section 1.4.3; [254], [195]).

In the hGALK1-galactose-ADP structure, the zeta carbon of the Arg228 sidechain moves 4.9 Å at a 45° angle (figure 5.6C, red arrow), relative to the hGALK1-galactose-AMPPNP structure, to point towards where the β-to- γ phosphoryl bond of ATP would be (substituted with a nitrogen atom in AMPPNP). The guanidino group of the Arg228 sidechain is now within 3 Å of the ATP β-to-γ phosphoryl bond (figure 5.6C, asterisk), supporting the role of this residue in hGALK1 catalysis suggested by molecular dynamics.

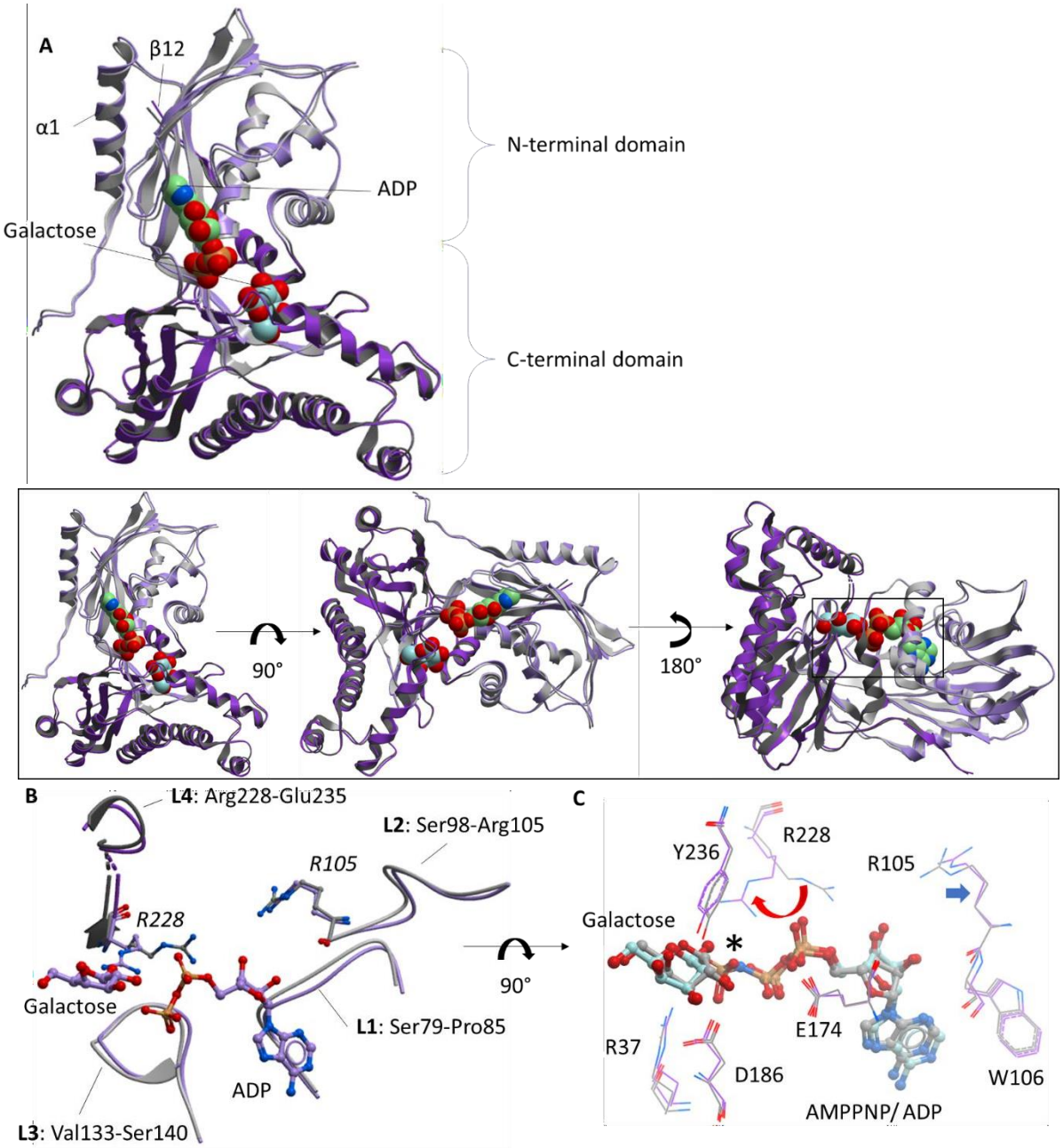


Figure 5.6: Comparison of hGALK1-galactose-AMPPNP and hGALK1-galactose-ADP crystal structures. *A. Ribbon diagram showing superposition of hGALK1-galactose-AMPPNP (PDB 1wuu, grey) and hGALK1-galactose-ADP (purple) structures. The C-terminal domains are in darker shades and the N-terminal domains in lighter shades. Ligands from the hGALK1-galactose-ADP structure are shown as spacefill with the carbon atoms of galactose and ADP in blue and green respectively. Inset: Ribbon diagrams illustrating orientation of panel B relative to panel A. B. Close-up view of hGALK1 active site showing ordered loops (L1 – L4) from the hGALK1-galactose-ADP (purple) and hGALK1- galactose-AMPPNP (grey) structures as ribbons. Galactose, ADP, Arg105 and Arg228 are shown as sticks. C. Close-up view of hGALK1 active site showing the position of catalytic residues in the hGALK1-galactose-ADP (purple lines) and hGALK1-galactose-AMPPNP (grey lines) structures. Galactose, AMPPNP and ADP are shown as sticks. The β-to-γ phosphoryl bond is marked with an asterisk. Key residues Arg105 and Arg228 are highlighted by blue and red arrows respectively.*

Secondly, molecular dynamics simulations indicated that the Arg105 sidechain moves in towards the bound ATP to form a hydrogen bond with the sidechain of Glu174 [195]. Conversely, the hGALK1-galactose-ADP structure shows displacement of the Arg105 guanidino group away from both ATP and Glu174. The Arg105 guanidino group is displaced by 1.3 Å so that it is now 5 Å from the ribose group, compared with 3.6 Å from the ribose group in the hGALK1-galactose-AMPPNP structure, and >10 Å from Glu174 (figure 5.6C, blue arrow). However, this structure represents a non-productive complex of one substrate (galactose) and one product (ADP) and may not reflect a true catalytic intermediate. Additional structures, particularly of the proposed 'open' apo form, prior to nucleotide binding and ordering of the active site loops L1-L4, are needed to fully elucidate the hGALK1 catalytic cycle [35], [254].

5.2.3 Structures of hGALK1 in complex with tool compounds **T1**, **T2**, **T3** and **T7**

The four tool compounds (**T1**, **T2**, **T3** and **T7**) that were co-crystallised with hGALK1 are very similar in chemical structure, bearing the backbone of a benzoxazole ring and a tetrahydroquinazolinone ring connected by an amide linker (figure 5.5A, table 5.2). Therefore, it was unsurprising that all four compounds were essentially superimposable in the crystal structures (figure 5.7A). The overall protein structures were also very similar to the hGALK1-galactose-AMPPNP (figure 5.7B, grey ribbon) and hGALK1- galactose-ADP (figure 5.7B, purple ribbon) structures with RMSDs from the hGALK1- galactose-ADP structure of 0.58, 0.90, 0.91 and 0.87 Å (over 389 Cα atoms) for hGALK1- galactose-**T1** (blue), hGALK1-galactose-**T2** (green), hGALK1-galactose-**T3** (yellow) and hGALK1-galactose-**T7** (pink) structures respectively (figure 5.7B).

The benzoxazole ring common to these compounds occupies a similar, though surprisingly not superimposable, position to the adenine moiety of ADP and AMPPNP and the amide linker and dihydro-pyrimidine constituents of the tool compounds occupy a similar position to the ribose moiety of ADP and AMPPNP (figure 5.7A, inset). While the phosphate groups of ADP and AMPPNP curve to interact with active site residues Arg37, Glu174, Asp186 and Tyr236, the

quinazolinone ring of the tool compounds is flat and rigid and thus points out towards the solvent.

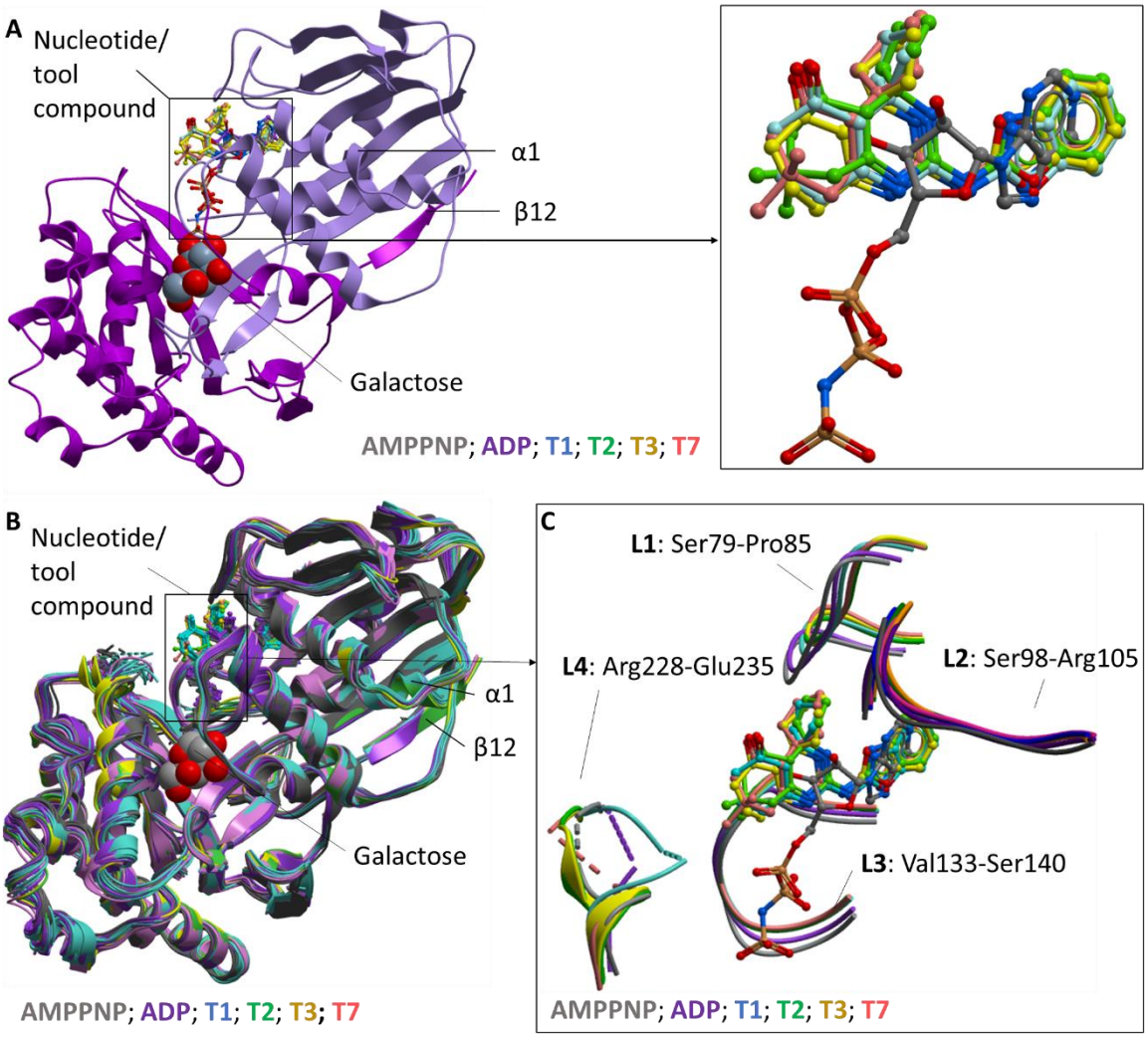


Figure 5.7: Comparison of hGALK1 structures in complex with ADP, AMPPNP or spiro-benzoxazole tool compounds. *A. Left: Ribbon diagram of hGALK1-galactose-ADP showing the position of all active site ligands relative to the overall hGALK1 monomer. Galactose is shown as grey spacefill. The N-terminal domain is lilac, and the C-terminal domain is purple. Inset: Close-up view of superimposed active site ligands, shown as sticks: AMPPNP (grey) and compounds **T1** (blue), **T2** (green), **T3** (yellow) and **T7** (pink). ADP was omitted for clarity. B. Superimposition of all chains in the asymmetric unit from hGALK1-galactose-AMPPNP (4 chains, grey), hGALK1-galactose-ADP (8 chains, purple), hGALK1-galactose-**T1** (8 chains, blue), hGALK1-galactose-**T2** (2 chains, green), hGALK1-galactose-**T3** (2 chains, yellow) and hGALK1- galactose-**T7** (4 chains, pink) structures. Galactose is shown as spacefill and AMPPNP, ADP and tool compounds are shown as sticks. The viewpoint shown in panel C is highlighted by a black rectangle. C. Close-up view of hGALK1 active site showing the position of loops L1 – L4 in the presence of AMPPNP (grey), ADP (purple), **T1** (blue), **T2** (green), **T3** (yellow) and **T7** (pink). For clarity, only one copy of each ligand is shown and galactose, though present in all structures, was omitted from close-up views.*

As mentioned, GALK1 protein chains from all solved structures were essentially superimposable indicating minimal conformational change, relative to nucleotide bound structures, upon binding of tool compounds (figure 5.7B). Similarly, analysis of the four loops stabilised by ATP binding (L1 – L4) showed that L2, closest to the adenine moiety of the nucleotides and the benzoxazole

moiety of tool compounds, superimposed well in all structures whereas the other 3 loops showed minor displacement upon ligand binding (figure 5.7C). L1, closest to the quinazolinone moieties of the tool compounds, is displaced slightly by the spiro- ring constituent of the tool compounds (figure 5.7C, L1).

L3, closest to the phosphate binding region of the pocket, is superimposable in all 4 structures containing a tool compound but is slightly closer to the phosphate binding region in these structures than in the hGALK1-galactose-AMPPNP and hGALK1-galactose-ADP structures (figure 5.7C, L3). L4, closest to the carbonyl end of the quinazolinone ring, shows considerable flexibility in the hGALK1-galactose-AMPPNP (grey), hGALK1-galactose-ADP (purple), hGALK1-galactose-**T1** (blue) and hGALK1-galactose-**T7** (pink) structures but is more ordered, with a partially helical structure, in the hGALK1-galactose-**T2** (green) and hGALK1-galactose-**T3** (yellow) structures (figure 5.7C, L4).

Closer inspection of these structures reveals conserved interactions between the protein and the shared backbone of the tool compounds. The benzoxazole ring is coordinated by π-π stacking with Trp106 and weak Van der Waals interactions with Val129 and Leu145 in the structures with compound **T1** (figure 5.8A), compound **T2** (figure 5.8B), compound **T3** (figure 5.8C) and compound **T7** (figure 5.8D). The pyrimidine moiety of the quinazolinone forms weak Van der Waals interactions with Leu135 and the cyclohexanone moiety displaces the sidechain of Arg228 (figure 5.8, red arrows). In the tool compounds with the larger spiro ring, the carbonyl moiety of the tool compound is close enough to hydrogen bond with the guanidino group of the Arg105 sidechain (compound **T3**, figure 5.8C and compound **T7**, figure 5.8D; blue arrows). The dimethyl groups of compounds **T2** and **T7** do not appear to contribute to binding to hGALK1 (figure 5.8B, D).

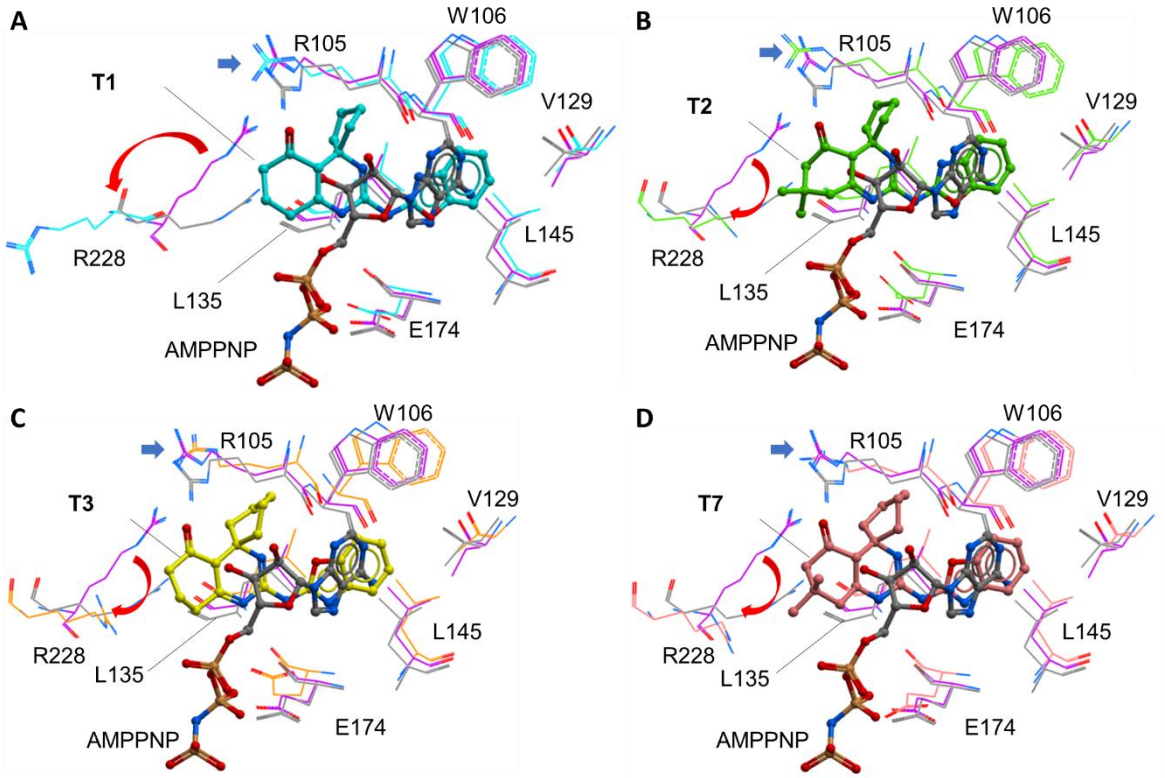


Figure 5.8: Close-up views of the hGALK1 ATP binding site with bound tool compounds. *Superposition of hGALK1-galactose-AMPPNP (grey) and hGALK1-galactose-ADP (purple) structures with A. hGALK1- galactose-**T1** (blue); B. hGALK1-galactose-**T2** (green); C. hGALK1-galactose-**T3** (yellow); and D. hGALK1-galactose-**T7** (pink). Structures are oriented the same as in figure 5.7A. Key residues that interact with, or are displaced by, the tool compounds are shown as lines. AMPPNP and tool compounds are shown as sticks. ADP and galactose have been omitted for clarity. Key residues Arg105 and Arg228 are highlighted by blue and red arrows respectively.*

5.3 Fragment screening campaign: from 180 structures to 3 binding pockets

As well as providing insight into their inhibition of hGALK1 through competition with ATP binding (explored further in chapter 6), co-crystallisation with tool compounds generated a robust crystal system that was suitable for fragment screening. For example, through setting up a 96-well plate, approximately 40 hGALK1-galactose-**T3** crystals, consistently diffracting below 2.5 Å, could be obtained within a week using the conditions described earlier (section 5.1).

5.3.1 Preparing hGALK1 crystals for fragment soaking and data collection

Optimised purification and crystallisation (section 5.1) parameters were used to grow approximately 300 crystals of hGALK1 for a fragment screening campaign (figure 5.9). These crystals were of the hGALK1-galactose-**T3** complex as both ligands were found to be necessary to achieve consistent atomic level resolution.

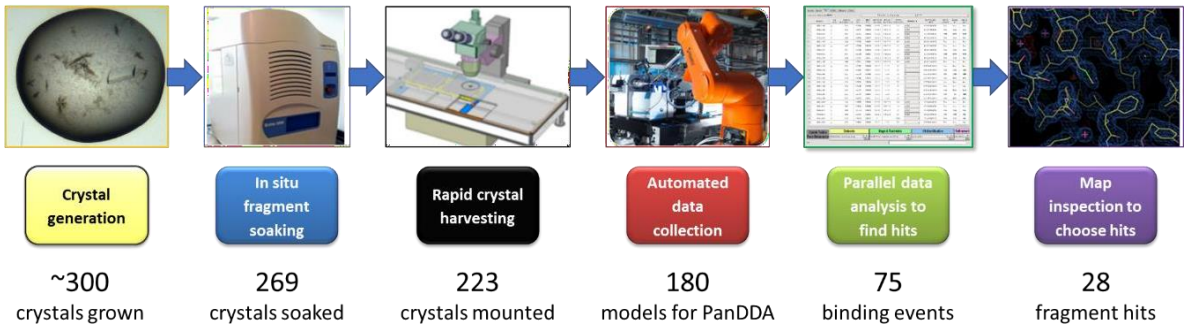


Figure 5.9: Overview of the fragment screening campaign against hGALK1. *Approximately 300 hGALK1-galactose-**T3** crystals were grown from 10 crystallisation plates. One fragment each was added at 20% final DMSO to each of 269 hGALK1-galactose-**T3** crystals by ECHO acoustic liquid dispensing. After an hour's incubation at room temperature, 223 crystals were successfully mounted using the Shifter crystal handling system. Automated data collection and integration yielded 180 models of suitable resolution and unit cell dimensions for PanDDA analysis. Inspection of PanDDA maps allowed placement of 75 fragments into binding event maps. Analysis of fragment-bound structures led to the selection of 28 fragment hits not resulting from artefactual or nonspecific binding.*

Crystallisation plates were set up from a protein solution containing 25 mg/mL hGALK1, purified in sodium phosphate (pH 7) buffer supplemented with 50 mM galactose, and 2.25 mM compound **T3** (final NMP concentration of 5%) and reservoir solutions forming a fine screen around the Morpheus-G08 condition. For each plate, a robotic liquid dispenser was used to set up 3 sitting drops of different protein: reservoir ratios around each of 96 reservoir solutions. As discussed in section 5.1.4, this crystal form had the additional advantage of favouring the identification of non-orthosteric ligands that could lead to novel, selective inhibitors of hGALK1.

In total, approximately 300 cuboid crystals of 150 – 200 µm size, from 10 crystallisation plates, were selected for fragment soaking. Crystals were soaked and mounted as described for the hHAO1 fragment screening campaign (section 3.3.1). In total, 223 out of 269 fragment-soaked crystals were successfully mounted. Of the 46 crystals that were soaked but not mounted, the majority (n=31) were melted after incubation with the fragment, whereas the remainder (n=15) were unsuccessful due to user error (loops broken or empty after mounting). Data collection was performed using the automated, unattended mode at the Diamond Light Source i04-1 beamline, with X- ray centring and the BART crystal loading robot facilitating rapid data collection. For 12 of the mounted crystals, automated X-ray centring failed, and no dataset could be collected. Thus, a total of 211 datasets were collected for high throughput, parallel processing.

5.3.2 Processing and overall analysis of the fragment data sets

All datasets were processed in parallel via the XChemExplorer workflow as described for hHAO1 (section 3.3). The search model used for molecular replacement was the hGALK1-galactose-**T3** structure described earlier in this chapter (section 5.2.3). Twelve datasets were excluded from PanDDA analysis based on their space group (7 datasets in P1 and 5 in P 1 2 1) and 19 were excluded due to low resolution (> 2.5 Å). Consequently, PanDDA analysis was performed on 180 datasets belonging to the monoclinic space group $P2_1$ (cell dimensions were a = 72-74 Å, b = 113-115 Å, C = 120-121 Å, $\alpha = 90°$, $\beta = 100°$, and $\gamma = 90°$) with 4 copies of hGALK1 per asymmetric unit and a resolution range of 1.8 to 2.5 Å. Data processing and structure refinement statistics are provided in table S2.2.

One benefit of having 4 protein chains – that is 4 hGALK1 monomers – within the asymmetric unit was that the same fragment, binding to the same pocket, could yield up to 4 binding events from a single dataset i.e. binding of fragment molecules to different protein chains are identified as separate binding events by the PanDDA software. PanDDA analysis of the 180 comparable hGALK1 models identified 286 binding events - weak electron densities not present in the ground state map – across 108 datasets. After visual inspection, fragments could be confidently built into 75 densities, across 39 datasets. Of these 39 datasets, fragments could be built into one hGALK1 chain of 16 datasets, two hGALK1 chains of 14 datasets, three hGALK1 chains of 5 datasets and four hGALK1 chains of 4 datasets.

As mentioned, each fragment-bound structure consisted of four copies of the hGALK1 monomer, referred to as chains A-D, as observed in the hGALK1-galactose-AMPPNP (PDB 1wuu, [24]) and hGALK1-galactose-**T7** (section 5.2.3) structures. Figure 5.10A shows the superimposition of the hGALK1-galactose-**T3** structure (grey ribbon) with the 4 protein chains of the fragment-bound structures.

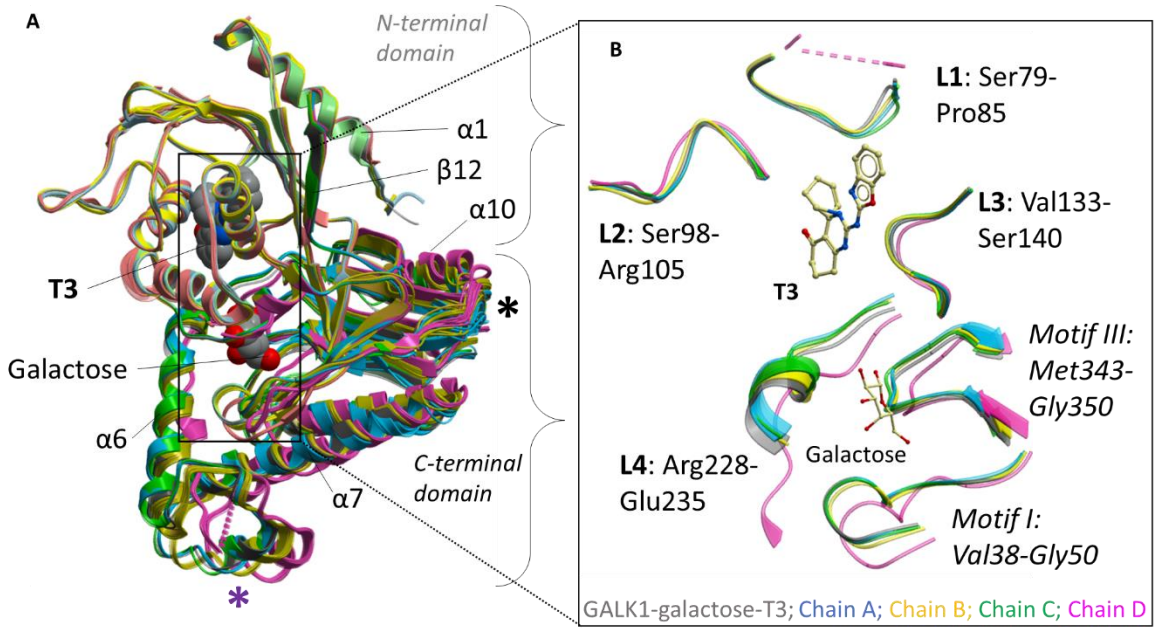


Figure 5.10: Comparison of protein chains A-D in fragment-bound structures and the hGALK1-galactose-T3 structure. *A. Ribbon diagram showing superimposition of selected chains A-D from fragment-bound structures with the hGALK1-galactose-**T3** structure (grey). Ribbon structures of the fragment-bound structures are coloured by the protein chain - blue for chain A, yellow for B, green for C and pink for D. The N-terminal domain of each chain is coloured in a lighter shade and the C-terminal domain in a darker shade. Galactose and compound **T3** are shown as grey spacefill. Asterisks highlight key loop regions that demonstrate chain-specific conformations, as described in the text. B. Close-up view of the active site loops in each chain of a representative fragment-bound structure superimposed with the hGALK1-galactose-**T3** structure. Loops in the fragment-bound structures are coloured by the protein chain and loops in the hGALK1-galactose-**T3** structure are grey. Galactose and compound **T3**, as positioned in the hGALK1-galactose-**T3** structure, are shown as sticks.*

Chains A, B and C (blue, yellow and green ribbons respectively in figure 5.10) superimpose well with the hGALK1- galactose-**T3** structure (RMSD of 0.81 Å over 385 Cα) and with each other (RMSD of 0.73 Å over 385 Cα), however chain D (pink ribbon in figure 5.10) is somewhat different from both chains A-C of the fragment bound structures and from the hGALK1-galactose-**T3** structure. Specifically, chain D of the fragment-bound datasets had RMSD values of 2.37 Å, 2.37 Å and 2.10 Å over 385 Cα, when compared to chains A, B and C respectively within the same dataset, and an RMSD of 2.25 Å over 385 Cα when compared with the hGALK1-galactose-**T3** structure. Comparison of the overall structures reveals the key differences in chain D relative to the others are displacement of helices α6, α7 and α10 and the loops between α6 and α7 (purple asterisk, figure 5.10A) and between β7 and β8 (black asterisk, figure 5.10A).

Comparison of the four loops that stabilise ATP binding (figure 5.10B, previously detailed in

section 5.2.2) show similar conformations in fragment-bound structures for L2 (Ser98-Arg105) and L3 (Val133- Ser140). However, the two loops predicted by molecular dynamics to be most involved in ATP- binding, L1 (Ser79-Pro85) and L4 (Arg228-Glu235), are significantly different in chain D relative to their structure in both the other chains of fragment-bound structures and in the hGALK1-galactose-**T3** structure. In chain D of the fragment-bound structures, no density was observed for galactose, compound **T3** or loop L1. Additionally, loop L4 no longer has the helical structure seen in chains A-C of fragment-bound structures and other structures described in this work. Additionally, the two key loops involved in the binding of galactose (Val38-Gly50 and Met343-Gly350), containing the conserved motifs I and II respectively, are displaced away from the galactose binding site in chain D of the fragment-bound structures, supporting molecular dynamics simulations of ATP-mediated conformational changes that increase the efficacy of galactose binding.

Of the 39 datasets containing one or more bound fragments, 11 datasets were excluded from further analysis as they represent various aspects of artefactual or nonspecific ligand binding. These include four datasets in which fragments bind at the crystallographic dimer interface, indicative of weak starting points to develop inhibitors of the hGALK1 monomer (figure 5.11A).

An additional three fragments displayed high promiscuity, binding to multiple pockets at the protein surface and were excluded as poor starting points for developing specific inhibitors of hGALK1 (figure 5.11B). The last four datasets were excluded because the fragments make few interactions with the protein, instead binding as a result of packing of protein molecules within the crystal (figure 5.11C).

After excluding the 11 hits resulting from artefactual or nonspecific binding, the remaining 28 fragments are found to cluster into three different pockets of the hGALK1 protein (figure 5.12). Binding events were observed in all 4 chains of the asymmetric unit and similar binding poses were observed when a fragment hit was observed binding in more than one chain. Two fragments (fragments **1** and **2**) bind at the entrance to the ATP binding site, near where compound **T3** is

bound. Three fragments bind close to where the N and C termini of the hGALK1 monomer meet (pocket at the protein termini; fragments **3-5**). The remaining 23 fragments occupy a binding hotspot, distal from the active site, and can be divided into upper (fragments **6-21**) and lower (fragments **22-28**) clusters.

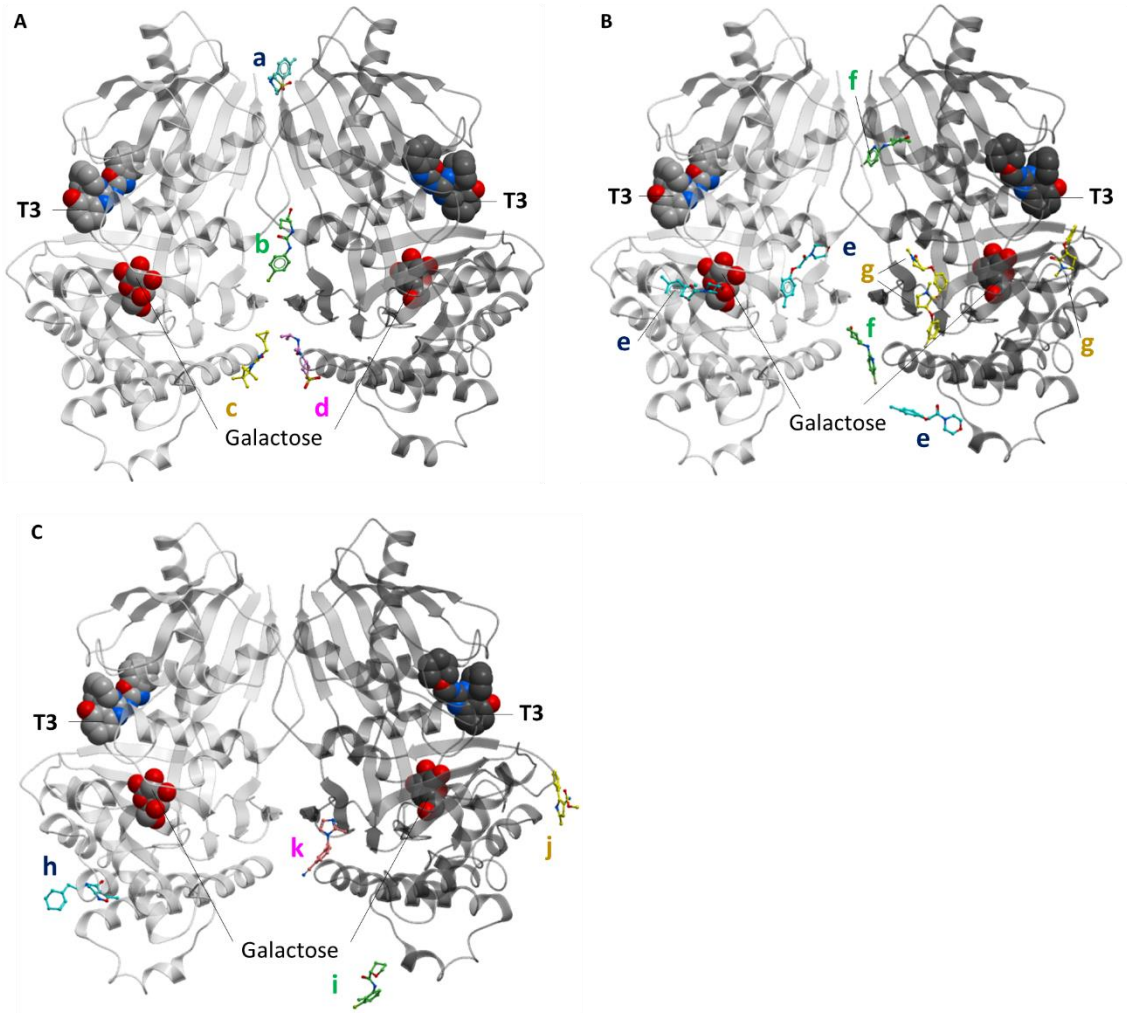


Figure 5.11: Fragment hits binding to hGALK1 that were excluded from further analysis. *A. Ribbon diagram of the hGALK1 crystallographic dimer showing the position of fragments (**a-d**) that bind at the crystallographic dimer interface. Galactose and compound **T3** are shown as spacefill. B. Ribbon diagram of the hGALK1 crystallographic dimer showing the position of fragments (**e-g**) each with multiple binding locations. C. Ribbon diagram of the hGALK1 crystallographic dimer showing the position of poorly coordinated fragments (**h-k**).*

The following sections provide a more detailed analysis of fragment binding events at the entrance to the ATP binding site (section 5.3.3), within the pocket at the protein termini (section 5.3.4) and at the binding hotspot (section 5.3.5).

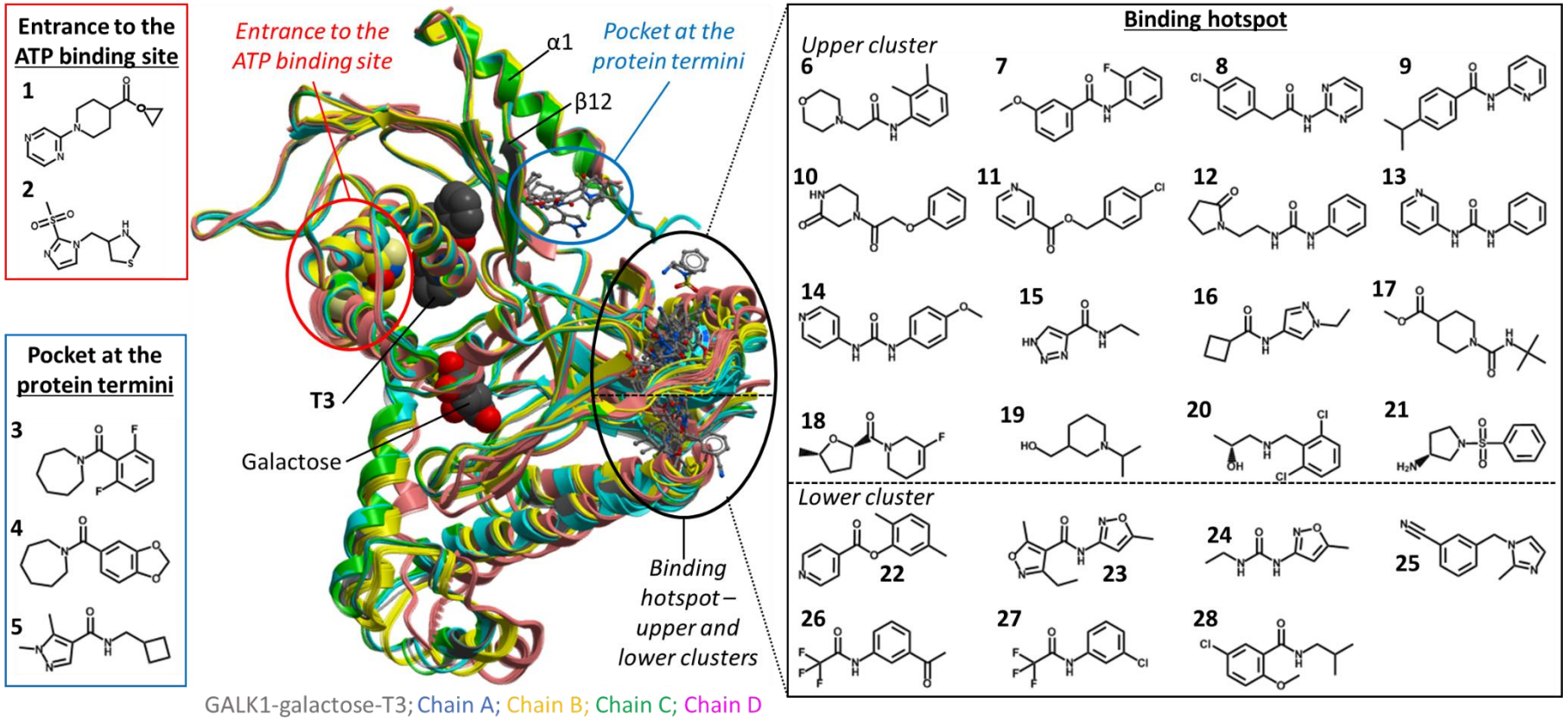


Figure 5.12: Fragment chemical structures and their binding sites identified in a screening campaign against hGALK1-galactose-T3 crystals. *The central ribbon diagram shows the superimposition of the 28 fragment-bound structures. Ribbons are coloured according to the protein chain – A (blue), B (yellow), C (green) and D (pink). Galactose and compound* ***T3*** *are shown as grey spacefill. Fragments are bound in 3 areas: at the entrance to the ATP binding site (red circle); in the pocket at the protein termini (blue circle) and at the binding hotspot (black circle) containing the upper and lower clusters. Fragments at the entrance to the ATP binding site are shown as spacefill to improve visibility and all other fragments are shown as grey sticks.*

5.3.3 Fragments **1** and **2** bind at the entrance to the ATP binding site

The two fragments (**1, 2**) at the entrance to the ATP binding pocket (figure 5.13A) are positioned within 5 Å of the compound **T3** carbonyl group (figure 5.13B). The two fragments do not share a common scaffold (figure 5.12, red box) or binding mode (figure 5.13B). It is of note that the presence of **T3** (or a similar ATP-mimetic) may be a prerequisite for the binding of these fragments. Both fragments interact with compound **T3**, Tyr109, Ala178 and Gly179.

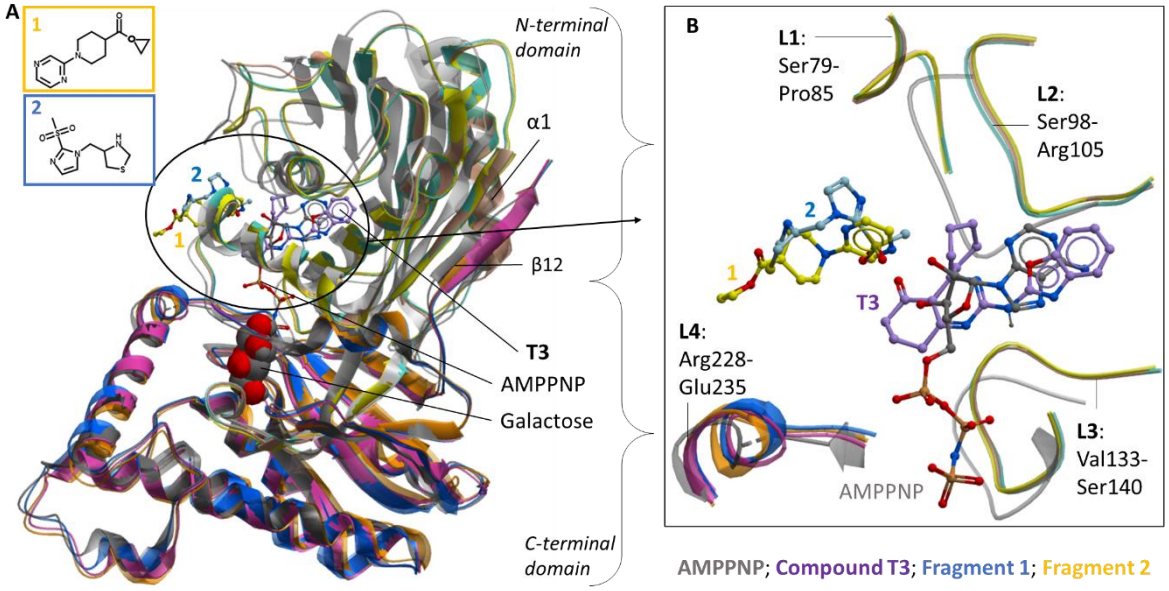


Figure 5.13: Two fragments bind close to T3, at the entrance to the ATP binding site. *A. Superimposition of hGALK1-galactose-AMPPNP (grey), hGALK1-galactose-**T3** (pink), hGALK1-galactose-**T3**-fragment **1** (yellow) and hGALK1-galactose-**T3**-fragment **2** (blue) structures. Galactose is shown as grey spacefill. AMPPNP, compound **T3** and fragments are shown as sticks. B. Close-up view showing the position of loops L1 – L4 at the ATP binding site in the presence of AMPPNP (grey), **T3** (lilac), fragment **1** (yellow) and fragment **2** (blue). C. Close-up view of fragment **1** and its coordinating residues in the hGALK1-galactose-**T3**-fragment **1** (yellow) structure relative to the hGALK1-galactose-AMPPNP (grey) and hGALK1-galactose-**T3** (purple) structures. Protein residues are shown as lines and ligands are shown as sticks. D. Close-up view of fragment **2** and its coordinating residues in the hGALK1-galactose-**T3**-fragment **2** (blue) structure relative to the hGALK1-galactose-AMPPNP (grey) and hGALK1-galactose-**T3** (purple) structures. AMPPNP and galactose have been omitted for clarity.*

Specifically, the N-6 amine of the pyrazine ring of fragment **1** is positioned to hydrogen bond with either the carbonyl group of compound **T3** or the sidechain hydroxyl of Tyr109 while the N-2 amine of the fragment hydrogen bonds with the backbone carboxyl of Ala178 (figure 5.13C). The piperidine ring of fragment **1** clashes with the conformation of the Arg105 sidechain in the hGALK1-galactose-**T3** structure and this sidechain is not visible in the hGALK1- galactose-**T3**-fragment **1** structure. Finally, the carbonyl and oxirane oxygen atoms are positioned to

hydrogen bond with Gly179. The sulfone group of fragment **2** is positioned to hydrogen bond with the sidechain amine groups of Arg105 and the sidechain hydroxyl of Tyr109 (figure 5.13D). The imidazole group of the fragment induces a minor displacement of the Arg105 sidechain, to avoid a steric clash, and also hydrogen bonds with the backbone carboxyl group of Ala178. Finally, the thiazolidine group of the fragment hydrogen bonds with the backbone amine groups of Ala178 and Gly179. The sidechain of Arg228, an essential active site residue, is not visible when fragments are bound, though the reason for the lack of density observed is not apparent from the examined structures.

5.3.4 Fragments **3-5** bind close to where the protein termini converge

Three fragments (**3- 5**) were identified in a hydrophobic pocket behind the galactose binding site, close to where the N and C termini of hGALK1 converge (figure 5.14A). This pocket is formed by β1 (Leu30-Ile41) and β2 (Leu51-Arg68), flanked on one side by β12 and the preceding (β11-β12) loop (Ala383-Leu390) and on the other side by α1 (Val9-Phe24), which is displaced slightly upon fragment binding (figure 5.14B; black arrows in figure 5.14D- F). The pocket is lined with hydrophobic residues that form weak interactions with the bound fragments but there are a few charged residues proximal to it that could form favourable interactions with polar groups of slightly larger ligands (blue and red surface representations in figure 5.14C indicate positively and negatively charged residues respectively).

The Val32Met variant of hGALK1, reported in galactosaemia type 2 patients, is inactive in patient red blood cell assays and insoluble in *E.coli*, indicating severely compromised hGALK1 structure and function [35], [184]. The backbone of Val32 is within 4 Å of all three fragments and hence fragment growth to further disrupt this residue may be a promising approach for hit development (red asterisks, figure 5.14D-F).

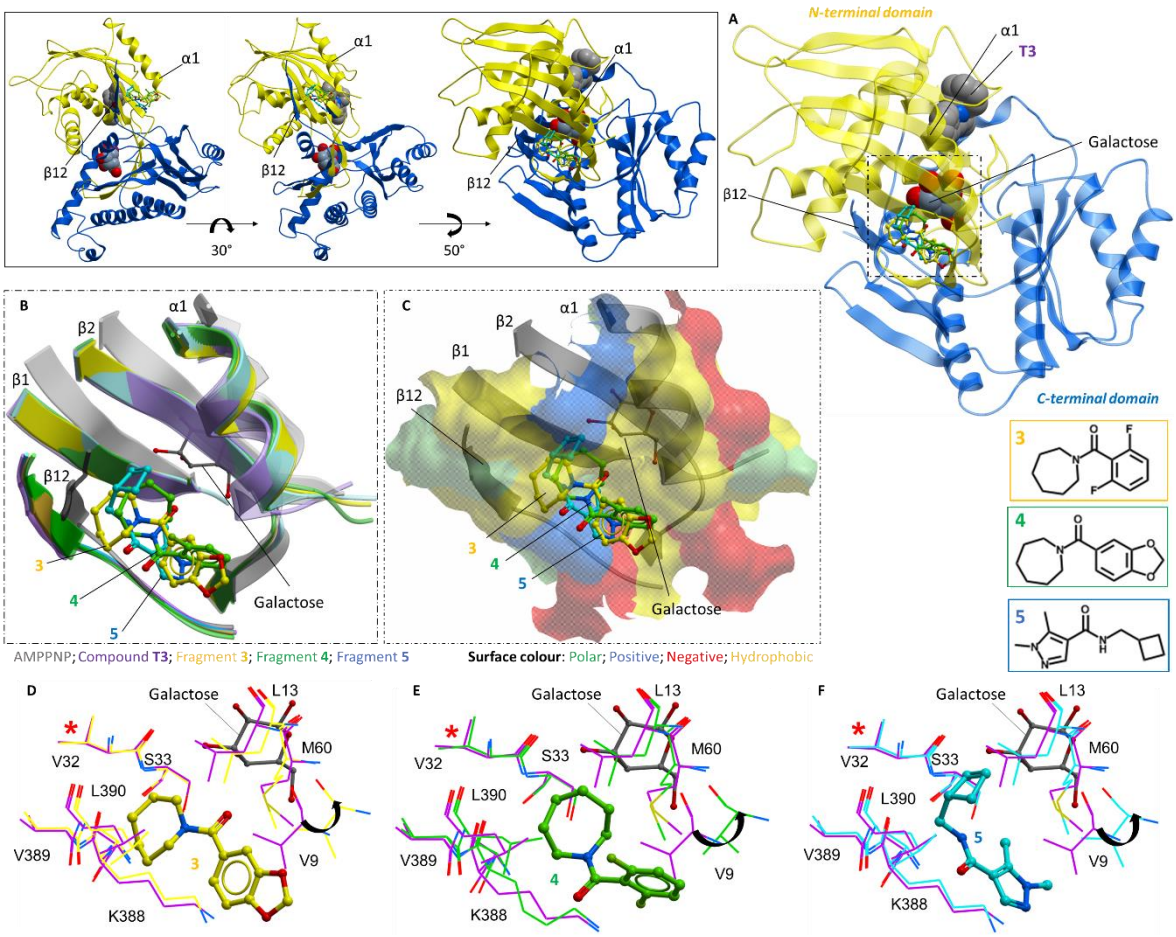


Figure 5.14: Three fragments bind in a pocket close to where the C and N termini of hGALK1 converge. *Inset: Ribbon diagram showing the rotation function applied to hGALK1 ribbon structure, relative to the default view (far left) used in earlier figures, to obtain the clearest view of the protein termini pocket (far right). A. Ribbon diagram of a representative fragment-bound structure illustrating the position of the protein termini pocket with respect to the N-terminal domain (yellow) and the C-terminal domain (blue) of hGALK1. Galactose and compound **T3** are shown in spacefill. Fragments are shown as sticks. The dashed box highlights the area of the protein shown in panels B and C. B. Close-up view of the binding pocket for fragments **3** – **5**, illustrating nearby secondary structure features. Ribbon representation is for superimposed hGALK1-galactose-AMPPNP (grey), hGALK1-galactose-**T3** (purple), hGALK1-galactose-**T3**-fragment **3** (yellow), hGALK1-galactose-**T3**-fragment **4** (green) and hGALK1-galactose-**T3**-fragment **5** (blue) structures. Galactose and fragments are shown as sticks. C. Surface representation of the protein termini pocket, illustrating positions of hydrophobic (yellow), polar (green), positively charged (blue) and negatively charged (red) residues. Galactose and fragments are shown as sticks. Secondary structures of this pocket are shown in grey ribbon. Inset: Chemical structures of fragments **3** (yellow box), **4** (green box) and **5** (blue box). D. Close-up view of fragment **3** and its coordinating residues in the hGALK1-galactose-**T3**-fragment **3** structure (yellow) relative to the hGALK1-galactose-**T3** structure (purple). E. Close-up view of fragment **4** and its coordinating residues in the hGALK1-galactose-**T3**-fragment **4** structure (green) relative to the hGALK1-galactose-**T3** structure (purple). F. Close-up view of fragment **5** and its coordinating residues in the hGALK1-galactose-**T3**-fragment **5** structure (blue) relative to the hGALK1-galactose-**T3** (purple) structure. In panels D-F, the black arrow indicates displacement of Val9, and the red asterisk highlights the position of residue Val32.*

Fragments **3** and **4** contain an azepane ring followed by a carbonyl linked to an aromatic ring, with the latter being a benzodioxole in **3** and a di-fluorobenzene in **4** (figure 5.14, inset). In both cases, the azepane ring is coordinated by weak van der Waals interactions with the sidechains of Leu13, Ser33, Met60, Lys388, Val389 and Leu390 (figure 5.14D and E for **3** and **4** respectively). The azepane nitrogen of fragment **3** is close enough to hydrogen bond with the sidechain hydroxyl of Ser33 while the carbonyl group of this fragment points toward the hydrophobic sidechains of Val9 and Met60. The azepane nitrogen of fragment **4** is similarly placed to hydrogen bond with the Ser33 hydroxyl, however the carbonyl group of this fragment points in the opposite direction to form a hydrogen bond with the sidechain amine of Lys388. The di-fluorobenzene of **4** and the benzodioxole of **3** can interact with sidechains of Val9, Leu13, Leu390; the former makes more extensive van der Waals interactions than the latter. The oxygen constituents of **3** are not close enough to hydrogen bond with nearby residues.

Fragment **5** is also an aliphatic amine and forms somewhat similar interactions with hGALK1 to fragments **3** and **4** (figure 5.14F). Specifically, fragment **5** consists of a cyclobutane ring rather than an azepane ring, which is connected to a dimethyl-pyrazole by a carboxamide linker (figure 5.14, inset). The cyclobutane ring and carboxamide linker are coordinated by similar residues to the azepane ring of fragments **2** and **3** (Leu13, Val389 and Leu390; figure 5.14D-E). Additionally, the amide of the linker is positioned to hydrogen bond with the hydroxyl of Ser33. The dimethyl- substituted pyrazole, occupying a similar position to the benzene ring of fragments **2** and **3**, is less well-coordinated, interacting only with the sidechains of Lys388 and Met60.

5.3.5 Fragments **6-28** bind in a binding hotspot distal to the active site

Twenty-three fragments clustered in a binding hotspot, on the opposite side of the protein to the substrate binding site (figure 5.15A). As illustrated in figure 5.15B, one side of the binding hotspot (left in this figure) is formed by β6 (His197-Asp202), the α5-β6 loop (Gly193-Gly196) and β7 (Thr208-Val211) and the other side (right in this figure) is formed by sections of β8 (Ala219-Ile222), β9 (Val338-Gly340), β10 (Val353-Glu357) and β11 (Phe378-Ser381). The

back of the pocket is formed by sections of β1 (Val38-Ile41), β2 (Val52-Glu58) and the β11-β12 loop (Gln382-Asp385) and the β7- β8 loop (Pro212-Leu218) runs in front of the binding pocket. The base of the binding hotspot is a U-shape formed by the C-terminal half of α8 (Thr288-Leu295), the N-terminal half of α9 (Arg301-Met307) and the turn connecting these two helices (Arg296-Tyr300) (figure 5.15B).

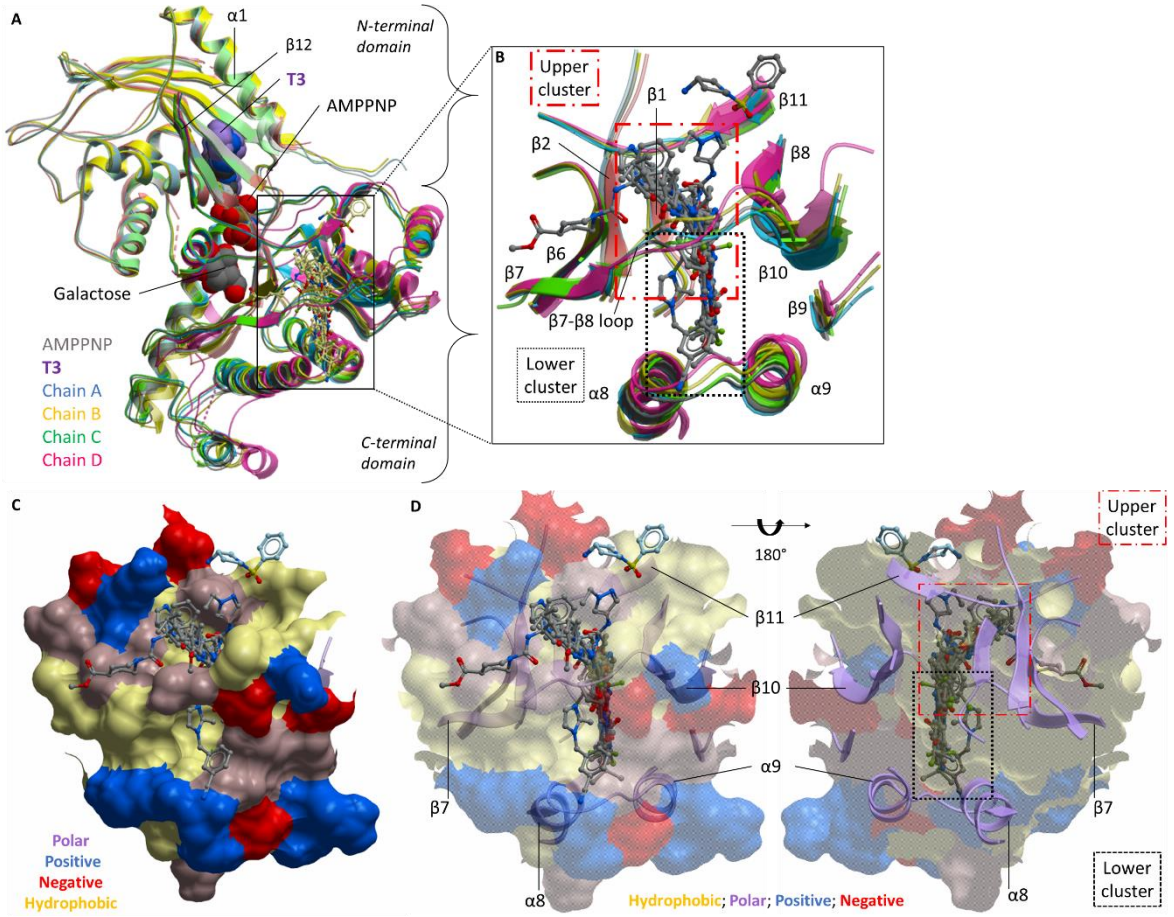


Figure 5.15: Twenty-two fragments bind at a binding hotspot distal to the active site. *A. Ribbon diagram showing superimposition of hGALK1-galactose-AMPPNP (grey) and hGALK1-galactose-**T3** (purple) with representative chain A (blue), chain B (yellow), chain C (green) and chain D (pink) structures of hGALK1 bound to hotspot fragments. All binding hotspot fragments are shown as sticks. AMPPMP (light grey), galactose (dark grey) and compound **T3** (purple) are shown in spacefill. B. Close-up view of the binding hotspot, illustrating nearby secondary structures. The upper cluster is outlined in red and the lower cluster in black. C. Close-up surface representation of the binding hotspot illustrating positions of hydrophobic (light yellow), polar (lilac), positively charged (blue) and negatively charged (red) residues. Fragments are shown as sticks. D. Close-up transparent surface representation of the front (left panel) and back (right panel) view of the binding hotspot, illustrating positions of hydrophobic (light yellow), polar (lilac), positively charged (blue) and negatively charged (red) residues within the binding hotspot. Fragments are shown as sticks. Secondary structure features are shown in lilac ribbon. The upper cluster is outlined in red and the lower cluster in black.*

The binding hotspot is a primarily hydrophobic pocket, partially protected from the solvent by a

polar surface loop (the β7- β8 loop, Pro212-Leu218) and sandwiched by charged residues above (Glu58, Lys195, Asp385) and below (Arg296, Arg297, Asp299, Arg301 and Arg305) (figure 5.15C-D). Fragment hits can be grouped according to whether they extend towards the charged residues above (upper cluster) or below (lower cluster) the hydrophobic core of the pocket. This hydrophobic binding region is the most buried part of the pocket, partially shielded from the solvent by a polar loop (the β7- β8 loop, Pro212-Leu218).

As illustrated in figure 5.16A, the most prevalent scaffold for hotspot-binding fragments, shared in 13 fragments, was two ring systems, one of which was a phenyl ring, separated by 1, 2, 3 or 5 bonds (fragments **6-14**, **18**, **21, 22**, **25**). Three fragments were excluded from further structural grouping because they are not well buried in the binding hotspot but rather interacted weakly with surface exposed residues (figure 5.16B; fragments **17**, **21**, **25**). In all other fragments, one of the constituent rings occupied the centre of the binding hotspot, coordinated by the hydrophobic residues of β1 and β8.This ring is referred to as the 'buried' ring in this work as it is located behind the β7- β8 loop (figure 5.16C).

The remaining 25 fragments were classified into four groups based on their binding pose relative to the centre of the binding hotspot (figure 5.16C-D). Group 1 fragments are the most buried of the upper cluster (fragments **8**, **11**, **12** and **14**; blue lines in figure 5.16C-D). Group 2 fragments are slightly less buried and generally shorter than the group 1 fragments (fragments **6**, **7**, **9**, **10**, **13**, **19** and **20**; green lines in figure 5.16C-D). Group 3 fragments do not share a binding pose and are less buried in the binding hotspot than the other groups (fragments **15**, **16** and **18**; lilac lines in figure 5.16C-D). Group 4 includes all the lower cluster of fragments (fragments **22-24** and **26-28**; yellow lines in figure 5.16C-D).

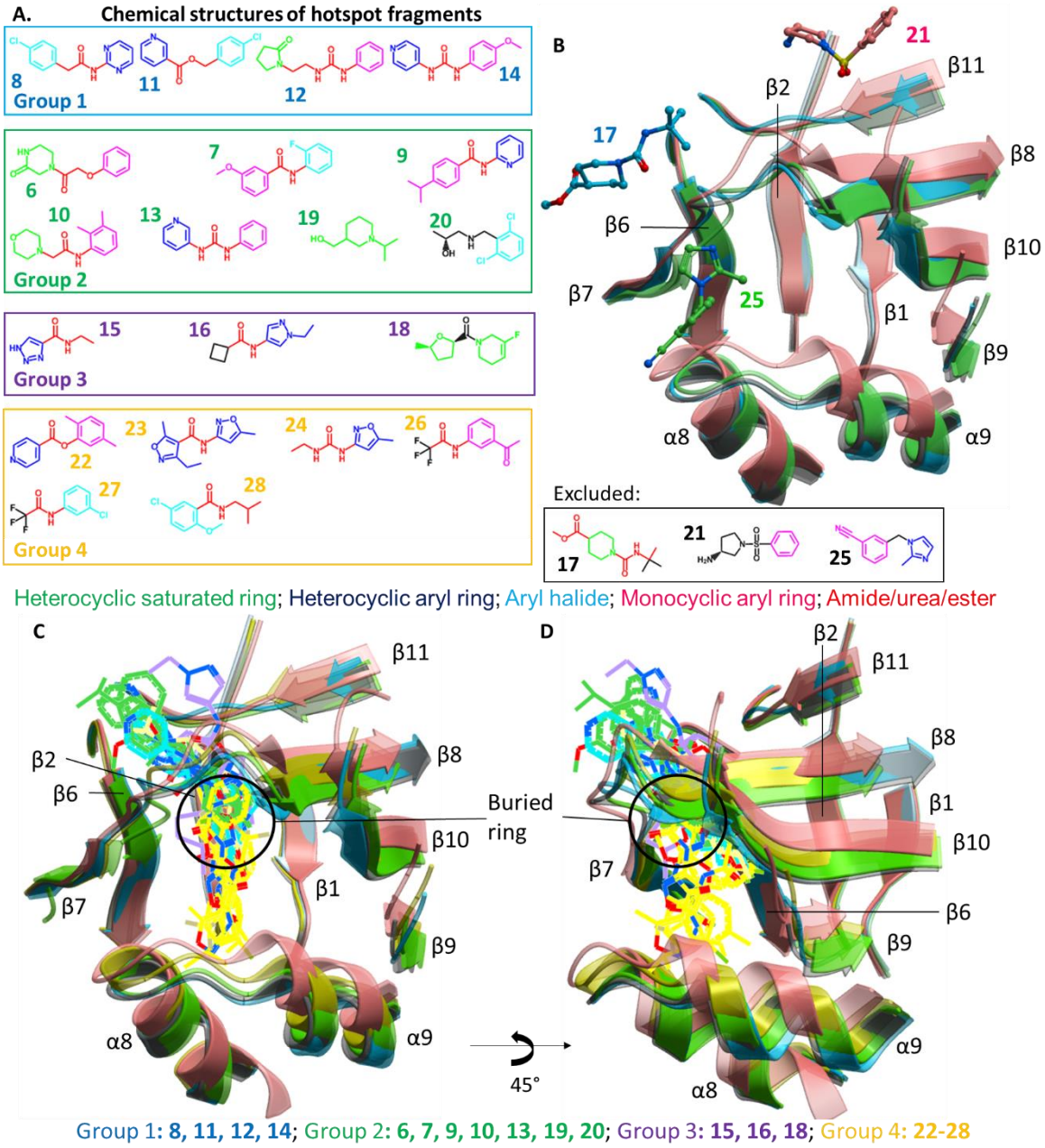


Figure 5.16: Fragments in the binding hotspot share common structural motifs and can be grouped to four binding poses. *A. Chemical structures of hotspot fragments and their structural motifs. Structures are coloured by shared motifs and grouped according to the observed fragment binding pose.*
*B. Ribbon diagram of the binding hotspot pocket, shown in the same orientation as figure 5.15, showing the location of three fragments (sticks) that were excluded from further analysis based on their lack of protein interactions. To account for the variability observed between protein chains, this ribbon diagram shows the superimposition of hGALK1-galactose-**T3** (grey) with a representative A (blue), B (yellow), C (green) and D (pink) protein chain from hGALK1 structures with bound hotspot fragments. C-D. Ribbon diagram of binding hotspot in 2 views, superimposing hGALK1-galactose-**T3** (grey) with a representative A (blue), B (yellow), C (green) and D (pink) protein chain from hGALK1 structures with bound hotspot fragments, to show the position of fragments within the hotspot. Fragments with similar binding poses are shown as lines of the same colour, forming a total of four groups. Galactose and **T3** are distal to the binding hotspot and have been omitted for clarity.*

Three of the four fragments in group 1 contain a substituted phenyl ring (chlorobenzene in **8** and **11** and methoxybenzene in **14**), bound to the centre of the hotspot, followed by a 3-carbon amide (**8**), ester (**11**) or urea (**14**) chain sandwiched between β7 and β8, and finally a pyridine (**11** and **14**) or pyrimidine (**8**) ring positioned above the β7- β8 loop (figure 5.17A-B; fragments **8**, **11** and **14** are shown as pink, blue and lilac lines respectively). The fourth fragment in group 1 (**12**) contains a smaller pyrrolidinone ring bound to the centre of the binding hotspot, followed by a 5-carbon urea chain that curves towards β8 and a benzene ring above the β7- β8 loop (figure 5.17A, **12** is shown as green lines).

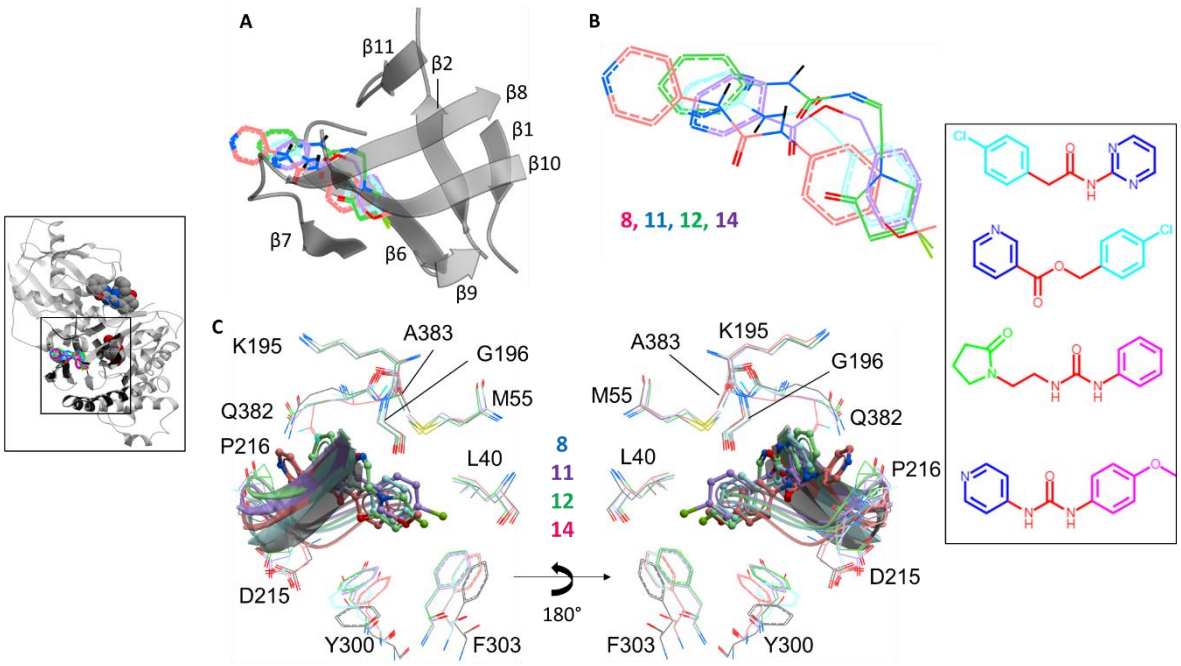


Figure 5.17: Binding interactions of group 1 fragments at the binding hotspot. *A. Side view of ribbon diagram showing secondary structure elements within binding hotspot that contribute to binding of group 1 fragments. Secondary structures are shown as grey ribbon and fragments as lines. B. Superimposed fragment structures shown as lines. C. Two side views of the binding hotspot as line and stick diagrams, showing protein-ligand interactions between group 1 fragments and hGALK1. Protein residues are shown as lines and fragments are shown as sticks. Lines and sticks from the same structure are shown in the same colour using the same shading as panel A. For clarity, Leu213-Val220 is shown primarily as ribbon rather than lines.*

Figure 5.17C shows the hGALK1 sidechains that coordinate binding of group 1 fragments. The buried ring at the centre of the binding hotspot is stabilised by hydrophobic interactions with Leu40, Met55, Tyr300 and Phe303. The aromatic side chains of Tyr300 and Phe303 are displaced inwards towards the buried ring of the fragments, compared to the hGALK1-galactose-

T3 structure, so that the sidechain hydroxyl of Tyr300 hydrogen bonds with the backbone amide of Ser214 in fragment-bound structures rather than hydrogen bonding with the sidechain of Asp215 as observed in the hGALK1-galactose-**T3** structure.

The amine and carbonyl groups of the 3- or 5-carbon chains of the group 1 fragments form hydrogen bonds with the backbone amide and carboxyl groups of residues within the β7- β8 loop (Ser214, Leu218) and with the backbone groups of Gly196 and Ala383. The second aromatic ring displaces and forms hydrophobic interactions with Pro216. For example, the benzene ring of fragment **12** is positioned to stack with the cyclic Pro216 sidechain. Additionally, the nitrogen ring substituents of fragments **8**, **11** and **14** hydrogen bond with the sidechain of Gln382, potentially disrupting its interaction with Lys195.

In group 2, five of the seven fragments consisted of two six-membered rings connected by a 2 or 3 carbon chain. These fragments consist of a benzene (**10**, **13**), pyridine (**9**), fluoro-benzene (**7**) or morpholine (**6**) ring, binding near the centre of the binding hotspot, followed by a 2- carbon amide (**9**) or 3-carbon amide (**6**, **7**), ester (**10**) or urea (**13**) chain sandwiched between β7 and β8, and finally either a substituted phenyl ring (dimethylbenzene in **6**, isopropyl benzene in **9** and methoxybenzene in **7**) or a nitrogen containing six-membered ring (piperazinone in **10** and pyridine in **13**) positioned above the β7- β8 loop (figure 5.18A-B).

The other two fragments of group 2 contain only one ring. One fragment is (1-isopropylpiperidin-3-yl)methanol, in which the methanol and piperidine groups occupy the centre of the binding spot and the isopropyl group extends behind the β7- β8 loop, the space occupied by the carbon chains of other fragments in this group (fragment **19**; lilac lines in figure 5.18A-B). The other fragment is ((2,6-dichlorobenzyl)amino)propan-2-ol, in which the dichlorobenzene group occupies the centre of the binding hotspot and the amino-propanol chain extending behind the β7- β8 loop (fragment **20**, yellow lines in figure 5.18A-B).

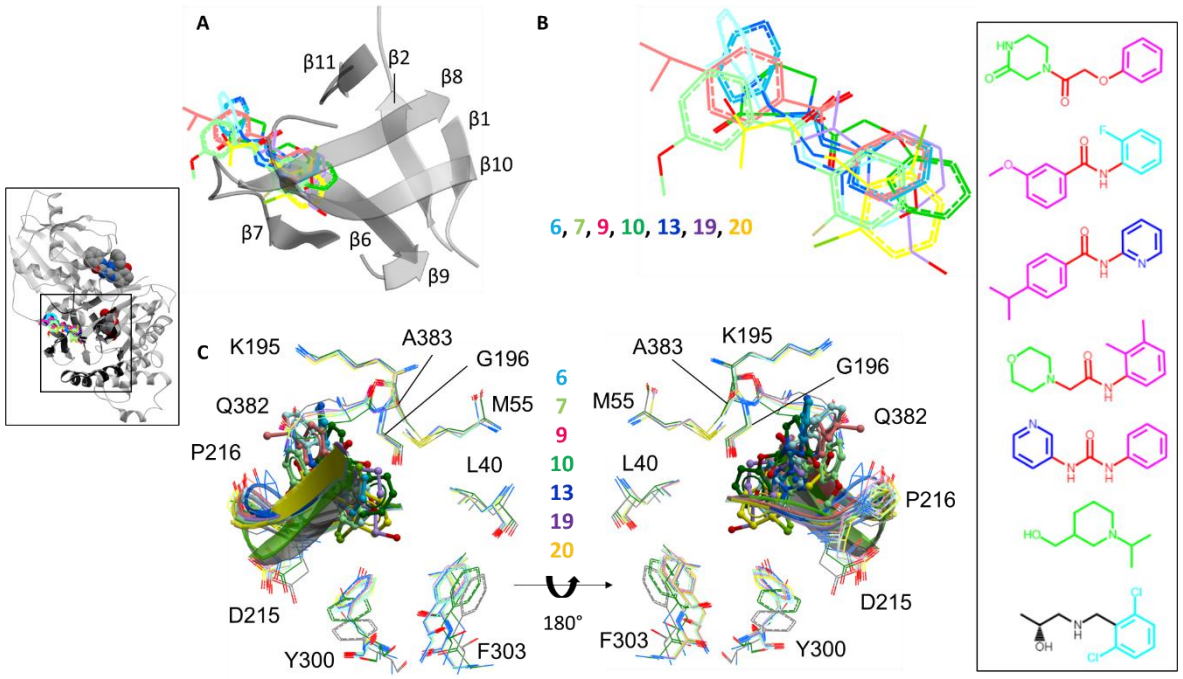


Figure 5.18: Binding interactions of group 2 fragments at the binding hotspot. *A. Side view of ribbon diagram showing secondary structure elements within binding hotspot that contribute to binding of group 2 fragments. Secondary structures are shown as grey ribbon and fragments as lines. B. Superimposed fragment structures shown as lines. C. Two side views of the binding hotspot as line and stick diagrams, showing protein-ligand interactions between group 2 fragments and hGALK1. Protein residues are shown as lines and fragments are shown as sticks. Lines and sticks from the same structure are shown in the same colour using the same shading as panel A. For clarity, Leu213-Val220 is shown primarily as ribbon rather than lines.*

Figure 5.18C shows the hGALK1 sidechains that coordinate group 2 fragments. The residues involved are identical to those for group 1 fragments, and only the relative distance between interacting residues and fragment moieties is changed. Specifically, the buried ring of group 2 fragments is further from the back of the pocket (β1 and β2) than in the group 1 fragments, resulting in weaker hydrophobic interactions with Leu40, Met55, Tyr300 and Phe303. A similar degree of displacement is seen, relative to the hGALK1-galactose-**T3** structure, of the aromatic sidechains of Tyr300 and Phe303. Hydrogen bonding is still observed between the amine and carbonyl groups of the 2 or 3 carbon chain of the group 2 fragments and the backbone amide and carboxyl groups of residues Gly196, Ser214, Leu218 and Ala383. Similar interactions are also seen with Pro216 and Gln382.

Group 3, also in the upper cluster, contains three fragments that are located within the binding hotspot but make fewer interactions than the group 1 and 2 fragments due to their small sizes (fragments **15**, **16** and **18**; figure 5.19). The triazole ring of **15** is positioned to form hydrogen bond with the backbone of residues within the β7- β8 loop (Ser214, Asp215, Pro216) and the sidechain of Gln382, while the amide chain of this fragment interacts with the backbone of residues Leu218, Ala219 and Val220 (figure 5.19C, green lines).

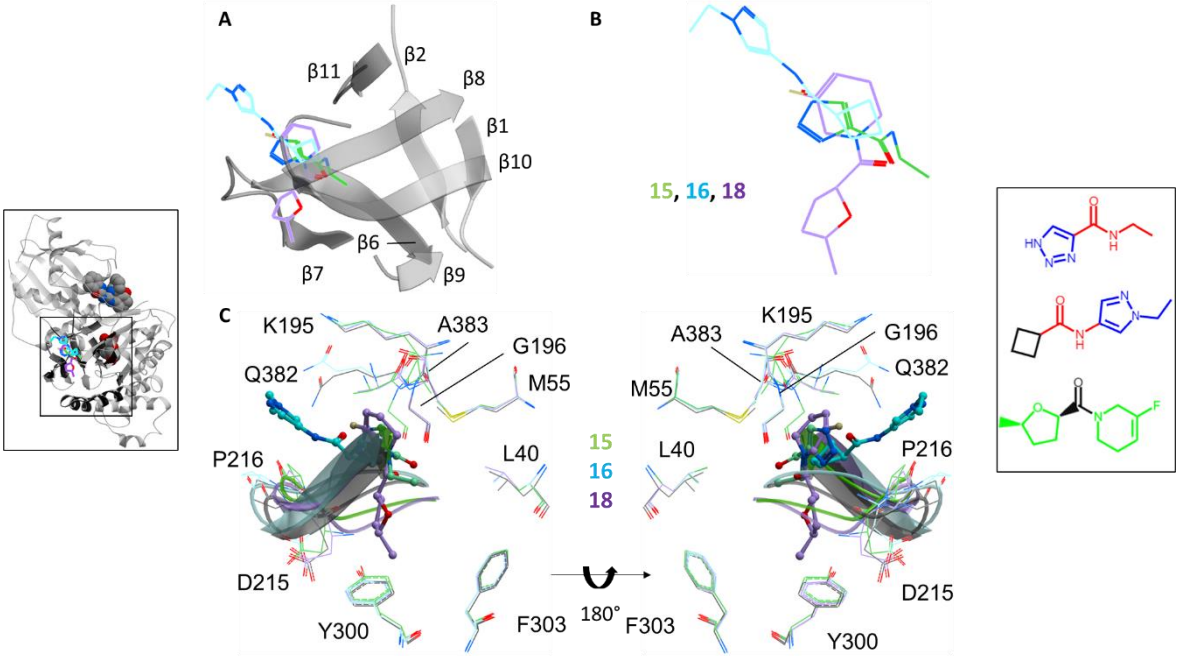


Figure 5.19: Binding interactions of group 3 fragments at the binding hotspot. *A. Side view of ribbon diagram showing secondary structure elements within binding hotspot that contribute to binding of group 3 fragments. Secondary structures are shown as grey ribbon and fragments as lines. B. Superimposed fragment structures shown as lines. C. Two side views of the binding hotspot as line and stick diagrams, showing protein-ligand interactions between group 3 fragments and hGALK1. Protein residues are shown as lines and fragments are shown as sticks. Lines and sticks from the same structure are shown in the same colour using the same shading as panel A. For clarity, Leu213-Val220 is shown primarily as ribbon rather than lines.*

The ethyl-pyrazole group of fragment **16** is coordinated by stacking with the sidechain ring of Pro216 and hydrogen-bonds with the sidechain of Gln382 while the carboxamide group of this fragment is positioned to form hydrogen bonds with the backbone of Ser214, Pro216 or Gln382, and the cyclobutane forms weak hydrophobic interactions with Leu213, Leu218 and Val220 (figure 5.17C, blue lines). The methyl-tetrahydrofuran moiety of fragment **18** forms weak

interactions with Leu213, Leu214 and Leu218 while the fluoropyridine group of this fragment is coordinated by hydrophobic interactions with Met55, Gly196 and Ala383 (figure 5.17C, lilac lines).

Four of the six fragments in group 4 (fragments **22**, **23**, **26** and **28**) contain a substituted phenyl ring that binds in the centre of the binding hotspot, similar to the buried ring constituent of the group 1-3 fragments, and a variable second moiety that points towards the α8-α9 turn (figure 5.20A).

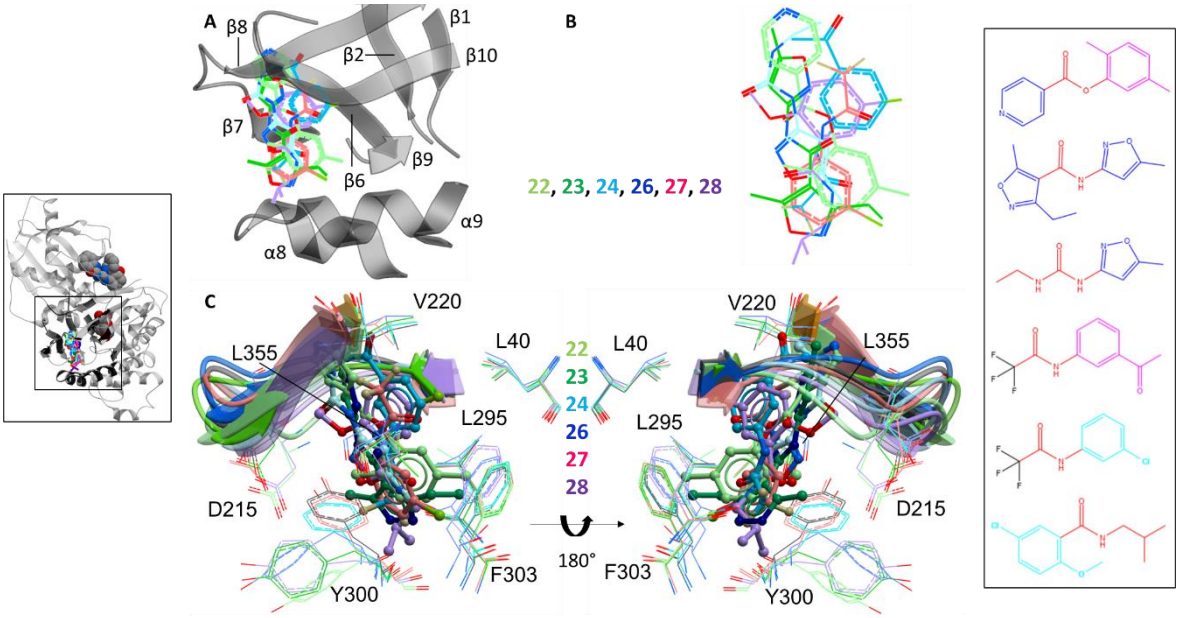


Figure 5.20: Binding interactions of group 4 fragments at the binding hotspot. *A. Side view of ribbon diagram showing secondary structure elements within binding hotspot that contribute to binding of group 4 fragments. Secondary structures are shown as grey ribbon and fragments as lines. B. Superimposed fragment structures shown as lines. C. Two side views of the binding hotspot as line and stick diagrams, showing protein-ligand interactions between group 4 fragments and hGALK1. Protein residues are shown as lines and fragments are shown as sticks. Lines and sticks from the same structure are shown in the same colour using the same shading as panel A. For clarity, Leu213-Val220 is shown primarily as ribbon rather than lines.*

The central ring constituent is xylene in **22**, acetophenone in **26**, chlorophenyl-methanol in **28** and methyl isoxazole in **23**. The variable second moiety is phenyl acetate (**22**), methyl isoxazole (**23**), trifluoroacetamide (**26**) or isopentyl acetamide (**28**). For the other two group 4 fragments, the ring constituent (isoxazole in **24** and chlorobenzene in **27**) superimposes with the second

moiety of the other four fragments, while the second moiety (ethyl urea in **24** and trifluoroacetamide in **27**) occupies the position in the centre of the binding hotspot (figure 5.20A-B, light blue and pink lines).

Figure 5.20C shows the sidechains that coordinate binding of the group 4 fragments. As observed for groups 1-3 in the upper cluster, group 4 fragments are coordinated in the centre of the binding hotspot through hydrophobic interactions with residues of the β7- β8 loop (Leu213, Leu218, Ala219), and with Leu295 and Leu355. There is also a similar displacement of the Phe303 sidechain, relative to the hGALK1-galactose-**T3** structure, to form hydrophobic interactions with the bound fragment. In the four fragments that have an aromatic ring bound in the centre of the hotspot (**22**, **23**, **26** and **28**), the second moiety occupies the space taken up by the sidechain of Tyr300 in the hGALK1-galactose-**T3** structure, and the fragment-bound structures show an ~90° displacement of this sidechain to avoid clashing with the fragment. In the other two group 4 fragments (**24** and **27**), minimal displacement of the Tyr300 sidechain was observed and the hydrogen bond between the sidechain hydroxyl group of Tyr300 and the sidechain of Asp215 is preserved suggesting that binding of an aromatic ring to the centre of the hotspot makes displacement of Tyr300 favourable.

In this chapter, optimised protocols for the production and crystallisation of recombinant hGALK1 were developed and used to solve crystals structures of hGALK1 with ADP and four ATP-competitive inhibitors (**T1**, **T2**, **T3** and **T7**). A fragment screening campaign performed against hGALK1-galactose-**T3** crystals identified 28 fragments clustering to three binding pockets. Chapter 6 describes the characterisation of both known inhibitors and fragment hits in solution and initial follow-up compounds derived from the binding hotspot fragments described in this chapter. Chapter 7 summarises the insights into hGALK1 function and inhibition gained from this work and discusses the limitations and future directions suggested by these data.

6. Characterisation and elaboration of hGALK1 fragment hits

Chapter 5 described the development of a robust hGALK1 crystal system (section 5.1) for structure determination with known inhibitors of hGALK1 (section 5.2, compounds **T1**-**T7**, table 5.2) and for a fragment screening campaign (section 5.3). The 28 hits identified bound in three distinct binding sites: at the entrance to the ATP binding site (fragments **1** and **2**), in a pocket close to the termini of the protein (fragments **3**-**5**) and in a binding hotspot on the opposite side of the protein to the active site (fragments **6**-**28**) (summarised in figure 5.12).

The next step was to investigate whether these fragments were able to affect the function of hGALK1 in solution and if so to characterise this effect. This is essential to translate the effect of high concentrations of fragment on solid phase protein crystals to dose-dependent behaviour in solution. An additional consideration for this particular system was that the protein crystals used for fragment screening contained both galactose and compound **T3**, blocking the active site. This meant that inhibition of hGALK1 by orthosteric binding of fragments could not be ruled out from the crystal structures and had to be investigated in solution.

Section 6.1 describes the development and validation of two hGALK1 activity assays to test fragment activity. Section 6.2 describes the identification and characterisation of two fragments (**4**, **8**) that weakly inhibit hGALK1 in both assays and do not compete with galactose or ATP. Section 6.3 describes the kinetic analysis of 5 compounds (**M1**-**M5**) designed by *in silico* merging of the most potent fragment (**8**) with nearby fragments (**22**-**28**).

6.1 Validation of two activity assays for the detection of hGALK1 inhibition

Two hGALK1 activity assays were set up in this work. The first assay, Kinase Glo, measures ATP concentration remaining after a set period of hGALK1 activity (figure 6.1A). The second assay, Galactose oxidase-Amplex Red, adopted to measure hGALK1 activity for the first time in this work, measures the remaining galactose concentration after a set period of hGALK1 activity (figure 6.1B).

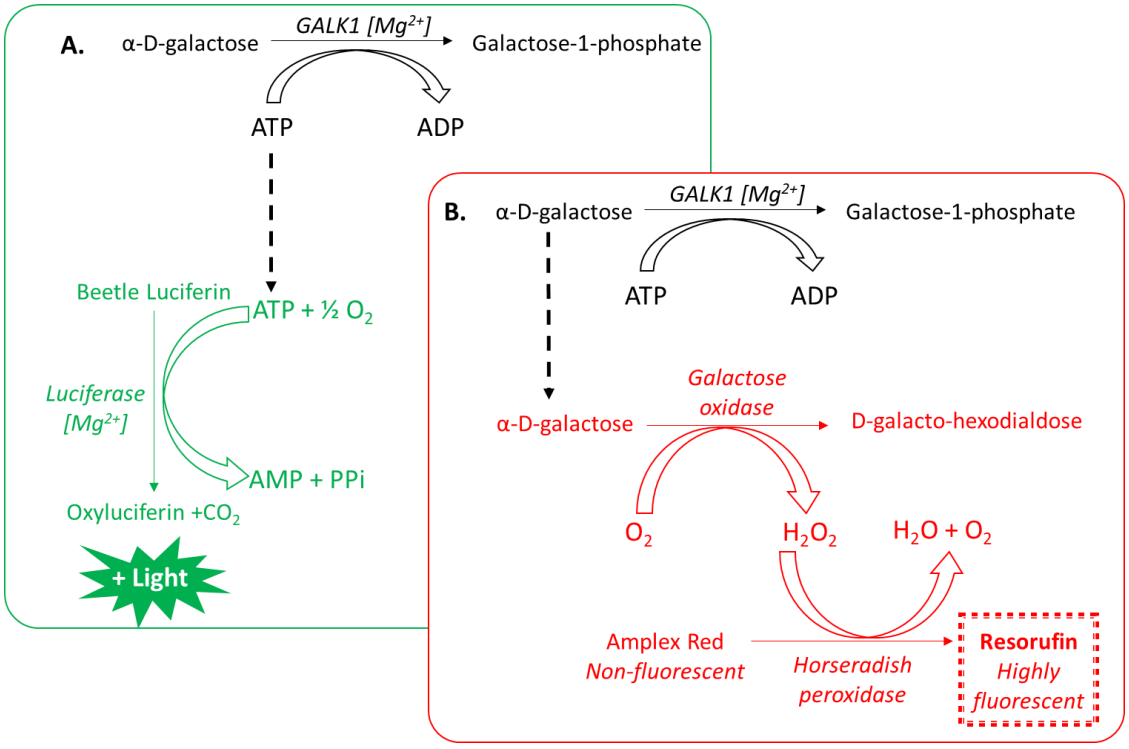


Figure 6.1: Principle of the hGALK1 activity assays used in this work. *A. Reaction scheme for the Kinase Glo assay. ATP remaining after phosphorylation of galactose by hGALK1 is used by luciferase to oxidise luciferin, generating oxyluciferin that emits light. B. Reaction scheme for the Amplex Red assay. Galactose remaining after phosphorylation of galactose by hGALK1 is oxidised by galactose oxidase, producing hydrogen peroxide. Hydrogen peroxide is used by horseradish peroxidase to reduce Amplex Red, generating highly fluorescent resorufin.*

6.1.1 Principle of the Kinase Glo activity assay

The Kinase Glo reagent contains luciferase (EC 1.13.12.7) and its substrate luciferin. Luciferase uses molecular oxygen to catalyse the oxidation of luciferin to oxyluciferin in an excited state, which emits visible light to return to the ground state. The luciferase reaction generates one photon of light per turnover so that the luminescence signal is directly proportional to the ATP concentration within the assay's measurement window. Addition of Kinase Glo reagent to a completed kinase reaction and subsequent quantification of the remaining ATP is a well-established method of measuring kinase activity, particularly in high throughput screening campaigns [276]–[278], including for hGALK1 [107], [133], [135], [139].

In this work, published Kinase Glo protocols were employed to first confirm that purified hGALK1 was active and behaved in agreement with published work with regards to substrate kinetics (section 6.1.3) and inhibition by tool compounds (section 6.1.5). Subsequently the

Kinase Glo assay was used to characterise inhibition of hGALK1 by fragments (6.2.1) and follow-up compounds (6.3).

6.1.2 Principle of the Amplex Red activity assay

The Amplex Red assay reagent contained horseradish peroxidase (EC 1.11.1.7), its substrate Amplex Red (10-acetyl-3,7-dihydroxyphenoxazine) and galactose oxidase (EC 1.1.3.9). Oxidation of galactose by galactose oxidase generates hydrogen peroxide which is used by horseradish peroxidase to reduce Amplex Red to its highly fluorescent product, resorufin (7-hydroxy-3-phenoxazin-3-one). The resultant fluorescence signal (measured with excitation at 530-560 nm and emission at 580-590 nm) is proportional to the amount of galactose present at the beginning of the galactose oxidase reaction, within the measurement window of the assay. As explained for the hHAO1 Amplex Red activity assay (section 4.1.1), at high concentrations of hydrogen peroxide, horseradish peroxidase further reduces resorufin to generate resazurin (7-hydroxy-10-oxidophenoxazin-10-ium-3-one), which does not fluoresce. Below this concentration, the fluorescence signal is inversely proportional to hGALK1 activity i.e. less hGALK1 activity results in more galactose present and subsequently higher fluorescence (figure 6.1B).

Amplex Red-based assays have previously been described for measuring the activity of a variety of oxidase enzymes that generate hydrogen peroxide (e.g. hHAO1, described in chapter 4 of this work and elsewhere [240]; monoamine oxidase [279]; and glucose formed by the activity of glucocerebrosidase can be oxidised by glucose oxidase to form hydrogen peroxide [8]). However, no studies to date have applied this type of assay to measure hGALK1 activity. In this work, Amplex Red assay parameters were established for hGALK1 including measuring hGALK1 substrate kinetics (section 6.1.4) and inhibition by tool compounds (section 6.1.5, 6.1.6). The optimised, validated assay was then used, in conjunction with the Kinase Glo assay, to determine the inhibition mode of promising fragments with respect to ATP (section 6.2) and to characterise inhibition of hGALK1 by the follow-up compounds (section 6.3).

6.1.3 Kinetic properties of hGALK1 determined by the Kinase Glo assay

The ATP concentration range over which luminescence signal remains linear varies between Kinase Glo reagents and is dependent on the luciferase enzyme, which is subject to product inhibition, and the buffer composition [280], [281]. In this work, the relationship between luminescence and ATP concentration was linear up to 60 µM ATP (R^2 = 0.987; figure 6.2A- B). Consequently, the ATP concentration reported in published screening campaigns (35 µM) was maintained in this work. Similarly, the hGALK1 concentration (10 nM) and galactose concentration (100 µM) described in the literature did not require optimisation. No significant change in luminescence signal was observed after incubating fixed amounts of the Kinase Glo reagent with up to 1 mM of galactose, confirming that galactose did not interfere with the readout of this assay (figure 6.2C).

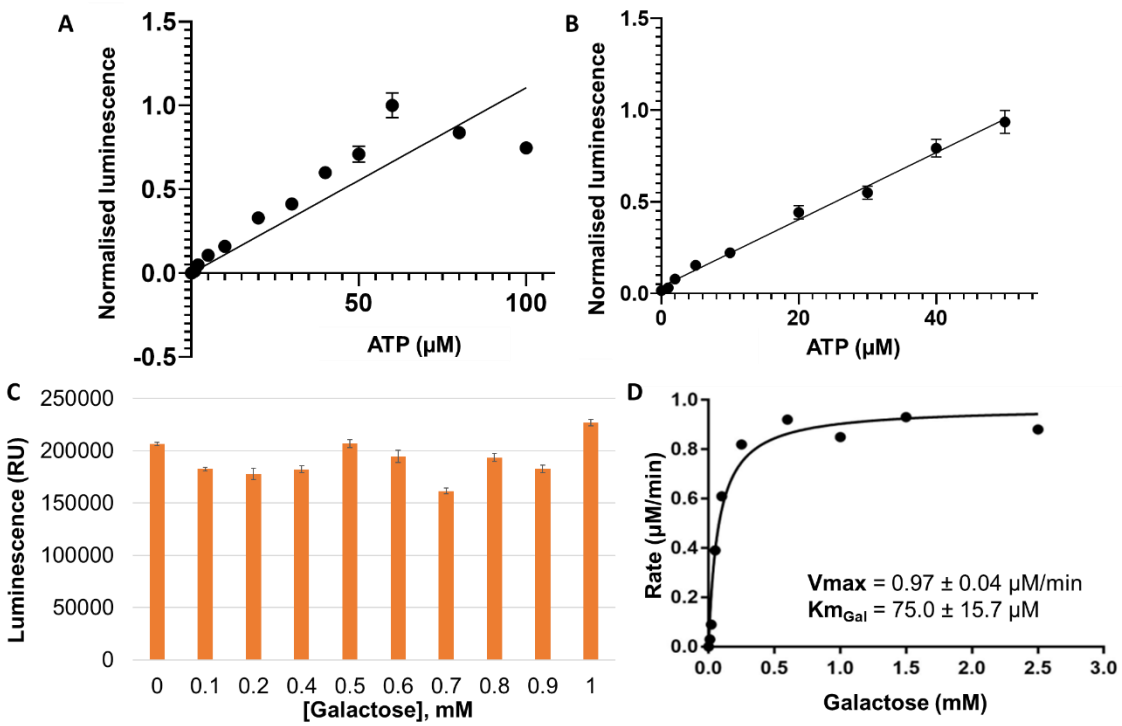


Figure 6.2: Confirmation of Kinase Glo assay parameters for the kinetic characterisation of hGALK1 in vitro activity. *A. Plot of normalised luminescence generated by the Kinase Glo reagent in the presence of increasing concentrations of ATP (0 – 100 µM) showing linearity is lost at higher ATP concentrations. B. Plot of normalised luminescence versus ATP concentration within the linear range of the Kinase Glo reagent (0 – 50 µM). C. Bar chart showing the relative luminescence generated by the Kinase Glo reagent in the presence of increasing concentrations of galactose, up to 1 mM. D. Michaelis-Menten plot of galactose concentration versus reaction rate for 10 nM hGALK1. Graph plotting and calculation of kinetic parameters were performed using GraphPad Prism software. Error bars are standard deviation of three technical replicates.*

Finally, galactose kinetics were determined for hGALK1 from Michaelis-Menten plots (figure 6.2D). The calculated maximal velocity ($Vmax^{apparent}$) was 0.97 μM ATP/minute, giving a kcat of 1.6 s^{-1}. This is a tenfold lower turnover than previously reported (17. 5 s^{-1} [133] and 8.7 s^{-1} [35]), attributable to a combination of the low ATP concentration (35 μM) and long reaction time (1 hour). The calculated Michaelis-Menten constant (Km) for galactose was 75 μM, which cannot be directly compared with literature because published values range from 135 μM [251] to 970 μM [35].

6.1.4 Kinetic properties of hGALK1 determined by the Amplex Red assay

Since this assay has not been previously reported for the detection of hGALK1 activity, several parameters have been established here before testing compounds. First, the amount of time required for complete conversion of galactose to resorufin was determined by measuring the fluorescence signal at 10, 20, 30, 40 and 50 minutes after addition of Amplex Red reagent to a range of galactose concentrations (figure 6.3A). No further increase in fluorescence signal was observed between 40 and 50 minutes, therefore the optimised protocol included a 40-minute incubation period between addition of the Amplex Red reagent and fluorescence measurement.

Next, the highest galactose concentration that still yielded a linear increase in fluorescence (50 μM) was selected as the galactose concentration for the hGALK1 reaction (figure 6.3B; R^2 = 0.967). Optimal hGALK1 and ATP concentrations were next determined by measuring the amount of galactose consumed by an array of seven hGALK1 concentrations (0-500 nM) against 12 ATP concentrations (0-500 μM). A clear ATP-dependent increase in galactose consumption was observed at 100, 250 and 500 nM hGALK1 (figure 6.3C, orange, black and brown lines respectively). hGALK1 activity at these three concentrations was high but not saturating in the presence of 50-100 μM ATP. Thus, 250 nM hGALK1 and 100 μM ATP were used in all subsequent assays.

There was minimal change in the rate of consumption of galactose up to 500 µM ATP in the absence of hGALK1 (figure 6.3C, dark blue line), indicating that ATP did not interfere with either horseradish peroxidase or galactose oxidase activity.

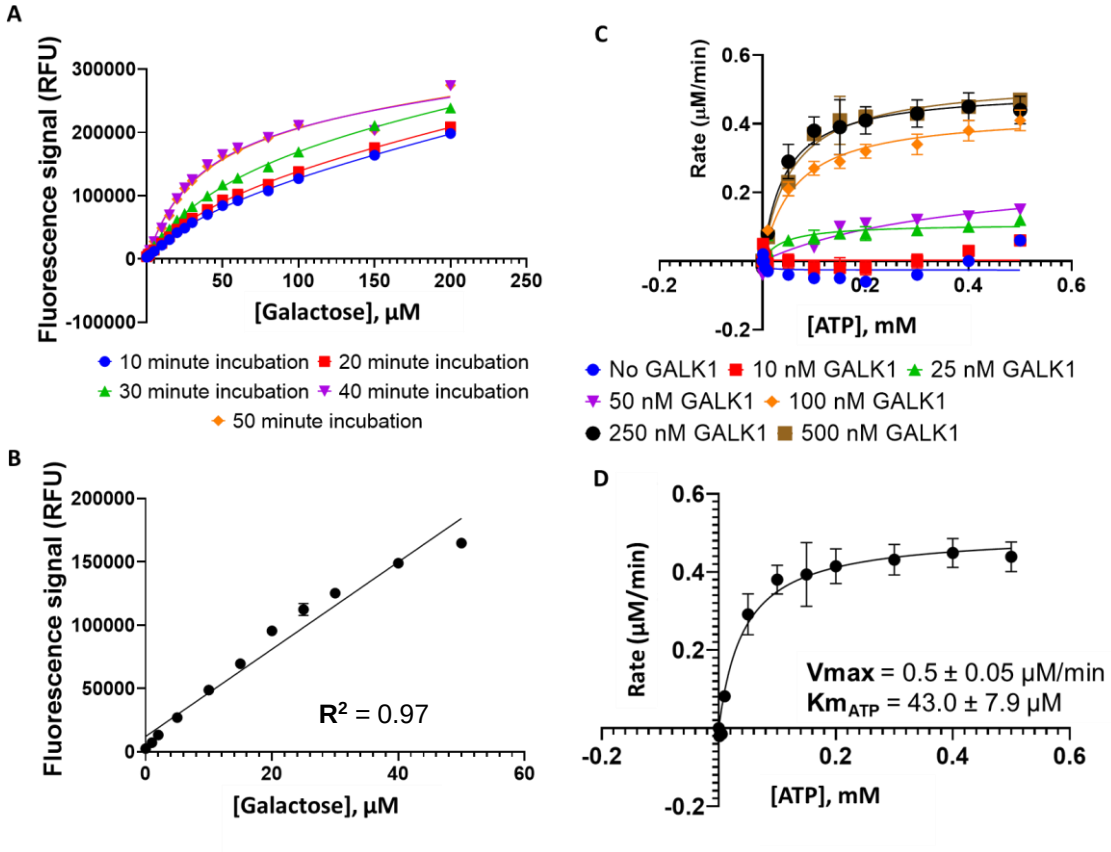


Figure 6.3: Optimisation of Amplex Red assay for the kinetic characterisation of hGALK1 in vitro activity. *A. Plot of fluorescence signal generated by the oxidation of increasing galactose concentrations by galactose oxidase. Curves are fitted to the fluorescence signal detected after 10 minutes (dark blue), 20 minutes (red), 30 minutes (green), 40 minutes (purple) or 50 minutes (orange) incubation. B. Plot of fluorescence signal versus galactose concentration, showing a linear relationship up to 50 µM galactose. C. Michaelis-Menten plot of ATP concentration versus hGALK1 reaction rate for different concentrations of hGALK1. Curves are fitted to the observed reaction rate with 10 nM (red), 25 nM (green), 50 nM (purple), 100 nM (orange), 250 nM (black) and 500 nM (brown) hGALK1 at a range of ATP concentrations. The control reaction without hGALK1 is shown in dark blue. D. Michaelis-Menten plot of ATP concentration versus reaction rate for 250 nM hGALK1. Graphs were plotted and* $Vmax^{apparent}$ *and* $Km^{apparent}$ *values calculated using GraphPad Prism software. Error bars are standard deviation of three technical replicates.*

The Michaelis-Menten plot of 250 nM hGALK1 over a range of ATP concentrations was used to calculate ATP kinetics for hGALK1 (figure 6.3D). The $Km^{apparent}$ of ATP was 43 ± 7.9 µM, in agreement with published values (ranging from 20.9 µM [134] to 42 µM [139]). The $Vmax^{apparent}$, 0.5 µM galactose/minute, is approximately half that observed in the Kinase Glo assay (figure 6.2D) but is also likely to be underestimated due to the low substrate

concentrations and long incubation times used. Based on figure 6.3, the optimal assay parameters were therefore determined to be: 250 nM hGALK1, 50 μM galactose and either 100 μM ATP (for single point assay) or an array spanning 0-250 μM ATP (for substrate competition).

6.1.5 Measurement of inhibition of hGALK1 by tool compounds

As an additional validation step, the inhibition of hGALK1 by the tool compounds introduced in chapter 5 (spiro-benzoxazole series, compounds **T1**-**T7**, table 5.2) was measured. Inhibition of three different purification batches of hGALK1 (referred to in this work as GALK1-1, GALK1-2 and GALK1-3; for reference SGC IDs are GALK1A-p031, p036 and p037) by 1 mM of each tool compound was determined in Kinase Glo and Amplex Red assays, using the assay parameters described above. Compounds **T4**, **T5** and **T6** were weak inhibitors at this concentration, demonstrating less than 20% inhibition across all three hGALK1 batches in both assays (figure 6.4). The weak potency of these compounds may explain the failure to achieve high resolution co-crystals of compounds **T4**, **T5** and **T6** with hGALK1 and galactose, as described in section 5.1.4.

In the Kinase Glo assay, the maximal inhibition observed was 92% with 1 mM compound **T1**, 51% with compound **T2**, 79% with compound **T3** and 61% with compound **T7** (figure 6.4A). Measurement in the Amplex Red assay showed very similar results (figure 6.4B). The low potency of compound **T2** is very surprising because it was previously shown to inhibit hGALK1 in a Kinase Glo assay, using the same parameters as this work, with a calculated IC_{50} of 6 μM [135]. To confirm that these four compounds inhibited hGALK1 in a target-specific, dose-dependent manner, inhibition of hGALK1 at 12 compound concentrations from 0-1 mM was measured in both assays and used to calculate IC_{50} values.

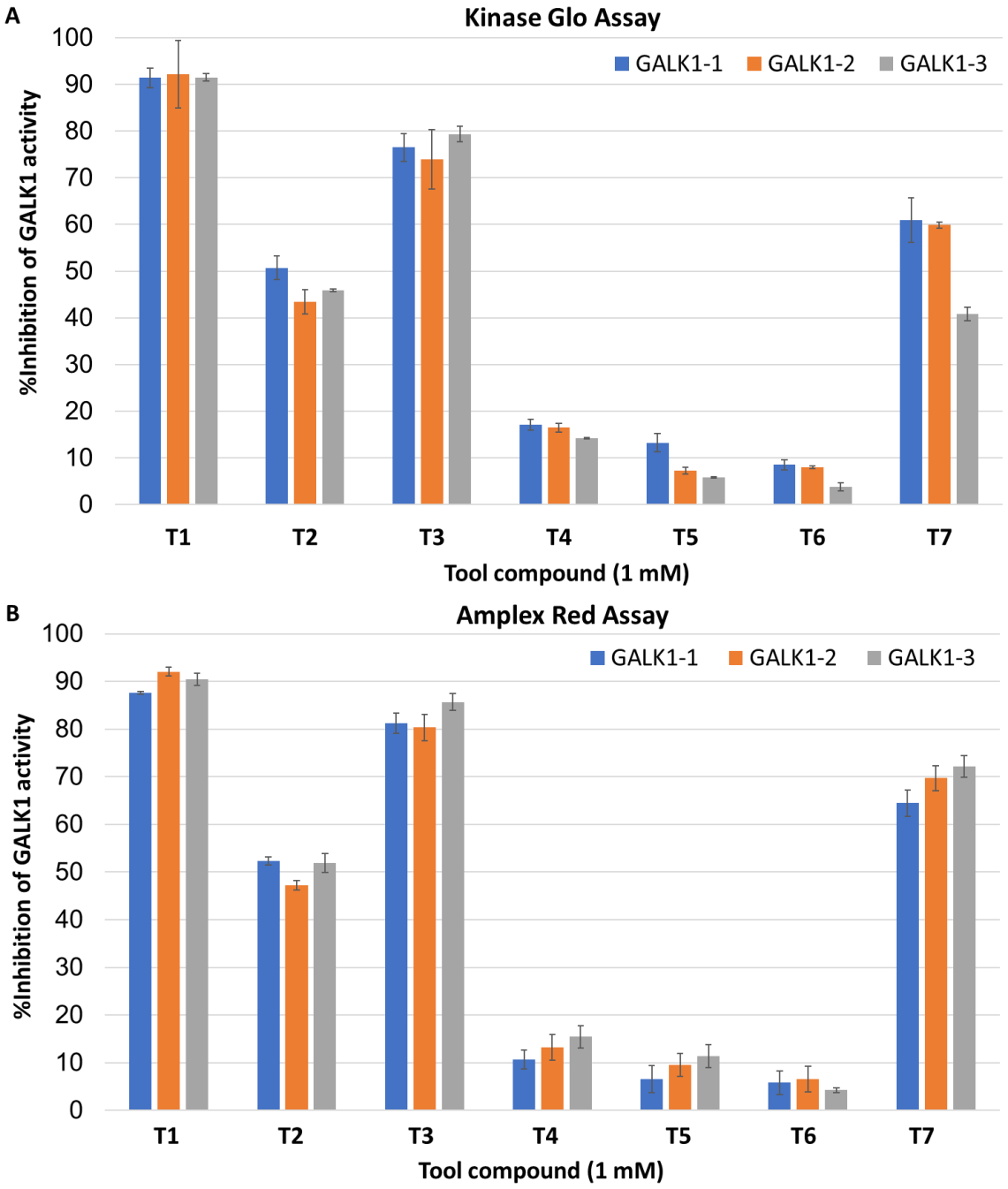


Figure 6.4: Inhibition of hGALK1 by spiro-benzoxazole tool compounds T1-T7. *A. Plot of % inhibition of three batches of hGALK1 by 1 mM of tool compound measured using the Kinase Glo assay. Errors bars are standard deviation of technical triplicates. B. Plot of % inhibition of three batches of hGALK1 by 1 mM of tool compound measured using the Amplex Red assay. Errors bars are standard deviation of technical triplicates.*

Dose-dependent inhibition of hGALK1 was observed with all four compounds in both assays (figure 6.5). Dose-response curves could be used to calculate IC50s for compounds **T1**, **T3** and **T7** but not for compound **T2** whose maximal inhibition was around 50%., as previously observed in the single concentrations screen. The compound IC50s (and LLEAT scores) calculated from Kinase Glo data were 12 μM for compound **T1** (figure 6.5A, blue; pIC50 4.9 ± 0.2; LLEAT

0.19 kcal/mol), 17 μM for compound **T3** (figure 6.5A, green; pIC50 4.8 ± 0.1; LLEAT 0.16 kcal/mol) and 98 μM for compound **T7** (figure 6.5A, purple; pIC50 4.0 ± 0.1; LLEAT 0.09 kcal/mol). In agreement with published data, compound **T1** was the most potent (published IC50 of 1 μM [135]). The calculated IC50 for compound **T3** is also similar to the published value (12 μM, [135]).

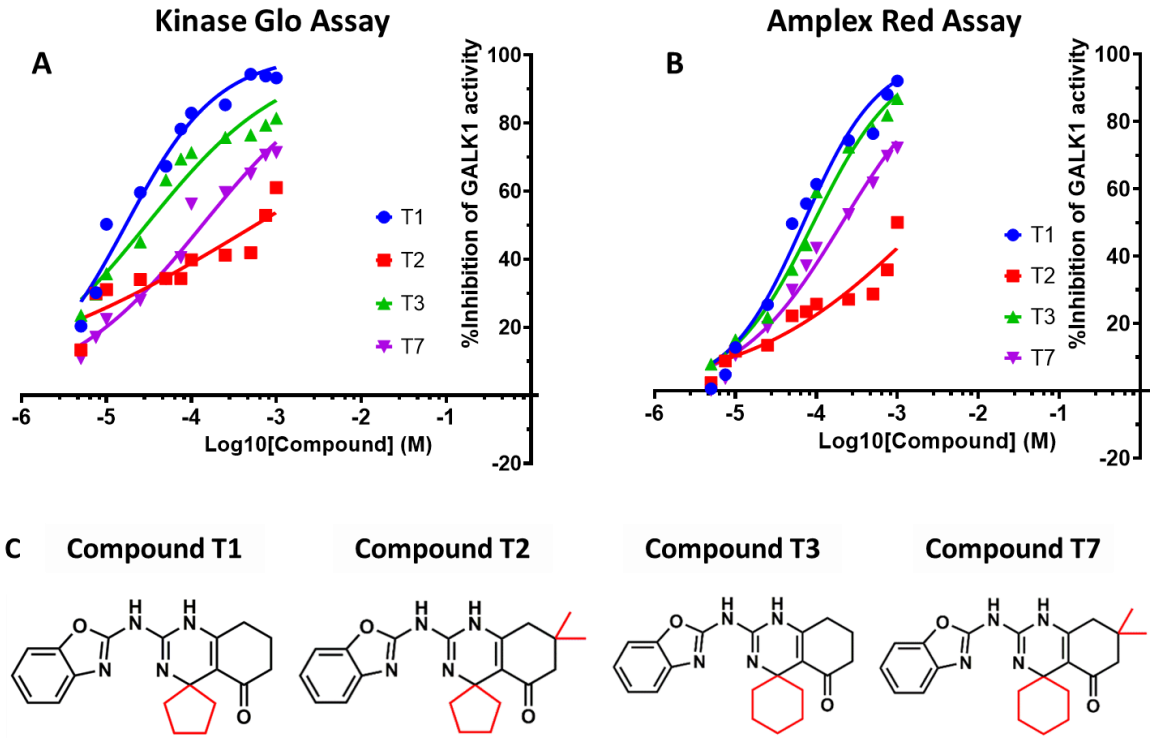


Figure 6.5: Characterisation of in vitro inhibition of hGALK1 by tool compounds. *A. Plot of log concentration of tool compounds versus %inhibition of hGALK1 activity observed in Kinase Glo assay. Inhibition values are a mean of three technical replicates. B. Plot of log concentration of tool compounds versus %inhibition of hGALK1 activity observed in Amplex Red assay. C. Chemical structures of tool compounds **T1**, **T2**, **T3** and **T7**. Variable chemical groups are coloured in red, and the shared scaffold is in black.*

The compound IC50s (and LLEAT scores) calculated from the Amplex Red data were 51 μM for compound **T1** (figure 6.5B, blue; pIC50 4.3 ± 0.2; LLEAT 0.16 kcal/mol), 82 μM for compound **T3** (figure 6.5B, green; pIC50 4.1 ± 0.1; LLEAT 0.12 kcal/mol) and 141 μM for compound **T7** (figure 6.5A, purple; pIC50 3.9 ± 0.4; LLEAT 0.08 kcal/mol). The same pattern of potency was observed in the Amplex Red assay with compound **T1** being the most potent, compound **T3** having a similar potency and compound **T7** having a markedly higher IC50.

The higher IC50 values calculated from the Amplex Red assay relative to the Kinase Glo assay

can be attributed to two factors: enzyme concentration and substrate concentration. hGALK1 concentration is 25-fold higher in the Amplex Red assay than in the Kinase Glo assay, therefore more inhibitor is required to inhibit half of the protein present. Additionally, ATP concentration is 4-fold higher in the Amplex Red assay and competes with the tool compounds for binding to GALK1. Comparison of the chemical structures of the tool compounds (figure 6.5C) indicate that the larger spiro ring (spiro-cyclohexane) is more favourable when the tetrahydroquinazolinone ring contains a dimethyl substituent, with compound **T7** showing better inhibition of hGALK1 than compound **T2**, whereas tool compounds without a dimethyl (**T1** and **T3**) demonstrate similar in potency. The superior potency of compounds **T1** and **T3** suggested that the dimethyl substituent reduces inhibitor efficacy.

6.1.6 Determination of the mode of hGALK1 inhibition for tool compound **T1**

After confirmation of hGALK1 inhibition by the tool compounds, the most potent compound (compound **T1**) was used to develop a suitable protocol for determining the inhibition mode of hGALK1 inhibitors with respect to the substrates, galactose and ATP. Compound **T1** was selected as a control compound as, from its co-crystal structure with hGALK1 and galactose (section 5.2.3), it is expected to be competitive with respect to ATP (i.e. binding of **T1** and ATP are mutually exclusive) but not with respect to galactose (i.e. hGALK1 can bind to **T1** and galactose concurrently, as observed in the crystal structure).

A compound's inhibition mode is determined by its binding affinity for the free enzyme and the enzyme-substrate complex. Competitive inhibitors, as the name suggests, compete directly with the substrate for binding to the active site of the free enzyme and therefore lose efficacy at high substrate concentrations. This is particularly relevant to the use of ATP-competitive compounds for hGALK1 inhibition because of the millimolar concentration of intracellular ATP [282]. Non-competitive inhibitors bind to both the free enzyme and the enzyme-substrate complex and decrease the turnover of the substrate without affecting the enzyme's affinity for the substrate. In the case of hGALK1, the enzyme-substrate complex could mean hGALK1-ATP, hGALK1-

galactose or hGALK1-galactose-ATP. Mixed inhibitors also bind to both the free enzyme and the enzyme- substrate complex, but they reduce both the catalytic efficiency (turnover) of the enzyme and its affinity for the substrate. Finally, uncompetitive inhibitors bind only to the enzyme-substrate complex, which leads to an overall decrease in turnover but also increased apparent binding affinity of free enzyme for the substrate.

These inhibition modes can be differentiated by comparison of the catalytic efficiency (turnover), in the form of changes in the reaction $Vmax^{apparent}$ at different inhibitor concentrations, and the substrate affinity, characterised as changes in the substrate $Km^{apparent}$ at different inhibitor concentrations. $Vmax^{apparent}$ and $Km^{apparent}$ values are determined through Michaelis-Menten plots of enzyme reaction rate versus substrate concentration, measured at several inhibitor concentrations, and several graphical methods have been developed to allow kinetics data to be plotted in linear form with the inhibition mode indicated by the intersection of the plotted lines [283], [284].

In this work, Michaelis-Menten plots were used to calculate $Vmax^{apparent}$ and $Km^{apparent}$ values, in combination with linear Cornish-Bowden plots, to determine inhibition mode. In Cornish-Bowden plots, [S]/v (concentration of substrate divided by reaction rate) is plotted against inhibitor concentration at different substrate concentrations so that each line represents a different substrate concentration [284]. In competitive inhibition, the reaction rate increases in proportion to the increase in substrate (as it out-competes the inhibitor) and therefore the lines of a Cornish-Bowden plot for a competitive inhibitor are parallel. For the other inhibition modes, reaction rate remains decreased at high substrate concentrations so that the plotted lines converge either above the x-axis (uncompetitive inhibitors), at the x-axis (non-competitive inhibitors) or below the x-axis (mixed inhibitors).

To determine the inhibition mode of compound **T1** with respect to galactose and ATP, an array of 12 inhibitor concentrations (0 – 1 mM) across one axis and 7 substrate concentrations (0 –

400 µM) down the other axis was prepared for each substrate (galactose or ATP). Concentrations of hGALK1 and the other substrate were kept constant. The array varying ATP was measured in the Amplex Red assay and the array varying galactose was measured in the Kinase Glo assay so that the initial concentration of the substrate being directly measured by the assay was kept constant, avoiding substrate interference with the linearity of the assay response.

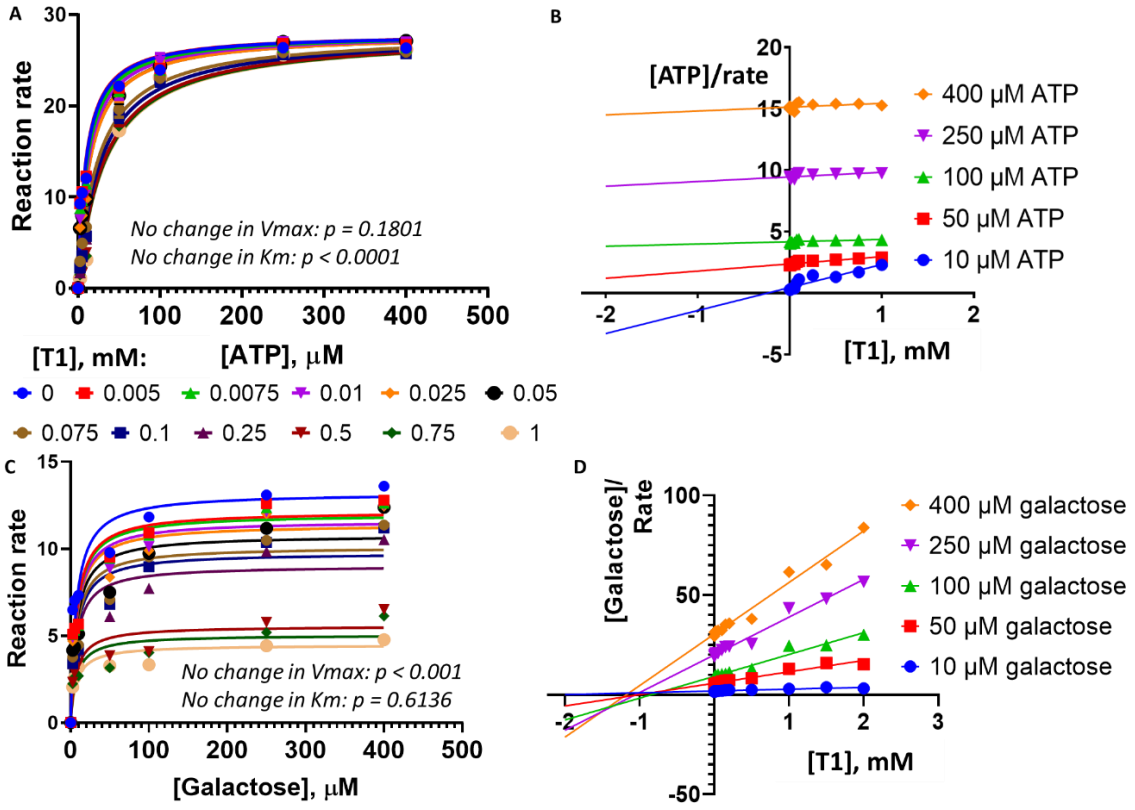


Figure 6.6: Kinetic characterisation of hGALK1 inhibition by compound T1. *A. Michaelis-Menten plot of hGALK1 reaction rate (µM galactose consumed by 250 nM hGALK1) against ATP concentration at twelve different compound* ***T1*** *concentrations, determined in the Amplex Red assay. Reactions were started by addition of 50 µM galactose. B. Cornish-Bowden plot of [ATP]/reaction rate (as defined by panel A) against compound* ***T1*** *concentration at five different ATP concentrations, determined in the Amplex Red assay. C. Michaelis-Menten plot of hGALK1 reaction rate (µM ATP consumed by 10 nM hGALK1) against galactose concentration at twelve different compound* ***T1*** *concentrations, determined in the Kinase Glo assay. Reactions were started by addition of 35 µM ATP. D. Cornish-Bowden plot of [galactose]/reaction rate (as defined in panel C) against compound* ***T1*** *concentration at five different galactose concentrations, determined in the Kinase Glo assay. All experiments were performed as technical triplicates and plotted data points are average values.*

As expected, competition of compound **T1** with ATP was confirmed by Michaelis-Menten (figure 6.6A) and Cornish-Bowden (figure 6.6B) plots. There was no significant change in $Vmax^{apparent}$ ($P = 0.1801$) and a significant increase in $Km^{apparent}$ ($P < 0.0001$) calculated from the

Michaelis- Menten plot and the lines of the Cornish-Bowden plot were parallel, indicating compound **T1** was ATP-competitive. Compound **T1** did not compete with galactose: a dose dependent decrease in $Vmax^{apparent}$ was observed in the presence of increasing compound **T1** concentration ($P < 0.001$; figure 6.6C) and there was no significant change in $Km^{apparent}$ ($P = 0.6136$). The Cornish- Bowden plot shows lines intersecting on the x-axis, indicative of non-competitive inhibition with respect to galactose and consistent with a decrease in $Vmax^{apparent}$ and no change in $Km^{apparent}$ (figure 6.6D).

6.2 Testing effects of fragments on hGALK1 activity

The two activity assays described in the previous section were next used to characterise 27 of the 28 fragment hits identified in chapter 5. One fragment was unavailable for commercial purchase (fragment **17**) and was not characterised at this time. Fragments were ordered from Enamine and readily dissolved at 100 mM in DMSO. Fragments were incubated with Kinase Glo reagent (luciferin, luciferase and ATP) to check for assay interference i.e. to identify fragments that cause changes in the luminescence signal produced in the absence of hGALK1. Three fragments (fragments **7**, **9** and **13**) reduced the luminescence signal by more than 25%, indicating either inhibition of luciferase or light absorption by the fragments or their impurities. Several similar compounds, consisting of aryl rings bridged by acyclic linkers, have been previously reported as inhibitors of luciferase [285] making the former interpretation reasonable. Consequently, fragments **7**, **9** and **13**, as potential promiscuous kinase inhibitors, were not characterised at this time.

6.2.1 Single concentration screen of fragments against hGALK1 by Kinase Glo

The remaining 24 fragments were tested at 5 mM final concentration, against three different purification batches of hGALK1 in the Kinase Glo assay. The effect on hGALK1 activity by fragments was generally very weak, as reflected by the high standard deviation of luminescence signal measurements within and between batches (figure 6.7A). Consequently, the change in hGALK1 activity in the presence of 5 mM fragment was converted to strictly standardised mean

difference (SSMD) values, relative to the DMSO only control, to facilitate determination of the statistical significance of fragment effects (figure 6.7B).

SSMD is a parameter that directly measures the magnitude of difference between two data groups, considering the data variability in both compared groups. It was originally proposed as an alternative to *P* value or Z score based metrics in RNAi HTS screening [286]–[288]. SSMD (in this case between hGALK1 activity in the absence and presence of each fragment) is defined as the mean of the log fold change between the two groups penalised by the standard deviation of the log fold change. A higher value of SSMD reflects a higher probability that the difference between the two groups is greater than zero i.e. is statistically significant [288]. In this work, $3 > SSMD \geq 2$ was interpreted as a strong inhibitory effect, $2 > SSMD \geq 1.6$ as a fairly strong effect, $1.6 > SSMD \geq 1$ as a moderate effect, and $1 > SSMD \geq 0.5$ as a weak inhibitory effect. Negative SSMD values indicate activating effects.

The fragments binding at the entrance to the ATP binding site (fragments **1** and **2**) did not inhibit hGALK1 (figure 6.7B, blue bars). Rather, fragment **2** demonstrated a weak activating effect (*P* = 0.04; SSMD = -0.9), causing a 7 ± 4.6% increase in hGALK1 activity (mean of n = 9). Consequently, the benefit of these hits for lead development lies primarily in their potential linkage with inhibitors occupying the ATP binding site.

Of the three fragments binding near the termini of hGALK1 (figure 6.7B, yellow bars), two fragments had no significant effect on hGALK1 activity (fragments **3** and **5**) and one (fragment **4**) demonstrated a fairly strong inhibitory effect (*P* =0.0024; SSMD = 1.63) with 17 ± 1.6 % inhibition of hGALK1 activity (mean of n = 9) at 5 mM fragment. Since in hGALK1 crystals fragment **4** binds to the termini groove when the active site is preloaded with galactose and compound 3, it was important to verify its inhibition mode relative to both galactose and ATP and this is described in section 6.2.2.

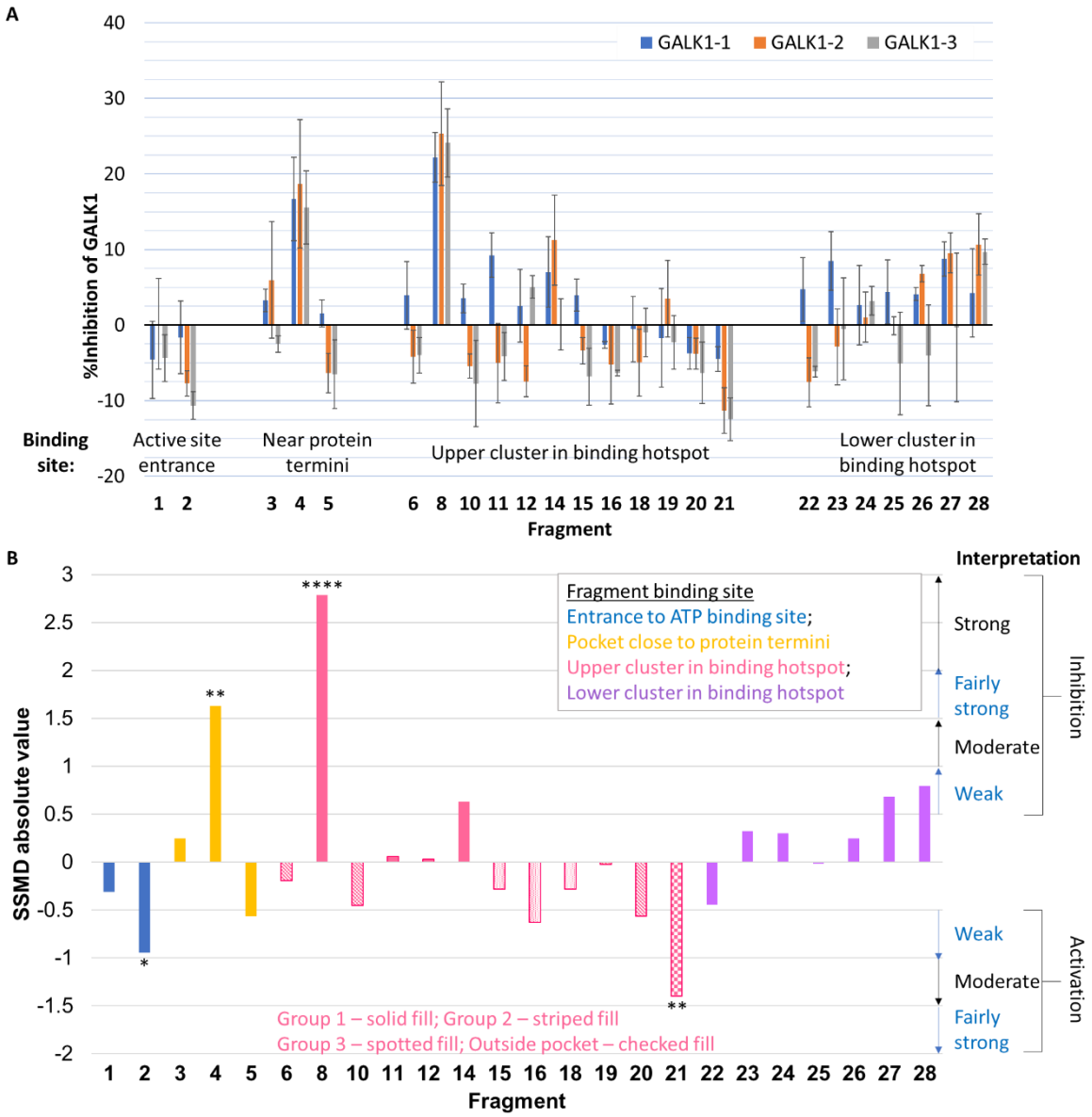


Figure 6.7: Change in hGALK1 activity in the presence of 5 mM fragment measured by Kinase Glo activity assay. *A. Bar chart showing %inhibition of three batches of hGALK1 (GALK1-1, GALK1-2 and GALK1-3) by 5 mM fragment. All reactions were carried out at 10 nM GALK1 with 35 µM ATP and 100 µM galactose with incubation for 1 hour before addition of the Kinase Glo reagent. Bars are coloured by protein batch and grouped by binding site. Error bars show standard deviation of technical triplicates. B. Bar chart showing SSMD of hGALK1 activity with 5 mM fragment relative to DMSO controls. Each bar represents the total SSMD calculated from technical triplicates of three protein batches (n=9). Bars are coloured by fragment binding site, as defined in section 5.3. Asterisks indicate statistical significance: * $P < 0.05$; ** $0.001 \leq P \leq 0.01$; **** $P \leq 0.0001$.*

As discussed in section 5.3.5, the 16 fragments belonging to the upper cluster of the binding hotspot can be further divided into three groups based on their binding mode within the pocket (figure 5.16). Group 1 fragments (**8**, **11**, **12**, **14**) are the most buried within the pocket and are the only fragments from the upper cluster that have positive SSMD values (figure 6.7B, solid pink bars). A weak, statistically non-significant inhibitory effect was observed with fragment **14** (*P*

=0.14; SSMD = 0.69) while **11** and **12** had no effect ($P > 0.5$; SSMD < 0.1) on hGALK1 activity. In contrast, a statistically significant, strong inhibitory effect was observed in the presence of **8** ($P < 0.0001$; SSMD = 2.79), with 24 ± 1.6 % inhibition of hGALK1 activity at 5 mM fragment. As mentioned for fragment **4**, investigation of the inhibition mode of fragment **8** was necessary to rule out orthosteric binding to free hGALK1 (section 6.2.2).

No statistically significant inhibition of hGALK1 was seen with group 2 (**6, 10, 19, 20**) or group 3 (**15**, **16**, **18**) fragments from the upper cluster. The SSMD values for these fragments ranged from -0.02 (fragment **19**) to -0.62 (fragment **16)** and *P* values were > 0.1 in all cases. The final fragment from the upper cluster, fragment **21**, is not part of the groupings described because it forms few direct interactions within the binding hotspot (figure 5.16B, pink sticks). Interestingly, it demonstrated a moderate activating effect (P =0.0059; SSMD = -1.40) with 9 ± 4.3 % increase in hGALK1 activity (mean of n = 9) at 5 mM fragment (figure 6.7B, pink checked bar).

Five of the seven fragments from the lower cluster of the binding hotspot fragments (fragments **22**-**26**) had SSMD values between -0.5 and 0.5, indicating minimal effect on hGALK1 activity (figure 6.7B, purple bars, $P > 0.2$). The remaining two fragments (fragments **27** and **28**) demonstrated a weak inhibitory effect on hGALK1 activity that was insufficient to reach statistical significance due to the observed variation between hGALK1 protein batches (SSMD 0.80 and 0.69 for **27** and **28** respectively with $P > 0.05$ in both cases).

6.2.2 Determination of inhibition mode for fragment hits **4** and **8**

The preclusion of the active site of hGALK1 in crystals used to soak fragments, in contrast to the free enzyme used in activity assays, made it essential to determine the inhibition mode of fragments **4** and **8** relative to galactose and ATP. The experimental approach used to determine the inhibition mode of fragments **4** and **8** was identical to that described for compound **T1** (section 6.1.6) except that 6 fragment concentrations (0 – 15 mM) were tested.

Fragment **4** was non-competitive with respect to galactose: there was a statistically significant,

dose-dependent decrease in $Vmax^{apparent}$ ($P < 0.0001$) but not $Km^{apparent}$ ($P = 0.2573$) in Michaelis-Menten plots (figure 6.8A), and lines of the Cornish-Bowden plot intersect on the x-axis (figure 6.8B). Fragment **4** was also non-competitive with respect to ATP: there was a statistically significant, dose-dependent decrease in $Vmax^{apparent}$ ($P < 0.0001$) but not $Km^{apparent}$ ($P = 0.9973$) in Michaelis-Menten plots (figure 6.8C) and most lines on the Cornish-Bowden plot intersect on the x-axis (figure 6.8D). Collectively, these results suggest that fragment **4** does not bind to the active site of hGALK1 nor impede binding of substrates to hGALK1.

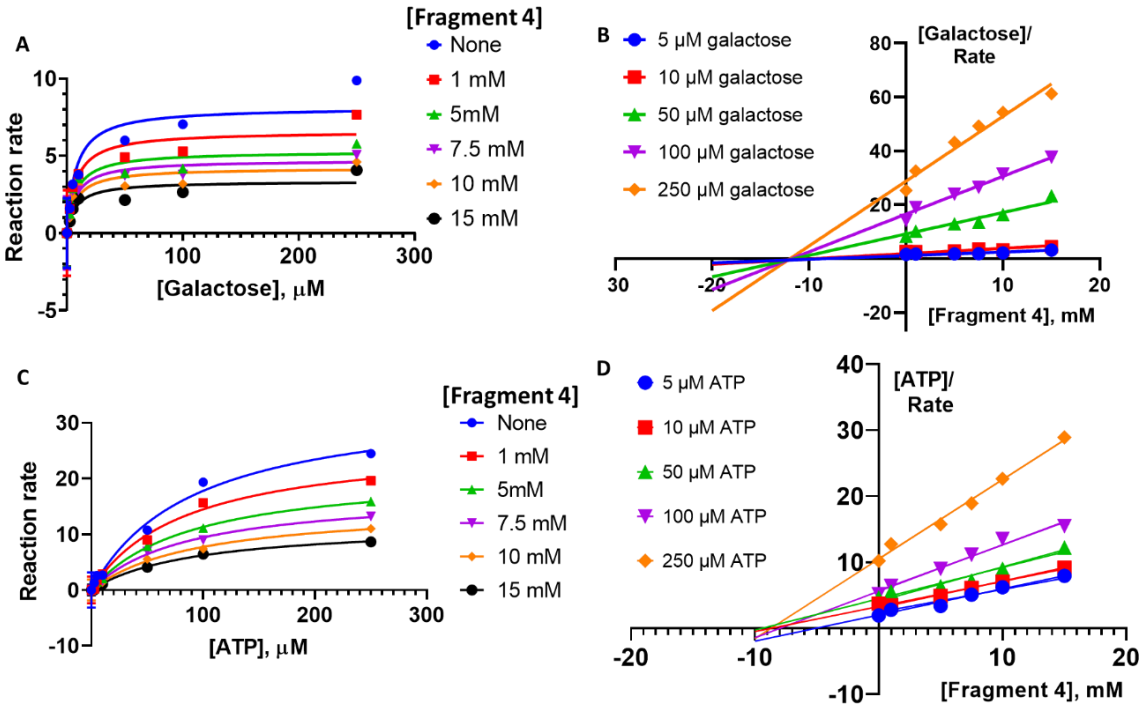


Figure 6.8: Kinetic characterisation of hGALK1 inhibition by fragment 4. *A. Michaelis-Menten plot of [galactose] against hGALK1 reaction rate (μM ATP consumed) without fragment* ***4*** *(blue) or in the presence of 1 mM (red), 5 mM (green), 7.5 mM (purple), 10 mM (orange) or 15 mM (black) fragment* ***4****. GALK1 reaction was performed at 10 nM GALK1 with 35 μM ATP for 1 hour and subsequently quantified using the described Kinase Glo reagent. B. Cornish-Bowden plot of [galactose]/GALK1 reaction rate (as defined in panel A) against fragment* ***4*** *concentration at 5 μM (blue), 10 μM (red), 50 μM (green), 100 μM (purple) and 250 μM (orange) galactose. Data are transformations from panel A. C. Michaelis- Menten plot of [ATP] against hGALK1 reaction rate (μM galactose consumed) without fragment* ***4*** *(blue) or in the presence of 1 mM (red), 5 mM (green), 7.5 mM (purple), 10 mM (orange) or 15 mM (black) of fragment* ***4****. GALK1 reaction was performed at 250 nM GALK1 with 50 μM galactose for 1 hour and subsequently quantified using the described Amplex Red reagent. D. Cornish-Bowden plot of [ATP]/GALK1 reaction rate (as defined in panel C) against fragment* ***4*** *concentration at 5 μM (blue), 10 μM (red), 50 μM (green), 100 μM (purple) and 250 μM (orange) ATP. Data are transformations from panel C. All experiments were performed as technical triplicates of biological duplicates and data points plotted were average values.*

Fragment **8** showed mixed inhibition with respect to galactose: there was a statistically significant, dose-dependent decrease in $Vmax^{apparent}$ ($P < 0.0001$) and a 6-fold increase in $Km^{apparent}$ ($P < 0.0001$) in Michaelis-Menten plots (figure 6.9A) and lines on the Cornish-Bowden plot intersect below the x-axis (figure 6.9B). Fragment **8** was non-competitive with respect to ATP: there was a statistically significant, dose-dependent decrease in $Vmax^{apparent}$ ($P < 0.0001$) but not $Km^{apparent}$ ($P = 0.4932$) in Michaelis-Menten plots (figure 6.9C) and most lines on the Cornish-Bowden plot intersect on the x-axis (figure 6.9D). Collectively, these results suggest that fragment **8** does not bind to the active site of hGALK1 and reduces the affinity of hGALK1 for galactose but not ATP.

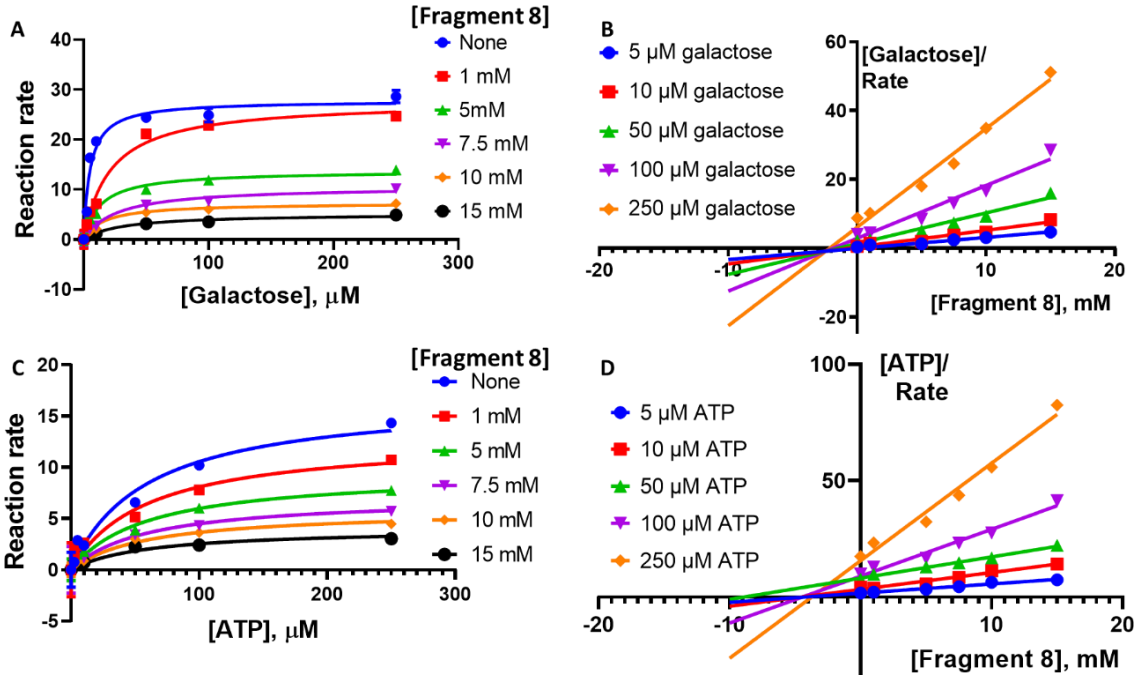


Figure 6.9: Kinetic characterisation of hGALK1 inhibition by fragment 8. *A. Michaelis-Menten plot of [galactose] against hGALK1 reaction rate (μM ATP consumed) without fragment* ***8*** *(blue) or in the presence of 1 mM (red), 5 mM (green), 7.5 mM (purple), 10 mM (orange) or 15 mM (black) of fragment* ***8****. GALK1 reaction was performed at 10 nM hGALK1 with 35 μM ATP for 1 hour and subsequently quantified using the described Kinase Glo reagent. B. Cornish-Bowden plot of [galactose]/GALK1 reaction rate against fragment* ***8*** *concentration at 5 μM (blue), 10 μM (red), 50 μM (green), 100 μM (purple) and 250 μM (orange) galactose. Data are transformations from panel A. C. Michaelis- Menten plot of [ATP] against hGALK1 reaction rate (μM galactose consumed) without fragment* ***8*** *(blue) or in the presence of 1 mM (red), 5 mM (green), 7.5 mM (purple), 10 mM (orange) or 15 mM (black) of fragment* ***8****. GALK1 reaction was performed at 250 nM hGALK1 with 50 μM galactose for 1 hour and subsequently quantified using the described Amplex Red reagent. D. Cornish-Bowden plot of [ATP]/GALK1 reaction rate against fragment* ***8*** *concentration at 5 μM (blue), 10 μM (red), 50 μM (green), 100 μM (purple) and 250 μM (orange) ATP. Data are transformations from panel C. All experiments were performed as technical triplicates of biological duplicates and data points plotted were average values.*

6.3 Merging of fragments from the binding hotspot

After confirming that some fragments inhibit hGALK1 by binding to a non-orthosteric site, the next step was to improve their weak potency, generating better compounds that can be more accurately analysed for structure-activity relationships.

Three approaches have been described for the optimisation of fragment hits: growing, linking and merging. Single fragment hits can be grown by addition of chemical groups that have complementarity with nearby protein residues (figure 6.10A). This can be achieved by screening of a library of compounds containing the fragment hit as their core scaffold (e.g. optimisation of HAO1 fragment hits as described in chapter 4) [289]. Alternatively, if two or more fragment hits are found in close proximity, they can be linked (e.g. by a substituted acyl chain) to form a larger molecule that preserves the interacting groups of both fragments (figure 6.10B) [290]. Finally, when the binding poses of two or more fragments overlap, they can be merged to generate a molecule incorporating both binding poses e.g. by incorporating ring substituents from different fragments with overlapping ring moieties into a single compound (figure 6.10C) [291].

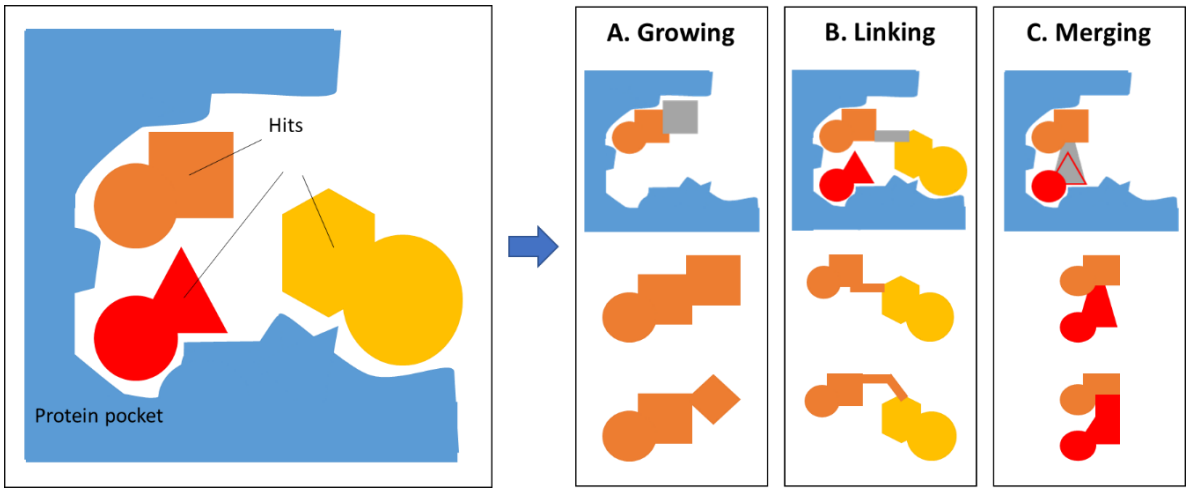


Figure 6.10: Illustration of approaches to fragment optimisation. *A. Optimisation by growing a fragment hit to interact with nearby complementary protein residues. B. Optimisation by linking two or more fragments located within the same binding pocket. C. Optimisation by merging two or more fragments that bind in close proximity within the same binding pocket.*

For hGALK1, the fragment merging approach was adopted to combine fragment hits from the upper and lower clusters of the binding hotspot, because many of these fragments overlap in their binding positions (as illustrated in figures 5.15-5.16). Importantly, the most potent

fragment, **8**, belongs to the upper cluster and represents a good subject for merging with fragments from the lower cluster.

6.3.1 Design and selection of merged compounds

The superimposed structures of hGALK1-galactose-**T3** bound to fragment **8** and to the lower cluster of fragments in the binding hotspot (fragments **22**-**28**) were inspected for shared chemical groups (figure 6.11A-B). Chemical structures were drawn to reflect combinations of these shared groups (figure 6.11C) and commercial libraries were searched to find and select available compounds with similar structures (figure 6.11D).

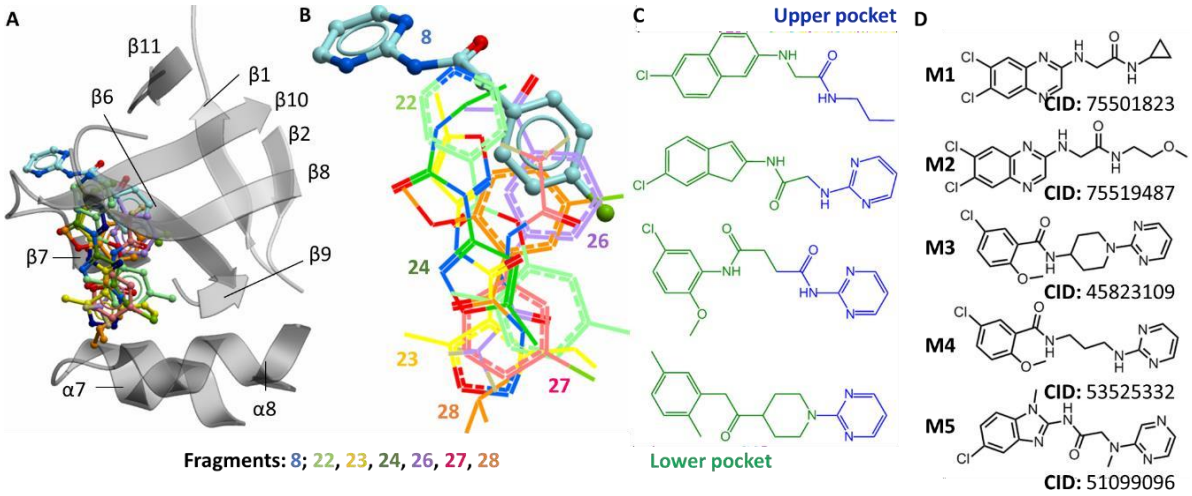


Figure 6.11: Merging of fragments from the binding hotspot. *A. Close-up view of the binding hotspot showing how the selected fragments overlap within the binding site. Fragments are shown as sticks. Secondary structures that form the binding hotspot are shown as grey ribbon. The selected viewpoint is that used in the figures of section 5.3.5. B. Close-up of overlaid fragments, in the same orientation as panel A, showing overlapping chemical groups suitable for merging. The upper cluster fragment,* ***8****, is shown as sticks and the six lower cluster fragments,* ***22-24****,* ***26-28****, are shown as lines. C. Chemical structures of theoretical 'merged' compounds that were used as search models to find commercially available analogues. The structures are coloured to reflect which cluster in the binding hotspot they originate from. D. Chemical structures of the five compounds purchased for in vitro testing (****M1-M5****).*

Five of these compounds (**M1**-**M5**) were purchased for testing *in vitro*. Three compounds (**M1**, **M2** and **M5**; figure 6.11D) consist of two fused, aromatic rings, to occupy the position where the lower cluster of fragments bind, and an acyl urea linkage extending towards where the upper cluster of fragments bound, ending in a cyclopropane (**M1**), methoxy (**M2**) or pyrimidine (**M5**). The fused ring of **M1** and **M2** is a dichloroquinoxaline whereas a smaller constituent, a 5- chloro-1-methyl-*1H*-benzoimidazole, is present in **M5**. The two other compounds (**M3** and **M4**) combine

the 1-chloro-4-methoxybenzene moiety of fragment **28** and the surface exposed pyrimidine of fragment **8** separated by either a piperidine ring (**M3**) or an amide linker (**M4**). The five compounds were first tested at 2 mM in both activity assays (section 6.3.2) before characterisation of the three hGALK1 inhibitors (**M1**-**M3**; section 6.3.3), one hGALK1 activator (**M4**; section 6.3.4) and one mixed effector of hGALK1 (**M5**; section 6.3.5).

6.3.2 Single concentration screen of merged compounds in hGALK1 activity assays

Inhibition of three different purification batches of hGALK1 (GALK1-1, GALK1-2 and GALK1-3) by 2 mM of each merged compound was determined in Kinase Glo and Amplex Red assays using the assay parameters described in the previous sections. Four of the five merged compounds (**M1**, **M2**, **M3** and **M5**) showed >30% inhibition of hGALK1 activity in both assays when tested at 2 mM, representing immediate improvement relative to fragment **8** (figure 6.12).

In the Kinase Glo assay, the maximal inhibition observed (at 2 mM fragment) was 80% for **M1**, 90% for **M2**, 62% for **M3** and 53% for **M5** (figure 6.12A). Measurement in the Amplex Red assay showed very similar results for these four compounds (figure 6.12B).

Surprisingly, one compound (**M4**) caused a >30% increase in hGALK1 activity. **M4** showed a maximal activation of 204% at 2 mM compound in the Kinase Glo assay (figure 6.12A). In contrast the maximal activation of hGALK1 by **M4** observed in the Amplex Red assay was only 36% (figure 6.12B). The 5.7-fold greater activation of hGALK1 by **M4** observed under the Kinase Glo assay conditions, compared to the activation observed under the Amplex Red assay conditions, may be attributed to either the 25-fold lower protein concentration, the 2.9-fold lower ATP concentration or the 2-fold higher galactose concentration in the Kinase Glo assay relative to the Amplex Red assay, or some combination of these factors. The latter two factors were explored further by characterising the substrate-dependence of hGALK1 activation by **M4** (section 6.3.4).

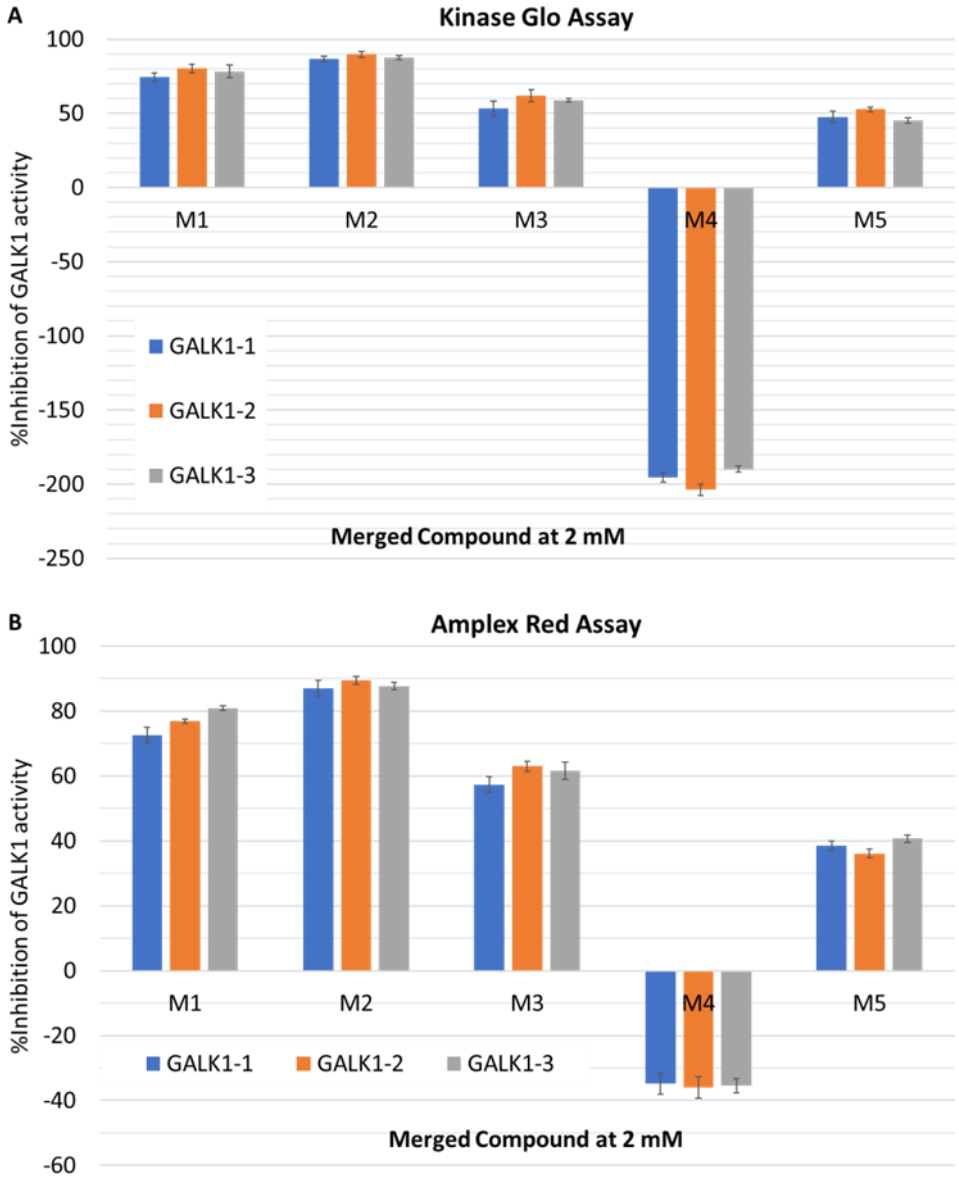


Figure 6.12: Measurement of hGALK1 activity in the presence of 2 mM merged compounds M1-M5. *A. Plot of Kinase Glo inhibition data showing the % inhibition of three batches of hGALK1 by 2 mM* ***M1- M5****. GALK1 reaction was performed at 10 nM hGALK1 with 35 µM ATP and 100 µM galactose for 1 hour before quantification using the described Kinase Glo assay. B. Plot of Amplex Red inhibition data showing the % inhibition of three batches of hGALK1 by 2 mM of* ***M1-M5****. GALK1 reaction was performed at 250 nM hGALK1 with 100 µM ATP and 50 µM galactose for 1 hour before quantification using the described Amplex Red assay. Three technical replicates were measured for each batch and error bars represent standard deviations.*

6.3.3 Characterisation of merged compounds **M1-M3** that inhibit hGALK1

To test whether merged compounds **M1**, **M2** and **M3** inhibited hGALK1 in a concentration-dependent manner, inhibition of two batches of hGALK1 (GALK1-1 and GALK1-2) at 12 compound concentrations from 0-2 mM was measured in both assays. Concentration-dependent inhibition of hGALK1 was observed by **M1**, **M2** and **M3** in both the Kinase Glo (figure 6.13A) and the Amplex Red (figure 6.13B) assays. Sigmoidal curves, consistent with hGALK1

inhibition proportional to compound concentration, were fitted to all six datasets (technical triplicates of two batches of hGALK1 tested for each of the three compounds) with high confidence ($R^2 > 0.95$).

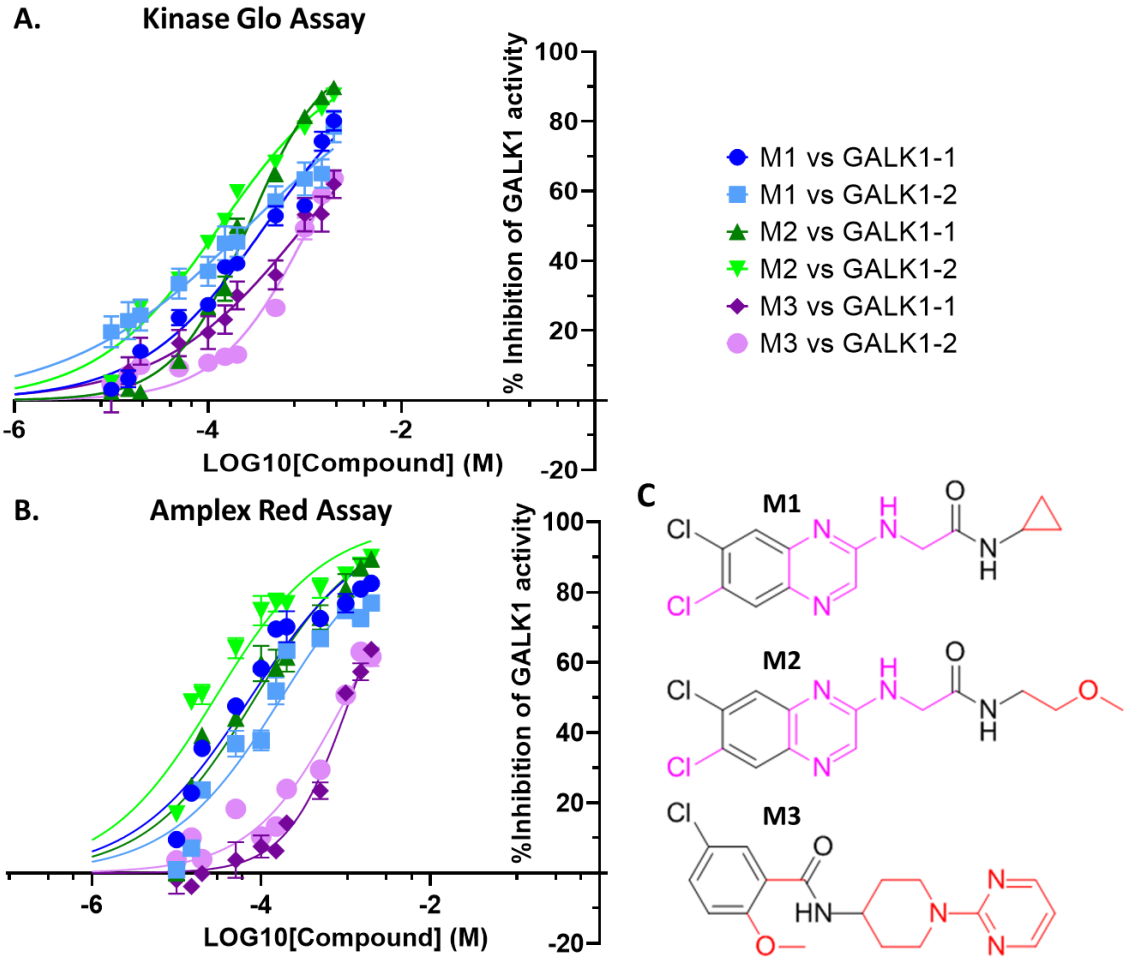


Figure 6.13: Characterisation of in vitro inhibition of hGALK1 by merged compounds M1-M3. *A. Plot of log concentration of **M1-M3** versus %inhibition of hGALK1 activity observed in the Kinase Glo assay (A) and in the Amplex Red assay (B). In the former, hGALK1 reaction was performed at 10 nM hGALK1 with 35 µM ATP and 100 µM galactose for 1 hour before addition of Kinase Glo reagent. In the latter, hGALK1 reaction was performed at 250 nM hGALK1 with 100 µM ATP and 50 µM galactose for 1 hour before addition of Amplex Red reagent. Concentration-response of **M1** is in blue, **M2** is in green and **M3** is in purple. Lighter shades depict inhibition of hGALK1-1 and darker shades depict inhibition of hGALK1-2. C. Chemical structures of merged compounds **M1-M3**. Chemical groups unique to a compound are coloured red, chemical groups shared by two compounds are pink and chemical groups shared by all three compounds are black. Error bars are standard deviation of technical triplicates.*

For compounds **M1** and **M2**, hGALK1 inhibition begins to plateau above 500 µM in the Kinase Glo assay with less than 15% additional inhibition at 4-fold higher concentrations (i.e. at 2 mM; figure 6.13A). This plateau is not observed in the curves fitted to the Amplex Red data (figure 6.13B), which is consistent with the curve plateau being caused by saturation of the available

hGALK1 binding sites, since the lower enzyme concentration in the Kinase Glo assay translates to a smaller number of inhibitor binding sites that must be occupied to achieve saturation. **M3** demonstrated a lower potency against hGALK1 than **M1** and **M2** in both assays and no plateau was observed after curve fitting.

The calculated IC_{50}s of **M1** against hGALK1-1 (dark blue) and GALK-2 (light blue) were 209 μM (pIC_{50} 3.7 ± 0.1) and 116 μM (pIC_{50} 3.9 ± 0.2) respectively in the Kinase Glo assay (figure 6.13A; LLE_{AT} 0.19 kcal/mol) and 31 μM (pIC_{50} 4.5 ± 0.1) and 65 μM (pIC_{50} 4.2 ± 0.1) respectively in the Amplex Red assay (figure 6.13B; LLE_{AT} 0.22 kcal/mol). The calculated IC_{50}s of **M2** against hGALK1-1 (dark green) and GALK-2 (light green) were 198 μM (pIC_{50} 3.7 ± 0.1) and 96 μM (pIC_{50} 4.0 ± 0.1) respectively in the Kinase Glo assay (figure 6.13A; LLE_{AT} 0.23 kcal/mol) and 45 μM (pIC_{50} 4.3 ± 0.1) and 25 μM (pIC_{50} 4.6 ± 0.1) respectively in the Amplex Red assay (figure 6.13B; LLE_{AT} 0.27 kcal/mol). Accurate calculation of IC_{50} values was not possible for **M3** due to its weaker potency, but 50% inhibition was achieved by approximately 1 mM **M3** in both assays (purple lines in figures 6.13A and B).

The inhibition mode of merged compounds **M1-M3** with respect to galactose and ATP were determined from Kinase Glo and Amplex Red assays respectively, using Michaelis-Menten and Cornish-Bowden plots as described previously (sections 6.1.6 and 6.2.2). A statistically significant, concentration-dependent decrease in $Vmax^{apparent}$ ($P < 0.001$) was observed in both galactose and ATP concentration arrays with increasing concentrations of **M1** (figure 6.14A, 6.14C), **M2** (figure 6.14E, 6.14G) and **M3** (figure 6.14I, 6.14K), suggesting that they were not competitive inhibitors and therefore do not bind to the active site of hGALK1.

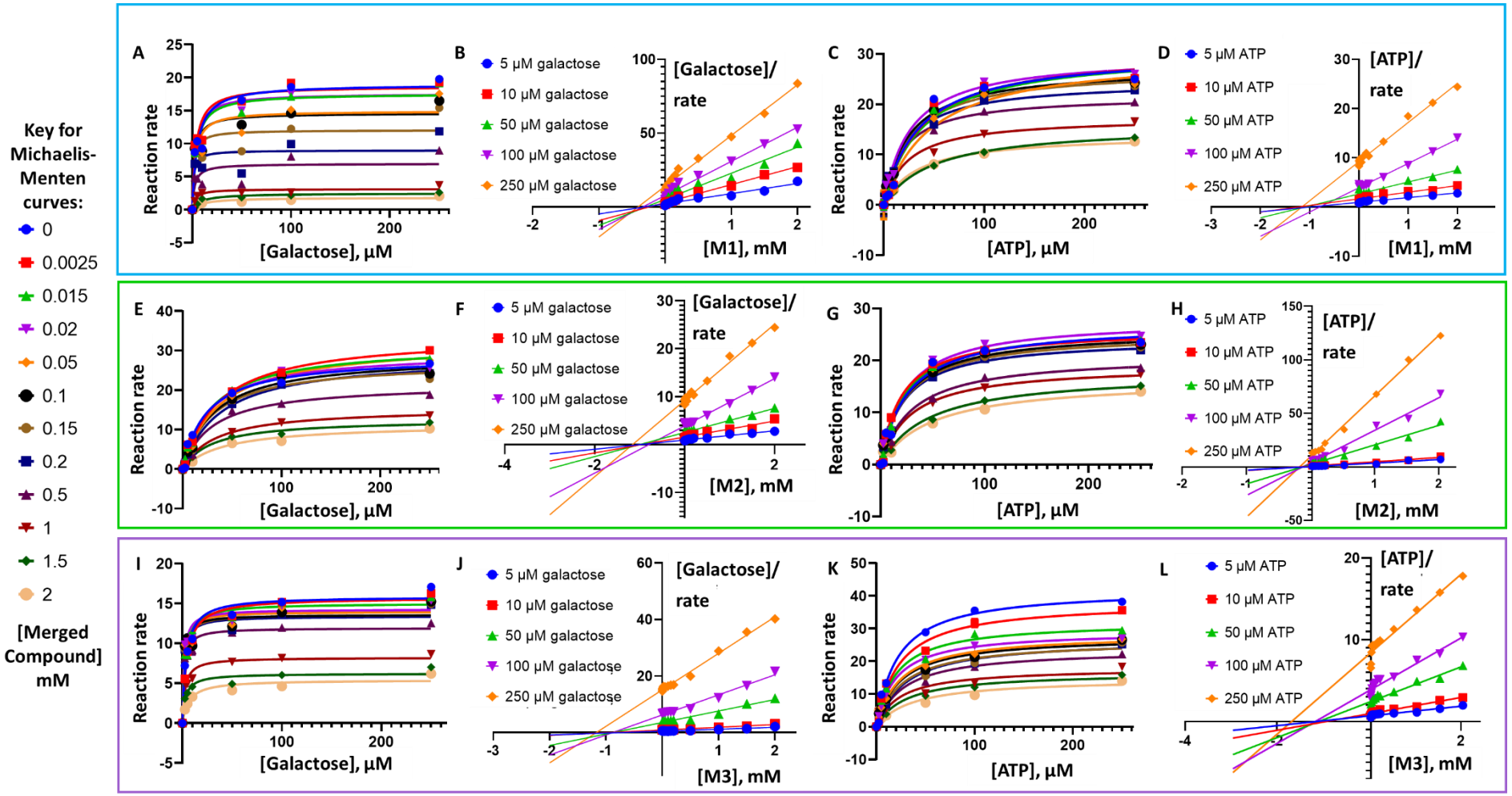


Figure 6.14: Kinetic characterisation of hGALK1 inhibition by merged compounds M1-M3. *Michaelis-Menten plots of hGALK1 reaction rate (µM ATP consumed) against [galactose] in the presence of increasing concentrations of* ***M1*** *(A),* ***M2*** *(E) or* ***M3*** *(I), determined in the Kinase Glo assay. Cornish-Bowden plots of [Galactose]/Rate (as defined in panels A, E and I) against concentration of* ***M1*** *(B),* ***M2*** *(F) or* ***M3*** *(J), determined in the Kinase Glo assay. Michaelis-Menten plots of hGALK1 reaction rate (µM galactose consumed) against [ATP] in the presence of increasing concentrations of* ***M1*** *(C),* ***M2*** *(G) or* ***M3*** *(K), determined in the Amplex Red assay. Cornish-Bowden plots of [ATP]/Rate (as defined in panels B, F and J) against concentration of* ***M1*** *(D),* ***M2*** *(H) or* ***M3*** *(L), determined in the Amplex Red assay. All experiments were performed as technical triplicates of biological duplicates and plotted data points are average values. Kinase Glo assay was performed with 10 nM hGALK1 and 35 µM ATP. Amplex Red assay was performed with 250 nM hGALK1 and 50 µM galactose.*

No statistically significant change in $Km^{apparent}$ for galactose (P >0.1) was observed in Michaelis Menten plots of reaction rate against galactose with increasing concentrations of **M1** (figure 6.14A) or **M2** (figure 6.14E), and lines on the Cornish-Bowden plot primarily intersected on the x-axis (figures 6.14B and 6.14F for compounds **M1** and **M2** respectively), suggesting that compounds **M1** and **M2** were non-competitive with respect to galactose. **M3** showed mixed type inhibition with respect to galactose, with a statistically significant, concentration-dependent increase in $Km^{apparent}$ for galactose and intersection below the x-axis for most lines on the Cornish-Bowden plot (P = 0.02; figure 6.14I, 6.14J).

No statistically significant change in $Km^{apparent}$ for ATP (P >0.1) was observed when increasing concentrations of **M1** (figure 6.14C), **M2** (figure 6.14G) or **M3** (figure 6.14K) and lines on the respective Cornish-Bowden plots primarily intersected on the x-axis, indicating that compounds **M1**, **M2** and **M3** were non- competitive with respect to ATP (figures 6.14D, 6.14H and 6.14L).

6.3.4 Characterisation of merged compound **M4** in hGALK1 activity assays

The identification of **M4** as an activator of hGALK1 in the single point assay (section 6.3.2) was puzzling, given its similar chemical structure to the inhibitory compounds **M1**-**M3**. In general terms, an activator may either increase the catalytic efficiency (substrate turnover) of its target, increase the target's affinity for the substrate, or some combination of the two [292]. In a dual substrate system such as hGALK1 an activator may affect the kinetic parameters of one or both substrates [275]. This section describes the characterisation of hGALK1 activation by **M4**.

To test whether **M4** activates hGALK1 in a concentration-dependent manner, the increase in activity of two batches of hGALK1 (GALK1-1 and GALK1-2) by 12 compound concentrations from 0-2 mM was measured in both assays. Concentration-dependent activation of hGALK1 was observed in both Kinase Glo (figure 6.15A) and Amplex Red (figure 6.15B) assays. An approximately 5-fold greater increase in hGALK1 activity was observed in the Kinase Glo assay

relative to the Amplex Red assay, as observed in the single point assay (section 6.3.2, figure 6.12).

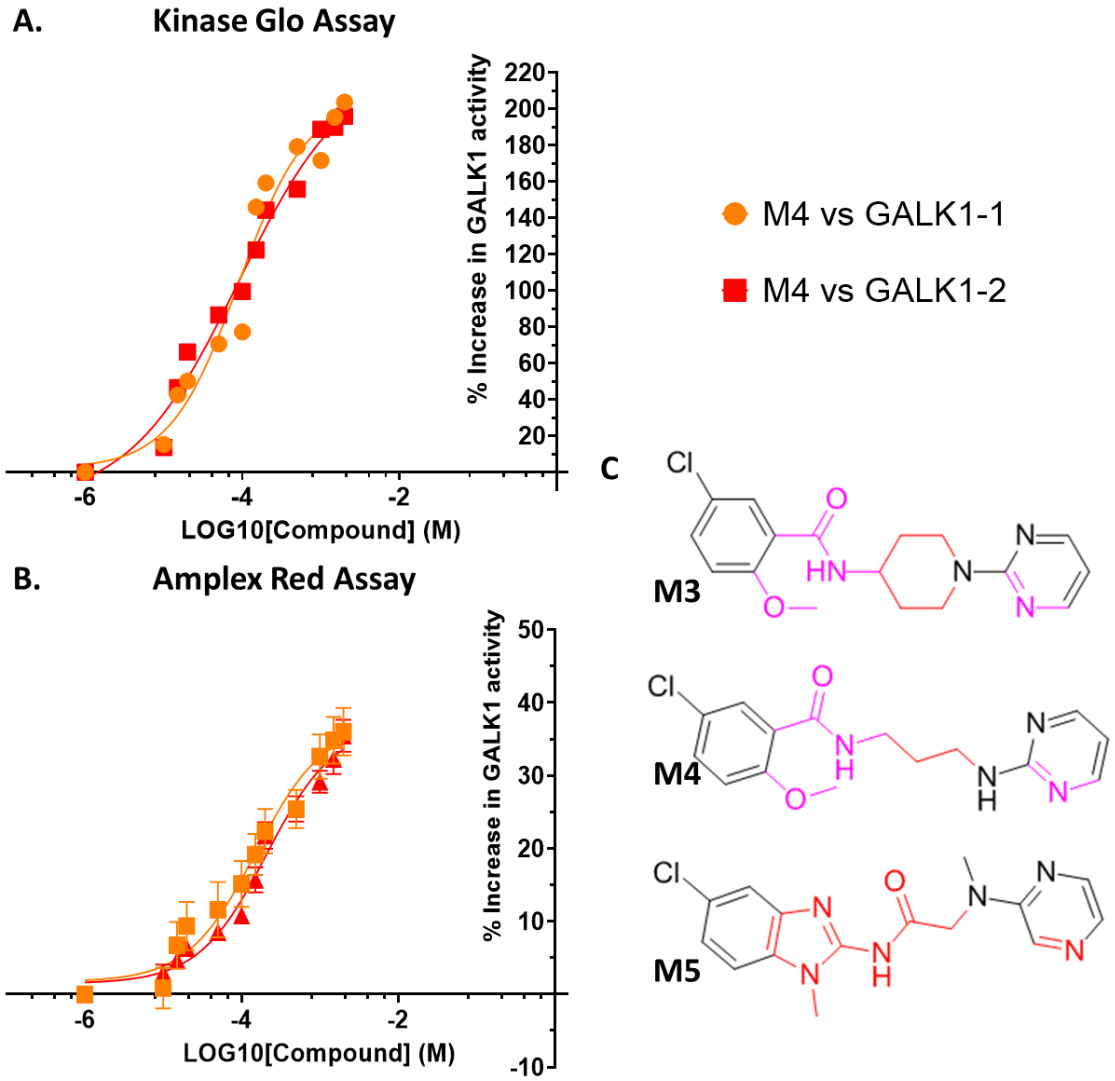


Figure 6.15: In vitro activation of hGALK1 by merged compound M4. *A. Plot of log concentration of* ***M4*** *versus %increase in hGALK1 activity observed in Kinase Glo assay. Reactions were carried out at 10 nM hGALK1, 35 µM ATP and 100 µM galactose. Concentration-response of GALK1-1 batch is in orange and concentration-response of GALK1-2 batch is in red. B. Plot of log concentration of* ***M4*** *versus %increase in hGALK1 activity observed in Amplex Red assay. Reactions were carried out at 250 nM hGALK1, 100 µM ATP and 50 µM galactose. Concentration-response of GALK1-1 batch is in orange and concentration -response of GALK1-2 batch is in red. Error bars are standard deviation of technical triplicates. C. Chemical structures of merged compounds* ***M3-M5****. Chemical groups unique to one compound are coloured red, chemical groups shared by two compounds are pink and chemical groups shared by all three compounds are black.*

Sigmoidal curves, consistent with hGALK1 activation proportional to compound concentration, were fitted to all datasets (two per compound) with high confidence ($R^2 > 0.96$). No plateau was observed in either assay, indicating that **M4** did not exert its maximal effect at 2 mM. The

calculated EC_{50}s of **M4** against hGALK1-1 (orange) and GALK-2 (red) were 90 μM and 82 μM respectively in the Kinase Glo assay (figure 6.15A) and 152 μM and 209 μM respectively in the Amplex Red assay (figure 6.15B). It is interesting to note the similarity in EC_{50}s between the two assays, despite the large difference in the observed magnitude of effect.

This difference in the observed magnitude of effect between the two assays could be due to the difference in protein concentration, galactose concentration or ATP concentration, as mentioned previously. The higher protein concentration in the Amplex Red assay means that the ratio of hGALK1 molecules to **M4** molecules in the reaction mixture is higher, so that the same number of binding events would represent a lower overall percentage of activated protein. This is supported by the lack of saturation observed in the dose-response curves (figure 6.15), indicating hGALK1 was not saturated with **M4** at the maximum tested concentration.

Reduced efficacy of an allosteric activator at high substrate concentration has previously been reported for glucokinase, for which a number of activators were found to be less active or not active at all at high glucose concentrations [293]. To determine whether this was true for hGALK1 activation by **M4**, the %increase in hGALK1 activity by **M4** was measured at different substrate concentrations. **M4** efficacy is significantly higher at lower ATP concentrations (figure 6.16A), whereas there is no apparent relationship between **M4** efficacy and galactose concentration (figure 6.16B).

These data were also analysed in Michaelis-Menten plots and by the complementary linear method, Lineweaver-Burk plots of 1/reaction rate against 1/substrate concentration. Lineweaver-Burk plots were chosen instead of Cornish-Bowden plots because Cornish- Bowden plots of these data were not linear. In the Lineweaver-Burk plot, the y-intercept is equal to 1/Vmax and the x-intercept is equal to -1/Km. Consequently, inhibition or activation mode can be determined from the comparison of x- and y- intercepts when lines are plotted at different compound

concentrations [294], [295].

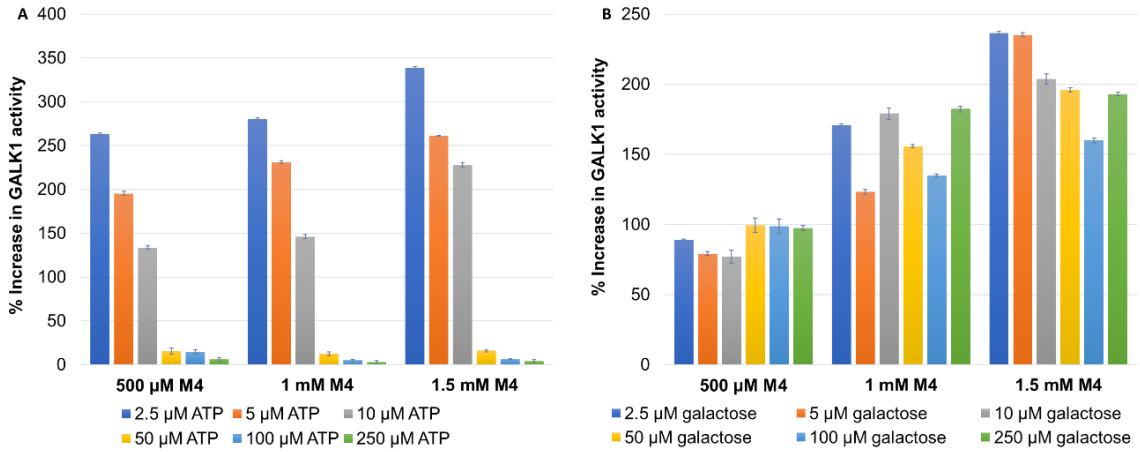


Figure 6.16: Characterisation of the substrate-dependence of hGALK1 activation by M4. *A. Bar chart showing %increase in hGALK1 activity by 500 µM, 1 mM or 1.5 mM* ***M4*** *in the presence of 2.5 (blue), 5 (orange), 10 (grey), 50 (yellow), 100 (light blue) or 250 (green) µM of ATP. hGALK1 reaction was performed at 250 nM hGALK1 with 50 µM galactose for 1 hour before quantification using the Amplex Red reagent as described. B. Bar chart showing %increase in hGALK1 activity by 500 µM, 1 mM or 1.5 mM* ***M4*** *in the presence of 2.5 (blue), 5 (orange), 10 (grey), 50 (yellow), 100 (light blue) or 250 (green) µM of galactose. hGALK1 reaction was performed at 10 nM hGALK1 with 35 µM ATP for 1 hour before quantification using the Kinase Glo reagent as described. Error bars are standard deviation of three technical replicates.*

When varying galactose concentrations, the Michaelis-Menten plot (figure 6.17A) show a dose-dependent increase in $Vmax^{apparent}$ (P = 0.005) and decrease in $Km_{Galactose}$ (P = 0.004) by compound **M4,** and the lines of the Lineweaver-Burk plot have different x- and y- intercepts (figure 6.17B). When varying ATP concentrations, Michaelis-Menten plots (figure 6.17C) show unchanged $Vmax^{apparent}$ values (P = 0.95) and a dose-dependent decrease in $Km^{apparent}$ for ATP (P = 0.004), and the lines of the Lineweaver-Burk plot have the same y-intercept but different x-intercepts (figure 6.17D). Collectively, these data suggest that **M4** activates hGALK1 by increasing the affinity of hGALK1 for both galactose and ATP (reflected in decreasing $Km^{apparent}$ values for both substrates), and by increasing the turnover of galactose (reflected in increasing $Vmax^{apparent}$ at saturating concentrations of galactose).

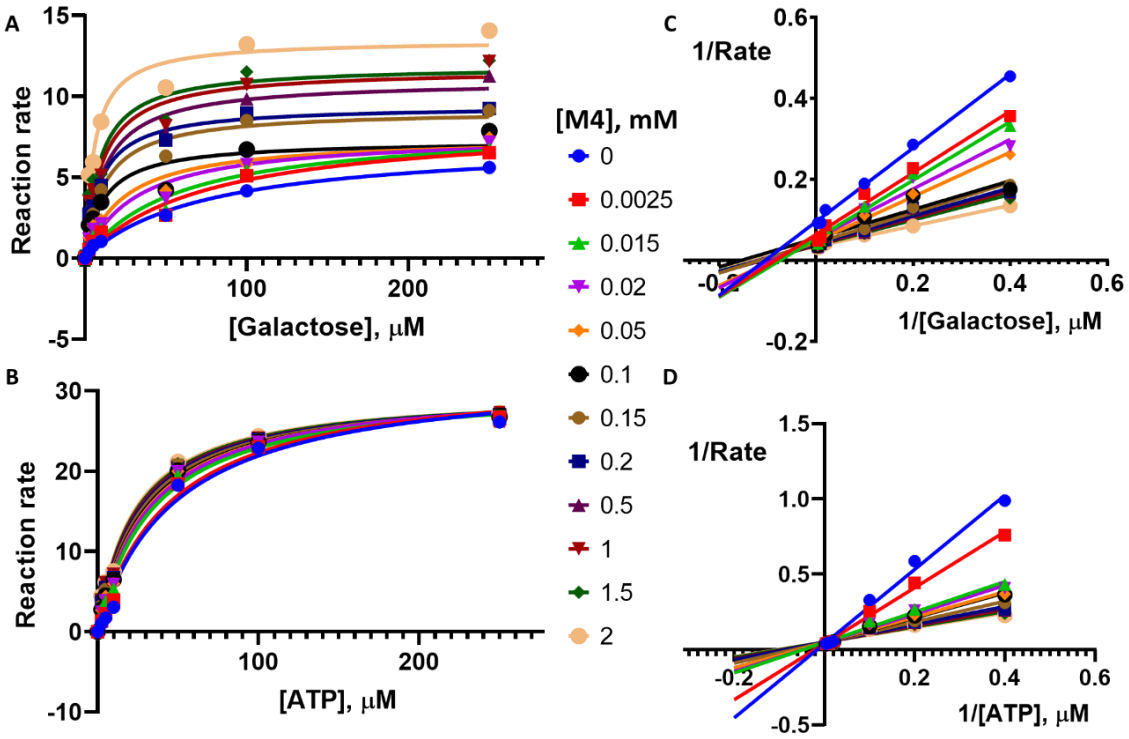


Figure 6.17: Kinetic characterisation of hGALK1 activation by M4. *A. Michaelis-Menten plot of hGALK1 reaction rate (μM ATP consumed) against [galactose] in the presence of increasing concentrations of* ***M4****. hGALK1 reaction was performed at 10 nM hGALK1 with 35 μM ATP for 1 hour before quantification using the Kinase Glo reagent as described. B. Lineweaver-Burk plot of 1/reaction rate (as defined in panel A) against 1/[Galactose] in the presence of increasing concentrations of* ***M4****. C. Michaelis-Menten plot of hGALK1 reaction rate (μM galactose consumed) against [ATP] in the presence of increasing concentrations of* ***M4****. hGALK1 reaction was performed at 250 nM hGALK1 with 50 μM galactose for 1 hour before quantification using the Amplex Red reagent as described. D. Lineweaver-Burk plot of 1/reaction rate (as defined in panel C) against 1/[ATP] in the presence of increasing concentrations of* ***M4****. All experiments were performed as technical triplicates of biological duplicates and plotted data points are average values.*

A possible interpretation of these results would be that binding of **M4** causes a conformational change of apo-GALK1 to the same, or a similar, conformation as the ATP-bound enzyme. This is supported by published kinetic studies showing that ATP binds to hGALK1 first, increasing the enzyme's affinity for galactose [136], [138], [249]. It is also consistent with the finding in this work that activation by **M4** declines to non-significant levels in the presence of saturating concentrations of ATP (figure 6.16A).

6.3.5 Characterisation of merged compound **M5** in hGALK1 activity assays

M5 behaves as an activator of hGALK1 at 10 – 100 μM, as a mixed activator and inhibitor at 150 and 200 μM, and as an inhibitor at ≥ 500 μM when concentration-response curves are fitted to

Kinase Glo (figure 6.18A) and Amplex Red (figure 6.18B) assay data.

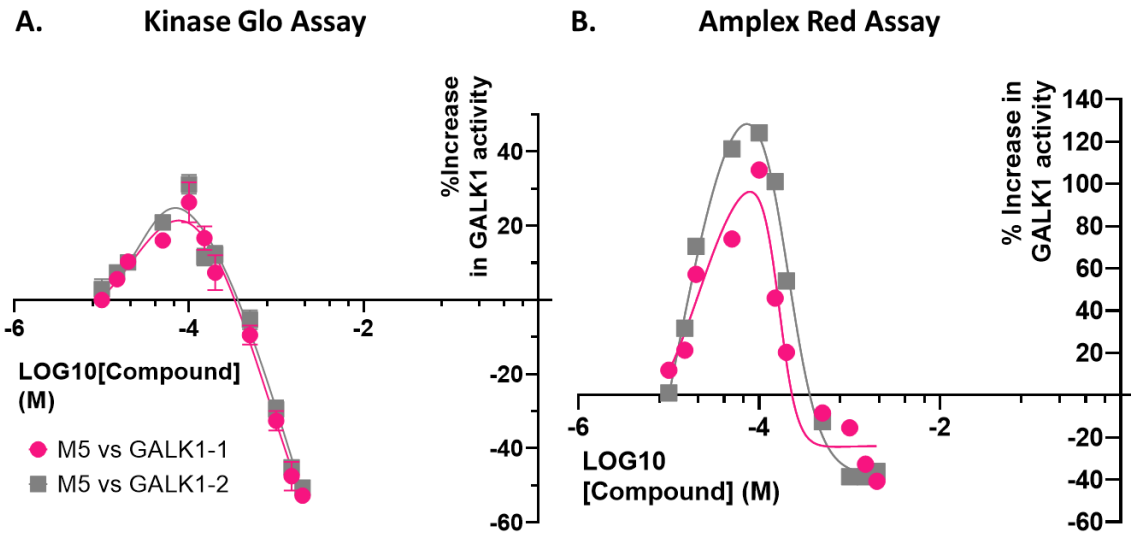


Figure 6.18: Characterisation of in vitro activation and inhibition of hGALK1 by merged compound M5. *A. Plot of log concentration of **M5** versus % increase of hGALK1 activity observed in Kinase Glo assay. Concentration-response of GALK1-1 is in pink and concentration-response of GALK1-2 is in grey. hGALK1 reaction was performed at 10 nM hGALK1 with 35 µM ATP and 100 µM galactose for 1 hour before quantification using the Kinase Glo reagent as described. B. Plot of log concentration of **M5** versus % increase of hGALK1 activity observed in Amplex Red assay. Concentration-response of GALK1-1 is in pink and concentration-response of GALK1- 2 is in grey. hGALK1 reaction was performed at 250 nM hGALK1 with 100 µM ATP and 50 µM galactose for 1 hour before quantification using the Amplex Red reagent as described. Error bars are standard deviation of technical triplicates.*

There are three mechanistic possibilities to explain an effector that stimulates an enzyme's activity at low concentrations and inhibits an enzyme's activity at high concentrations. Firstly, this pattern can occur where the compound represses an endogenous inhibition mechanism when present at low concentrations e.g. by competing with substrate to prevent non-productive substrate binding [296]. Secondly, sub-inhibitory concentrations of competitive inhibitors can bind to one subunit of an oligomeric protein to increase the affinity of the other subunits for the substrate [297]. These two possibilities are not applicable to hGALK1 as it is a monomer with no known feedback inhibition [24], [298], [299].

The third mechanism is through binding of the effector to two different sites, usually an allosteric site leading to activation and an orthosteric site leading to inhibition [300]. In this mechanism, binding of the compound to the activating site is either of higher affinity or more rapid than binding to the inhibitory site, so that at low concentrations of the effector only activation is

observed. In terms of hGALK1 and **M5**, this mechanism could involve **M5** binding at the hotspot to activate hGALK1 when present at low concentrations and binding at the active site when present at higher concentration, resulting in inhibition of hGALK1. Binding at the active site could be envisaged with the adenine-like moiety of **M5** mediating binding in the ATP binding site. This was investigated by determining the inhibition mode at high **M5** concentrations (figure 6.19).

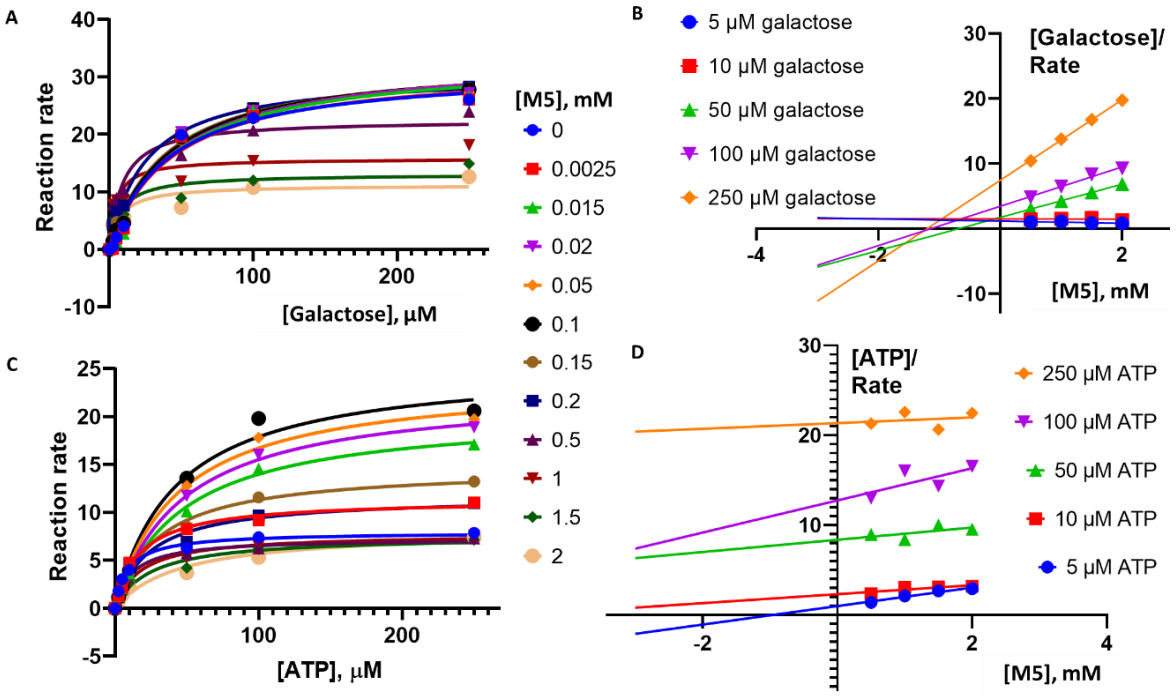


Figure 6.19: Kinetic characterisation of GALK1 inhibition by M5. *A. Michaelis-Menten plot of hGALK1 reaction rate (µM ATP consumed) against [galactose] in the presence of increasing concentrations of **M5**, determined in the Kinase Glo assay. hGALK1 reaction was performed at 10 nM hGALK1 with 35 µM ATP and 100 µM galactose for 1 hour before quantification using the Kinase Glo reagent as described. B. Cornish-Bowden plot of [Galactose]/Rate (as defined in panel A) against concentration of **M5** at 5 µM (blue), 10 µM (red), 50 µM (green), 100 µM (purple) and 250 µM (orange) galactose, determined in the Kinase Glo assay. C. Michaelis-Menten plot of hGALK1 reaction rate (µM galactose consumed) against [ATP] in the presence of increasing concentrations of **M5**, determined in the Amplex Red assay. hGALK1 reaction was performed at 250 nM hGALK1 with 100 µM ATP and 50 µM galactose for 1 hour before quantification using the Amplex Red reagent as described. D. Cornish-Bowden plot of [ATP]/Rate (as defined in panel C) against concentration of M5 at 5 µM (blue), 10 µM (red), 50 µM (green), 100 µM (purple) and 250 µM (orange) galactose, determined in the Amplex Red assay. Error bars are standard deviation of technical triplicates.*

When varying galactose levels in the presence of high concentrations (500 µM – 2 mM) of compound **M5**, Michaelis-Menten plots (figure 6.19A) show a concentration-dependent

decrease in $Vmax^{apparent}$ ($P < 0.0001$) and increase in $Km^{apparent}$ ($P < 0.0001$), and the lines of the Cornish-Bowden plot (figure 6.19B) mostly intersect below the x-axis, consistent with mixed type inhibition. When varying ATP levels in the presence of high concentrations (500 μM – 2 mM) of **M5**, Michaelis-Menten plots (figure 6.19C) show a concentration-dependent increase in $Km^{apparent}$ ($P < 0.0001$) and unchanged $Vmax^{apparent}$ values, and the lines of the Cornish-Bowden plots are parallel (figure 6.19D), indicating competitive inhibition. This analysis therefore supports binding of high concentrations of **M5** to the ATP binding site, resulting in inhibition of hGALK1.

In this chapter, two activity assays were developed, validated and used to screen for inhibition of hGALK1 by the fragment hits described in the previous chapter. Fragment **4** from the protein termini pocket and fragment **8** from the hotspot demonstrated weak inhibition of hGALK1 in both assays and were not competitive with respect to either galactose or ATP. Next, fragment **8** was optimised by merging with other fragments from the hotspot pocket and the resultant merged compounds (**M1**-**M5**) characterised in the two activity assays. Three of the merged compounds inhibited hGALK1, one activated hGALK1 and one was a mixed effector. The most potent inhibitor, **M2**, demonstrating fair potency ($IC_{50} < 100\mu M$) and good ligand efficiency ($LLE_{AT} > 0.20$) in both assays. The significance of these findings and how they may be applied to the development of new therapeutic modalities for classic galactosaemia is discussed in chapter 7.

7. Discussion and Conclusions

This thesis extends the available structural data of hHAO1 and hGALK1 by describing the characterisation and optimisation of small molecule effectors discovered by crystallography-based fragment screening.

7.1 Structural biology and small molecule effectors of hHAO1

The development of new therapeutic approaches for primary hyperoxaluria is essential because of the high associated morbidity of the current standard of care, namely combined kidney and liver transplantation (section 1.2.2). Knockdown of hHAO1 expression by siRNA has shown therapeutic efficacy in clinical trials (NCT02706886; Lumasiran/ALN-GO1) but this treatment modality is associated with high cost, toxicity stemming from immune response, difficulties in delivery to the target organ and limited *in vivo* stability (section 1.2.4) [146], [154], [301]– [303].

These limitations support the development of small molecule inhibitors, via the substrate reduction approach, as adjunct or alternate therapies. Therefore, this work pursued a structure-guided approach to screen a diverse fragment library and identify new chemotypes for hHAO1 inhibitors that can be developed as novel therapeutics for primary hyperoxaluria. Fragment screening by X-ray crystallography (section 3.3) identified fragments at the glycolate binding site (fragments **1** and **2**), close to the gating loop (fragment **5**), and at the subunit interface (fragment **6**).

New chemotypes from active site fragments

To date, all known hHAO1 inhibitors bind at the active site and contain a polar head group and a hydrophobic secondary group (section 1.5.4). The polar group generally contains a carboxylic acid or di-carbonyl moiety that mimics the glycolate-binding interactions. The hydrophobic group may be composed of either an aromatic ring or a long hydrocarbon chain, and this group interacts with nonpolar side-chains proximal to the active site. It has been hypothesised that the latter interaction takes advantage of the ability of hHAO1 to bind substrates with a carbon chain up to 16-carbons long (section 1.5.3).

The active site fragment hits identified in this work (**1** and **2**) extend the available polar groups beyond the known inhibitor repertoire to include a carboxamide group and extend the known hydrophobic groups to include a fused aromatic ring. Optimisation of the fragment **1** (IC_{50} 420 μM; LLE_{AT} 0.39 kcal/mol) scaffold achieved a 5-fold increase in potency (compound **26**; IC_{50} 81 μM) and concurrent increase in ligand efficiency (LLE_{AT} 0.47 kcal/mol). Optimisation of the fragment **2** (37% inhibition of 100 nM hHAO1 by 10 mM fragment) scaffold identified 4 compounds (**27**, **32**, **34** and **35**) with IC_{50} < 100 μM and very high LLE_{AT} scores (0.71-0.89 kcal/mol). Compounds **27** – **35** from the fragment **2** scaffold also provide the first evidence that good inhibition of hHAO1 can be achieved by the polar moiety of active site inhibitors alone, supporting development of potent inhibitors lacking the solubility and permeability issues that arise from the hydrophobic moiety of published inhibitors.

Future optimisation of active site inhibitors would be facilitated by extensive structure-activity relationships of the relative efficacy of carboxylic acid and carboxamide polar groups. A simple amide coupling can be used to convert the carboxylic acids to carboxamide equivalents (figure 7.1A). Carboxylic acid equivalents of the inhibitors described here – e.g. 1-methyl-indazole-3-carboxylate (CID 689105, equivalent to compound **8**) and 5-bromo-indazole-3-carboxylate (CID 7157358, equivalent to compound **15**) – are commercially available (figure 7.1B).

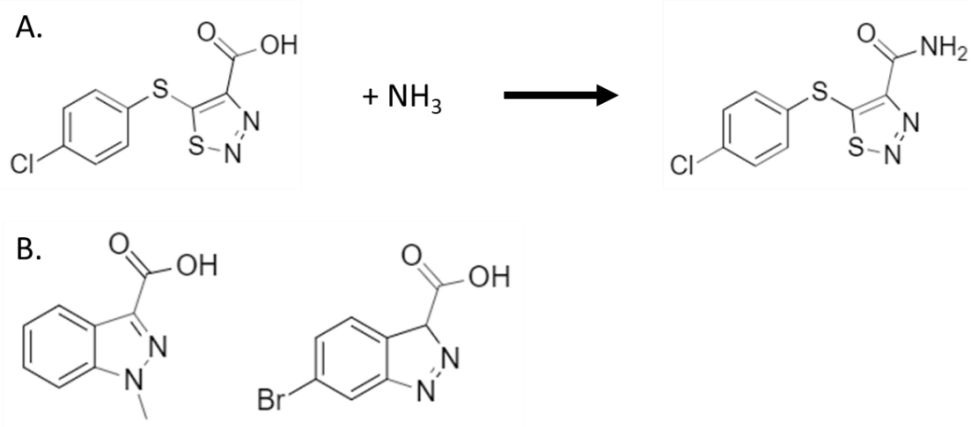


Figure 7.1: Potential strategies to optimise active site inhibitors of hHAO1. *A. Reaction mechanism to exchange carboxylic acid and carboxamide groups. B. Commercially available carboxylic acid-containing analogues of the indazole-carboxamide compounds.*

A hHAO1-unique pocket at the gating loop

This work also uncovers inhibitory fragments with an IC_{50} < 100 μM and LLE_{AT} > 0.2, that target a novel binding pocket at the gating loop outside the active site e.g. compound **60** had an IC_{50} of 40 μM and an LLE_{AT} of 0.31 kcal/mol.

Movement of the gating loop has been directly linked to binding of glycolate and inhibitors to hHAO1 [203]. However, the role of this loop in catalysis, and as such its suitability for inhibitor development, has not previously been investigated. The inhibitors at the gating loop described in this work demonstrate the utility of this novel binding site for inhibitor development and could also be applied as tool compounds to characterise the role of the gating loop in catalysis and substrate specificity. Both the gating loop itself and the neighbouring residues show low sequence conservation among 2-hydroxyacid oxidases (section 3.3.4, figure 3.14), indicating this is a good target for the development of selective hHAO1 inhibitors. The sequence variability of the gating loop also relates to the ability of hHAO1, but not other glycolate oxidases, to oxidise both long and short chain 2-hydroxyacids [203].

In future work, the activity assay should be employed to determine the mode of action i.e. whether the gating loop compounds inhibit hHAO1 by preventing substrate binding (competitive inhibition), by affecting the affinity of hHAO1 for substrate and its catalytic efficacy (uncompetitive/ mixed competitive inhibition) or by reducing the catalytic efficacy only (non-competitive inhibition). This assay could also be used to compare the relative efficacy of the gating loop compounds in inhibition of hHAO1 activity with substrates of differing lengths. Collectively, these experiments would provide insight into the contribution of the gating loop to hHAO1 activity. If paired with biophysical/structural observation of the gating loop during addition of the inhibitor, this would help determine the contribution of the gating loop to both binding and turnover of long and short chain substrates.

For example, movements of the gating loop could be measured by mutating a residue central to the gating loop to allow attachment of a fluorescent probe (e.g. Ser192Cys or Ser194Cys). After attachment of a fluorescent probe such as fluorescein, the motion of the probe and therefore the gating loop can be monitored by fluorescence anisotropy. Direct observation of loop movements in solution would also be useful in determining whether the gating loop observed in crystal structures (e.g. PDB 2rdu, 6gmb) represents the only functional conformation of the gating loop, or if conformational dynamics of the gating loop are essential for hHAO1 catalysis. This distinction could be useful in prioritising medicinal chemistry efforts to optimise lead compound interactions with hHAO1 residues that contribute to forming functional gating loop conformation(s).

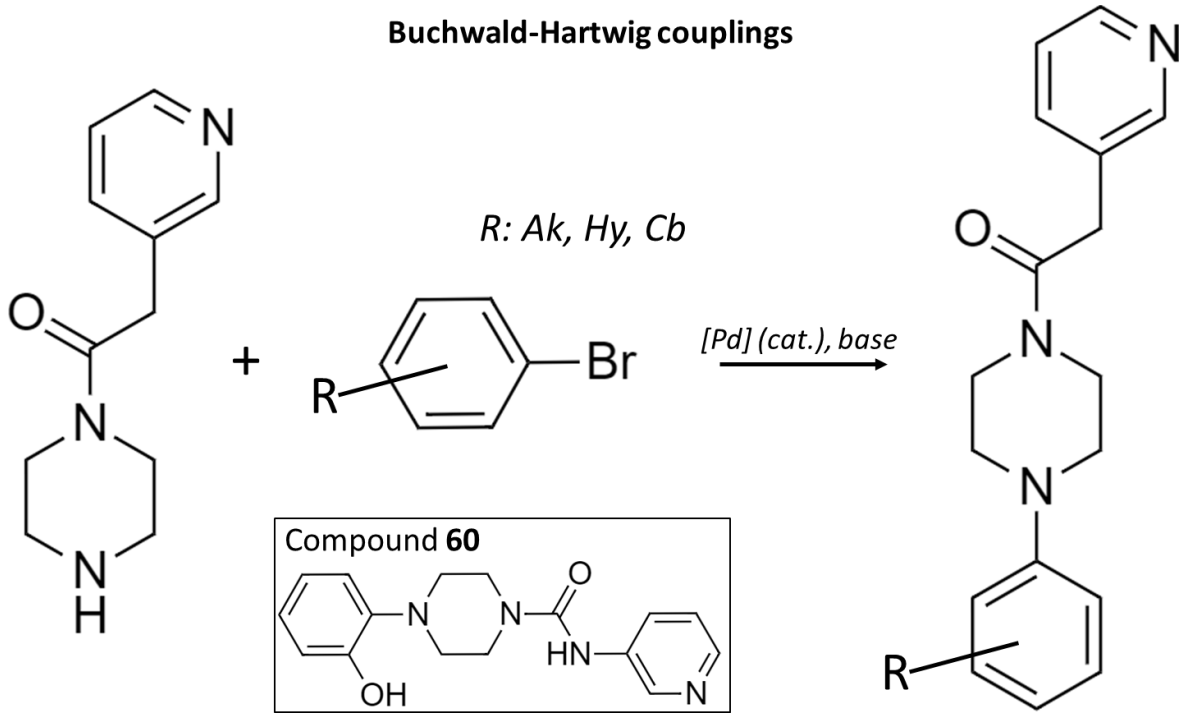


Figure 7.2: Proposed synthetic strategy to optimise gating loop inhibitors of hHAO1. *Reaction mechanism to couple compound **60** scaffold and substituted phenyl building blocks via Buchwald-Hartwig coupling. This reaction requires the presence of a catalytic base and a palladium catalyst ([Pd]). Inset: Chemical structure of starting compound, **60**.*

The initial gating loop follow-up compounds primarily extended the fluorophenyl-piperazine scaffold from the piperazine end of the molecule and demonstrated high lipophilicity and molecular weight relative to their inhibitory activity. Therefore, the next chemical series to investigate would explore SAR of compound **60** (figure 7.2, inset) derivatives.

For example, by coupling the (pyridin-3-yl)piperazine-1-carboxamide scaffold to a set of substituted phenyl building blocks (Buchwald-Hartwig coupling, figure 7.2), one can investigate the effect of substituents at different positions of the phenyl ring [304]. This is supported by the presence of halide, hydroxyl, methyl and methoxy groups at various positions of the phenyl ring in many of the gating loop compounds (e.g. all 9 compounds with sufficient potency to calculate IC_{50}s had one or two substitutions to the phenyl ring; section 4.3.2, figure 4.12). This coupling reaction could enable the compounds described in this work to be grown into the pocket occupied by the hydrophobic moiety of existing inhibitors and/or facilitate merging with active site inhibitors.

Is hHAO1 subunit interface a druggable region?

This work further describes the first known ligands binding at the subunit-subunit interface of the hHAO1 tetramer (fragment **6**). The most potent follow-up compound from this pocket was compound **132** with IC_{50} 82 μM and LLE_{AT} 0.25 kcal/mol. The tetrameric nature of hHAO1 is well established, but no studies have explored the relationship between hHAO1 activity and oligomeric state. While studies describing inhibitors that target subunit-subunit interfaces are relatively scarce, the elucidated mechanisms are diverse and include disruption of the oligomer leading to non-functional monomers or lower order oligomers [305]–[307], conformational changes that reduce the function of the oligomer [308], and increasing the rate of dissociation of the functional oligomer to non-functional monomers or lower order oligomers [309]. The active site of individual hHAO1 subunits are distal from each other in the tetramer, indicating that oligomerisation is unlikely to play a direct role in hHAO1 activity.

The good agreement between activity assay and native MS results for the follow-up compounds derived from fragment **6** is encouraging, although a number of questions remain unanswered. For example, one compound (**125**) demonstrated the ability to disrupt the hHAO1 oligomer yet showed poor inhibition in the activity assay (21 ± 3%), whereas no binding or conformational

change to the tetramer was observed upon addition of the 2nd most potent inhibitor in the activity assay (compound **124**, IC_{50} 180 μM).

The main limitations in the native MS data relate to the mass and charge radii changes caused by binding of a variable number of DMSO molecules to the observed protein species, and the limited resolution achieved in the initial experiment. The variable binding of DMSO has previously been reported to hamper interpretation of native MS experiments, and changing solvent (i.e. to a solvent that does not stick to protein or change its charge state, such as ethanol) could address this [310]. Additional optimisation of the desalting procedure and inclusion of additives previously shown to improve MS resolution such as amino acids, imidazole or ammonium bromide/iodide [311] should also be explored. Binding of these compounds to hHAO1 should also be quantified in future by another technique such as SPR.

Correlation studies, exploring how ligand binding at the subunit interface translates to inhibition of hHAO1 activity, are also necessary. Firstly, it is important to rule out the possibility that binding at the interface is secondary to inhibition achieved through fragment binding at the active site. This can be explored by determining the inhibition mode of the compounds in the activity assay, a similar approach to that applied to hGALK1 in this work. Secondly, a reasonable hypothesis would be that these compounds either destabilise the hHAO1 tetramer, promoting its dissociation, or cause conformational changes that result in a stable but less active tetrameric structure.

This hypothesis could be tested using size exclusion chromatography coupled with multi angle light scattering (SEC-MALS), a technique that is similar in principle to SAXS and also provides information about oligomer stability. Alternatively, analytical ultracentrifugation (AUC), a potentially higher resolution method based on the movement of protein molecules through a high gravitational field, could be applied to determine the size, shape and solution association constants of hHAO1 subunits in the absence and presence of inhibitors targeting the subunit

interface. Collectively, these methods would help establish the utility of this binding pocket and guide medicinal chemistry efforts to improve potency of these compounds.

To summarise, this work describes 8 compounds, optimised from fragments binding to three binding sites of hHAO1, that have IC_{50}s < 100 μM and LLEAT > 0.2. These compounds have the potential to overcome the limitations of existing hHAO1 inhibitor chemotypes and address fascinating questions about the role of the gating loop and oligomerisation in the activity of hHAO1 and its homologues. For these compounds, determination of their inhibition mode and selectivity over homologous enzymes, as well as validation of their affinity and potency using orthogonal assays, should be prioritised. For the next steps, the 10- to 100-fold increase in potency that is required to develop these compounds into lead candidates should be pursued through structure-guided medicinal chemistry efforts. Drug-like compounds emerging from hit-to-lead optimisation can then be tested in cellular and animal models (described in section 1.1.4) for their *in vivo* efficacy as well as pre-clinical evaluation of pharmacokinetic and pharmacodynamic properties.

7.2 Structural biology and small molecule effectors of hGALK1

While acute manifestations of classic galactosaemia can be avoided by dietary management through galactose restriction, current treatment modalities do not prevent the long-term complications associated with damage to the central nervous and reproductive systems [29]. Consequently, substrate reduction therapy, employing small molecules to inhibit hGALK1, has gained considerable traction as a promising alternative approach [107], [134]. However, current small molecule inhibitors of hGALK1 are ATP-competitive or are predicted to bind in the active site and demonstrate limited scaffold diversity, despite screening campaigns involving >500'000 compounds (section 1.4.4, figure 1.11) [107], [133]–[139]. Development of potent inhibitors from these scaffolds has also been hampered by the lack of data on the protein-ligand interactions that facilitate hGALK1 inhibition as there are currently no published inhibitor-bound structures to guide drug design.

This study first describes the co-crystallisation of hGALK1 with galactose and four ATP-competitive inhibitors (**T1**, **T2**, **T3** and **T7**) derived from the previously reported spiro-benzoxazole series [133], [135]. The solved structures confirm that these compounds bind in the ATP pocket of the active site as predicted from kinetic experiments, and elucidate the protein-ligand interactions involved, including changes in three of the four loops involved in ATP binding (section 5.2.3). Longer term experiments to optimise the spiro-benzoxazole series of ATP- competitive inhibitors should expand the lead compound towards the ATP-phosphate and galactose binding pockets of hGALK1. For example, addition of hydrophilic groups opposite the spiro-ring substituent of the quinazolinone ring to interact with polar residues in these pockets would have the corollary benefit of reducing the overall lipophilicity of the lead compounds, thus improving their pharmacokinetic properties.

An alternative approach that has been shown to improve solubility and permeability of lipophilic lead compounds is the disruption of molecular planarity and/or symmetry [312]. For example, one or more nitrogen atoms or a substituent such as fluorine, which has a small Van der Waals radius and so would cause minimal steric restriction, could be added to the unsubstituted aromatic ring of the benzoxazole moiety of compound **T1**.

Crystals pre-loaded with the substrate galactose and an orthosteric inhibitor (the spiro-benzoxazole compound **T3**) were employed in the second part of this work, performing a fragment soaking campaign to identify surface accessible pockets outside the active site. Screening of ~200 fragments against such crystals identified a total of 28 fragments clustering to three binding sites (section 5.3). The high hit rate observed in the fragment screening campaign against hGALK1 (>10%), relative to that observed against hHAO1 (~1%) despite the similar size of the two proteins and the higher number of fragments screened against hHAO1, illustrates the high druggability of hGALK1, as described for other kinases [313]. An additional contributory factor may be that, because hGALK1 is monomeric, a greater surface area is

available for fragment binding; i.e. hHAO1 and hGALK1 monomers have similar total surface areas (14767.1 and 15515.6 Å^2 respectively, calculated using the PISA server [315]) but >20% of this area is buried in the hHAO1 tetramer.

Fragments to merge with active site inhibitors

The two fragments (**1** and **2**) bound at the entrance to the ATP binding site (section 5.3.3) showed no inhibition of hGALK1 activity when tested *in vitro*. Consequently, their key utility lies in their potential for merging with larger inhibitor scaffolds at the ATP binding site, such as the spiro-benzoxazole series. The carbonyl group of spiro-benzoxazole compound **T3** is 3.2 Å from the proximal oxygen and 3.5 Å from the methyl of the thioether group of fragment **1** (figure 7.3A) and is 3.6 Å from the proximal nitrogen of the pyrimidine group of fragment **2** (figure 7.3B). There is potential for linking of these fragments to compound **T3** to expand protein-ligand interactions.

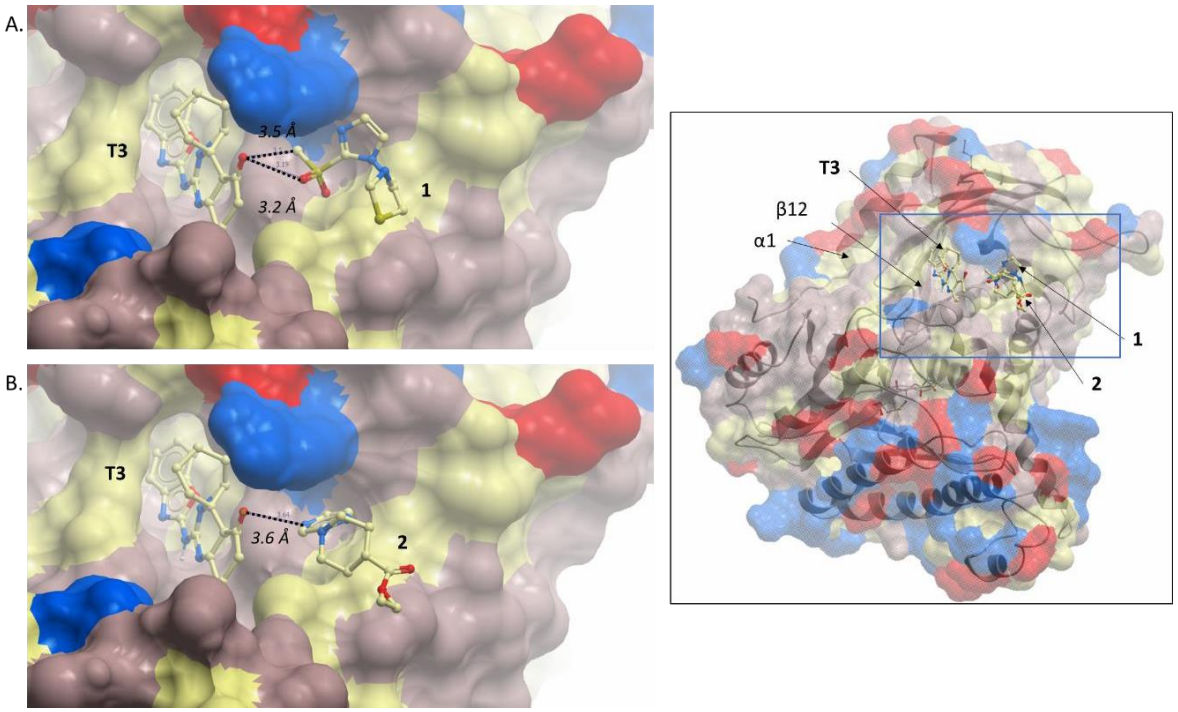


Figure 7.3: Potential strategies to link fragments at the entrance to the ATP binding site of hGALK1. *A. Close up surface view of hGALK1 ATP binding site showing the binding pose of fragment **1** with respect to **T3**. The tool compound **T3**, and fragments are shown as sticks. Dotted lines indicate the distance in Ångstrom between proximal atoms. B. Close up surface view of hGALK1 ATP binding site showing the binding pose of fragment **2** with respect to **T3**. Dotted lines indicate the distance in Angstrom between proximal atoms. Inset: Surface and ribbon diagram of hGALK1 monomer indicating the orientation and location from which the close-up views are derived.*

Fragments in a region containing activity variants

Three fragments (**3** – **5**) were bound in a hydrophobic pocket close to the protein N- and C-termini (section 5.3.4). Fragment **4** displayed weak inhibition of hGALK1 (17 ± 2 % inhibition at 5 mM, $P = 0.0024$) and was not competitive with either galactose or ATP. The rationale for targeting this pocket to inhibit hGALK1 activity is supported by severe mutations reported in some of the nonpolar residues lining or close to this binding pocket. For example, the hGALK1 Pro28Thr, Val32Met and Ala384Pro variants (figure 7.4A) are insoluble in *E. coli* and patient red blood cells harbouring these mutations have no detectable hGALK1 activity [35], [184], [252].

Protein residues lining this pocket are generally similar across the GHMP kinase family (figure 7.4C), indicating a conserved function of this pocket, and consequently inhibitors targeting this pocket may also inhibit GALK2 and MVK, the other two GHMP kinase family members found in humans. The N-terminus of α1 and the C- terminus of β12 are more variable than the other secondary structures of this pocket (figure 7.4C, underlined in green) and the development of compounds that interact with these structures may provide selectivity over GALK2 and MVK.

While no adverse effects have been reported for inhibition of GALK2, inherited loss-of-function variants of MVK are associated with a spectrum of disorders, characterised by recurrent inflammatory attacks, that can be fatal in infancy and for which there are no known therapies [315], [316]. Since both GALK2 and MVK consume ATP to phosphorylate their sugar substrates, inhibition of these targets by compounds binding to this pocket can be detected using the same Kinase Glo assay set-up as described for hGALK1 in chapter 6.

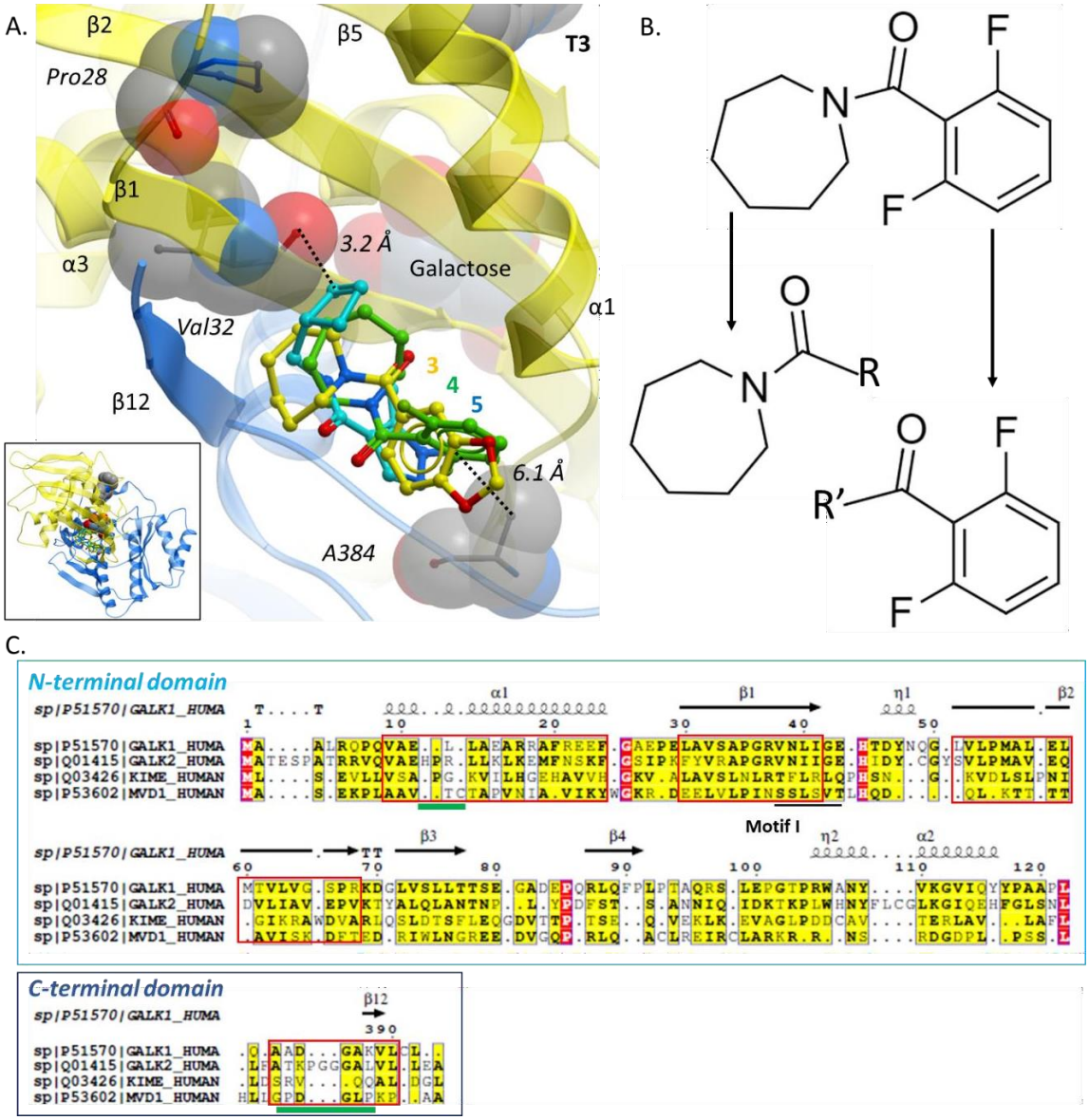


Figure 7.4: Potential strategies to optimise protein termini site inhibitors of hGALK1. *A. Close up view showing hGALK1 protein termini binding site as ribbon. Fragments* ***3-5*** *are shown as sticks. Tool compound* ***T3*** *and galactose are shown as grey spacefill. Residues with reported insoluble variants – Pro28, Val32 and Ala384 – are shown as sticks and spacefill. Ribbon is yellow for the N-terminal domain and blue for the C-terminal domain. Dotted lines indicate the distance in Angstroms between proximal atoms. Inset: Overall structure from which the close-up view was extracted, shown as ribbon diagram. Orientation is the same as that used in section 5.3.4 figures. B. Proposed synthons from fragment* ***4*** *for a substructure search to select follow-up compounds for further testing. C. Sequence alignment (extracted from figure 1.9C). Invariant residues are shown as white characters on a red background, residues that are similar in >70% of sequences are shown as bold characters on a yellow background and residues that are similar in ≥50% of sequences also have a yellow background. The GHMP kinase motifs are underlined in black and secondary structure features of the protein termini site are shown in red boxes. Variable regions are underlined in green.*

An important avenue for future work is to investigate whether small molecules targeting this pocket exert similar effects on hGALK1 as the disease variants. To achieve this, chemical improvement of fragment **4** potency should be performed, to generate more advanced tool

compounds that allow unambiguous determination of their mechanism of action. Since this fragment came from the DSi-poised library, a similar approach to that described for hHAO1 can be used to disconnect the aromatic and cycloalkyl substituents and search a commercial library to identify follow-up compounds (figure 7.4B).

Inhibitory fragments at a binding hotspot

Fragments **6** – **28** are bound in two adjacent clusters (upper and lower) within a hotspot region that is distant from the active site (section 5.3.5). Among them, fragment **8** from the upper cluster demonstrated weak but statistically significant inhibition of hGALK1 (24 ± 2 % inhibition at 5 mM, $P < 0.0001$) and did not compete with either galactose or ATP. Fragment **8** was selected for initial SAR-by-catalogue experiments as the fragment-bound crystal structures provide a structural basis for optimisation by fragment merging (section 6.3.1). Fragment **8** was combined with fragments from the lower cluster of the hotspot by *in silico* design, resulting in three follow-up inhibitors (**M1** – **M3**; section 6.3.2), which did not compete with either galactose or ATP.

Among them, the two more potent inhibitors have a shared 6,7-dichloro-quinoxaline scaffold and similar average IC_{50}s of 105 ± 70 μM (**M1**; LLEAT 0.20 kcal/mol) and 91 ± 70 μM (**M2**; LLEAT 0.24 kcal/mol). Altogether, the emerging kinetics data indicate that fragment **8** and follow-up compounds **M1**-**M3** bind away from the active site in solution, as shown in the crystal structures. The β7 – β8 loop (Pro212 – Leu218) and the α8 – α9 turn (Arg296 – Tyr300) that form the surface-exposed face of the binding hotspot differ in conformation between protein molecules within the asymmetric unit (figure 5.10A), indicating conformational plasticity to be exploited for drug development at this pocket.

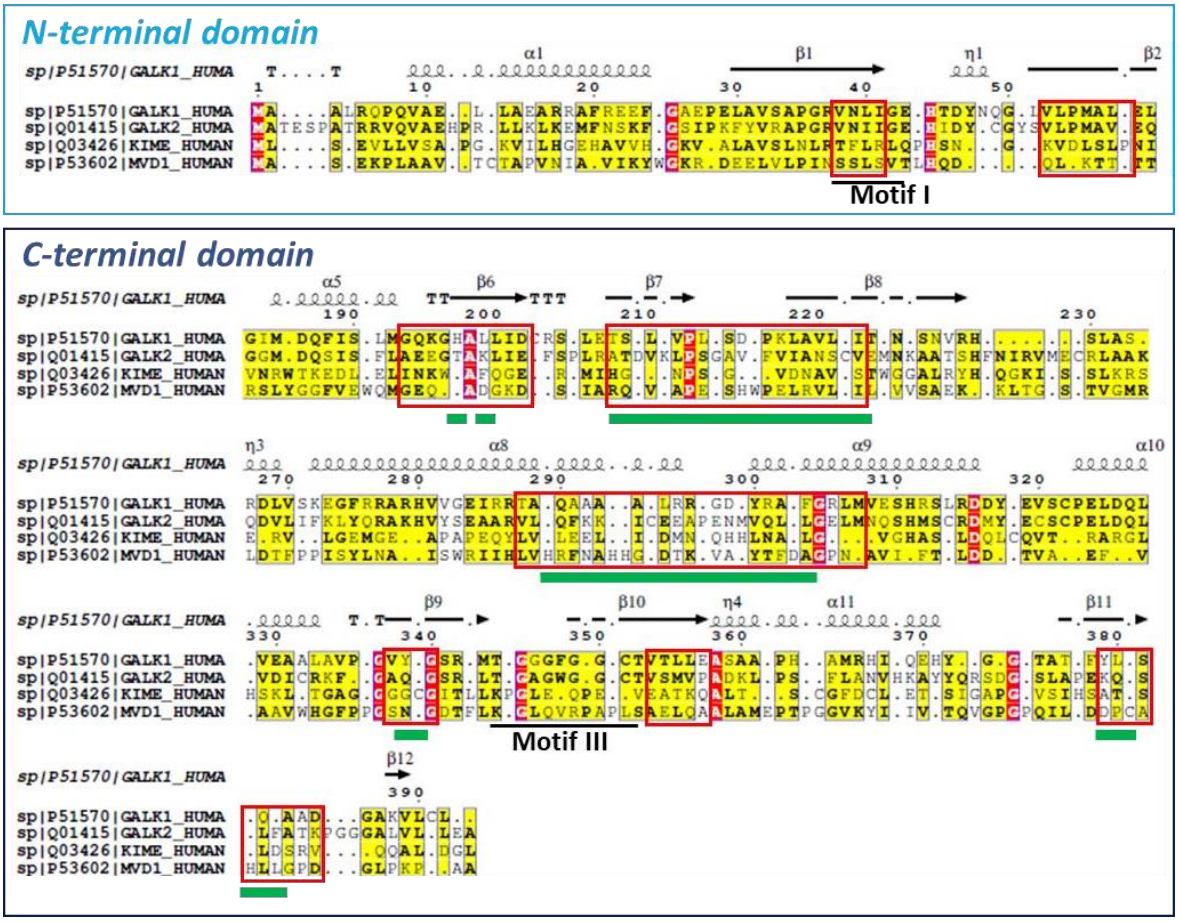


Figure 7.5: Sequence alignment of GHMP kinase family members comparing secondary structures at the binding hotspot of hGALK1. *A. Sequence alignment (extracted from figure 1.9C). Invariant residues are shown as white characters on a red background, residues that are similar in >70% of sequences are shown as bold characters on a yellow background and residues that are similar in ≥50% of sequences also have a yellow background. The GHMP kinase motifs are shown underlined in black and secondary structure features of the hotspot site are shown in red boxes. Variable regions are underlined in green.*

An additional required experiment for these compounds is the determination of their specificity over the other two GHMP kinase family members found in humans, GALK2 and MVK, which can be achieved as described for the protein termini pocket. Many hGALK1 residues in the binding hotspot demonstrate low conservation across the GHMP kinase family, particularly the flexible β7-β8 loop (Pro212-Leu218) that forms the front of the binding pocket, and the helical secondary structures that form the base of the pocket (α8 helix, α8-α9 turn and α9 helix; Thr288-Phe303) (highlighted green in figure 7.5).

GALK1 activators at the hotspot region

Interestingly, *in silico* fragment merging at the hotspot yielded not only inhibitors **M1**-**M3** described above, but also the first described activators for hGALK1. Activation of hGALK1

by **M4** (EC_{50} 133 ± 55 μM) is unaffected by galactose concentration but is abolished by high concentrations of ATP (section 6.3.4). Taken together with the known substrate binding mechanism of hGALK1, in which ATP binds to apo-GALK1 and induces conformational changes that increase the enzyme's affinity for galactose, these data suggest that binding of **M4** activates hGALK1 by mimicking the conformational changes induced by ATP binding.

Consequently, **M4** has limited use in biological settings as it would have no effect in the presence of the millimolar levels of ATP found intracellularly. By contrast, activation of hGALK1 by low concentrations of **M5** was constant across the tested galactose and ATP ranges, indicating a different, and as yet uncharacterised, mechanism of activation. High concentrations of **M5** demonstrated competitive inhibition with respect to ATP, indicating binding in the ATP binding site (section 6.3.5). Because hGALK1 activation by **M5** is not abolished by high ATP concentrations, optimised compounds based on **M5** could serve as tool compounds to investigate hGALK1 activity in cellular and animal models.

Further experiments to explore mechanisms of action for hGALK1 inhibitors

Optimisation of inhibitors targeting the novel pockets identified in this work would benefit from several lines of investigation. From a structural point-of-view, direct observation of protein-ligand interactions by co-crystallisation with the merged compounds **M1**-**M5**, to determine if they indeed bind to the hotspot region as predicted, would be essential. A better understanding of the mechanisms of action for these inhibitors is equally important, and should include analysis of changes to protein stability, flexibility and/or conformation upon inhibitor binding.

Changes in protein stability can be assessed by comparing thermodynamic and kinetic stability in the presence of the allosteric ligand relative to the unliganded enzyme. For example, thermodynamic stability can be assessed by comparing the degree of thermal denaturation using DSF (as described in section 5.1.3) or circular dichroism [317], [318], and kinetic stability can be determined by comparing the rate of chaotrope-induced unfolding using equilibrium differential scanning calorimetry [319]. Studies have shown that both increasing [320] and

decreasing [321] protein flexibility can reduce protein function, and consequently measurement of changes in protein flexibility upon fragment binding can provide insight into the mechanism of action of inhibitors. For example, a recent study described the combination of molecular dynamics simulations with SAXS experiments to quantify a protein's flexibility [322].

An alternate approach to investigate an inhibitor's mechanism of action is to characterise the binding pocket of the protein. Mutagenesis studies can be employed to mimic fragment binding e.g. by introducing aromatic sidechains in positions equivalent to where a fragment's aromatic moiety is bound in the co-crystal structure or by sequential mutations to assess the relative importance of each residue within a binding pocket.

For example, substitution of Val9 to an aromatic residue such as tryptophan would result in a hydrophobic moiety (the tryptophan indole) occupying the centre of the protein termini pocket (figure 7.6A), and substitution of Leu218 to an aromatic residue such as phenylalanine would provide similar information at the binding hotspot (figure 7.6B). Similarly, alanine scanning of the binding pocket residues, by generating constructs in which one or more of the identified binding site residues is mutated to alanine, could be used to explore the function of these residues, that is modulated by fragment binding, and the relative importance of specific residues in coordinating fragment binding.

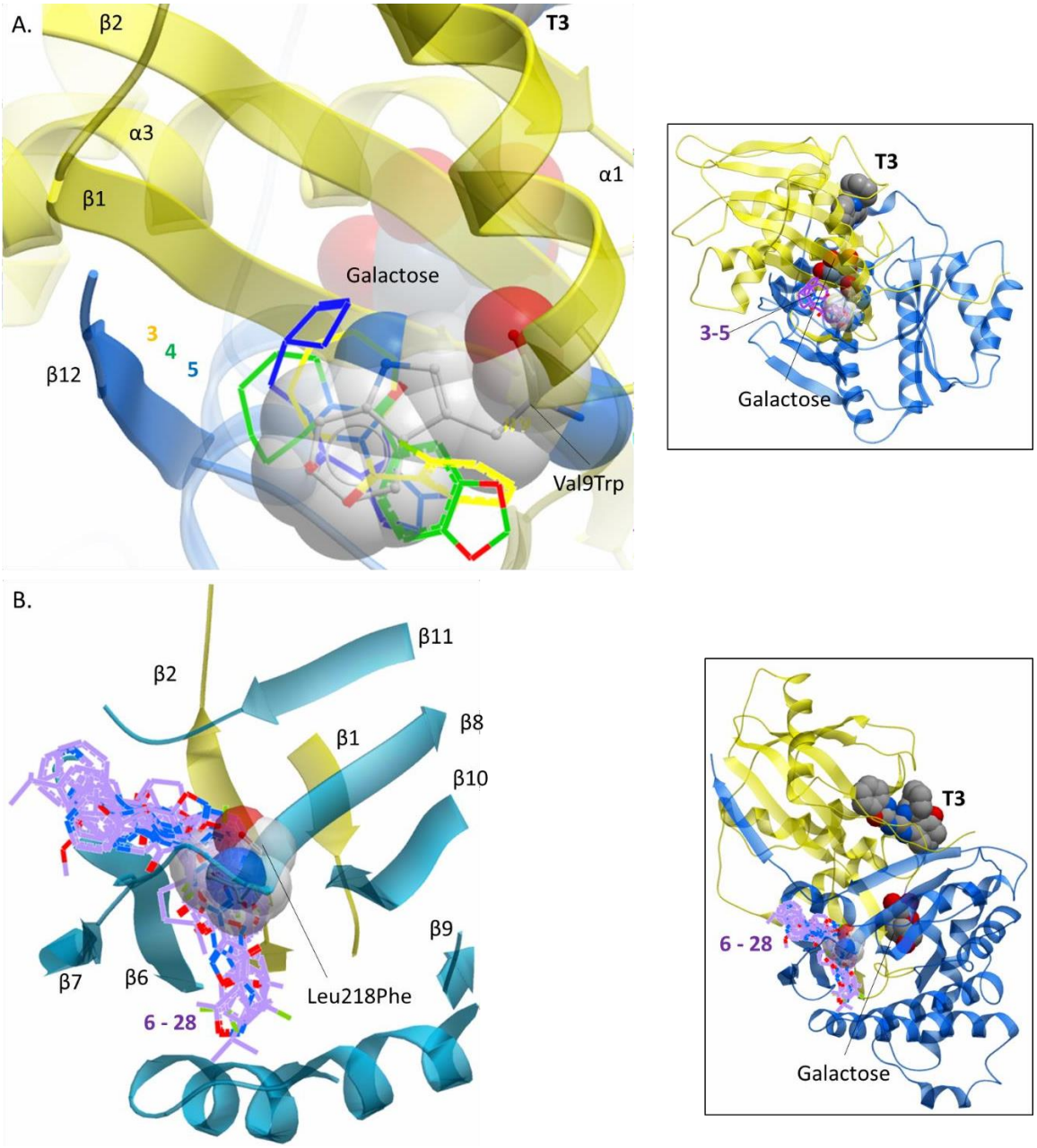


Figure 7.6: Structural models of hGALK1 variants proposed to investigate the novel binding pockets described in this work. *A. Close up view showing hGALK1 protein termini binding site as ribbon. The Val9Trp variant suggested in the text was modelled by in silico mutagenesis with side chain optimisation and is shown as sticks and spacefill. Fragments* ***3-5*** *are shown as lines and galactose and compound* ***T3*** *are shown as grey spacefill. Protein ribbon displayed is of the hGALK1-galactose-****T3*** *structure with N-terminal domain coloured yellow and the C-terminal domain coloured blue. Inset: Overall ribbon diagram illustrating the orientation of the hGALK1 structure in panel A. B. Close up view showing hGALK1 hotspot binding site as ribbon. The Leu218Phe variant suggested in the text was modelled by in silico mutagenesis with side chain optimisation and is shown as sticks and spacefill. Fragments* ***6 - 28*** *are shown as lines. Inset: Overall ribbon diagram illustrating the orientation of the hGALK1 structure in panel B.*

Potential mechanism of action for novel hGALK1 binding pockets

A final investigative direction is based on the observation that non-conserved but essential structural features can contribute to protein-specific functions such as interaction with binding

partners (reviewed in [323]). There is some evidence for intracellular interaction between Leloir pathway enzymes (GALM, GALK1, GALT and GALE) in yeast [49] and humans [324], and the possibility of these enzymes forming a single 'galactose-metabolising metabolon' has been suggested but not investigated [325]. Proteins that interact with hGALK1 could be identified by using immobilised hGALK1 (e.g. attached to agarose beads via an affinity tag) to capture interacting proteins from whole cell lysates and subsequent identification of the captured interacting partners by mass spectrometry (affinity purification-mass spectrometry [326]). After identification of interacting partners, several *in vitro* methods are available to characterise their interaction with hGALK1 and measure changes in these interactions caused by ligand binding or mutagenesis at the novel sites described in this work. For example, surface plasmon resonance or a similar binding assay could be used to measure the binding affinity between hGALK1 and other proteins from the Leloir pathway.

The hypothetical galactose-metabolising metabolon would be of sufficient size for structural determination using cryo-electron microscopy. If such a metabolon is formed, then evaluation of the interactions of hGALK1 with other members could elucidate not only an additional *in vivo* mode of action for the compounds described in this work but also contribute to the understanding of both hGALK1 mutations that are distal to the active site and the basic science underpinning galactose metabolism, thus providing a deeper insight into the biochemical phenotype of classic galactosaemia.

In summary, this work presents the first hGALK1 structures to date bound with non-endogenous ligands, characterises the first known inhibitors of hGALK1 that bind outside of the active site (fragment **4**, fragment **8** and merged compounds **M1** – **M3**), and reports the first activators of hGALK1 (compounds **M4** and **M5**). Collectively, these compounds are promising starting points for the development of next-generation hGALK1 effectors aimed at improving scaffold diversity and potentially solubility, metabolic stability and cellular toxicity profiles, relative to previously

published, orthosteric inhibitors. Optimised compounds could then be tested for rescue of galactosaemia phenotypes in cellular models (e.g. galactose toxicity in yeast cells or patient fibroblasts with reduced/abolished GALT function) before translation to whole organism disease models such as the fruit fly, zebrafish and mouse models.

7.3 Potential of substrate reduction therapy for other metabolic disorders

A subset of IEMs could be addressed by substrate reduction therapy.

Of the 1239 IEMs reported to date, 97% are characterised as loss-of-function disorders and 87% are described as well-characterised, meaning that the causative gene product and the metabolic step it facilitates are known. However, high risk is associated with drug development programmes for IEMs as the rarity of individual disorders often translates to insufficient patients' materials for testing, and limited financial incentive for drug marketing. Additionally, therapies to address loss-of-function disorders by repairing or bypassing the defective or deficient enzyme are considerably less explored in the drug development industry at present. Rather, most therapeutic molecules developed through drug discovery projects are inhibitors addressing gain-of-function pathologies [327]. Consequently, approved therapies are available for less than 10% of IEMs [91], [93], which constitutes a significant unmet therapeutic need and resultant clinical burden because the collective occurrence of IEMs may be as high as 1:500 live births in high risk populations [2]-[4].

The therapeutic approach described in this work, substrate reduction therapy, holds considerable promise for IEMs because it plays to the strengths of the drug development industry in developing inhibitors, and is supported by the interlinked nature of metabolic enzymes within a pathway, meaning that targeting an early step in a given pathway can be an effective therapeutic approach for several disorders caused by loss-of-function in downstream enzymes. According to recent nosology studies [1], [8], IEMs in which substrate accumulation is the primary pathogenic driver occur in 7 of the 9 broad categories of IEMs (excluding only defects of energy metabolism and disorders of glycosylation) and in 43 of the 86 pathway groups within these categories. This

translates to a minimum of 50 well- characterised disorders for which there is a sound theoretical basis for substrate reduction therapy.

Indeed, proof of concept studies and development of a few clinically approved molecules support substrate reduction therapy for various categories of IEM disorders. Specific examples include: inhibition of 4-hydroxyphenylpyruvate dioxygenase by Nitisinone, an approved treatment for hereditary tyrosinaemia 1 (disorder of tyrosine metabolism), animal and patient cell models supporting inhibition of aminoadipate semi-aldehyde synthase to treat pyridoxine- dependent epilepsy (disorder of lysine metabolism), animal disease models supporting inhibition of glycogen synthase to treat Pompe disease (glycogen storage disorder), inhibition of 5'aminolevulinate synthase 1 by RNAi ALN-AS1, an approved treatment for acute intermittent porphyria (disorder of haeme biosynthesis), and inhibition of glucosylceramide synthase by Miglustat or Eliglustat, approved treatments for Gaucher disease and Niemann- Pick type C disease (lysosomal storage disorder) (reviewed in [92]).

Druggability of metabolic enzymes can be addressed by crystallography-based fragment screening

The majority (~70%) of approved small molecule drugs target only four protein groups – GPCRs, ion channels, protein kinases and nuclear receptors – with few drugs targeting metabolic enzymes [328], [329]. There are significantly fewer tool molecules available for metabolic enzymes, to serve as starting templates for focussed screening and chemistry optimisation. This work highlights the utility of X-ray crystallography in identifying chemical starting points for targets where druggable pockets have not previously been characterised. The maximal concentration accessible by X-ray crystallography is at least an order of magnitude greater than that used in traditional biochemical screening assays and thus can identify allosteric ligands that have very weak potency but high ligand efficiency. Co-crystallisation with a known ligand prior to fragment soaking can also be employed to exclude undesirable binding pockets

(e.g. active site occlusion). The key benefit of screening fragments is the ease with which hits can be grown, linked or merged to improve hit potency. This work illustrates how a single round of testing of low-cost, fairly simple molecules can provide good starting points for inhibitor development with particular emphasis on the identification of readily evolved compounds binding to previously uncharacterised sites.

All together this work adds two inborn errors of metabolism to the growing body of evidence demonstrating that fragment screening by X-ray crystallography can de-risk the early stages of drug development for rare diseases, which are collectively estimated to affect 3.5 – 5.9 % of the global population [330]. Through this and similar approaches, orphan diseases are being rehomed one by one to receive the attention they deserve – an eminently worthy goal that not even the most capricious of crystals can dull my enthusiasm for.

8. References

[1] C. R. Ferreira, C. D. M. van Karnebeek, J. Vockley, and N. Blau, (2019), A proposed nosology of inborn errors of metabolism., *Genet. Med.*, **21**, pp. 102–106.

[2] D. Waters, D. Adeloye, D. Woolham, E. Wastnedge, S. Patel, and I. Rudan, (2018), Global birth prevalence and mortality from inborn errors of metabolism: A systematic analysis of the evidence, *J. Glob. Health*, **8(2)**, article 021102.

[3] M. Alfadhel, M. Benmeakel, M. A. Hossain, F. Al Mutairi, A. Al Othaim, A. A. Alfares, M. Al Balwi, A. Alzaben, and W. Eyaid, (2016), Thirteen year retrospective review of the spectrum of inborn errors of metabolism presenting in a tertiary center in Saudi Arabia, *Orphanet J. Rare Dis*., **11**, article 126.

[4] S. Sanderson, A. Green, M. A. Preece, and H. Burton, (2006), The incidence of inherited metabolic disorders in the West Midlands, UK, *Arch. Dis. Child.*, **91**, pp. 896–899.

[5] A. E. Garrod, (1902), The Incidence of Alkaptonuria: A Study in Chemical Individuality, *Lancet*, **2**, pp. 1616–1620.

[6] A. E. Garrod and H. Harris, *Inborn errors of metabolism.* Oxford University Press, 1909.

[7] C. R. Ferreira and N. Blau, Nosology of Inborn Errors of Metabolism, 2019. [Online]. Available: http://www.iembase.org/nosology/index.asp. [Accessed: 30-Sep-2019].

[8] J. M. Saudubray, F. Mochel, F. Lamari, and A. G. Cazorla, (2019), Proposal for a simplified classification of IMD based on a pathophysiological approach: A practical guide for clinicians, *J Inherit. Metab. Dis.*, **42 (4)**, 706 – 727.

[9] A. I. Coelho, G. T. Berry, and M. E. Rubio-Gozalbo, (2015), Galactose metabolism and health, *Curr. Opin. Clin. Nutr. Metab. Care*, **18**, pp. 422–427.

[10] H. O. Kim, C. Hartnett, and C. H. Scaman, (2007), Free galactose content in selected fresh fruits and vegetables and soy beverages, *J. Agric. Food Chem.*, **55**, pp. 8133–8137.

[11] A. I. Coelho, M. E. Rubio-Gozalbo, J. B. Vicente, and I. Rivera, (2017), Sweet and sour: an update on classic galactosemia, *J. Inherit. Metab. Dis.*, **40**, pp. 325–342.

[12] P. Schadewaldt, L. Kamalanathan, H. W. Hammen, J. Kotzka, and U. Wendel, (2014), Endogenous galactose formation in galactose-1-phosphate uridyltransferase deficiency, *Arch. Physiol. Biochem.*, **120**, pp. 228–239.

[13] G. T. Berry, P. J. Moate, R. A. Reynolds, C. T. Yager, C. Ning, R. C. Boston, and S. Segal, (2004), The rate of de novo galactose synthesis in patients with galactose-1-phosphate uridyltransferase deficiency, *Mol. Genet. Metab.*, **81**, pp. 22–30.

[14] G. T. Berry, (2011), Is prenatal myo-inositol deficiency a mechanism of CNS injury in galactosemia?, *J. Inherit. Metab. Dis*., **34**, pp. 345–355.

[15] J. B. Thoden, J. Kim, F. M. Raushel, and H. M. Holden, (2003), The catalytic mechanism of galactose mutarotase, *Protein Sci.*, **12**, pp. 1051–1059.

[16] P. A. Frey, (1996), The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose., *FASEB J*., **10**, pp. 461–70.

[17] G. T. Berry, (1995), The role of polyols in the pathophysiology of hypergalactosemia, *Eur. J. Pediatr.*, **154**, pp. 53–64.

[18] V. N. Reddy, L. R. Lin, F. J. Giblin, B. Chakrapani, and T. Yokoyama, (1992), Study of the polyol pathway and cell permeability changes in human lens and retinal pigment epithelium in tissue culture, *Investig. Ophthalmol. Vis. Sci*., **33**, pp. 2334–2339.

[19] P. Cuatrecasas and S. Segal, (1966), Galactose Conversion to d-Xylulose: An Alternate Route of Galactose Metabolism, *Science*, **153**, pp. 549–551.

[20] S. Segal and P. Cuatrecasas, (1968), The oxidation of ^{C14}galactose by patients with congenital galactosemia, *Am. J. Med.*, **44**, pp. 340–347.

[21] C. Ning, R. Reynolds, J. Chen, C. Yager, G. T. Berry, N. Leslie, and S. Segal, (2001), Galactose metabolism in mice with galactose-1-phosphate uridyltransferase deficiency: Sucklings and 7-week-old animals fed a high-galactose diet, *Mol. Genet. Metab.*, **72**, pp. 306– 315.

[22] J. I. Führing, J. T. Cramer, J. Schneider, P. Baruch, R. Gerardy-Schahn, and R. Fedorov, (2015), A Quaternary Mechanism Enables the Complex Biological Functions of Octameric Human UDP-glucose Pyrophosphorylase, a Key Enzyme in Cell Metabolism, *Sci. Rep*., **5**, pp. 1–11.

[23] K. L. Ross, C. N. Davis, and J. L. Fridovich-Keil, (2004), Differential roles of the Leloir pathway enzymes and metabolites in defining galactose sensitivity in yeast, *Mol. Genet. Metab.*, **83**, pp. 103–116.

[24] H. M. Holden, J. B. Thoden, D. J. Timson, and R. J. Reece, (2004), Galactokinase: Structure, function and role in type II galactosemia, *Cell. Mol. Life Sci*., **61**, pp. 2471–2484.

[25] T. J. McCorvie and D. J. Timson, (2011), The structural and molecular biology of type I galactosemia: Enzymology of galactose 1-phosphate uridylyltransferase, *IUBMB Life*, **63**, pp. 694–700.

[26] P. A. Frey and A. D. Hegeman, (2013), Chemical and stereochemical actions of UDP-galactose 4-epimerase, *Acc. Chem. Res*., **46**, pp. 1417–1426.

[27] J. B. Thoden, T. M. Wohlers, J. L. Fridovich-Keil, and H. M. Holden, (2001), Human UDP galactose 4-epimerase. Accommodation of UDP-N-acetylglucosamine within the active site, *J. Biol. Chem*., **276**, pp. 15131–15136.

[28] D. Demirbas, A. I. Coelho, M. E. Rubio-Gozalbo, and G. T. Berry, (2018), Hereditary galactosemia, *Metabolism*., **83**, pp. 188–196.

[29] W. G. Ng, W. R. Bergen, and G. N. Donnel, (1964), Galactose-1-Phosphate Uridylyl Transferase Activity in Galactosaemia, *Nature*, **203**, pp. 845–847.

[30] K. Nelson and D. Y. Y. Hsia, (1967), Screening for galactosemia and glucose-6-phosphate dehydrogenase deficiency in newborn infants, *J. Pediatr.*, **71**, pp. 582–585.

[31] M. E. Rubio-Gozalbo, M. Haskovic, A. M. Bosch, B. Burnyte, A. I. Coelho, D. Cassiman, M. L. Couce, C. Dawson, D. Demirbas, T. Derks, F. Eyskens, M. T. Forga, S. Grunewald, J. Häberle, M. Hochuli, A. Hubert, H. H. Huidekoper, P. Janeiro, J. Kotzka, I. Knerr, P. Labrune, Y. E. Landau, J. G. Langendonk, D. Möslinger, D. Müller-Wieland, E. Murphy, K. Õunap, D. Ramadza, I. A. Rivera, S. Scholl-Buergi, K. M. Stepien, A. Thijs, C. Tran, R. Vara, G. Visser, R. Vos, M. De Vries, S. E. Waisbren, M. M. Welsink-Karssies, S. B. Wortmann, M. Gautschi, E. P. Treacy, and G. T. Berry, (2019), The natural history of classic galactosemia: Lessons from the GalNet registry, *Orphanet J. Rare Dis*., **14**, pp. 1–11.

[32] G.T. Berry, Classic Galactosemia and Clinical Variant Galactosemia, (2000), In: M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, LJ.H. Bean and K. Stephens, editors, *GeneReviews® [Internet]*, Seattle (WA): University of Washington, Seattle, 1993-2020, Available from: https://www.ncbi.nlm.nih.gov/books/NBK1518/

[33] M. Murphy, B. McHugh, O. Tighe, P. Mayne, C. O'Neill, E. Naughten, and D. T. Croke, (1999), Genetic basis of transferase-deficient galactosaemia in Ireland and the population history of the Irish Travellers, *Eur. J. Hum. Genet.*, **7**, pp. 549–554.

[34] R. Gitzelmann, (1967), Hereditary galactokinase deficiency, a newly recognized cause of juvenile cataracts, *Pediatr. Res.*, **1**, pp. 14–23.

[35] D. J. Timson and R. J. Reece, (2003), Functional analysis of disease-causing mutations in human galactokinase, *Eur. J. Biochem*., **270**, pp. 1767–1774.

[36] L. Kalaydjieva, A. Perez-Lezaun, D. Angelicheva, S. Onengut, D. Dye, N. U. Bosshard, A. Jordanova, A. Savov, P. Yanakiev, I. Kremensky, B. Radeva, J. Hallmayer, A. Markov, V. Nedkova, I. Tournev, L. Aneva, and R. Gitzelmann, (1999), A founder mutation in the GK1 gene is responsible for galactokinase deficiency in Roma (Gypsies), *Am. J. Hum. Genet.*, **65**, pp. 1299–1307.

[37] A. M. Bosch, H. D. Bakker, A. H. van Gennip, J. V. van Kempen, R. J. A. Wanders, and F. A. Wijburg, (2003), Clinical features of galactokinase deficiency: A review of the literature, *J. Inherit. Metab. Dis*., **25**, pp. 629–634.

[38] J. B. Holton, M. G. Gillett, R. MacCaul, and R. Young, (1981), Galactosaemia: A new severe variant due to uridine diphosphate galactose-4-epimerase deficiency, *Arch. Dis. Child*., **56**, pp. 885–887.

[39] I. B. Sardharwalla, J. E. Wraith, C. Bridge, B. Fowler, and S. A. Roberts, (1988), A patient with severe type of epimerase deficiency galactosaemia, *J. Inherit. Metab. Dis*., **11**, pp. 249– 251.

[40] T. M. Wohlers, N. C. Christacos, M. T. Harreman, and J. L. Fridovich-Keil, (1999), Identification and characterization of a mutation, in the human UDP-galactose-4-epimerase gene, associated with generalized epimerase-deficiency galactosemia, *Am. J. Hum. Genet*., **64**, pp. 462–470.

[41] B. B. Quimby, A. Alano, S. Almashanu, A. M. DeSandro, T. M. Cowan, and J. L. Fridovich-Keil, (1997), Characterization of two mutations associated with epimerase-deficiency galactosemia, by use of a yeast expression system for human UDP-galactose-4- epimerase, *Am. J. Hum. Genet.*, **61**, pp. 590–598.

[42] A. Alano, S. Almashanu, J. M. Chinsky, P. Costeas, M. G. Blitzer, E. A. Wulfsberg, and T. M. Cowan, (1998), Molecular characterization of a unique patient with epimerase-deficiency galactosaemia, *J. Inherit. Metab. Dis.*, **21**, pp. 341–350.

[43] Y. Wada, A. Kikuchi, N. Arai-Ichinoi, O. Sakamoto, Y. Takezawa, S. Iwasawa, T. Niihori, H. Nyuzuki, Y. Nakajima, E. Ogawa, M. Ishige, H. Hirai, H. Sasai, R. Fujiki, M. Shirota, R. Funayama, M. Yamamoto, T. Ito, O. Ohara, K. Nakayama, Y. Aoki, S. Koshiba, T. Fukao, and S. Kure, (2019), Biallelic GALM pathogenic variants cause a novel type of galactosemia, *Genet. Med.*, **21**, pp. 1286–1294.

[44] H.M. Cheng, H. Xiong, J. Xiong, and K. Tsubota, (1990), Metabolic studies of galactosaemic cataract, *Exp. Eye Res*., **51**, pp. 345–349.

[45] K. Lai and L. J. Elsas, (2000), Overexpression of human UDP-glucose pyrophosphorylase rescues galactose-1-phosphate uridyltransferase-deficient yeast, *Biochem. Biophys. Res. Commun*., **271**, pp. 392–400.

[46] D. V. Mehta, A. Kabir, and P. J. Bhat, (1999), Expression of human inositol monophosphatase suppresses galactose toxicity in Saccharomyces cerevisiae: Possible implications in galactosemia, *Biochim. Biophys. Acta - Mol. Basis Dis.*, **1454**, pp. 217–226.

[47] J. Charlwood, P. Clayton, G. Keir, N. Mian, and B. Winchester, (1998), Defective galactosylation of serum transferrin in galactosemia, *Glycobiology*, **8**, pp. 351–357.

[48] J. L. Fridovich-Keil and S. Jinks-Robertson, (1993), A yeast expression system for human galactose-1-phosphate uridylyltransferase, *Proc. Natl. Acad. Sci. U. S. A.*, **90**, pp. 398–402.

[49] N. C. Christacos, M. J. Marson, L. Wells, K. Riehman, and J. L. Fridovich-Keil, (2000), Subcellular localization of galactose-1-phosphate uridylyltransferase in the yeast Saccharomyces cerevisiae, *Mol. Genet. Metab*., **70**, pp. 272–280.

[50] P. Jumbo-Lucioni, W. Parkinson, and K. Broadie, (2014), Over-elaborated synaptic architecture and reduced synaptomatrix glycosylation in a Drosophila classic galactosemia disease model, *DMM Dis. Model. Mech*., **7**, pp. 1365–1378.

[51] J. M. I. Daenzer, P. P. Jumbo-Lucioni, M. L. Hopson, K. R. Garza, E. L. Ryan, and J. L. Fridovich-Keil, (2016), Acute and long-term outcomes in a Drosophila melanogaster model of classic galactosemia occur independently of galactose-1-phosphate accumulation, *DMM Dis. Model. Mech*., **9**, pp. 1375–1382.

[52] J. M. Vanoevelen, B. van Erven, J. Bierau, X. Huang, G. T. Berry, R. Vos, A. I. Coelho, and M. E. Rubio-Gozalbo, (2018), Impaired fertility and motor function in a zebrafish model for classic galactosemia, *J. Inherit. Metab. Dis*., **41**, pp. 117–127.

[53] N. D. Leslie, K. L. Yager, P. D. McNamara, and S. Segal, (1996), A mouse model of galactose1-phosphate uridyl transferase deficiency, *Biochem. Mol. Med.*, **59**, pp. 7–12.

[54] B. Balakrishnan, A. Siddiqi, J. Mella, A. Lupo, E. Li, J. Hollien, J. Johnson, and K. Lai, (2019), Salubrinal enhances eIF2α phosphorylation and improves fertility in a mouse model of Classic Galactosemia, *Biochim. Biophys. Acta - Mol. Basis Dis.*, **1865**, p. 165516.

[55] R. P. Holmes and D. G. Assimos, (1998), Glyoxylate synthesis, and its modulation and influence on oxalate synthesis, *J. Urol.*, **160**, pp. 1617–1624.

[56] C. B. Summitt, L. C. Johnson, T. J. Jönsson, D. Parsonage, R. P. Holmes, W. T. Lowther, C. Summitt, L. Johnson, T. Jönsson, and R. Holmes, (2015), Proline dehydrogenase 2 (PRODH2) is a hydroxyproline dehydrogenase (HYPDH) and molecular target for treating primary hyperoxaluria, *Biochem J*., **466(2)**, pp. 273–281.

[57] D. Srivastava, R. K. Singh, M. A. Moxley, M. T. Henzl, D. F. Becker, and J. J. Tanner, (2012), The Three-Dimensional Structural Basis of Type II Hyperprolinaemia, *J. Mol. Biol*., **420**, pp. 176–189.

[58] X. Jiang, J. Wang, H. Chang, and Y. Zhou, (2016), Recombinant expression, purification and crystallographic studies of the mature form of human mitochondrial aspartate aminotransferase, *Biosci. Trends*, **10**, pp. 79–84.

[59] T. J. Riedel, L. C. Johnson, J. Knight, R. R. Hantgan, R. P. Holmes, and W. Todd, (2011),

Structural and Biochemical Studies of Human 4-hydroxy- 2-oxoglutarate Aldolase : Implications for Hydroxyproline Metabolism in Primary Hyperoxaluria, 6.
[60] G. Rumsby and D. P. Cregeen, (1999), Identification and expression of a cDNA for human hydroxypyruvate/glyoxylate reductase, *Biochim. Biophys. Acta - Gene Struct. Expr*., **1446**, pp. 383–388.
[61] M. P. S. Booth, R. Conners, G. Rumsby, and R. L. Brady, (2006), Structural Basis of Substrate Specificity in Human Glyoxylate Reductase / Hydroxypyruvate Reductase, *J. Mol. Biol.*, **360(1)**, pp. 178–189.
[62] J. M. Jones, J. C. Morrell, and S. J. Gould, (2000), Identification and characterization of HAOX1, HAOX2, and HAOX3, three human peroxisomal 2-hydroxy acid oxidases., *J. Biol. Chem*., **275**, pp. 12590–7.
[63] M. Nagata, A. Ichiyama, T. Takayama, T. Oda, S. Mugiya, and S. Ozono, (2009), Assay of alanine:glyoxylate aminotransferase in human liver by its serine: glyoxylate aminotransferase activity., *Biomed. Res*., **30**, pp. 295–301.
[64] S. B. Strauss, T. Waltuch, W. Bivin, F. Kaskel, and T. L. Levin, (2017), Primary hyperoxaluria: spectrum of clinical and imaging findings, *Pediatr. Radiol.*, **47**, pp. 96–103.
[65] T. Ermer, K. Eckardt, P. S. Aronson, and F. Knauf, (2016), Oxalate, inflammasome, and progression of kidney disease, *Curr. Opin. Nephrol. Hypertens.*, **25**, pp. 363–371.
[66] E. L. Williams, C. Acquaviva, A. Amoroso, F. Chevalier, M. Coulter-Mackie, C. G. Monico, D. Giachino, T. Owen, A. Robbiano, E. Salido, H. Waterham, and G. Rumsby, (2009), Primary hyperoxaluria type 1: Update and additional mutation analysis of the AGXT gene, *Hum. Mutat*., **30**, pp. 910–917.
[67] S. D. Cramer, P. M. Ferree, K. Lin, D. S. Milliner, and R. P. Holmes, (1999), The gene encoding hydroxypyruvate reductase (GRHPR) is mutated in patients with primary hyperoxaluria type II, *Hum. Mol. Genet.*, **8**, pp. 2063–2069.
[68] J. R. Macdonald, A. D. Huang, and K. M. Loomes, (2016), Cellular degradation of 4-hydroxy2-oxoglutarate aldolase leads to absolute deficiency in primary hyperoxaluria type 3, *FEBS Lett.*, **590**, pp. 1467–1476.
[69] P. Cochat, A. Deloraine, M. Rotily, F. Olive, I. Liponski, and N. Deries, (1995), Epidemiology of primary hyperoxaluria type 1. Societe de Nephrologie and the Societe de Nephrologie Pediatrique., *Nephrol. Dial. Transplant*, **10 Suppl. 8**, pp. 3–7.
[70] C. S. van Woerden, J. W. Groothoff, R. J. A. Wanders, J.-C. Davin, and F. A. Wijburg, (2003), Primary hyperoxaluria type 1 in The Netherlands: prevalence and outcome., *Nephrol. Dial. Transplant*, **18**, pp. 273–279.
[71] K. Hopp, A. G. Cogal, E. J. Bergstralh, B. M. Seide, J. B. Olson, A. M. Meek, J. C. Lieske, D. S. Milliner, and P. C. Harris, (2015), Phenotype-Genotype Correlations and Estimated Carrier Frequencies of Primary Hyperoxaluria, *J. Am. Soc. Nephrol.*, **26**, pp. 2559–2570.
[72] E. L. Williams, C. Acquaviva, A. Amoroso, F. Chevalier, M. Coulter-Mackie, C. G. Monico, D. Giachino, T. Owen, A. Robbiano, E. Salido, H. Waterham, and G. Rumsby, (2009), Primary hyperoxaluria type 1: Update and additional mutation analysis of the AGXT gene, *Hum. Mutat*., **30**, pp. 910–917.
[73] P. E. Purdue, Y. Takada, and C. J. Danpure, (1990), Identification of mutations associated with peroxisome-to-mitochondrion mistargeting of alanine/glyoxylate aminotransferase in primary hyperoxaluria type 1, *J. Cell Biol*., **111**, pp. 2341–2351.
[74] M. B. Coulter-Mackie and Q. Lian, (2006), Consequences of missense mutations for dimerization and turnover of alanine:glyoxylate aminotransferase: Study of a spectrum of mutations, *Mol. Genet. Metab*., **89**, pp. 349–359.
[75] M. D. Lage, A. M. C. Pittman, A. Roncador, B. Cellini, and C. L. Tucker, (2014), Allelespecific characterization of alanine: Glyoxylate aminotransferase variants associated with primary hyperoxaluria, *PLoS One*, **9**, pp. 1–9.
[76] E. D. Hopper, A. M. C. Pittman, M. C. Fitzgerald, and C. L. Tucker, (2008), In vivo and in vitro examination of stability of primary hyperoxaluria-associated human alanine:glyoxylate aminotransferase, *J. Biol. Chem*., **283**, pp. 30493–30502.
[77] B. Cellini, A. Lorenzetto, R. Montioli, E. Oppici, and C. B. Voltattorni, (2010), Human liver peroxisomal alanine:glyoxylate aminotransferase: Different stability under chemical stress of

the major allele, the minor allele, and its pathogenic G170R variant, *Enferm. Infecc. Microbiol. Clin.*, **28**, pp. 1801–1811.
[78] M. J. Kemper, S. Conrad, and D. E. Müller-Wiefel, (1997), Primary hyperoxaluria type 2, *Eur. J. Pediatr.*, **156**, pp. 509–512.
[79] D. S. Milliner, D. M. Wilson, and L. H. Smith, (2001), Phenotypic expression of primary hyperoxaluria: Comparative features of types I and II, *Kidney Int.*, **59**, pp. 31–36.
[80] T. Dhondup, E. C. Lorenz, D. S. Milliner, and J. C. Lieske, (2018), Combined Liver-Kidney Transplantation for Primary Hyperoxaluria Type 2: A Case Report, *Am. J. Transplant.*, **18**, pp. 253–257.
[81] E. L. Williams, D. Bockenhauer, W. G. Van Hoff, N. Johri, C. Laing, M. D. Sinha, R. Unwin, A. Viljoen, and G. Rumsby, (2012), The enzyme 4-hydroxy-2-oxoglutarate aldolase is deficient in primary hyperoxaluria type 3, *Nephrol Dial Transplant*, **27(8)**, pp. 3191–3195.
[82] S. Yamaguchi, J. H. Wiessner, A. T. Hasegawa, L. Y. Hung, G. S. Mandel, and N. S. Mandel, (2005), Study of a rat model for calcium oxalate crystal formation without severe renal damage in selected conditions, *Int. J. Urol.*, **12**, pp. 290–298.
[83] Y. Li and K. E. McMartin, (2009), Strain differences in urinary factors that promote calcium oxalate crystal formation in the kidneys of ethylene glycol-treated rats, *Am. J. Physiol. - Ren. Physiol.*, **296**, pp. 1080–1087.
[84] M. L. Green, M. Hatch, and R. W. Freel, (2005), Ethylene glycol induces hyperoxaluria without metabolic acidosis in rats, *Am. J. Physiol. - Ren. Physiol.*, **289**, pp. 536–543.
[85] S. R. Khan, P. A. Glenton, and K. J. Byer, (2006), Modeling of hyperoxaluric calcium oxalate nephrolithiasis: Experimental induction of hyperoxaluria by hydroxy-L-proline, *Kidney Int.*, **70**, pp. 914–923.
[86] H. Yang, M. Male, Y. Li, N. Wang, C. Zhao, S. Jin, J. Hu, Z. Chen, Z. Ye, and H. Xu, (2018), Efficacy of Hydroxy-L-proline (HYP) analogs in the treatment of primary hyperoxaluria in Drosophila Melanogaster., *BMC Nephrol.*, **19**, p. 167.
[87] E. C. Salido, X. M. Li, Y. Lu, X. Wang, A. Santana, N. Roy-Chowdhury, A. Torres, L. J. Shapiro, and J. Roy-Chowdhury, (2006), Alanine-glyoxylate aminotransferase-deficient mice, a model for primary hyperoxaluria that responds to adenoviral gene transfer, *Proc. Natl. Acad. Sci. U. S. A.*, **103**, pp. 18249–18254.
[88] A. Roncador, E. Oppici, R. Montioli, F. Maset, and B. Cellini, (2013), TAT-mediated delivery of human alanine: Glyoxylate aminotransferase in a cellular model of primary hyperoxaluria type I, *Int. J. Pept. Res. Ther.*, **19**, pp. 175–184.
[89] E. Oppici, R. Montioli, M. Dindo, L. MacCari, V. Porcari, A. Lorenzetto, S. Chellini, C. B. Voltattorni, and B. Cellini, (2015), The Chaperoning Activity of Amino-oxyacetic Acid on Folding-Defective Variants of Human Alanine:Glyoxylate Aminotransferase Causing Primary Hyperoxaluria Type I, *ACS Chem. Biol.*, **10**, pp. 2227–2236.
[90] J. T. Behnam, E. L. Williams, S. Brink, G. Rumsby, and C. J. Danpure, (2006), Reconstruction of human hepatocyte glyoxylate metabolic pathways in stably transformed Chinese-hamster ovary cells, *Biochem. J.*, **394**, pp. 409–416.
[91] V. M. Ginocchio and N. Brunetti-Pierri, (2016), Progress toward improved therapies for inborn errors of metabolism, *Hum. Mol. Genet.*, **25**, pp. R27–R35.
[92] W. W. Yue, S. Mackinnon, and G. A. Bezerra, (2019), Substrate reduction therapy for inborn errors of metabolism, *Emerg. Top. Life Sci.*, **3**, pp. 63–73.
[93] M. Alfadhel, K. Al-Thihli, H. Moubayed, W. Eyaid, and M. Al-Jeraisy, (2013), Drug treatment of inborn errors of metabolism: A systematic review, *Arch. Dis. Child.*, **98**, pp. 454–461.
[94] N. Blau, (2016), Genetics of Phenylketonuria: Then and Now, *Hum. Mutat.*, **37**, pp. 508–515.
[95] A. M. J. Van Wegberg, A. MacDonald, K. Ahring, A. Bélanger-Quintana, N. Blau, A. M. Bosch, A. Burlina, J. Campistol, F. Feillet, M. Giżewska, S. C. Huijbregts, S. Kearney, V. Leuzzi, F. Maillot, A. C. Muntau, M. Van Rijn, F. Trefz, J. H. Walter, and F. J. Van Spronsen, (2017), The complete European guidelines on phenylketonuria: Diagnosis and treatment, *Orphanet J. Rare Dis.*, **12**, pp. 1–56.
[96] L. I. Woolf, R. Griffiths, and A. Moncrieff, (1955), Treatment of phenylketonuria with a diet low in phenylalanine, *Br. Med. J.*, **1**, pp. 57–64.
[97] V. M. Bruinenberg, D. Van Vliet, E. van der Goot, D. S. Counotte, M. Kuhn, F. J. Van

Spronsen, and E. A. van der Zee, (2019), Long-term dietary intervention with low Phe and/or a specific nutrient combination improve certain aspects of brain functioning in phenylketonuria (PKU), *PLoS One*, **14**, pp. 1–20.
[98] P. T. Clayton, (2006), B6-responsive disorders: A model of vitamin dependency, *J. Inherit. Metab. Dis.*, **29**, pp. 317–326.
[99] F. Porta, M. Spada, and A. Ponzone, (2017), Early screening for tetrahydrobiopterin responsiveness in phenylketonuria, *Pediatrics*, **140**, pp. 2015–2019.
[100] S. Harvey Mudd, (1974), Inborn errors of metabolism: Vitamin-responsive genetic disease, *J. Clin. Pathol.*, **8**, pp. 38–47.
[101] S. Adam, R. Akroyd, S. Bernabei, S. Bollhalder, S. Boocock, A. Burlina, T. Coote, K. Corthouts, J. Dalmau, S. Dawson, S. Defourny, A. De Meyer, A. Desloovere, Y. Devlin, M. Diels, K. Dokoupil, S. Donald, S. Evans, I. Fasan, C. Ferguson, S. Ford, M. Forga, G. Gallo, S. C. Grünert, M. Heddrich-Ellerbrok, C. Heidenborg, C. Jonkers, K. Lefebure, K. Luyten, A. MacDonald, U. Meyer, A. Micciche, E. Müller, P. Portnoi, S. Ripley, M. Robert, L. V. Robertson, S. Rosenbaum-Fabian, K. Sahm, S. Schultz, K. Singleton, E. Sjöqvist, L. Stoelen, A. Terry, S. Thompson, C. Timmer, K. Vande Kerckhove, L. van der Ploeg, M. Van Driessche, M. van Rijn, A. van Teeffelen-Heithoff, I. Vitoria, C. Voillot, J. Wenz, M. Westbrook, J. Wildgoose, and H. Zweers, (2015), How strict is galactose restriction in adults with galactosaemia? International practice, *Mol. Genet. Metab.*, **115**, pp. 23–26.
[102] B. Hoffmann, N. Dragano, and S. Schweitzer-Krantz, (2012), Living situation, occupation and health-related quality of life in adult patients with classic galactosemia, *J. Inherit. Metab. Dis.*, **35**, pp. 1051–1058.
[103] A. E. Ten Hoedt, H. Maurice-Stam, C. C. A. Boelen, M. E. Rubio-Gozalbo, F. J. Van Spronsen, F. A. Wijburg, A. M. Bosch, and M. A. Grootenhuis, (2011), Parenting a child with phenylketonuria or galactosemia: Implications for health-related quality of life, *J. Inherit. Metab. Dis.*, **34**, pp. 391–398.
[104] P. Cochat, S. A. Hulton, C. Acquaviva, C. J. Danpure, M. Daudon, M. De Marchi, S. Fargue, J. Groothoff, J. Harambat, B. Hoppe, N. V. Jamieson, M. J. Kemper, G. Mandrile, M. Marangella, S. Picca, G. Rumsby, E. Salido, M. Straub, C. S. van Woerden, and OxalEurope, (2012), Primary hyperoxaluria Type 1: indications for screening and guidance for diagnosis and treatment., *Nephrol. Dial. Transplant*, **27**, pp. 1729–1736.
[105] H. Brzica, D. Breljak, B. C. Burckhardt, G. Burckhardt, and I. Sabolić, (2013), Oxalate: From the Environment to Kidney Stones, *Arch. Ind. Hyg. Toxicol.*, **64**, pp. 609–630.
[106] M. L. Schilsky, (2013), Transplantation for inherited metabolic disorders of the liver, *Transplant. Proc.*, **45**, pp. 455–462.
[107] M. Tang, S. I. Odejinmi, H. Vankayalapati, K. J. Wierenga, and K. Lai, (2012), Innovative therapy for Classic Galactosemia - Tale of two HTS, Mol. Genet. Metab., 105, pp. 44–55.
[108] J. I. Malone, A. Diaz-Thomas, and K. Swan, (2011), Problems with the newborn screen for galactosaemia, *BMJ Case Rep.*, pp. 9–11.
[109] G. Bollée, P. Cochat, and M. Daudon, (2015), Recurrence of crystalline nephropathy after kidney transplantation in APRT deficiency and primary hyperoxaluria, *Can. J. Kidney Heal. Dis.*, **2:31**, pp. 1–13.
[110] P. Cochat, J. M. Gaulier, P. C. Koch Nogueira, J. Feber, N. V. Jamieson, M.-O. Rolland, P. Divry, D. Bozon, and L. Dubourg, (1999), Combined liver-kidney transplantation in primary hyperoxaluria type 1, *Eur. J. Pediatr.*, **158**, pp. S075–S080.
[111] M. J. Kemper, (2005), The role of pre-emptive liver transplantation in primary hyperoxaluria type 1., *Urol. Res.*, **33**, pp. 376–9.
[112] B. Balakrishnan, D. An, V. Nguyen, C. DeAntonis, P. G. V. Martini, and K. Lai, (2019), Novel mRNA-Based Therapy Reduces Toxic Galactose Metabolites and Overcomes Galactose Sensitivity in a Mouse Model of Classic Galactosemia, *Mol. Ther.*, **28(1)**, pp. 304-312.
[113] A. I. Coelho, S. Lourenço, M. Trabuco, M. J. Silva, A. Oliveira, A. Gaspar, L. Diogo, I. Tavares De Almeida, J. B. Vicente, and I. Rivera, (2015), Functional correction by antisense therapy of a splicing mutation in the GALT gene, *Eur. J. Hum. Genet.*, **23**, pp. 500–506.
[114] E. Salido, M. Rodriguez-Pena, A. Santana, S. G. Beattie, H. Petry, and A. Torres, (2011), Phenotypic correction of a mouse model for primary hyperoxaluria with adeno-associated virus

gene transfer, *Mol. Ther.*, **19**, pp. 870–875.
[115] R. Castello, R. Borzone, S. D'Aria, P. Annunziata, P. Piccolo, and N. Brunetti-Pierri, (2016), Helper-dependent adenoviral vectors for liver-directed gene therapy of primary hyperoxaluria type 1, *Gene Ther.*, **23**, pp. 129–134.
[116] T. Slepak, M. Tang, F. Addo, and K. Lai, (2005), Intracellular galactose-1-phosphate accumulation leads to environmental stress response in yeast model, *Mol. Genet. Metab.*, **86**, pp. 360–371.
[117] P. Jumbo-Lucioni, W. Parkinson, and K. Broadie, (2014), Over-elaborated synaptic architecture and reduced synaptomatrix glycosylation in a Drosophila classic galactosemia disease model, *Dis. Model. Mech.*, **7**, pp. 1365–1378.
[118] B. Balakrishnan, C. Nicholas, A. Siddiqi, W. Chen, E. Bales, M. Feng, J. Johnson, and K. Lai, (2017), Reversal of aberrant PI3K/Akt signalling by Salubrinal in a GalT-deficient mouse model, *Biochim. Biophys. Acta - Mol. Basis Dis.*, **1863**, pp. 3286–3293.
[119] T. I. Slepak, M. Tang, V. Z. Slepak, and K. Lai, (2007), Involvement of endoplasmic reticulum stress in a novel Classic Galactosemia model, *Mol. Genet. Metab.*, **92**, pp. 78–87.
[120] A. d'Acierno, B. Scafuri, A. Facchiano, and A. Marabotti, (2018), The evolution of a Web resource: The Galactosemia Proteins Database 2.0, *Hum. Mutat.*, **39**, pp. 52–60.
[121] T. J. McCorvie, T. J. Gleason, J. L. Fridovich-Keil, and D. J. Timson, (2013), Misfolding of galactose 1-phosphate uridylyltransferase can result in type I galactosemia, *Biochim. Biophys. Acta - Mol. Basis Dis.*, **1832**, pp. 1279–1293.
[122] A. I. Coelho, M. Trabuco, R. Ramos, M. J. Silva, I. Tavares de Almeida, P. Leandro, I. Rivera, and J. B. Vicente, (2014), Functional and structural impact of the most prevalent missense mutations in classic galactosemia, *Mol. Genet. Genomic Med.*, **2**, pp. 484–496.
[123] A. I. Coelho, M. Trabuco, M. J. Silva, I. T. de Almeida, P. Leandro, I. Rivera, and J. B. Vicente, (2015), Arginine Functionally Improves Clinically Relevant Human Galactose-1-Phosphate Uridylyltransferase (GALT) Variants Expressed in a Prokaryotic Model, *JIMD Reports*, **23**, pp. 1–6.
[124] M. Haskovic, B. Derks, L. van der Ploeg, J. Trommelen, J. Nyakayiru, L. J. C. van Loon, S. Mackinnon, W. W. Yue, R. W. A. Peake, L. Zha, D. Demirbas, W. Qi, X. Huang, G. T. Berry, J. Achten, J. Bierau, M. E. Rubio-Gozalbo, and A. I. Coelho, (2018), Arginine does not rescue p.Q188R mutation deleterious effect in classic galactosemia, *Orphanet J. Rare Dis.*, **13**, p. 212.
[125] N. Miyata, J. Steffen, M. E. Johnson, S. Fargue, C. J. Danpure, and C. M. Koehler, (2014), Pharmacologic rescue of an enzyme-trafficking defect in primary hyperoxaluria 1., *Proc. Natl. Acad. Sci. U. S. A.*, **111**, pp. 14406–11.
[126] S. Hou, F. Madoux, L. Scampavia, J. A. Janovick, P. M. Conn, and T. P. Spicer, (2017), Drug Library Screening for the Identification of Ionophores That Correct the Mistrafficking Disorder Associated with Oxalosis Kidney Disease, *SLAS Discov.*, **22**, pp. 887–896.
[127] N. Mesa-Torres, I. Fabelo-Rosa, D. Riverol, C. Yunta, A. Albert, E. Salido, and A. L. Pey, (2013), The Role of Protein Denaturation Energetics and Molecular Chaperones in the Aggregation and Mistargeting of Mutants Causing Primary Hyperoxaluria Type I, *PLoS One*, **8(8)**, e71963.
[128] M. J. Lumb, G. M. Birdsey, and C. J. Danpure, (2003), Correction of an enzyme trafficking defect in hereditary kidney stone disease *in vitro*, *Biochem. J.*, **87**, pp. 79–87.
[129] D. A. Gibbs and R. W. Watts, (1970), The action of pyridoxine in primary hyperoxaluria., *Clin. Sci.*, **38**, pp. 277–286.
[130] C. G. Monico, J. B. Olson, and D. S. Milliner, (2005), Implications of Genotype and Enzyme Phenotype in Pyridoxine Response of Patients with Type I Primary Hyperoxaluria, *Am. J. Nephrol.*, **25(2)**, pp. 183–188.
[131] C.S. van Woerden, J.W. Groothoff, F.A. Wijburg, C. Annink, R.J. Wanders, H.R. Waterham. (2004), Clinical implications of mutation analysis in primary hyperoxaluria type 1, *Kidney Int.*, **66**, pp. 746–752.
[132] S. Fargue, J. Knight, R. P. Holmes, G. Rumsby, and C. J. Danpure, (2016), Effects of alanine:glyoxylate aminotransferase variants and pyridoxine sensitivity on oxalate metabolism in a cell-based cytotoxicity assay, *Biochim. Biophys. Acta - Mol. Basis Dis.*, **1862**, pp. 1055–1062.

[133] M. B. Boxer, M. Shen, C. Tanega, M. Tang, K. Lai, and D. S. Auld, (2010), Toward Improved Therapy for Classic Galactosemia, *Genome Res*., **152**, pp. 1–19.
[134] M. Tang, K. Wierenga, L. J. Elsas, and K. Lai, (2010), Molecular and biochemical characterization of human galactokinase and its small molecule inhibitors, *Chem. Biol. Interact*., **188**, pp. 376–385.
[135] L. Liu, M. Tang, M. J. Walsh, K. R. Brimacombe, R. Pragani, C. Tanega, J. M. Rohde, H. L. Baker, E. Fernandez, B. Blackman, J. M. Bougie, W. H. Leister, D. S. Auld, M. Shen, K. Lai, and M. B. Boxer, (2015), Structure activity relationships of human galactokinase inhibitors, *Bioorganic Med. Chem. Lett*., **25**, pp. 721–727.
[136] F. Chiappori, I. Merelli, L. Milanesi, and A. Marabotti, (2013), Static and dynamic interactions between GALK enzyme and known inhibitors: Guidelines to design new drugs for galactosemic patients, *Eur. J. Med. Chem*., **63**, pp. 423–434.
[137] S. I. Odejinmi, R. G. Rascon, M. Tang, H. Vankayalapati, and K. Lai, (2011), Structureactivity analysis and cell-based optimization of human galactokinase inhibitors, *ACS Med. Chem. Lett.*, **2**, pp. 667–672.
[138] X. Hu, Y. Q. Zhang, O. W. Lee, L. Liu, M. Tang, K. Lai, M. B. Boxer, M. D. Hall, and M. Shen, (2019), Discovery of novel inhibitors of human galactokinase by virtual screening, *J. Comput. Aided. Mol. Des.*, **33**, pp. 405–417.
[139] K. J. Wierenga, K. Lai, P. Buchwald, and M. Tang, (2008), High-throughput screening for human galactokinase inhibitors., *J. Biomol. Screen*., **13**, pp. 415–23.
[140] M. J. Robinson, I. S. Menzies, and I. Sloan, (1980), Hydroxyprolineaemia with normal development, *Arch. Dis. Child*., **55**, pp. 484–486.
[141] C. Staufner, T. B. Haack, P. Feyh, G. Gramer, D. E. Raga, C. Terrile, S. Sauer, J. G. Okun, J. Fang-Hoffmann, E. Mayatepek, H. Prokisch, G. F. Hoffmann, and S. Kölker, (2016), Genetic cause and prevalence of hydroxyprolinemia, *J. Inherit. Metab. Dis*., **39**, pp. 625–632.
[142] Y. Frishberg, A. Zeharia, R. Lyakhovetsky, R. Bargal, and R. Belostotsky, (2014), Mutations in HAO1 encoding glycolate oxidase cause isolated glycolic aciduria., *J. Med. Genet*., **51**, pp. 526–529.
[143] T. L. McGregor, K. A. Hunt, P. Nioi, D. Mason, S. Ticau, M. Pelosi, L. Perry, S. Finer, C. Griffiths, D. MacArthur, R. C. Trembath, D. Oglesbee, J. Lieske, J. Wright, D. Erbe, and D. van Heel, (2020), Deep phenotyping of a healthy human HAO1 knockout informs therapeutic development for primary hyperoxaluria type 1., *eLife*, **9**, e54363.
[144] Y. Takahashi, H. Miyajima, and E. Kaneko, (1995), Genetic Analysis of a Family of Lactate Dehydrogenase a Subunit Deficiency, *Intern. Med*., **34**, pp. 326–329.
[145] T. Kanno, K. Sudo, M. Maekawa, Y. Nishimura, M. Ukita, and K. Fukutake, (1988), Lactate dehydrogenase M-subunit deficiency: a new type of hereditary exertional myopathy, *Clin. Chim. Acta*, **173**, pp. 89–98.
[146] K. Gavrilov and W. M. Saltzman, (2012), Therapeutic siRNA: Principles, challenges, and strategies, Yale J. Biol. Med., 85, pp. 187–200.
[147] X. Li, J. Knight, S. Fargue, B. Buchalski, Z. Guan, E. W. Inscho, A. Liebow, K. Fitzgerald, W. Querbes, W. Todd Lowther, and R. P. Holmes, (2016), Metabolism of (13)C5-hydroxyproline in mouse models of Primary Hyperoxaluria and its inhibition by RNAi therapeutics targeting liver glycolate oxidase and hydroxyproline dehydrogenase., *Biochim. Biophys. Acta*, **1862**, pp. 233–9.
[148] C. Lai, N. Pursell, J. Gierut, U. Saxena, W. Zhou, M. Dills, R. Diwanji, C. Dutta, M. Koser, N. Nazef, R. Storr, B. Kim, C. Martin-Higueras, E. Salido, W. Wang, M. Abrams, H. Dudek, and B. D. Brown, (2018), Specific Inhibition of Hepatic Lactate Dehydrogenase Reduces Oxalate Production in Mouse Models of Primary Hyperoxaluria, *Mol. Ther.*, **26**, pp. 1983–1995.
[149] C. Dutta, N. Avitahl-Curtis, N. Pursell, M. Larsson Cohen, B. Holmes, R. Diwanji, W. Zhou, L. Apponi, M. Koser, B. Ying, D. Chen, X. Shui, U. Saxena, W. A. Cyr, A. Shah, N. Nazef, W. Wang, M. Abrams, H. Dudek, E. Salido, B. D. Brown, and C. Lai, (2016), Inhibition of Glycolate Oxidase With Dicer-substrate siRNA Reduces Calcium Oxalate Deposition in a Mouse Model of Primary Hyperoxaluria Type 1, *Mol. Ther.*, **24**, pp. 770–778.
[150] A. Liebow, X. Li, T. Racie, J. Hettinger, B. R. Bettencourt, N. Najafian, P. Haslett, K. Fitzgerald, R. P. Holmes, D. Erbe, W. Querbes, and J. Knight, (2017), An Investigational RNAi

Therapeutic Targeting Glycolate Oxidase Reduces Oxalate Production in Models of Primary Hyperoxaluria, *J. Am. Soc. Nephrol.*, **28**, pp. 494–503.
[151] N. Zabaleta, M. Barberia, C. Martin-Higueras, N. Zapata-Linares, I. Betancor, S. Rodriguez, R. Martinez-Turrillas, L. Torella, A. Vales, C. Olagüe, A. Vilas-Zornoza, L. Castro-Labrador, D. Lara-Astiaso, F. Prosper, E. Salido, G. Gonzalez-Aseguinolaza, and J. R. Rodriguez-Madoz, (2018), CRISPR/Cas9-mediated glycolate oxidase disruption is an efficacious and safe treatment for primary hyperoxaluria type I., *Nat. Commun.*, **9**, p. 5454.
[152] X. Chi, P. Gatti, and T. Papoian, (2017), Safety of antisense oligonucleotide and siRNA-based therapeutics, *Drug Discov. Today*, **22**, pp. 823–833.
[153] J. Domingos, V. Ricotti, A. E. Martinez, and F. Muntoni, (2018), Severe persistent injection site reactions after subcutaneous 2′-O-methyl phosphorothioate oligonucleotide therapy for Duchenne muscular dystrophy, *Neuromuscul. Disord.*, **28**, pp. 176–177.
[154] R. Titze-de-Almeida, C. David, and S. S. Titze-de-Almeida, (2017), The Race of 10 Synthetic RNAi-Based Drugs to the Pharmaceutical Market, *Pharm. Res.*, **34**, pp. 1339–1363.
[155] C. I. Wooddell, D. B. Rozema, M. Hossbach, M. John, H. L. Hamilton, Q. Chu, J. O. Hegge, J. J. Klein, D. H. Wakefield, C. E. Oropeza, J. Deckert, I. Roehl, K. Jahn-Hofmann, P. Hadwiger, H. P. Vornlocher, A. McLachlan, and D. L. Lewis, (2013), Hepatocyte-targeted RNAi therapeutics for the treatment of chronic hepatitis B virus infection, *Mol. Ther.*, **21**, pp. 973–985.
[156] J. E. Zuckerman and M. E. Davis, (2015), Clinical experiences with systemically administered siRNA-based therapeutics in cancer, *Nat. Rev. Drug Discov.*, **14**, pp. 843–856.
[157] K. K. Hou, H. Pan, P. H. Schlesinger, and S. A. Wickline, (2015), A role for peptides in overcoming endosomal entrapment in siRNA delivery - A focus on melittin., *Biotechnol. Adv.*, **33**, pp. 931–40.
[158] M. Congreve, R. Carr, C. Murray, and H. Jhoti, (2003), A "Rule of Three" for fragment-based lead discovery?, *Drug Discov. Today*, **8**, pp. 876–877.
[159] S. Schultes, C. De Graaf, E. E. J. Haaksma, I. J. P. De Esch, R. Leurs, and O. Krämer, (2010), Ligand efficiency as a guide in fragment hit selection and optimization, *Drug Discov. Today Technol.*, **7**, pp. 157–162.
[160] R. S. Bohacek, C. McMartin, and W. C. Guida, (1996), The art and practice of structure-based drug design: A molecular modeling perspective, *Med. Res. Rev.*, **16**, pp. 3–50.
[161] L. Ruddigkeit, R. Van Deursen, L. C. Blum, and J. L. Reymond, (2012), Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17, *J. Chem. Inf. Model.*, **52**, pp. 2864–2875.
[162] A. R. Bradley, A. Echalier, M. Fairhead, C. Strain-Damerell, P. Brennan, A. N. Bullock, N. A. Burgess-Brown, E. P. Carpenter, O. Gileadi, B. D. Marsden, W. H. Lee, W. Yue, C. Bountra, and F. Von Delft, (2017), The SGC beyond structural genomics: redefining the role of 3D structures by coupling genomic stratification with fragment-based discovery, *Essays Biochem.*, **61**, pp. 495–503.
[163] D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke, and H. Jhoti, (2016), Twenty years on: The impact of fragments on drug discovery, *Nat. Rev. Drug Discov.*, **15**, pp. 605–619.
[164] A. Bradley, P. E. Brennan, T. Krojer, B. D. Marsden, O. B. Cox, P. Collins, J. Spencer, O. Monteiro, R. Talon, J. Amin, O. Fedorov, F. von Delft, 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, (2015), A poised fragment library enables rapid synthetic expansion yielding the first reported inhibitors of PHIP(2), an atypical bromodomain, *Chem. Sci.*, **7**, pp. 2322–2330.
[165] Q. Cao, R. Goodnow, J. Patel, A. McKenzie, F. Edfeldt, H. Schönherr, S. Cowen, L. Spadola, A. Rabow, and N. Fuller, (2016), An improved model for fragment-based lead generation at AstraZeneca, *Drug Discov. Today*, **21**, pp. 1272–1283.
[166] D. Joseph-McCarthy, A. J. Campbell, G. Kern, and D. Moustakas, (2014), Fragment-based lead discovery and design, *J. Chem. Inf. Model.*, **54**, pp. 693–704.
[167] J. Barker, S. Courtney, T. Hesterkamp, D. Ullmann, and M. Whittaker, (2006), Fragment screening by biochemical assay., *Expert Opin. Drug Discov.*, **1**, pp. 225–236.
[168] J.P. Renaud, C. Chung, U. H. Danielson, U. Egner, M. Hennig, R. E. Hubbard, and H. Nar, (2016), Biophysics in drug discovery: impact, challenges and opportunities, *Nat. Rev. Drug*

Discov., **15**, pp. 679–698.
[169] T. Sugiki, K. Furuita, T. Fujiwara, and C. Kojima, (2018), Current NMR techniques for structure-based drug discovery, *Molecules*, **23**, pp. 1–27.
[170] C. A. Shepherd, A. L. Hopkins, and I. Navratilova, (2014), Fragment screening by SPR and advanced application to GPCRs, *Prog. Biophys. Mol. Biol*., **116**, pp. 113–123.
[171] T. G. Davies, W. E. Wixted, J. E. Coyle, C. Griffiths-Jones, K. Hearn, R. McMenamin, D. Norton, S. J. Rich, C. Richardson, G. Saxty, H. M. G. Willems, A. J. A. Woolford, J. E. Cottom, J. P. Kou, J. G. Yonchuk, H. G. Feldser, Y. Sanchez, J. P. Foley, B. J. Bolognese, G. Logan, P. L. Podolin, H. Yan, J. F. Callahan, T. D. Heightman, and J. K. Kerns, (2016), Monoacidic Inhibitors of the Kelch-like ECH-Associated Protein 1: Nuclear Factor Erythroid 2-Related Factor 2 (KEAP1:NRF2) Protein-Protein Interaction with High Cell Potency Identified by Fragment-Based Discovery, *J. Med. Chem.*, **59**, pp. 3991–4006.
[172] J. C. Spurlino, Fragment screening purely with protein crystallography, 1st ed., 493. Elsevier Inc., 2011.
[173] W. W. Yue, (2016), From structural biology to designing therapy for inborn errors of metabolism, *J. Inherit. Metab. Dis.*, **39**, pp. 489–498.
[174] S. Cheek, H. Zhang, and N. V. Grishin, (2002), Sequence and structure classification of kinases, *J. Mol. Biol.*, **320**, pp. 855–881.
[175] S. Cheek, K. Ginalski, H. Zhang, and N. V. Grishin, (2005), A comprehensive update of the sequence and structure classification of kinases, *BMC Struct. Biol*., **5**, pp. 1–19.
[176] P. Bork, C. Sander, and A. Valencia, (1993), Convergent evolution of similar enzymatic function on different protein folds: The hexokinase, ribokinase, and galactokinase families of sugar kinases, *Protein Sci.*, **2**, pp. 31–40.
[177] C. H. Trinh, A. Asipu, D. T. Bonthron, and S. E. V. Phillips, (2009), Structures of alternatively spliced isoforms of human ketohexokinase, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **65**, pp. 201–211.
[178] J. Park, P. van Koeverden, B. Singh, and R. S. Gupta, (2007), Identification and characterization of human ribokinase and comparison of its properties with *E. coli* ribokinase and human adenosine kinase, *FEBS Lett*., **581**, pp. 3211–3216.
[179] J. A. Sigrell, A. D. Cameron, T. A. Jones, and S. L. Mowbray, (1998), Structure of *Escherichia coli* ribokinase in complex with ribose and dinucleotide determined to 1.8 A resolution: insights into a new family of kinase structures., *Structure*, **6**, pp. 183–93.
[180] Y. Zhang, M. Dougherty, D. M. Downs, and S. E. Ealick, (2004), Crystal structure of an aminoimidazole riboside kinase from Salmonella enterica: Implications for the evolution of the ribokinase superfamily, *Structure*, **12**, pp. 1809–1821.
[181] J. H. Hurley, (1996), The sugar kinase/heat shock protein 70/actin superfamily: Implications of conserved structure for mechanism, Annu. Rev. Biophys. Biomol. Struct., 25, pp. 137–162.
[182] J. L. Andreassi and T. S. Leyh, (2004), Molecular functions of conserved aspects of the GHMP kinase family, *Biochemistry*, **43**, pp. 14594–14601.
[183] A. Agnew and D. Timson, (2010), Mechanistic studies on human N-acetylgalactosamine kinase, *J. Enzyme Inhib. Med. Chem*., **25**, pp. 370–376.
[184] J. B. Thoden, D. J. Timson, R. J. Reece, and H. M. Holden, (2005), Molecular Structure of Human Galactokinase: Implications for Type II Galactosemia, *J. Biol. Chem.*, **280**, pp. 9662–9670.
[185] J. B. Thoden and H. M. Holden, (2005), The molecular architecture of human N-acetylgalactosamine kinase, *J. Biol. Chem.*, **280**, pp. 32784–32791.
[186] Z. Fu, N. E. Voynova, T. J. Herdendorf, H. M. Miziorko, and J. J. P. Kim, (2008), Biochemical and structural basis for feedback inhibition of mevalonate kinase and isoprenoid metabolism, *Biochemistry*, **47**, pp. 3715–3724.
[187] N. E. Voynova, Z. Fu, K. P. Battaile, T. J. Herdendorf, J.-J. P. Kim, and H. M. Miziorko, (2008), Human mevalonate diphosphate decarboxylase: Characterization, investigation of the mevalonate diphosphate binding site, and crystal structure, *Arch. Biochem. Biophys*., **480**, pp. 58–67.
[188] D. Timson, (2007), GHMP Kinases - Structures, Mechanisms and Potential for Therapeutically Relevant Inhibition, *Curr. Enzym. Inhib*., **3**, pp. 77–94.

[189] P. N. Mortenson and C. W. Murray, (2011), Assessing the lipophilicity of fragments and early hits, *J. Comput. Aided. Mol. Des*., **25**, pp. 663–667.
[190] C. P. Austin, L. S. Brady, T. R. Insel, and F. S. Collins, (2004), NIH Molecular Libraries Initiative., *Science*, **306**, pp. 1138–9.
[191] T. I. Oprea, C. G. Bologa, S. Boyer, R. F. Curpan, R. C. Glen, A. L. Hopkins, C. A. Lipinski, G. R. Marshall, Y. C. Martin, L. Ostopovici-halip, G. Rishton, O. Ursu, R. J. Vaz, and C. Waller, (2013), A Crowdsourcing Evaluation of the NIH Chemical Probes, *Nat. Chem. Biol*., **5**, pp. 441– 447.
[192] F. Corpet, (1988), Multiple sequence alignment with hierarchical clustering, *Nucleic Acids Res*., **16**, pp. 10881–10890.
[193] X. Robert and P. Gouet, (2014), Deciphering key features in protein structures with the new ENDscript server, *Nucleic Acids Res*., **42**, pp. 320–324.
[194] M. McAuley, M. Huang, and D. J. Timson, (2017), Insight into the mechanism of galactokinase: Role of a critical glutamate residue and helix/coil transitions, *Biochim. Biophys. Acta - Proteins Proteomics*, **1865**, pp. 321–328.
[195] M. Huang, X. Li, J. W. Zou, and D. J. Timson, (2013), Role of Arg228 in the phosphorylation of galactokinase: The mechanism of GHMP kinases by quantum mechanics/molecular mechanics studies, Biochemistry, 52, pp. 4858–4868.
[196] P. Singh, B. J. Harden, B. J. Lillywhite, and P. M. Broad, (2004), Identification of kinase inhibitors by an ATP depletion method, Assay Drug Dev. Technol., 2, pp. 161–169.
[197] A. Miyanaga, S. Fujisawa, N. Furukawa, K. Arai, M. Nakajima, and H. Taguchi, (2013), The crystal structure of d-mandelate dehydrogenase reveals its distinct substrate and coenzyme recognition mechanisms from those of 2-ketopantoate reductase, *Biochem. Biophys. Res. Commun.*, **439**, pp. 109–114.
[198] Y. Umena, K. Yorita, T. Matsuoka, A. Kita, K. Fukui, and Y. Morimoto, (2006), The crystal structure of l-lactate oxidase from Aerococcus viridans at 2.1 Å resolution reveals the mechanism of strict substrate recognition, *Biochem. Biophys. Res. Commun*., **350**, pp. 249–256.
[199] A. Belmouden and F. Lederer, (1996), The role of a β barrel loop 4 extension in modulating the physical and functional properties of long-chain 2-hydroxy-acid oxidase isozymes, *Eur. J. Biochem.*, **238**, pp. 790–798.
[200] R. R. Copley and P. Bork, (2000), Homology among (βα)8 barrels: Implications for the evolution of metabolic pathways, *J. Mol. Biol.*, **303**, pp. 627–641.
[201] Y. Lindqvist, (1989), Refined structure of spinach glycolate oxidase at 2 A resolution., *J. Mol. Biol.*, **209**, pp. 151–166.
[202] C. Vignaud, N. Pietrancosta, E. L. Williams, G. Rumsby, and F. Lederer, (2007), Purification and characterization of recombinant human liver glycolate oxidase, *Arch. Biochem. Biophys.*, **465**, pp. 410–416.
[203] M. S. Murray, R. P. Holmes, and W. T. Lowther, (2008), Active site and loop 4 movements within human glycolate oxidase: implications for substrate specificity and drug design., *Biochemistry*, **47**, pp. 2439–49.
[204] W. D. Lienhart, V. Gudipati, and P. MacHeroux, (2013), The human flavoproteome, *Arch. Biochem. Biophys*., **535**, pp. 150–162.
[205] P. MacHeroux, B. Kappes, and S. E. Ealick, (2011), Flavogenomics - A genomic and structural view of flavin-dependent proteins, *FEBS J*., **278**, pp. 2625–2634.
[206] Z. Chen, C. Vignaud, A. Jaafar, B. Lévy, F. Guéritte, D. Guénard, F. Lederer, F. S. Mathews, B. Lévy, F. Lederer, Z. Chen, F. S. Mathews, and A. Jaafar, (2012), High resolution crystal structure of rat long chain hydroxy acid oxidase in complex with the inhibitor 4-carboxy-5-[(4chlorophenyl)sulfanyl]-1, 2, 3-thiadiazole. Implications for inhibitor specificity and drug design, *Biochimie*, **94**, pp. 1172–1179.
[207] Y. Dellero, C. Mauve, E. Boex-Fontvieille, V. Flesch, M. Jossier, G. Tcherkez, and M. Hodges, (2015), Experimental evidence for a hydride transfer mechanism in plant glycolate oxidase catalysis, *J. Biol. Chem.*, **290**, pp. 1689–1698.
[208] P. Macheroux, V. Kieweg, V. Massey, E. Soderlind, K. Stenberg, and Y. Lindqvist, (1993), Role of tyrosine 129 in the active site of spinach glycolate oxidase., *Eur. J. Biochem*., **213**, pp. 1047–1054.

[209] G. A. Reid, S. White, M. T. Black, F. Lederer, F. S. Mathews, and S. K. Chapman, (1988), Probing the active site of flavocytochrome b2 by site-directed mutagenesis., *Eur. J. Biochem.*, **178**, pp. 329–333.
[210] L. M. Cunane, J. D. Barton, Z. W. Chen, K. H. D. Lê, D. Amar, F. Lederer, and F. S. Mathews, (2005), Crystal structure analysis of recombinant rat kidney long chain hydroxy acid oxidase, *Biochemistry*, **44**, pp. 1521–1531.
[211] J. M. Bourhis, C. Vignaud, N. Pietrancosta, F. Guéritte, D. Guénard, F. Lederer, and Y. Lindqvist, (2009), Structure of human glycolate oxidase in complex with the inhibitor 4carboxy-5-[(4-chlorophenyl)sulfanyl]-1,2,3-thiadiazole, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, **65**, pp. 1246–1253.
[212] K. Stenberg and Y. Lindqvist, (1997), Three-dimensional structures of glycolate oxidase with bound active-site inhibitors., *Protein Sci.*, **6**, pp. 1009–1015.
[213] M.D. Moya-Garzon, C.M. Higueras, P. Penalver, M. Romera, M.X. Fernandes, F. Franco-Montalban, J.A. Gomez-Vidal, E. Salido and M. Diaz-Gavilan, (2018), Salicyclic acid derivatives inhibit oxalate production in mouse hepatocytes with primary hyperoxaluria type 1, *J. Med. Chem.*, **61**, pp. 7144-7167.
[214] C. Martin-Higueras, S. Luis-Lima, and E. Salido, (2016), Glycolate Oxidase Is a Safe and Efficient Target for Substrate Reduction Therapy in a Mouse Model of Primary Hyperoxaluria Type I, *Mol. Ther.*, **24**, pp. 719–725.
[215] C. S. Rooney, W. C. Randall, K. B. Streeter, C. Ziegler, E. J. Cragoe, H. Schwam, S. R. Michelson, H. W. Williams, E. Eichler, D. E. Duggan, E. H. Ulm, and R. M. Noll, (1983), Inhibitors of glycolic acid oxidase. 4-Substituted 3-hydroxy-1H-pyrrole-2,5-dione derivatives., *J. Med. Chem.*, **26**, pp. 700–14.
[216] M. R. Wilkins, E. Gasteiger, A. Bairoch, J. C. Sanchez, K. L. Williams, R. D. Appel, and D. F. Hochstrasser, (1999), Protein identification and analysis tools in the ExPASy server., *Methods Mol. Bio*l., **112**, pp. 531–552.
[217] J. Newman, D. Egan, T. S. Walter, R. Meged, I. Berry, M. Ben Jelloul, J. L. Sussman, D. I. Stuart, and A. Perrakis, (2005), Towards rationalization of crystallization screening for small-To medium-sized academic laboratories: The PACT/JCSG+ strategy, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **61**, pp. 1426–1431.
[218] A. Chaikuad, S. Knapp, and F. Von Delft, (2015), Defined PEG smears as an alternative approach to enhance the search for crystallization conditions and crystal-quality improvement in reduced screens, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **71**, pp. 1627–1639.
[219] F. Gorrec, (2009), The MORPHEUS protein crystallization screen, *J. Appl. Crystallogr.*, **42**, pp. 1035–1042.
[220] J. T. Ng, C. Dekker, M. Kroemer, M. Osborne, and F. Von Delft, (2014), Using textons to rank crystallization droplets by the likely presence of crystals, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **70**, pp. 2702–2718
[221] N.D. Wright, P. Collins, R. Talon, E. Nelson, L. Koekemoer, M. Ye, R. Nowak, J. Newman, J.T. Ng, N. Mitrovich, H. Wiggers, and F. von Delft, (2019), The Low-Cost, Semi-Automated Shifter Microscope Stage Transforms Speed and Robustness of Manual Protein Crystal Harvesting, *bioRxiv*, p. 2019.12.20.875674.
[222] G. Winter, (2010), Xia2: An expert system for macromolecular crystallography data reduction, *J. Appl. Crystallogr.*, **43**, pp. 186–190.
[223] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read, (2007), Phaser crystallographic software, *J. Appl. Crystallogr.*, **40**, pp. 658–674.
[224] 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, (2011), REFMAC5 for the refinement of macromolecular crystal structures, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **67**, pp. 355–367.
[225] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan, (2010), Features and development of Coot, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**, pp. 486–501.
[226] T. Krojer, R. Talon, N. Pearce, P. Collins, A. Douangamath, J. Brandao-Neto, A. Dias, B. Marsden, and F. Von Delft, (2017), The XChemExplorer graphical workflow tool for routine or large-scale protein-ligand structure determination, *Acta Crystallogr. Sect. D Struct. Biol.*, **73**, pp. 267–278.

[227] M. Wojdyr, R. Keegan, G. Winter, and A. Ashton, (2015), DIMPLE - a pipeline for the rapid generation of difference maps from protein crystals with putatively bound ligands, *Acta Crystallogr. Sect. A Found. Crystallogr.*, **69**, pp. s299–s299.
[228] N. M. Pearce, T. Krojer, A. R. Bradley, P. Collins, R. P. Nowak, R. Talon, B. D. Marsden, S. Kelm, J. Shi, C. M. Deane, and F. von Delft, (2017), A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density., *Nat. Commun.*, **8**, p. 15123.
[229] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson, (2010), MolProbity: All-atom structure validation for macromolecular crystallography, Acta Crystallogr. Sect. D Biol. Crystallogr., 66, pp. 12–21. [225] R. Abagyan, M. Totrov, and D. Kuznetsov, (1994), ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation, *J. Comput. Chem.*, **15**, pp. 488–506.
[230] R. Abagyan, M. Totrov, and D. Kuznetsov, (1994), ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation, *J. Comput. Chem.*, **15**, pp. 488–506.
[231] M. Totrov and R. Abagyan, (2008), Flexible ligand docking to multiple receptor conformations: a practical alternative, *Curr. Opin. Struct. Biol.*, **18**, pp. 178–184.
[232] S. Forster, L. Apostol and W. Bras, (2010), Scatter: software for the analysis of nano- and mesoscale small-angle scattering, *J. Appl. Cryst.*, **43**, pp. 639-646.
[233] D Franke, M.V Petoukhov, P.V. Konarev, A. Panjkovich, A. Tuukkanen, H.D.T. Mertens, A.G. Kikhney, N.R. Hajizadeh, J.M.Franklin, C.M. Jeffries, and D.I. Svergun, (2017), ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions, *J. Appl. Cryst.*, **50**, pp. 1212-1225.
[234] E. Williams, D. Cregeen, and G. Rumsby, (2000), Identification and expression of a cDNA for human glycolate oxidase., *Biochim. Biophys. Acta*, **1493**, pp. 246–8.
[235] P. Savitsky, J. Bray, C. D. O. Cooper, B. D. Marsden, P. Mahajan, N. A. Burgess-Brown, and O. Gileadi, (2010), High-throughput production of human proteins for crystallization: The SGC experience, *J. Struct. Biol.*, **172**, pp. 3–13.
[236] M. V. Petoukhov and D. I. Svergun, (2013), Applications of small-angle X-ray scattering to biomacromolecular solutions, *Int. J. Biochem. Cell Biol.*, **45**, pp. 429–437.
[237] A. W. Malaby, S. Chakravarthy, T. C. Irving, S. V. Kathuria, O. Bilsel, and D. G. Lambright, (2015), Methods for analysis of size-exclusion chromatography-small-angle X-ray scattering and reconstruction of protein scattering, *J. Appl. Crystallogr.*, **48**, pp. 1102–1113.
[238] P. Bernadó, N. Shimizu, G. Zaccai, H. Kamikubo, and M. Sugiyama, (2018), Solution scattering approaches to dynamical ordering in biomolecular systems, *Biochim. Biophys. Acta - Gen. Subj.*, **1862**, pp. 253–274.
[239] S. Mattu, F. Fornari, L. Quagliata, A. Perra, M. M. Angioni, A. Petrelli, S. Menegon, A. Morandi, P. Chiarugi, G. M. Ledda-Columbano, L. Gramantieri, L. Terracciano, S. Giordano, and A. Columbano, (2016), The metabolic gene HAO2 is downregulated in hepatocellular carcinoma and predicts metastasis and poor survival., *J. Hepatol.*, **64**, pp. 891–8.
[240] M. Wang, M. Xu, Y. Long, S. Fargue, N. Southall, X. Hu, J. C. McKew, C. J. Danpure, and W. Zheng, (2016), High throughput cell-based assay for identification of glycolate oxidase inhibitors as a potential treatment for Primary Hyperoxaluria Type 1, *Sci. Rep.*, **6**, p. 34060.
[241] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, and E. E. Bolton, (2019), PubChem 2019 update: Improved access to chemical data, *Nucleic Acids Res.*, **47**, pp. D1102–D1109.
[242] J. B. Hendrickson, P. Huang, and A. G. Toczko, (1987), Molecular Complexity: A Simplified Formula Adapted to Individual Atoms, *J. Chem. Inf. Comput. Sci.*, **27**, pp. 63–67.
[243] A. Kumar, A. Voet, and K. Y. J. Zhang, (2012), Fragment Based Drug Design: From Experimental to Computational Approaches, *Curr. Med. Chem.*, **19**, pp. 5128–5147.
[244] T. H. Keller, A. Pichota, and Z. Yin, (2006), A practical view of "druggability," *Curr. Opin. Chem. Biol.*, **10**, pp. 357–361.
[245] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, (2001), Experimental and computational approaches to estimate solubility and permeability in drug discovery and

development settings 1PII of original article: S0169-409X(96)00423-1. The article was originally published in Advanced Drug Delivery Reviews 23 (1997), *Adv. Drug Deliv. Rev*., **46**, pp. 3–26.

[246] S. H. Bertz, (1981), The First General Index of Molecular Complexity, J. Am. Chem. Soc., 103, pp. 3599–3601. [241] A. C. Salzberg, K. T. Smyth, Q. Cao, D. R. Caffrey, P. Soulard, R. G. Coleman, A. C. Cheng, and E. S. Huang, (2007), Structure-based maximal affinity model predicts small-molecule druggability, *Nat. Biotechnol*., **25**, pp. 71–75.

[247] M. Orita, K. Ohno, and T. Niimi, (2009), Two "Golden Ratio" indices in fragment-based drug discovery, *Drug Discov. Today*, **14**, pp. 321–328.

[248] D. F. Veber, S. R. Johnson, H. Y. Cheng, B. R. Smith, K. W. Ward, and K. D. Kopple, (2002), Molecular properties that influence the oral bioavailability of drug candidates, *J. Med. Chem.*, **45**, pp. 2615–2623.

[249] M. McAuley, M. Huang, and D. J. Timson, (2018), Modulation of the mobility of a key region in human galactokinase: Impacts on catalysis and stability, *Bioorg. Chem.*, **81**, pp. 649–657.

[250] D. Stambolian, Y. Ai, D. Sidjanin, K. Nesburn, G. Sathe, M. Rosenberg, and D. J. Bergsma, (1995), Cloning of the galactokinase cDNA and identification of mutations in two families with cataracts, *Nat. Genet*., **10**, pp. 307–312.

[251] C. F. Megarity, M. Huang, C. Warnock, and D. J. Timson, (2011), The role of the active site residues in human galactokinase: Implications for the mechanisms of GHMP kinases, *Bioorg. Chem*., **39**, pp. 120–126.

[252] P. Sneha, E. A. Ebrahimi, S. A. Ghazala, D. Thirumal Kumar, R. Siva, G. Priya Doss C, and H. Zayed, (2018), Structural analysis of missense mutations in galactokinase 1 (GALK1) leading to galactosemia type-2, *J. Cell. Biochem.*, **119**, pp. 7585–7598.

[253] D. Stambolian, V. Scarpino-Myers, and H. Harris, (1985), Purification of human galactokinase and evidence for its existence as a monomer form, *Biochim. Biophys. Acta Protein Struct. Mol*., **831**, pp. 306–312.

[254] X. Hu, Y. Q. Zhang, O. W. Lee, L. Liu, M. Tang, K. Lai, M. B. Boxer, M. D. Hall, and M. Shen, (2019), Discovery of novel inhibitors of human galactokinase by virtual screening, *J. Comput. Aided. Mol. Des*., **33**, pp. 405–417.

[255] D. R. Cooper, T. Boczek, K. Grelewska, M. Pinkowska, M. Sikorska, M. Zawadzki, and Z. Derewenda, (2007), Protein crystallization by surface entropy reduction: optimization of the SER strategy, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **63**, pp. 636–645

[256] D. Baker and D. A. Agard, (1994), Kinetics versus Thermodynamics in Protein Folding, *Biochemistry*, **33**, pp. 7505–7509.

[257] J. M. Sanchez-Ruiz, (2010), Protein kinetic stability, *Biophys. Chem*., **148**, pp. 1–15.

[258] D. J. Hart and G. S. Waldo, (2013), Library methods for structural biology of challenging proteins and their complexes, *Curr. Opin. Struct. Biol*., **23**, pp. 403–408.

[259] W. N. Price, Y. Chen, S. K. Handelman, H. Neely, P. Manor, R. Karlin, R. Nair, J. Liu, M. Baran, J. Everett, S. N. Tong, F. Forouhar, S. S. Swaminathan, T. Acton, R. Xiao, J. R. Luft, A. Lauricella, G. T. DeTitta, B. Rost, G. T. Montelione, and J. F. Hunt, (2009), Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data, *Nat. Biotechnol.*, **27**, pp. 51–57.

[260] M. C. Deller, L. Kong, and B. Rupp, (2016), Protein stability: a crystallographer's perspective, *Acta Crystallogr. Sect. F Struct. Biol. Commun*., **72**, pp. 72–95.

[261] F. Dupeux, M. Röwer, G. Seroul, D. Blot, and J. A. Márquez, (2011), A thermal stability assay can help to estimate the crystallization likelihood of biological samples, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **67**, pp. 915–919.

[262] M. Vedadi, F. H. Niesen, A. Allali-Hassani, O. Y. Fedorov, P. J. Finerty, G. A. Wasney, R. Yeung, C. Arrowsmith, L. J. Ball, H. Berglund, R. Hui, B. D. Marsden, P. Nordlund, M. Sundstrom, J. Weigelt, A. M. Edwards, M. Vedadi, C. Arrowsmith, L. J. Ball, G. A. Wasney, R. Hui, J. Weigelt, B. D. Marsden, F. H. Niesen, R. Yeung, A. M. Edwards, M. Sundstrom, P. J. Finerty, A. Allali-Hassani, O. Y. Fedorov, and P. Nordlund, (2006), Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination, *Proc. Natl. Acad. Sci.*, **103**, pp. 15835–15840.

[263] F. H. Niesen, H. Berglund, and M. Vedadi, (2007), The use of differential scanning fluorimetry

to detect ligand interactions that promote protein stability, *Nat. Protoc*., **2**, pp. 2212–2221.
[264] G. A. Senisterra and P. J. Finerty, (2009), High throughput methods of assessing protein stability and aggregation, *Mol. Biosyst.*, **5**, pp. 217–223.
[265] M. W. Pantoliano, E. C. Petrella, J. D. Kwasnoski, V. S. Lobanov, J. Myslik, E. Graf, T. Carver, E. Asel, B. A. Springer, P. Lane, and F. R. Salemme, (2001), High-Density Miniaturized Thermal Shift Assays as a General Strategy for Drug Discovery, *J. Biomol. Screen*., **6**, pp. 429–440.
[266] M. C. Lo, A. Aulabaugh, G. Jin, R. Cowling, J. Bard, M. Malamas, and G. Ellestad, (2004), Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery, *Anal. Biochem.*, **332**, pp. 153–159.
[267] L. Reinhard, H. Mayerhofer, A. Geerlof, J. Mueller-Dieckmann, and M. S. Weiss, (2013), Optimization of protein buffer cocktails using Thermofluor, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun*., **69**, pp. 209–214.
[268] T. M. Mezzasalma, J. K. Kranz, W. Chan, G. T. Struble, C. Schalk-Hihi, I. C. Deckman, B. A. Springer, and M. J. Todd, (2007), Enhancing recombinant protein quality and yield by protein stability profiling, *J. Biomol. Screen*., **12**, pp. 418–428.
[269] G. Mlynek, *Development and establishment of platforms for the production and crystallization of recalcitrant proteins and protein complexes*., 2016, University of Vienna, PhD dissertation.
[270] U. B. Ericsson, B. M. Hallberg, G. T. DeTitta, N. Dekker, and P. Nordlund, (2006), Thermofluor-based high-throughput stability optimization of proteins for structural studies, *Anal. Biochem.*, **357**, pp. 289–298.
[271] A. McPherson and B. Cudney, (2014), Optimization of crystallization conditions for biological macromolecules, *Acta Crystallogr. Sect. F Struct. Biol. Commun*., **70**, pp. 1445–1467.
[272] R. Nussinov and C. J. Tsai, (2013), Allostery in disease and in drug discovery, *Cell*, **153**, pp. 293–305.
[273] A. F. Abdel-Magid, (2015), Allosteric Modulators: An Emerging Concept in Drug Discovery, *ACS Med. Chem. Lett*., **6**, pp. 104–107.
[274] A. K. Grover, (2013), Use of allosteric targets in the discovery of safer drugs, *Med. Princ. Pract.*, **22**, pp. 418–426.
[275] S. Liu, M. J. Ammirati, X. Song, J. D. Knafels, J. Zhang, S. E. Greasley, J. A. Pfefferkorn, and X. Qiu, (2012), Insights into mechanism of glucokinase activation: Observation of multiple distinct protein conformations, *J. Biol. Chem*., **287**, pp. 13598–13610.
[276] J. R. Veloria, A. K. Devkota, E. J. Cho, and K. N. Dalby, (2017), Optimization of a Luminescence-Based High-Throughput Screening Assay for Detecting Apyrase Activity, *J. Biomol. Screen*., **22**, pp. 94–101.
[277] J. K. Jiang, M. B. Boxer, M. G. Vander Heiden, M. Shen, A. P. Skoumbourdis, N. Southall, H. Veith, W. Leister, C. P. Austin, H. W. Park, J. Inglese, L. C. Cantley, D. S. Auld, and C. J. Thomas, (2010), Evaluation of thieno[3,2-b]pyrrole[3,2-d]pyridazinones as activators of the tumor cell specific M2 isoform of pyruvate kinase, *Bioorganic Med. Chem. Lett.*, **20**, pp. 3387–3393.
[278] T. C. Coombs, C. Tanega, M. Shen, J. L. Wang, D. S. Auld, S. W. Gerritz, F. J. Schoenen, C. J. Thomas, and J. Aubé, (2013), Small-molecule pyrimidine inhibitors of the cdc2-like (Clk) and dual specificity tyrosine phosphorylation-regulated (Dyrk) kinases: Development of chemical probe ML315, *Bioorganic Med. Chem. Lett.*, **23**, pp. 3654–3661.
[279] T. Herraiz, A. Flores, and L. Fernández, (2018), Analysis of monoamine oxidase (MAO) enzymatic activity by high-performance liquid chromatography-diode array detection combined with an assay of oxidation with a peroxidase and its application to MAO inhibitors from foods and plants, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci*., **1073**, pp. 136–144.
[280] R. Somberg, B. Pferdehirt, and K. Kupcho, (2003), A Universal Kinase Assay for a World of Kinases, *Promega Notes*, **83**.
[281] S. R. Ford, K. H. Chenault, L. S. Bunton, G. J. Hampton, J. McCarthy, M. S. Hall, S. J. Pangburn, L. M. Buck, and F. R. Leach, (1996), Use of firefly luciferase for ATP measurement: Other nucleotides enhance turnover, *J. Biolumin. Chemilumin*., **11**, pp. 149–167.
[282] T.W. Traut, (1994), Physiological concentrations of purines and pyrimidines., *Mol. Cell. Biochem.*, **140**, pp. 1–22.

[283] M. Yoshino and K. Murakami, (2009), A graphical method for determining inhibition constants, *J. Enzyme Inhib. Med. Chem*., **24**, pp. 1288–1290.
[284] A. Cornish-Bowden, (1974), A simple graphical method for determining the inhibition constants of mixed, uncompetitive and non-competitive inhibitors, *Biochem. J*., **137**, pp. 143–144.
[285] N. Thorne, M. Shen, W. A. Lea, A. Simeonov, S. Lovell, D. S. Auld, and J. Inglese, (2012), Firefly luciferase in chemical biology: a compendium of inhibitors, mechanistic evaluation of chemotypes, and suggested use as a reporter., *Chem. Biol.*, **19**, pp. 1060–72.
[286] X. D. Zhang, M. Ferrer, A. S. Espeseth, S. D. Marine, E. M. Stec, M. A. Crackower, D. J. Holder, J. F. Heyse, and B. Strulovici, (2007), The use of strictly standardized mean difference for hit selection in primary RNA interference high-throughput screening experiments, *J. Biomol. Screen*., **12**, pp. 497–509.
[287] X. D. Zhang, (2007), A pair of new statistical parameters for quality control in RNA interference high-throughput screening assays, *Genomics*, **89**, pp. 552–561.
[288] X. D. Zhang, (2011), Illustration of SSMD, z score, SSMD*, z* score, and t statistic for hit selection in RNAi high-throughput screens, *J. Biomol. Screen*., **16**, pp. 775–785.
[289] L. Hoffer, Y. V. Voitovich, B. Raux, K. Carrasco, C. Muller, A. Y. Fedorov, C. Derviaux, A. Amouric, S. Betzi, D. Horvath, A. Varnek, Y. Collette, S. Combes, P. Roche, and X. Morelli, (2018), Integrated Strategy for Lead Optimization Based on Fragment Growing: The Diversity Oriented-Target-Focused-Synthesis Approach, *J. Med. Chem*., **61**, pp. 5719–5732.
[290] A. W. Hung, H. L. Silvestre, S. Wen, A. Ciulli, T. L. Blundell, and C. Abell, (2009), Application of fragment growing and fragment linking to the discovery of inhibitors of mycobacterium tuberculosis pantothenate synthetase, *Angew. Chemie - Int. Ed*., **48**, pp. 8452– 8456.
[291] A. J. A. Woolford, P. J. Day, V. Bénéton, V. Berdini, J. E. Coyle, Y. Dudit, P. Grondin, P. Huet, L. Y. W. Lee, E. S. Manas, R. L. McMenamin, C. W. Murray, L. W. Page, V. K. Patel, F. Potvain, S. J. Rich, Y. Sang, D. O. Somers, L. Trottet, Z. Wan, and X. Zhang, (2016), Fragment-Based Approach to the Development of an Orally Bioavailable Lactam Inhibitor of Lipoprotein-Associated Phospholipase A2 (Lp-PLA2), *J. Med. Chem*., **59**, pp. 10738–10749.
[292] J. A. Zorn and J. A. Wells, (2010), Turning enzymes on with small molecules, *Nat. Chem. Biol*., **6**, pp. 179–188.
[293] M. Lang, M. H. J. Seifert, K. K. Wolf, A. Aschenbrenner, R. Baumgartner, T. Wieber, V. Trentinaglia, M. Blisse, N. Tajima, T. Yamashita, D. Vitt, and H. Noda, (2011), Discovery and hit-to-lead optimization of novel allosteric glucokinase activators, *Bioorganic Med. Chem. Lett*., **21**, pp. 5417–5422.
[294] L. Qin, Y. Wu, Y. Liu, Y. Chen, and P. Zhang, (2014), Dual effects of alpha-arbutin on monophenolase and diphenolase activities of mushroom tyrosinase, *PLoS One*, **9**, pp. 1–7.
[295] B. A. K. Chibber and F. J. Castellino, (1986), Regulation of the streptokinase-mediated activation of human plasminogen by fibrinogen and chloride ions, *J. Biol. Chem*., **261**, pp. 5289–5295.
[296] S. Kuusk and P. Väljamäe, (2017), When substrate inhibits and inhibitor activates: Implications of β-glucosidases, *Biotechnol. Biofuels*, **10**, pp. 1–15.
[297] Q. Wang, Y. Qi, N. Yin, and L. Lai, (2014), Discovery of novel allosteric effectors based on the predicted allosteric sites for Escherichia coli D-3-phosphoglycerate dehydrogenase, *PLoS One*, **9**, pp. 3–10.
[298] M. McAuley, H. Kristiansson, M. Huang, A. L. Pey, and D. J. Timson, (2016), Galactokinase promiscuity: a question of flexibility?, *Biochem. Soc. Trans*., **44**, pp. 116–22.
[299] L. P. Conway and J. Voglmeir, (2016), Functional analysis of anomeric sugar kinases, *Carbohydr. Res.*, **432**, pp. 23–30.
[300] S. Caramia, A. G. M. Gatius, F. dal Piaz, D. Gaja, and A. Hochkoeppler, (2017), Dual role of imidazole as activator/inhibitor of sweet almond (Prunus dulcis) β-glucosidase, *Biochem. Biophys. Reports*, **10**, pp. 137–144.
[301] L. Aagaard and J. J. Rossi, (2007), RNAi therapeutics: Principles, prospects and challenges, *Adv. Drug Deliv. Rev.*, **59**, pp. 75–86.
[302] C. A. Sledz, M. Holko, M. J. De Veer, R. H. Silverman, and B. R. G. Williams, (2003), Activation of the interferon system by short-interfering RNAs, *Nat. Cell Biol*., **5**, pp. 834–839.

[303] J. T. Marques and B. R. G. Williams, (2005), Activation of the mammalian immune system by siRNAs, *Nat. Biotechnol*., **23**, pp. 1399–1405.
[304] A. Biffis, P. Centomo, A. Del Zotto, and M. Zecca, (2018), Pd Metal Catalysts for Cross Couplings and Related Reactions in the 21st Century: A Critical Review, *Chem. Rev*., **118**, pp. 2249–2295.
[305] G. Chandramouly, S. McDevitt, K. Sullivan, T. Kent, A. Luz, J. F. Glickman, M. Andrake, T. Skorski, and R. T. Pomerantz, (2015), Small-Molecule Disruption of RAD52 Rings as a Mechanism for Precision Medicine in BRCA-Deficient Cancers., *Chem. Biol.*, **22**, pp. 1491–1504.
[306] J. E. Gable, G. M. Lee, T. M. Acker, K. R. Hulce, E. R. Gonzalez, P. Schweigler, S. Melkko, C. J. Farady, and C. S. Craik, (2016), Fragment-Based Protein-Protein Interaction Antagonists of a Viral Dimeric Protease., *ChemMedChem*, **11**, pp. 862–9.
[307] K. McMillan, M. Adler, D. S. Auld, J. J. Baldwin, E. Blasko, L. J. Browne, D. Chelsky, D. Davey, R. E. Dolle, K. A. Eagen, S. Erickson, R. I. Feldman, C. B. Glaser, C. Mallari, M. M. Morrissey, M. H. J. Ohlmeyer, G. Pan, J. F. Parkinson, G. B. Phillips, M. A. Polokoff, N. H. Sigal, R. Vergona, M. Whitlow, T. A. Young, and J. J. Devlin, (2000), Allosteric inhibitors of inducible nitric oxide synthase dimerization discovered via combinatorial chemistry, *Proc. Natl. Acad. Sci. U. S. A.*, **97**, pp. 1506–1511.
[308] L. F. Silvian, J. E. Friedman, K. Strauch, T. G. Cachero, E. S. Day, F. Qian, B. Cunningham, A. Fung, L. Sun, G. W. Shipps, L. Su, Z. Zheng, G. Kumaravel, and A. Whitty, (2011), Small molecule inhibition of the TNF family cytokine CD40 ligand through a subunit fracture mechanism., *ACS Chem. Biol*., **6**, pp. 636–47.
[309] M. M. He, A. S. Smith, J. D. Oslob, W. M. Flanagan, A. C. Braisted, A. Whitty, M. T. Cancilla, J. Wang, A. A. Lugovskoy, J. C. Yoburn, A. D. Fung, G. Farrington, J. K. Eldredge, E. S. Day, L. A. Cruz, T. G. Cachero, S. K. Miller, J. E. Friedman, I. C. Choong, and B. C. Cunningham, (2005), Small-molecule inhibition of TNF-alpha., *Science*, **310**, pp. 1022–5.
[310] D. S. H. Chan, D. Matak-Vinković, A. G. Coyne, and C. Abell, (2016), Insight into Protein Conformation and Subcharging by DMSO from Native Ion Mobility Mass Spectrometry, *ChemistrySelect*, **1**, pp. 5686–5690.
[311] D. J. Clarke and D. J. Campopiano, (2015), Desalting large protein complexes during native electrospray mass spectrometry by addition of amino acids to the working solution, *Analyst*, **140**, pp. 2679–2686.
[312] M. Ishikawa and Y. Hashimoto, (2011), Improvement in aqueous solubility in small molecule drug discovery programs by disruption of molecular planarity and symmetry, *J. Med. Chem.*, **54**, pp. 1539–1554.
[313] C. Yueh, J. Rettenmaier, B. Xia, D. R. Hall, A. Alekseenko, K. A. Porter, K. Barkovich, G. Keseru, A. Whitty, J. A. Wells, S. Vajda, and D. Kozakov, (2019), Kinase atlas: Druggability analysis of potential allosteric sites in kinases, *J. Med. Chem*., **62**, pp. 6512–6524.
[314] E. Krissinel and K. Henrick, (2007), Inference of macromolecular assemblies from crystalline state., *J. Mol. Biol.*, **372**, pp. 774–797.
[315] S. H. L. Mandey, M. S. Schneiders, J. Koster, and H. R. Waterham, (2006), Mutational spectrum and genotype-phenotype correlations in mevalonate kinase deficiency., *Hum. Mutat.*, **27**, pp. 796–802.
[316] S. Zhang, (2016), Natural history of mevalonate kinase deficiency: A literature review, *Pediatr. Rheumatol.*, **14**, pp. 1–10.
[317] C. L. Araya, D. M. Fowler, W. Chen, I. Muniez, J. W. Kelly, and S. Fields, (2012), A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function, *Proc. Natl. Acad. Sci. U. S. A.*, **109**, pp. 16858–16863.
[318] C. Alfano, D. Sanfelice, S. R. Martin, A. Pastore, and P. A. Temussi, (2017), An optimized strategy to measure protein stability highlights differences between cold and hot unfolded states, *Nat. Commun*., **8**, pp. 1–9.
[319] R. Godoy-Ruiz, F. Ariza, D. Rodriguez-Larrea, R. Perez-Jimenez, B. Ibarra-Molero, and J. M. Sanchez-Ruiz, (2006), Natural Selection for Kinetic Stability Is a Likely Origin of Correlations between Mutational Effects on Protein Energetics and Frequencies of Amino Acid Occurrences in Sequence Alignments, *J. Mol. Biol*., **362**, pp. 966–978.

[320] M. Amaral, D. B. Kokh, J. Bomke, A. Wegener, H. P. Buchstaller, H. M. Eggenweiler, P. Matias, C. Sirrenberg, R. C. Wade, and M. Frech, (2017), Protein conformational flexibility modulates kinetics and thermodynamics of drug binding, *Nat. Commun*., **8**., article: 2276.
[321] C. Zhou, Y. Xue, and Y. Ma, (2010), Enhancing the thermostability of α-glucosidase from Thermoanaerobacter tengcongensis MB4 by single proline substitution, *J. Biosci. Bioeng*., **110**, pp. 12–17.
[322] V. M. Burger, D. J. Arenas, and C. M. Stultz, (2016), A Structure-free Method for Quantifying Conformational Flexibility in proteins, *Sci. Rep*., **6**, pp. 1–9.
[323] G. M. Cooper and C. D. Brown, (2008), Qualifying the relationship between sequence conservation and molecular function, *Genome Res*., **18**, pp. 201–205.
[324] I. Prodan-Zitnik, N. Karas-Kuzelicki, and J. Lukac-Bajalo, (2009), Positive correlation between galactose-1-phosphate uridyltransferase (GALT) and UDP-galactose-4′-epimerase (GALE) activities, *Clin. Biochem.*, **42**, pp. 1561–1564.
[325] D. J. Timson, (2016), The molecular basis of galactosemia — Past, present and future, *Gene*, **589**, pp. 133–141.
[326] W. H. Dunham, M. Mullin, and A. C. Gingras, (2012), Affinity-purification coupled to mass spectrometry: Basic principles and strategies, *Proteomics*, **12**, pp. 1576–1590.
[327] B. Chen and R. B. Altman, (2017), Opportunities for developing therapies for rare genetic diseases: Focus on gain-of-function and allostery, *Orphanet J. Rare Dis*., **12**, pp. 1–7.
[328] M. Rask-Andersen, S. Masuram, and H. B. Schiöth, (2014), The Druggable Genome: Evaluation of Drug Targets in Clinical Trials Suggests Major Shifts in Molecular Class and Indication, *Annu. Rev. Pharmacol. Toxicol*., **54**, pp. 9–26.
[329] R. Santos, O. Ursu, A. Gaulton, A. P. Bento, R. S. Donadi, C. G. Bologa, A. Karlsson, B. AlLazikani, A. Hersey, T. I. Oprea, and J. P. Overington, (2017), A comprehensive map of molecular drug targets., *Nat. Rev. Drug Discov.*, **16**, pp. 19–34.
[330] S. Nguengang Wakap, D. M. Lambert, A. Olry, C. Rodwell, C. Gueydan, V. Lanneau, D. Murphy, Y. Le Cam, and A. Rath, (2019), Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database, *Eur. J. Hum. Genet.*, **28**, pp. 165-173.

Appendix 1: Publications

Classic galactosaemia/GALK1

Bezerra GA, Mackinnon S, McCorvie T, Zhang M, Tang M, Foster W, Fairhead M, Tumber A, Strain-Damerell C, Kopec J, Fitzpatrick F, Whitby F, Hill CP, Lai K and Yue WW. (2019). Human Galactose-1-phosphate Uridylyltransferase (GALT), Galactokinase 1 (GALK1): A Target Enabling Package (Version 2) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.3234909

Mackinnon S, Bezerra GA, Krojer T, Foster WR, Tang M, von Delft F, Lai K and Yue WW. Fragment screening reveals starting points for rational design of galactokinase 1 inhibitors to treat classic galactosemia. *Manuscript ready to be submitted (ACS Chem. Bio.)*

Primary hyperoxaluria/HAO1

Mackinnon S, Bezerra GA, Krojer T, Bradley AR, Talon R, Brandao-Neto J, Douangamath A, Oppermann U, von Delft F, Brennan PE and Yue WW. (2018). Human Hydroxyacid Oxidase (HAO1); A Target Enabling Package (Version 1) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.1344541

Mackinnon S, Bezerra GA, Krojer T, Tzommer T, von Delft F, Brennan PE and Yue WW. Identification and characterisation of novel starting points for development of glycolate oxidase (HAO1) inhibitors by fragment screening x-ray crystallography. *In preparation.*

Open lab notebook, including 10 blog posts and 3 Zenodo depositions: https://openlabnotebooks.org/category/disease-foundation/primary-hyperoxaluria/

Other

REVIEW: Yue WW, Mackinnon S, Bezerra GA. (2019). Substrate reduction therapy for inborn errors of metabolism. *Emerg. Top Life Sci.*, **3**, 63.

IN SILICO ANALYSIS OF DISEASE VARIANTS: Ghosh A, Mercer J, Mackinnon S, Yue WW, Church H, Beesley CE, Broomfield A, Jones SA and Tylee K. (2017). IDUA mutational profile and genotype-phenotype relationships in UK patients with Mucopolysaccharidosis Type I. *Hum. Mutat.* **38**, 1555.

IN SILICO ANALYSIS OF DISEASE VARIANTS: Demain LA, Urquhart JE, O'Sullivan J, Williams SG, Bhaskar SS, Jenkinson EM, Lourenco CM, Heiberg A, Pearce SH, Shalev SA, Yue WW, Mackinnon S, Munro KJ, Newbury-Ecob R, Becker K, Kim MJ, O-Keefe RT and

Newman WG. (2017). Expanding the genotypic spectrum of Perrault syndrome. *Clin. Genet.*, **91**, 302.

IN SILICO ANALYSIS OF DISEASE VARIANTS: Knerr I, Colombo R, Urquhart J, Morais A, Merinero B, Oyarzabal A, Perez B, Jones SA, Perveen R, Preece MA, Rogers Y, Treacy EP, Mayne P, Zampino G, Mackinnon S, Wassmer E, Yue WW, Robinson I, Rodriguez-Pombo P, Olpin SE and Banka S. (2019). Expanding the genetic and phenotypic spectrum of branched-chain amino acid transferase 2 deficiency. *J. Inherit. Metab. Dis.*, **42**, 809.

DSF OF PURIFIED PROTEIN ± ARGININE: Haskovic M, Derks B, van der Ploeg L, Trommelen J, Nyakayiru J, van Loon LJC, Mackinnon S, Yue WW, Peake RWA, Zha L, Demirbas D, Qi W, Huang X, Berry GT, Achten J, Bierau J, Rubio-Gozalbo ME and Coelho AI. (2018). Arginine does not rescue p.Q188R mutation deleterious effect in classic galactosemia. *Orphanet J Rare Dis.*, **13**, 212.

PROTEIN PURIFICATION, *IN SILICO* PREDICTION OF LIGAND BINDING: Sokolovskaya OM, Plessl T, Bailey H, Mackinnon S, Baumgartner MR, Yue WW, Froese DS and Taga ME. (2020) Naturally occurring cobalamin (B12) analogs can function as cofactors for human methylmalonyl-CoA mutase. *BioRxiv:* 2020.03.20.997551. *Submitted to Biochimie.*

Appendix 2: Crystallographic refinement statistics

Table S2.1: Crystallographic refinement statistics for hHAO1 co-crystal structures. *Data for the highest resolution shell are shown in parentheses. All structures also contain FMN. hHAO1-FMN-CCPST and hHAO1-FMN-glycolate structures were collected at the Diamond Light Source synchrotron beamline i03 at a wavelength of 0.97625. All other structures were collected at the i04-1 beamline with a wavelength of 0.91587.*

Ligand:		**CCPST**	**Glycolate**	**Fragment 1**	**Fragment 2**	**Fragment 3**
Unit cell parameters	(Å)	97.43 97.43 80.44	97.42 97.42 80.35	97.46 97.46 80.52	98.00 98.00 81.74	97.54 97.54 80.65
	(°)	90 90 90	90 90 90	90 90 90	90 90 90	90 90 90
Space group		I4				
Resolution (Å)		31.02 - 1.2	30.81 - 1.35	38.33 - 1.33	26.43 - 1.48	26.21 - 1.42
Observed/Unique reflections		476163/115886 (45385/11609)	501802/81457 (27696/7622)	437306/84166 (14118/7122)	386723/63919 (26506/6325)	408774/71153 (25375/7008)
R-merge		0.0777 (0.6564)	0.04117 (0.1033)	0.04526 (0.6312)	0.03872 (0.3422)	0.05126 (0.4915)
CC (1/2)		0.993 (0.705)	0.999 (0.986)	0.999 (0.552)	0.999 (0.929)	0.999 (0.801)
I/sig(I)		8.96 (1.90)	28.77 (9.68)	17.09 (1.18)	23.46 (3.40)	17.35 (2.62)
Completeness		99.47 (99.28)	99.10 (93.12)	97.65 (82.88)	99.45 (99.14)	99.83 (98.89)
Multiplicity		4.1 (3.9)	6.2 (3.6)	5.2 (2.0)	6.1 (4.2)	5.7 (3.6)
Rwork (%)		13.4	12.8	18.6	19.0	20.0
Rfree (%)		15.5	13.1	19.9	20.7	20.8
Rmsd bond length (Å)		0.009	0.008	0.017	0.014	0.013
Rmsd bond angle (°)		1.37	1.31	1.86	1.58	1.56
Clashscore		10.48	4.37	2.95	1.84	1.39

Table S2.1: Crystallographic refinement statistics for hHAO1 co-crystal structures. *Continued.*

Ligand:		**Fragment 4**	**Fragment 5**	**Fragment 6**	**Fragment 7**	**Compound 19**
Data collection and processing						
Unit cell parameters	(Å)	97.78 97.78 81.268	97.59 97.59 80.87	97.49 97.49 80.39	97.68 97.68 80.41	97 97 80.64
	(°)	90 90 90	90 90 90	90 90 90	90 90 90	90 90 90
Space group		I4				
Resolution (Å)		26.33 - 1.42	48.79 - 1.34	26.16 - 1.45	26.2 - 1.34	62.01 - 1.46
Observed/Unique reflections		412150/71894 (26272/7167)	440078/81913 (15596/6324)	392224/66359 (25722/6602)	432780/80067 (15258/5578)	381858/64628 (25914/6439)
R-merge		0.03548 (0.1027)	0.0472 (0.6642)	0.05521 (0.4594)	0.05306 (0.5608)	0.05746 (0.9356)
CC (1/2)		0.999 (0.987)	0.999 (0.623)	0.999 (0.831)	0.999 (0.785)	0.999 (0.601)
I/sig(I)		32.49 (10.19)	18.12 (1.28)	16.04 (2.62)	18.06 (1.65)	14.53 (1.19)
Completeness		99.84 (99.79)	96.50 (74.94)	99.93 (99.83)	94.95 (66.08)	99.85 (99.77)
Multiplicity		5.7 (3.7)	5.4 (2.5)	5.9 (3.9)	5.4 (2.8)	5.9 (4.0)
Refinement						
Rwork (%)		18.6	19.3	18.8	17.3	21.3
Rfree (%)		20.3	20.7	20.3	20.1	23.6
Rmsd bond length (Å)		0.014	0.017	0.017	0.018	0.019
Rmsd bond angle (°)		1.63	1.78	1.78	1.84	1.44
Clashscore		0.53	2.55	1.24	0.71	6.78

Table S2.1: Crystallographic refinement statistics for hHAO1 co-crystal structures. *Continued.*

Ligand:		**Compound 47**	**Compound 79**	**Compound 123**
Data collection and processing				
Unit cell parameters	(Å)	97.13 97.13 80.84	97.06 97.06 80.22	97.03 97.03 80.12
	(°)	90 90 90	90 90 90	90 90 90
Space group		I4		
Resolution (Å)		62.13 - 1.37	40.11 - 1.29	38.16 - 1.35
Observed/Unique reflections		424651/77167 (19553/6662)	442261/88616 (9902/6345)	425610/80253 (16750/7273)
R-merge		0.03251 (0.7318)	0.04631 (0.4941)	0.03856 (0.6551)
CC (1/2)		1 (0.549)	0.999 (0.596)	1 (0.595)
I/sig(I)		20.29 (1.14)	15.60 (1.04)	17.84 (1.09)
Completeness		98.11 (84.84)	94.97 (68.19)	98.70 (89.83)
Multiplicity		5.5 (2.9)	5.0 (1.6)	5.3 (2.3)
Refinement				
Rwork (%)		14.6	14.5	14.5
Rfree (%)		17.1	16.5	16.7
Rmsd bond length (Å)		0.014	0.015	0.008
Rmsd bond angle (°)		1.51	1.54	1.17
Clashscore		5.51	6.42	1.41

Table S2.2: Crystallographic refinement statistics for hGALK1 co-crystal structures. *Data for the highest resolution shell are shown in parentheses. All structures contain galactose. All fragment-bound structures contain **T3**. Structures with ADP, **T1** and **T7** were collected at the i03 beamline at a wavelength of 0.97619. Structures with **T2** and **T3** were collected at the i04 beamline at a wavelength of 0.97950. All other structures were collected at the i04-1 beamline with a wavelength of 0.91587.*

Ligand:		**ADP**	**T1**	**T2**	**T3**	**T7**
Data collection and processing						
Unit cell parameters	(Å)	78.21 110.32 118.23	119.9 97.17 144.88	107.7 151.3 230.08	107.73 151.68 230.58	115.51 219.18 151.24
	(°)	113.85 102.3 101.29	90 98.51 90	90 90 90	90 90 90	90 90 90
Space group		P 1	P 1 21 1	F 2 2 2	F 2 2 2	C 2 2 21
Resolution (Å)		47.78 - 2.49	65.78 - 2.5	57.52 - 2.12	57.84 - 2.1	54.8 - 1.83
Observed/Unique reflections		141747/61792 (12777/5809)	749430/113385 (71274/2599)	353912/53149 (34504/5265)	355089/54819 (26154/5332)	4396587/167966 (427524/16582)
R-merge		0.09817 (0.4781)	0.2717 (2.329)	0.1526 (1.401)	0.07543 (0.2962)	0.1414 (3.188)
CC (1/2)		0.973 (0.481)	0.989 (0.591)	0.997 (0.43)	0.998 (0.92)	0.904 (1)
I/sig(I)		10.05 (3.34)	6.35 (1.22)	7.72 (1.05)	14.83 (4.03)	16.09 (1.43)
Completeness		97.29 (96.17)	99.10 (98.73)	99.91 (99.83)	99.35 (89.98)	99.85 (99.37)
Multiplicity		2.3 (2.2)	6.6 (6.4)	6.7 (6.6)	6.5 (4.9)	26.2 (25.8)
Refinement						
Rwork (%)		22.2	18.4	19.8	18.8	21.5
Rfree (%)		25.5	26.2	25.1	22.2	24.3
Rmsd bond length (Å)		1.00	1.00	1.00	0.015	0.008
Rmsd bond angle (°)		0.004	0.015	0.009	1.85	1.22
Clashscore		1.11	1.79	1.25	13.37	9.53

Table S2.2: Crystallographic refinement statistics for hGALK1 co-crystal structures. *Continued.*

Ligand:		**Fragment 1**	**Fragment 2**	**Fragment 3**	**Fragment 4**	**Fragment 5**
Data collection and processing						
Unit cell parameters	(Å)	73.26 114.81 120.82	73.70 114.87 121.04	73.31 114.42 120.9	73.45 114.55 120.86	73.59 114.4 120.84
	(°)	90 100.49 90	90 100.56 90	90 100.55 90	90 100.55 90	90 100.57 90
Space group		P 1 21 1				
Resolution (Å)		82.56 - 1.96	82.65 - 2.4	82.43 - 1.81	82.47 - 2.11	52.71 - 2.4
Observed/Unique reflections		472391/ 139631 (48402/13992)	264639/77146 (26708/7657)	603464/177671 (60403/17736)	377059/112695 (34477/11041)	258207/76605 (26770/7678)
R-merge		0.1168 (1.957)	0.06401 (0.4706)	0.05142 (0.8445)	0.09342 (0.8126)	0.03059 (0.1188)
CC (1/2)		0.996 (0.387)	0.998 (0.894)	0.997 (0.691)	0.986 (0.681)	0.999 (0.99)
I/sig(I)		7.15 (0.84)	11.70 (2.22)	10.55 (1.01)	6.46 (1.05)	23.80 (8.83)
Completeness		99.31 (99.50)	99.46 (99.45)	99.64 (99.74)	99.42 (97.70)	99.46 (99.79)
Multiplicity		3.4 (3.5)	3.4 (3.5)	3.4 (3.4)	3.3 (3.1)	3.4 (3.5)
Refinement						
Rwork (%)		23.5	22.13	21.9	21.3	22.6
Rfree (%)		25.5	26.9	25.3	24.9	26.7
Rmsd bond length (Å)		0.004	0.004	0.015	0.015	0.017
Rmsd bond angle (°)		0.92	0.88	1.98	1.98	1.68
Clashscore		6.16	4.79	5.06	8.66	8.33

Table S2.2: Crystallographic refinement statistics for hGALK1 co-crystal structures. *Continued.*

Ligand:		**Fragment 6**	**Fragment 7**	**Fragment 8**	**Fragment 9**	**Fragment 10**
Data collection and processing						
Unit cell parameters	(Å)	73.85 115.05 121.2	73.66 115.01 120.99	73.56 114.46 121.01	73.57 114.74 120.65	73.49 114.48 120.94
	(°)	90 100.71 90	90 100.55 90	90 100.39 90	90 100.57 90	90 100.7 90
Space group		P 1 21 1				
Resolution (Å)		67.82 - 2.89	72.41 - 2.3	72.35 - 2.3	67.5 - 2.18	82.45 - 1.91
Observed/Unique reflections		149088/44368 (15170/4407)	295617/85698 (30891/8557)	297390/86745 (30702/8620)	348380/102577 (33277/10157)	509078/148597 (53000/14780)
R-merge		0.04978 (0.1727)	0.05198 (0.3983)	0.0762 (0.5393)	0.1053 (0.928)	0.06974 (0.8254)
CC (1/2)		0.998 (0.972)	0.996 (0.964)	0.997 (0.834)	0.992 (0.66)	0.997 (0.713)
I/sig(I)		18.12 (7.48)	11.44 (3.22)	10.26 (2.55)	7.27 (1.16)	9.33 (1.43)
Completeness		99.19 (99.75)	98.01 (97.23)	99.07 (98.90)	99.72 (99.23)	97.66 (97.55)
Multiplicity		3.4 (3.4)	3.4 (3.6)	3.4 (3.6)	3.4 (3.3)	3.4 (3.6)
Refinement						
Rwork (%)		20.5	21.6	25.5	19.7	21.8
Rfree (%)		24.8	25.3	28.6	24.1	25.2
Rmsd bond length (Å)		0.015	0.018	0.018	0.015	0.018
Rmsd bond angle (°)		2.09	2.28	2.25	2.13	2.25
Clashscore		9.75	9.1	9.47	9.31	8.42

Table S2.2: Crystallographic refinement statistics for hGALK1 co-crystal structures. *Continued.*

Ligand:		**Fragment 11**	**Fragment 12**	**Fragment 13**	**Fragment 14**	**Fragment 15**
Data collection and processing						
Unit cell parameters	(Å)	73.67 114.56 120.93	73.77 114.45 120.92	73.84 114.71 121.04	73.55 114.51 120.81	73.56 114.36 120.51
	(°)	90 100.49 90	90 100.55 90	90 100.56 90	90 100.33 90	90 100.41 90
Space group		P 1 21 1				
Resolution (Å)		82.5 - 2.3	67.67 - 2.3	82.58 - 1.98	67.41 - 2.4	67.4 - 2.16
Observed/Unique reflections		299490/87629 (30596/8727)	299485/87557 (30630/8711)	462289/137618 (45639/13581)	261047/76815 (26798/7687)	359781/103043 (34726/10183)
R-merge		0.08892 (0.6217)	0.06551 (0.1866)	0.07299 (0.9069)	0.05221 (0.3622)	0.1335 (4.397)
CC (1/2)		0.994 (0.814)	0.99 (0.966)	0.996 (0.635)	0.999 (0.93)	0.996 (0.4)
I/sig(I)		7.22 (1.83)	11.16 (5.01)	8.71 (1.19)	15.30 (3.64)	6.56 (0.34)
Completeness		99.84 (99.62)	99.80 (99.54)	99.78 (98.96)	99.64 (99.79)	98.05 (97.46)
Multiplicity		3.4 (3.5)	3.4 (3.5)	3.4 (3.4)	3.4 (3.5)	3.5 (3.4)
Refinement						
Rwork (%)		21.3	20.9	21.6	21.9	19.8
Rfree (%)		25.0	26.1	25.3	24.8	24.4
Rmsd bond length (Å)		0.019	0.017	0.016	0.015	0.017
Rmsd bond angle (°)		2.06	2.21	2.19	1.94	2.42
Clashscore		8.02	8.82	8.67	8.42	12.4

Table S2.2: Crystallographic refinement statistics for hGALK1 co-crystal structures. *Continued.*

Ligand:		**Fragment 16**	**Fragment 17**	**Fragment 18**	**Fragment 19**	**Fragment 20**
Data collection and processing						
Unit cell parameters	(Å)	73.48 114.39 120.67	73.31 114.48 120.78	73.50 114.45 120.86	73.65 114.14 120.84	73.36 114.21 120.92
	(°)	90 100.56 90	90 100.51 90	90 100.44 90	90 100.55 90	90 100.47 90
Space group		P 1 21 1				
Resolution (Å)		67.43 - 1.87	72.08 - 2.27	67.42 - 2.3	67.57 - 1.96	82.37 - 2.4
Observed/Unique reflections		544195/161433 (55500/15980)	301826/89739 (31212/8964)	299970/86468 (30872/8576)	474927/140358 (43419/13682)	258863/76648 (25874/7639)
R-merge		0.06946 (1.044)	0.127 (1.204)	0.04279 (0.2317)	0.1008 (0.9718)	0.06186 (0.2262)
CC (1/2)		0.991 (0.544)	0.985 (0.641)	0.999 (0.964)	0.943 (0.568)	0.98 (0.953)
I/sig(I)		8.78 (1.19)	7.71 (1.00)	16.60 (4.66)	5.62 (0.96)	11.98 (4.22)
Completeness		99.68 (98.74)	98.72 (98.85)	99.00 (98.85)	99.59 (97.63)	99.90 (99.97)
Multiplicity		3.4 (3.5)	3.4 (3.5)	3.5 (3.6)	3.4 (3.2)	3.4 (3.4)
Refinement						
Rwork (%)		23.1	24.2	21.8	22.8	23.2
Rfree (%)		27.4	26.5	25.5	25.9	26.9
Rmsd bond length (Å)		0.026	0.028	0.019	0.016	0.017
Rmsd bond angle (°)		2.66	2.31	1.96	2.14	2.22
Clashscore		12.88	12.76	8.43	9.33	10.99

Table S2.2: Crystallographic refinement statistics for hGALK1 co-crystal structures. *Continued.*

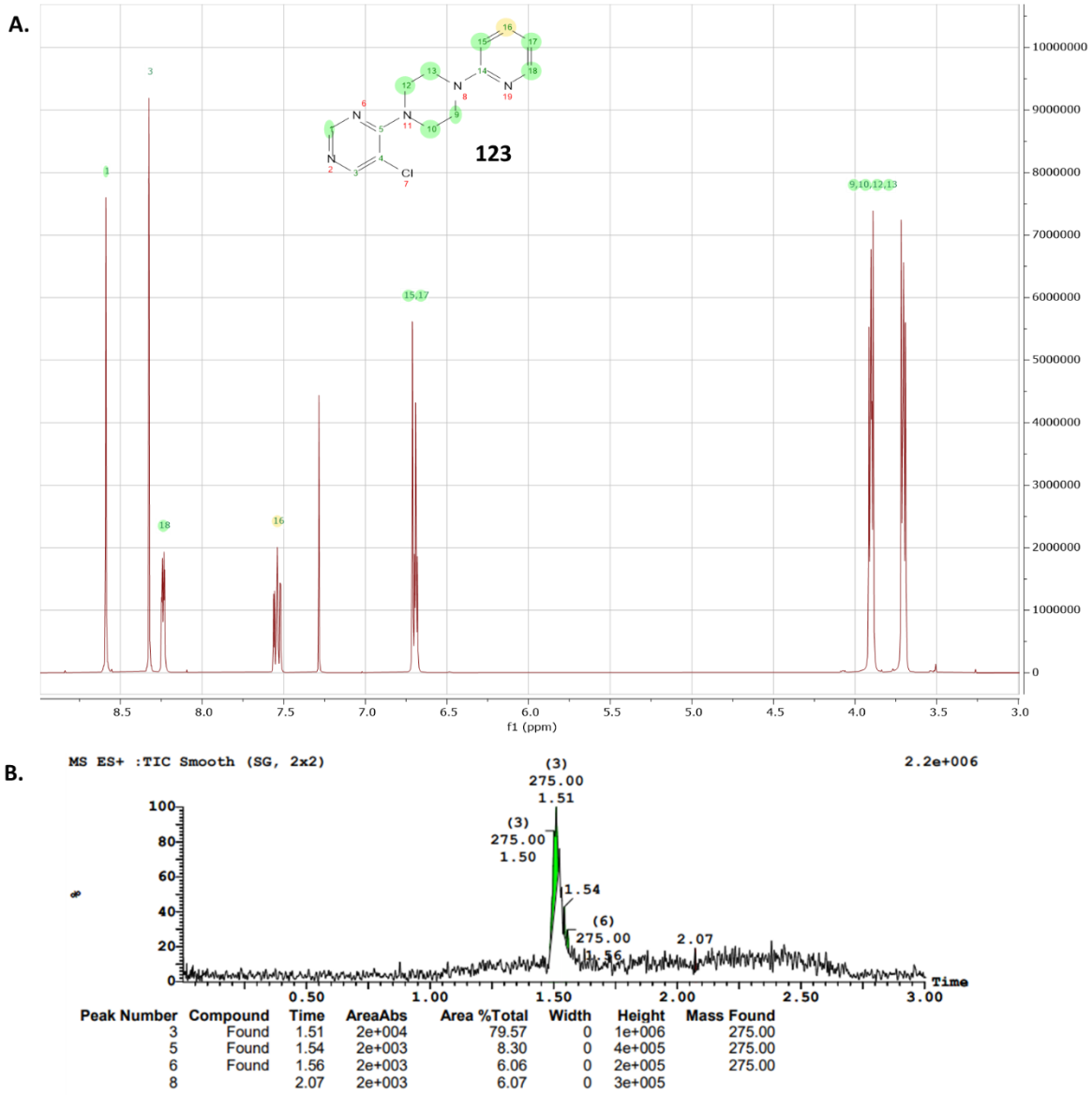
Ligand:		**Fragment 21**	**Fragment 22**	**Fragment 23**	**Fragment 24**	**Fragment 25**
Data collection and processing						
Unit cell parameters	(Å)	73.65 114.63 120.83	73.31 114.63 120.82	73.7 114.89 120.92	73.62 114.99 120.98	73.50 114.44 120.76
	(°)	90 100.49 90	90 100.59 90	90 100.62 90	90 100.62 90	90 100.58 90
Space group		P 1 21 1				
Resolution (Å)		118.8 - 2.23	118.8 - 1.93	72.44 - 2.3	59.45 - 2.3	67.46 - 2.0
Observed/Unique reflections		327578/95514 (32743/9299)	510715/147891 (49894/14701)	296206/87464 (30489/8724)	301686/87548 (30887/8716)	448241/132140 (41549/12978)
R-merge		0.1219 (1.191)	0.108 (2.476)	0.07513 (0.8277)	0.098 (1.273)	0.07664 (0.9772)
CC (1/2)		0.994 (0.592)	0.996 (0.335)	0.998 (0.728)	0.998 (0.582)	0.997 (0.624)
I/sig(I)		6.72 (1.20)	5.77 (0.28)	10.32 (1.37)	5.14 (0.60)	8.30 (1.07)
Completeness		99.20 (96.97)	99.61 (99.57)	99.48 (99.91)	99.44 (99.70)	99.06 (98.41)
Multiplicity		3.4 (3.5)	3.5 (3.4)	3.4 (3.5)	3.4 (3.5)	3.4 (3.2)
Refinement						
Rwork (%)		24.1	22.4	23.3	23.1	21.5
Rfree (%)		27.7	26.2	26.3	26.9	25.2
Rmsd bond length (Å)		0.015	0.016	0.017	0.022	0.015
Rmsd bond angle (°)		2.08	2.07	2.13	1.97	1.94
Clashscore		10.64	16.21	10.01	10.35	5.01

Table S2.2: Crystallographic refinement statistics for hGALK1 co-crystal structures. *Continued.*

Ligand:		**Fragment 26**	**Fragment 27**	**Fragment 28**
Unit cell parameters	(Å)	73.21 114.30 120.81	73.53 115.00 120.85	73.32 114.96 120.92
	(°)	90 100.54 90	90 100.65 90	90 100.79 90
Space group		P 1 21 1		
Resolution (Å)		82.36 - 2.3	118.8 - 2.27	82.61 - 2.4
Observed/Unique reflections		296178/86820 (30380/8665)	316652/90900 (32559/9086)	263015/76670 (26611/7668)
R-merge		0.1111 (0.6936)	0.1291 (1.941)	0.09289 (0.7412)
CC (1/2)		0.991 (0.773)	0.996 (0.409)	0.997 (0.763)
I/sig(I)		7.60 (1.94)	4.68 (0.48)	5.39 (1.15)
Completeness		99.80 (99.55)	99.68 (99.81)	99.36 (99.77)
Multiplicity		3.4 (3.5)	3.5 (3.6)	3.4 (3.5)
Rwork (%)		20.9	21.9	21.9
Rfree (%)		24.0	24.8	24.7
Rmsd bond length (Å)		0.016	0.016	0.015
Rmsd bond angle (°)		2.14	2.29	2.05
Clashscore		9.01	11.39	10.11

Appendix 3: Characterisation of synthesised compounds

Section 2.5 describes the synthesis of 13 follow-up compounds derived from fragment **6**. The identity of these compounds was confirmed from their spectra in ^{1}H-NMR and mass in analytical LC-MS. Representative NMR and LC-MS spectra, for compound **123**, are shown in figures S3.1A and 3.1B, respectively, and multiplet reports for all compounds are listed in table S3.1.



Peak Number	Compound	Time	AreaAbs	Area %Total	Width	Height	Mass Found
3	Found	1.51	2e+004	79.57	0	1e+006	275.00
5	Found	1.54	2e+003	8.30	0	4e+005	275.00
6	Found	1.56	2e+003	6.06	0	2e+005	275.00
8		2.07	2e+003	6.07	0	3e+005	

Figure S3.1: Characterisation of compound 123. *A. ^{1}H-NMR spectrum of compound **123** showing assignment of hydrogen atoms to the compound structure. B. LC-MS spectrum of final product from synthesis of compound **123** showing the expected molecular weight of 275 Daltons.*

Table S3.1: ^{1}H-NMR spectra of synthesised compounds. *Data were analysed using MestReNova software.*

Compound	1H NMR spectrum
123	^{1}H NMR (400 MHz, $CDCl_3$) δ 8.57 (s, 1H), 8.30 (s, 1H), 8.25 – 8.18 (m, 1H), 7.56 – 7.47 (m, 1H), 6.71 – 6.63 (m, 2H), 3.91 – 3.84 (m, 4H), 3.72 – 3.64 (m, 4H).
124	^{1}H NMR (400 MHz, $CDCl_3$) δ 8.51 (s, 1H), 8.27 – 8.20 (m, 2H), 7.64 (s, 1H), 7.24 – 7.16 (m, 2H), 7.16 – 7.07 (m, 1H), 7.02 – 6.94 (m, 2H), 6.85 (tt, *J* = 7.4, 1.1 Hz, 1H), 6.82 – 6.69 (m, 1H), 6.69 – 6.63 (m, 2H), 6.63 – 6.49 (m, 1H), 5.64 (s, 1H), 3.80 (s, 2H), 3.73 (s, 1H), 3.59 (s, 1H).
125	^{1}H NMR (400 MHz, $CDCl_3$) δ 8.56 (s, 1H), 8.35 (s, 1H), 7.54 – 7.45 (m, 1H), 7.08 (s, 1H), 7.00 – 6.92 (m, 1H), 4.55 (s, 5H), 4.35 – 4.24 (m, 0H), 4.11 (dd, *J* = 9.4, 4.9 Hz, 1H), 3.95 (dd, *J* = 9.4, 6.6 Hz, 1H), 2.97 (dd, *J* = 12.7, 10.1 Hz, 1H), 2.84 (dd, *J* = 12.7, 3.0 Hz, 1H), 2.63 (s, 4H), 2.06 (s, 3H).
126	^{1}H NMR (400 MHz, $CDCl_3$) δ 8.58 (s, 1H), 8.24 (s, 1H), 7.69 (s, 2H), 7.31 – 7.22 (m, 2H), 5.90 (d, *J* = 8.0 Hz, 1H), 5.57 (ddd, *J* = 9.2, 8.0, 6.1 Hz, 1H), 3.52 (s, 2H), 2.28 (ddd, *J* = 14.3, 8.3, 6.1 Hz, 1H), 2.23 – 2.12 (m, 1H), 1.81 (ddt, *J* = 13.0, 8.3, 6.5 Hz, 1H), 1.04 (d, *J* = 6.6 Hz, 3H), 0.99 (d, *J* = 6.6 Hz, 3H).
127	^{1}H NMR (400 MHz, $CDCl_3$) δ 8.55 (s, 1H), 8.25 (s, 1H), 7.27 (d, *J* = 3.5 Hz, 4H), 7.21 (s, 1H), 4.79 (s, 1H), 4.25 (s, 2H), 4.00 (s, 1H), 3.53 – 3.39 (m, 1H), 3.07 – 2.98 (m, 3H), 2.93 (dd, *J* = 13.1, 5.2 Hz, 1H), 1.66 (s, 2H), 1.53 (s, 9H).
128	^{1}H NMR (400 MHz, DMSO) δ 12.28 (s, 1H), 8.44 (s, 1H), 8.26 (s, 1H), 7.71 (t, *J* = 5.7 Hz, 1H), 7.58 – 7.51 (m, 1H), 7.43 (d, *J* = 7.2 Hz, 1H), 7.12 (dq, *J* = 9.3, 5.7, 4.7 Hz, 2H), 3.88 (td, *J* = 7.2, 5.7 Hz, 2H), 3.13 (t, *J* = 7.3 Hz, 2H).
129	^{1}H NMR (400 MHz, $CDCl_3$) δ 8.48 (s, 1H), 8.17 (s, 1H), 6.77 (s, 1H), 4.14 (tdd, *J* = 5.6, 3.4, 2.1 Hz, 1H), 3.73 (t, *J* = 4.7 Hz, 4H), 2.70 (ddd, *J* = 10.3, 7.2, 5.6 Hz, 1H), 2.51 (p, *J* = 6.8, 6.3 Hz, 4H), 2.18 – 1.91 (m, 3H), 1.81 (dddd, *J* = 12.0, 9.4, 7.1, 2.4 Hz, 1H), 1.76 – 1.58 (m, 2H).
130	^{1}H NMR (500 MHz, DMSO) δ 9.65 (s, 1H), 8.96 (d, *J* = 1.6 Hz, 1H), 8.52 (d, *J* = 1.6 Hz, 1H), 8.45 (s, 1H), 8.10 – 8.03 (m, 2H), 7.57 – 7.49 (m, 2H), 7.44 (ddt, *J* = 8.3, 6.8, 1.5 Hz, 1H).
131	^{1}H NMR (400 MHz, $CDCl_3$) δ 8.55 (s, 1H), 8.33 (s, 1H), 7.51 – 7.43 (m, 1H), 7.23 (s, 3H), 7.08 (s, 1H), 7.01 – 6.91 (m, 1H), 3.32 – 3.25 (m, 3H), 2.80 (t, *J* = 5.1 Hz, 3H), 2.46 (s, 2H), 2.05 (s, 3H).
132	^{1}H NMR (400 MHz, DMSO) δ 8.54 (s, 1H), 8.38 (s, 1H), 7.25 (t, *J* = 6.1 Hz, 2H), 5.24 (s, 4H).
133	^{1}H NMR (400 MHz, DMSO) δ 12.98 (s, 1H), 12.82 (s, 0H), 9.20 (s, 1H), 8.56 (s, 1H), 8.50 (s, 1H), 8.05 – 7.97 (m, 2H), 7.76 – 7.60 (m, 1H), 7.40 – 7.32 (m, 1H), 3.18 (s, 1H).
134	^{1}H NMR (400 MHz, $CDCl_3$) δ 8.61 (s, 1H), 8.32 (s, 1H), 8.09 (s, 1H), 7.67 (d, *J* = 0.7 Hz, 1H), 7.34 (td, *J* = 8.0, 5.8 Hz, 1H), 7.04 (ddt, *J* = 8.7, 5.4, 1.7 Hz, 3H), 6.94 (dt, *J* = 9.5, 2.1 Hz, 1H), 5.33 (s, 2H).
135	^{1}H NMR (400 MHz, $CDCl_3$) δ 8.65 (s, 1H), 8.36 (s, 1H), 7.66 (s, 1H), 7.42 (d, *J* = 8.3 Hz, 1H), 7.23 (d, *J* = 2.1 Hz, 1H), 6.96 (dd, *J* = 8.2, 2.1 Hz, 1H), 6.78 (d, *J* = 0.8 Hz, 1H), 5.16 (s, 2H), 2.29 (d, *J* = 0.7 Hz, 3H).