UNIVERSITY OF OXFORD



Design and Utilisation of a Poised Fragment Library against Epigenetic Proteins

Oakley Cox

Keble College,

Systems Approaches to Biomedical Sciences - Industrial Doctoral Centre,

Structural Genomics Consortium,

and Diamond Light Source

A thesis submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy

Trinity Term 2016

Abstract

Fragment based drug discovery has become an important method in biomedical research. In particular, it has enabled academic researchers to develop drug-like compounds and chemical probes which can be used for target validation. Recent developments at Diamond Light Source have made X-ray crystallography a viable primary fragment screen. The hits identified by X-ray screening are in general too weak to be verified by orthogonal solution assays, yet such information is a common requirement in medicinal chemistry operations. Progressing weak hits to a potent lead series requires changes not only to the periphery, where conventional SAR operates, but also to the core of the compound. To overcome the problem, I introduce the concept of poised fragment screening. Poised fragments can be synthesised in one step from commercially available starting materials using robust, high yielding reactions. As a result, a library of analogues can be synthesised quickly and at relatively low cost.

In this thesis I report the design and purchase of a 776-compound, chemically diverse poised fragment library (DSPL) for screening using X-ray crystallography, which is in regular use at Diamond. To aid researchers performing follow-up campaigns of poised hits, I developed a computational tool, XPoise, which systematically builds virtual libraries of follow-up compounds using poised reactions.

Follow-up campaigns against four epigenetic proteins revealed the benefit of using poised chemistry over traditional methods such as SAR by catalogue. Weak hits were improved into lead compounds with modest potency alongside structural data. Finally, a lead series of PHIP(2) inhibitors was elaborated using bespoke organic synthesis and late stage biotransformation to build complex, chiral compounds in the search for a potent and selective chemical probe.

Dedicated to Ben Probert

Acknowledgements

Thank you to my supervisors, Paul Brennan and Frank von Delft, for giving me the opportunity to work on such a rewarding project. In particular to Paul, for his support in my pursuit of the many opportunities Oxford has to offer.

I would like to thank the incredible scientists I have had the privilege to work with and who have contributed to my project. Above all, the work of Tobias, Patrick and Anthony, as well as the tireless efforts of Saleta, Filippo, Moses and Dave who have helped bring my thesis together. And thank you to the Med Chem group who have always made me feel welcome and given me a friendly and fun place to work for over three years.

A special mention to all those at Keble College who have made my time at Oxford so memorable. During my time at Keble, college life has always had an energy and generosity that has made me feel at home. I will always treasure the many memories and friendships it has given me.

And finally to my family: Mum, Dad, Silas, Grandpa, Grampy and Nan. You have always encouraged me to be true to myself and to embrace new opportunities and challenges. Thank you for all you have given me. I look forward to our next adventure.

Table of Contents

1 Introduction 1

1.1 Overview 1

1.2 Drug Discovery 2

1.2.1 Target Validation 4

1.2.2 Hit Discovery 6

1.2.3 Hit-to-Lead Development 9

1.2.4 Lead Optimisation and Preclinical Development 11

1.2.5 A Faltering Pipeline 15

1.3 Chemical Probes 17

1.4 Epigenetic Proteins 21

1.4.1 Bromodomains 23

1.4.2 Lysine Demethylases 25

1.5 Fragment Based Drug Discovery 27

1.5.1 History of FBDD 28

1.5.2 Hit Discovery in FBDD 29

1.5.3 Hit-to-Lead Optimisation in FBDD 33

1.6 X-ray Crystallography and Structure Based Drug Design 34

1.7 Medicinal Chemistry 38

1.8 Project Aims 42

2 Design of a Poised Fragment Library 44

2.1 Overview 44

2.2 Background 44

2.2.1 Principles of Fragment Library Design 44

2.2.2 Literature Review of Fragment Libraries 46

2.2.3 Chemical Diversity Metrics 49

2.2.4 Project Aim 53

2.3 Drug-Like Commercial Fragments 55

2.3.1 ZINC Database 55

2.3.2 Rapid Elimination of Swill 58

2.4 Poised FBDD 59

2.4.1 Defining poised chemistry 59

2.4.2 Identifying Poised Fragment Space 62

2.5 Diversity Filter 64

2.6 Purchase of DSPL 66

2.6.1 Properties of DSPL 67

2.6.2 Comparison of DSPL to other fragment libraries 69

2.7 Ongoing Work and Conclusions 72

3 XPoise - Fragment Elaboration Software 74

3.1 Overview 74

3.2 Background 74

3.2.1 Reaction Cheminformatics 74

3.2.2 LLOOMMPPAA 77

3.2.3 Project Aim 79

3.3 Methods and Materials 80

3.3.1 Overview 80

3.3.2 Building Blocks library 80

3.3.3 RDKit for KNIME 82

3.4 Results 84

3.4.1 Overview 84

3.4.2 The XPoise Workflow 85

3.5 Ongoing Work and Conclusions 93

4 Poised Fragments for Epigenetic Targets 96

4.1 Overview 96

4.2 Background 97

4.2.1 Fragment Soaking 97

4.2.2 Fragment Screening of Epigenetic Proteins 99

4.2.3 AlphaScreen 100

4.2.4 Project Aim 102

4.3 Brd BRD1 103

4.3.1 BRD1 Fragment Soaking Campaign 103

4.3.2 BRD1 Follow-up Campaigns 106

4.3.3 Re-Synthesis of BRD1 Fragment Hits 111

4.4 Brd PHIP(2) 114

4.4.1 PHIP(2), an Atypical Bromodomain 114

4.4.2 PHIP(2) Fragment Soaking Campaign 115

4.4.3 Thioureas as PHIP(2) Inhibitors 118

4.4.4 Computational Design of Thiourea Analogues 121

4.4.5 *N*-Benzyl Amides as PHIP(2) Inhibitors 127

4.4.6 PHIP(2) Overview 130

4.5 KDM JMJD2D 130

4.5.1 Background 130

4.5.2 JMJD2D Fragment Soaking Campaign 131

4.5.3 Using SAR by Catalogue to Study JMJD2D Hits 133

4.5.4 SAR by Catalogue vs Poised Elaboration 135

4.6 Tudor Domain of SETDB1 137

4.6.1 SETDB1 Fragment Soaking Campaign 137

4.6.2 SETDB1 Follow-up Libraries 139

4.7 Ongoing Work and Conclusions 141

5 Synthesis of *N*-Dimethoxybenzyl Lactam Analogues as PHIP(2) Inhibitors 143

5.1 Overview 143

5.2 Microbial Biotransformation 144

5.2.1 Characterisation of Hydroxylated Lactams 144

5.3 PHIP(2) Lead Optimisation 147

5.3.1 Chirality 147

5.3.2 Nitro-Mannich Lactamisation Cascade in the Synthesis of PHIP(2) Inhibitors 149

5.4 Multi-Component Reactions 155

5.5 Ongoing Work and Concluding Remarks 158

6 Conclusions and Future Plans 159

6.1 Evolution of DSPL 160

6.2 Development of XPoise 161

6.3 PHIP(2) Chemical Probe 162

6.4 Concluding Remarks 162

Bibliography 163

Appendix A: REOS Filter used in Chapters 2 and 3 177

Appendix B: KNIME Workflows used in Chapters 2 and 3 181

Appendix C: List of DSPL Fragments as 2D Structures 186

Appendix D: Poised Reaction SMARTS used in Chapter 3 192

Appendix E: Experimental Procedures 196

Appendix F: Microbial Biotransformation Methods and Materials 277

List of Figures

Figure 1.1: The drug discovery pipeline 4

Figure 1.2: Biological understanding assembled from target validation 5

Figure 1.3: Methods used in hit discovery 8

Figure 1.4: The iterative process used in hit-to-lead development 10

Figure 1.5: The challenges of multi-parameter lead optimisation 12

Figure 1.6: Formulae used to calculate ligand efficiency (LE) and lipophilic efficiency (LipE) 15

Figure 1.7: Significant factors affecting the profits of many companies in the pharmaceutical industry 16

Figure 1.8: Property requirements for a chemical probe 20

Figure 1.9: The three epigenetic protein classes 22

Figure 1.10: Bromodomain family tree with sub-families as defined by Filippakopoulos *et al* (2012) 24

Figure 1.11: Typical Brd binding site 25

Figure 1.12: JmjC domain containing demethylation mechanism 25

Figure 1.13: The KDM4D trimethyl lysine binding site 26

Figure 1.14: History of FBDD 28

Figure 1.15: Success of clinical compounds developed by FBDD 29

Figure 1.16: Hit-to-lead optimisation in FBDD: fragment linking, growing and merging 34

Figure 1.17: The iterative process of SBDD 37

Figure 1.18: Typical synthetic route employed by medicinal chemists 39

Figure 1.19: Amide coupling schematic and selected coupling reagents 40

Figure 1.20: Synthesis of thioureas 41

Figure 1.21: Project overview within the context of SBDD 43

Figure 2.1: The principles of Fragment Library Design 45

Figure 2.2: Circular fingerprint schematic 50

Figure 2.3: USRCAT fingerprint schematic 52

Figure 2.4: The framework used to design the Diamond-SGC Poised fragment library, DSPL 54

Figure 2.5: Schematic depicted the pipeline concept used in the KNIME Analytics Platform 55

Figure 2.6: Schematic illustrating how a poised fragment library can be used to rapidly synthesise a follow-up library of analogues 60

Figure 2.7: Scaffolds used to identify poised fragments 61

Figure 2.8: Schematic of KNIME workflow used to identify poised fragments 62

Figure 2.9: Synthetic reactions used to describe each of the selected poised reactions 63

Figure 2.10: Distributions of poised reactions in poised fragment space and DSPL_1000 66

Figure 2.11: DSPL poised reaction distribution 68

Figure 2.12: Physicochemical Properties of DSPL 69

Figure 2.13: Number of poised disconnections per fragment in SGC fragment libraries 71

Figure 2.14: Comparison of the distribution of poised reaction types between four fragment libraries. 72

Figure 2.15: Overview of the DSPL design process 73

Figure 3.1: Workflow used by the XPoise algorithm to generate fragment follow-up libraries 74

Figure 3.2: Reaction query strings for the Paal-Knorr pyrrole synthesis 76

Figure 3.3: The LLOOMMPPAA method 78

Figure 3.4: Proposed schematic for the design of follow-up libraries for poised crystal hits using the self-contained computational tool, XPoise 79

Figure 3.5: Price distribution of the first version of the building blocks library 81

Figure 3.6: Use of SMARTS in XPoise 82

Figure 3.7: Use of a phosphorus dummy-atom in reaction SMARTS 83

Figure 3.8: Reaction regioselectivity was ensured by removing building blocks which afforded two or more reaction products 84

Figure 3.9: Schematic representation of the XPoise workflow 86

Figure 3.10: 2D representations of the SMARTS strings used to generate libraries of amides 87

Figure 3.11: 2D representations of the SMARTS strings used to generate libraries of benzimidazoles 89

Figure 3.12: How atom numbering results in the formation of two equivalent benzimidazole products 91

Figure 3.13: SAR by catalogue using XPoise 92

Figure 3.14: Poised expansion of Suzuki coupling building blocks using XPoise 94

Figure 4.1: Comparison of conventional fragment screening with the DLS beamline facility.....98

Figure 4.2: Schematic depicting the AlphaScreen peptide displacement assay.....101

Figure 4.3: Hit-to-lead development of poised fragment hits.....102

Figure 4.4: BRD1 in complex to OF-1.....104

Figure 4.5: Fragment hits of BRD1 identified by fragment soaking experiments performed at DLS 105

Figure 4.6: Fragment hits of BRD1 selected for follow-up campaigns.....106

Figure 4.7: Retrosynthesis of BRD1 hits **17** and **25**.....107

Figure 4.8: Crystal structures of follow-up pyrimidine compound **51** bound to BRD1.....109

Figure 4.9: Crystal structures of follow-up amide compounds bound to BRD1 111

Figure 4.10: PHIP(2) in complex with NMP.....114

Figure 4.11: Fragment hits of PHIP(2) identified by fragment soaking experiments performed at DLS 115

Figure 4.12: PHIP(2) in complex with selected fragment hits.....116

Figure 4.13: PHIP(2) in complex with fragment hits selected for follow-up campaigns.....117

Figure 4.14: Synthesis of thiourea follow-up library 118

Figure 4.15: Structures of thiourea analogues bound to PHIP(2).....121

Figure 4.16: Design of a thiourea follow-up library using the XPoise/LLOOMMPPAA method 121

Figure 4.17: Alternative reaction conditions for the synthesis of terminal thioureas.....122

Figure 4.18: Thiourea library built using LLOOMMPPAA.....124

Figure 4.19: Comparing activity and LE of thiourea follow-up libraries 126

Figure 4.20: Synthesis of the *N*-benzylamide follow-up library.....127

Figure 4.21: PHIP(2) in complex with *N*-benzylacetamide analogues 129

Figure 4.22: JMJD2D in complex to KDOPP7 131

Figure 4.23: Fragment hits of JMJD2D binding in an allosteric site 132

Figure 4.24: Crystal structures of the JMJD2D allosteric binders 133

Figure 4.25: JMJD2D follow-ups designed by scaffold hopping 135

Figure 4.26: Comparison of SAR by catalogue versus poised elaboration of JMJD2D fragment hits 136

Figure 4.27: Fragment hits in complex to the SETDB1 Tudor domain 138

Figure 5.1: PHIP(2) binders **177** and **178** used as starting points in lead optimisation 143

Figure 5.2: NMR characterisation of hydroxylated lactams **234** and **235** 146

Figure 5.3: Chirality in amino acids 148

Figure 5.4: Development of the BRD7/9 chemical probe, LP99 149

Figure 5.5: Synthesis of amines **246** and **247** using a nitro-Mannich lactamisation cascade 150

Figure 5.6: Synthesis of δ-lactam **249** 151

Figure 5.7: Assignment of the single diastereomer of amine **247**, using NMR 153

Figure 5.8: Synthesis of amine **247** derivatives 153

Figure 6.1: The poised paradigm in drug discovery 159

List of Tables

Table 1.1: Hit discovery techniques commonly used in FBDD....30

Table 2.1: Selected list of commercially available fragment libraries....47

Table 2.2: Vendor catalogues used to build the DSPL....57

Table 3.1: Filters and SMARTS reactions used to generate amide libraries....88

Table 3.2: Filters and SMARTS reactions used to generate benzimidazole libraries....90

Table 4.1: Synthesis of structural analogues to BRD1 hit **17**....108

Table 4.2: Synthesis of structural analogues to BRD1 hit **25**....110

Table 4.3: Resynthesis of BRD1 fragment hits containing amide bonds....112

Table 4.4: Comparing the BRD1 inhibition of fragment hits **22** and **29** to three chemical probes....113

Table 4.5: Synthesis and screening of thiourea follow-up series....120

Table 4.6: Synthesis and screening of thiourea follow-up library suggested by LLOOMMPPAA....125

Table 4.7: Synthesis and screening of amide follow-up series....128

Table 4.8: JMJD2D follow-up library designed by SAR by catalogue....134

Table 4.9: Design of follow-up libraries for fragment hits of SETDB1....140

Table 5.1: PHIP(2) inhibition of the *N*-dimethoxybenzyl lactam series....154

Table 5.2: Synthesis of 4,5-substituted pyrimidines....156

Table 5.3: Synthesis of tetrazoles using a variation of the Ugi-reaction....157

Abbreviations

2-OG	2-Oxoglutarate
3DFC	3D Fragment Consortium
ADME	Absorption, distribution, metabolism and excretion
AlphaScreen	Amplified luminescent proximity homogeneous assay
Asn	Asparagine
APT	Attached proton test
ATP	Adenosine triphosphate
BAZ2	Bromodomain adjacent to zinc finger domain 2
BB	Building blocks
BCL2	B-cell lymphoma 2
BET	Bromodomain and extra terminal domain
Brd	Bromodomain
BRD	Bromodomain containing protein
BRPF	Bromodomain and PHD finger-containing protein
CAS	Chemical abstracts service
CBP	CREB-binding protein
CDI	Carbonyldiimidazole
CDMT	2-Chloro-4,6-dimethoxy-1,3,5-triazine
CECR2	Cat eye syndrome critical region protein 2
CNS	Central nervous system
COSY	Homonuclear correlation spectroscopy
CRISPR	Clustered regularly interspaced short palindromic repeats
CRO	Contract research organisation
CSV	Comma separated values file
Da	Dalton

DCC	*N*,*N'*-Dicyclohexylcarbodiimide
DLS	Diamond Light Source
DMB	Dimethoxybenzyl
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DOGS	Design of genuine structures
DSPL	Diamond SGC Poised Library
EC_{50}	Concentration of 50% maximum effect
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
ESI	Electrospray ionisation
FBDD	Fragment based drug discovery
FDA	Food and Drug Administration
FRET	Förster resonance energy transfer
fsp^3	Fraction of atoms with sp^3 hybridisation
GBP	Great British pound
GPCR	G-protein-coupled receptor
HA	Heavy atom
HAT	Histone acetyltransferase
HATU	1-[Bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxid hexafluorophosphate
HBA	Hydrogen bond acceptor
HBD	Hydrogen bond donor
HDAC	Histone deacetylase
HKDM	Histone lysine demethylase
HMBC	Heteronuclear multiple-bond correlation spectroscopy
HMT	Histone methyltransferase
HPLC	High pressure liquid chromatography

HSP90	Heat shock protein 90
HSQC	Heteronuclear single quantum coherence spectroscopy
HTS	High throughput screening
IC_{50}	Concentration of 50% inhibition
IP	Intellectual property
J	Coupling constant (Hz)
JmjC	Jumonji *C*-terminal domain
JMJD	Jumonji domain containing
K_d	Dissociation constant
KAc	Acetyl lysine
KDM	Lysine demethylase
Kme	Methyl lysine
Kme2	Dimethyl lysine
Kme3	Trimethyl lysine
LCMS	Liquid chromatography–mass spectrometry
LDV	Low dead volume
LE	Ligand efficiency
LipE	Lipophilic efficiency
LXR	Liver X receptor
MACCS	Molecular access system structural
MCR	Multi-component reaction
MDS	Multi-dimensional scaling
MMP	Matched molecular pair
MQL	Molecular query language
mRNA	Messenger RNA
MS	Mass spectrometry
MW	Molecular weight

NCI/NIH	National Cancer Institute/ National Institutes of Health
NME	New molecular entity
NMP	*N*-Methyl pyrrolidone
NMR	Nuclear magnetic resonance
NRB	Number of rotatable bonds
PAINS	Pan-assay interference compounds
PANDDA	Pan-Dataset Density Analysis
PB1	Polymerase basic protein 1
PCR	Polymerase chain reaction
PD	Pharmacodynamics
PDB	Protein Data Bank
Ph	Phenyl
PHIP	Pleckstrin homology domain interacting protein
PK	Pharmacokinetics
PLIFs	Protein-ligand interaction fingerprints
PPI	Protein-protein interaction
PRMT	Protein arginine methyltransferase
PW	PHIP water
PyBOP	(Benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate
Quant.	Quantitative
REOS	Rapid elimination of swill
RNA	Ribonucleic acid
RNAi	RNA interference
Ro3	Rule of three
Ro5	Rule of five
SAR	Structure-activity relationship
SBDD	Structure based drug design

SCUBIDOO	Screenable Chemical Universe Based on Intuitive Data Organization
SDF	Structure data file
SETDB1	Su(var)3-9 and 'Enhancer of zeste' domain bifurcated 1
SGC	Structural Genomics Consortium
SMARCA	SWI/SNF-related, matrix-associated, actin-dependent regulator chromatin A
SMARTS	SMILES arbitrary target specification
SMILES	Simplified molecular-input line-entry system
S_N2	Second order nucleophilic substitution
S_NAr	Nucleophilic aromatic substitution
SPR	Surface plasmon resonance
SYNOPSIS	SYNthesize and OPtimize System In Silico
T3P	Propylphosphonic anhydride
TBTU	*O*-(Benzotriazol-1-yl)-*N*,*N*,*N*',*N*'-tetramethyluronium tetrafluoroborate
TCDI	1,1'-Thiocarbonyldiimidazole
TEA	Triethylamine
TFFH	Fluoro-*N*,*N*,*N*',*N*'-tetramethylformamidinium hexafluorophosphate
THP	Tetrahydropyran
TMS	Trimethylsilyl
TSA	Thermal shift assay
USRCAT	Ultrafast shape recognition with Credo atom types
ZINC	Zinc Is Not Commercial

1 Introduction

1.1 Overview

In the past twenty years, discovery of small molecule drugs for the treatment of human disease has been remarkably successful. The pharmaceutical industry invests billions of dollars and employs thousands of people in the search for new treatments. However rates of late stage failure, one of the biggest costs incurred by big pharma, remain high. The biggest cause of drug attrition is a lack of efficacy in the clinic, often the result of a shortage of target understanding when designing preclinical disease models. Consequently more attention is being paid to target validation, and more emphasis placed on the effective use of chemical probes to understand a target's role in disease.

In the first part of this chapter I outline the drug discovery pipeline used by big pharma and how chemical probes can be utilised to improve target validation. The SGC (Structural Genomics Consortium) at the University of Oxford has had widespread success in the development of potent and selective chemical probes targeting epigenetic proteins. The epigenome plays a critical role in the expression of DNA, but how it does this is still largely unknown. Chemical probes are helping to unravel epigenetic mechanisms and identify potential targets for the treatment of disease. In the second part of this chapter I summarise epigenetic mechanisms, with a focus on two protein classes: bromodomains and lysine demethylases.

In the final part of the chapter I describe how fragment based drug discovery can be utilised to develop novel drug candidates and chemical probes in academic labs. X-ray crystallography at DLS (Diamond Light Source) can be used as a primary screening method for hit discovery, enabling the

detection of weak fragment hits against challenging target classes. Structure based drug design can then be employed for hit-to-lead development by combining computational tools with rational design. The biggest bottleneck in hit-to-lead development is the design and synthesis of analogues using medicinal chemistry, especially if hits are not synthetically tractable.

This thesis describes the development of a novel hit-to-lead method which utilises parallel synthesis and fragment based drug discovery. First I describe the design and purchase of a fragment library, the Diamond-SGC Poised Library, which was built by selecting compounds amenable to parallel synthesis (Chapter 2). Next I explain the development of a computational tool to design follow-up compounds for hits obtained from the Diamond-SGC Poised Library (Chapter 3). I then explore the utility of using poised fragments in hit-to-lead development by designing and synthesising analogues to fragment hits against the bromodomains PHIP(2) and BRD1, and the lysine demethylase, JMJD2D (Chapter 4). Finally I explore how modern organic chemistry can be utilised in hit-to-lead development, including multi-component reactions, microbial biotransformations and diastereoselective cyclisations (Chapter 5).

1.2 Drug Discovery

A pharmaceutical drug is a chemical or biological substance used to prevent, treat, cure or diagnose disease. Drugs can be extracted from natural sources such as plants or developed synthetically using organic chemistry or bioengineering. Most drugs are orally bioavailable, but can also be injected, inhaled or absorbed via a patch. The majority of drugs are classified as small molecule entities (<900 Da in molecular weight), however an increasing number are so-called biologics. Biologics are biological medicines which include vaccines, blood components, cell or gene therapy, and recombinant proteins including antibodies. They can be living entities such as cells and tissues, or composed of sugars, proteins, polypeptides and/or nucleic acids. Applications

to CDER (Center for Drug Evaluation and Research) for the approval of biologics as medincines (MW > 1000 Da) have been steadily increasing over the past two decades. These so called Biologics License Applications made up less than a tenth of approvals in the decade 1993 to 2002, but comprised a quarter of approvals in the past two years (Mullard 2015; Mullard 2016).

Despite the success of biologics, small molecules make up the majority of New Molecular Entities (NMEs) approved as medicines. Small molecules have a number of advantages (Laird 2013). Their well defined structures mean they are easily characterised. Potential modifications are well defined making IP (intellectual property) protection more straightforward. The majority of small molecules are stable, can be easily stored and transported, and can be produced on a large-scale at low-cost and reproducibly in different laboratories. The majority of small molecule drugs are developed and produced using organic synthesis. In this thesis I will focus on how organic synthesis can be utilised to improve the discovery and development of small molecule drugs.

The drug discovery process can be broadly split into two categories: preclinical and clinical (Figure 1.1). The preclinical research pipeline can be broken down into five steps. Firstly, a pharmacological link between a protein and a disease is identified (target validation). Next, small molecules which bind to the target of interest are detected (hit discovery). Third, hits are triaged for favourable properties (hit-to-lead development). The chemical structure of the selected molecule is then optimised into a drug candidate (lead optimisation) and finally feasibility, safety and dosage data are collected ready for clinical trials (preclinical development).

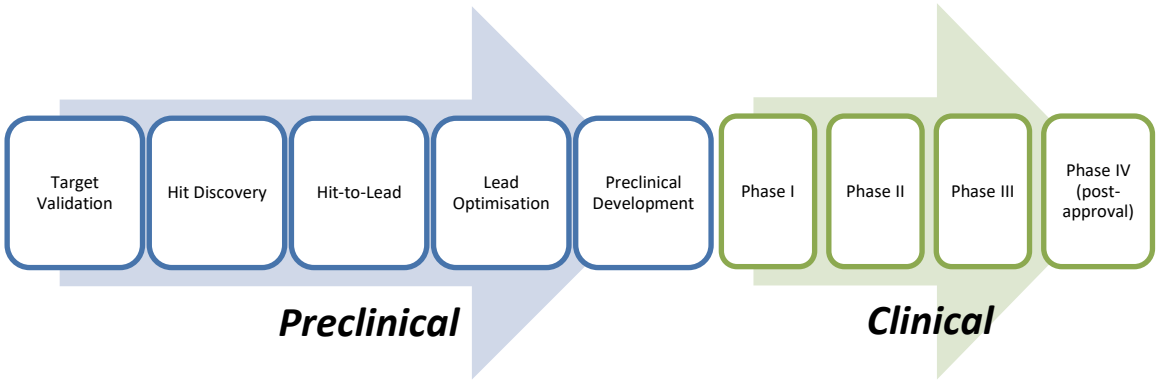


Figure 1.1: The drug discovery pipeline. Preclinical research identifies how to target disease using a drug candidate. Clinical trials assess the efficacy and safety of the drug candidate in humans both before and after approval.

Clinical research aims to understand the effect of a drug candidate in humans. Typically this is achieved using four phases of clinical trials (Phase I to Phase IV). Phase I tests the toxicological effect of the drug candidate in healthy humans. Phase II and III quantify the therapeutic effect of the drug on patient populations and determine a safe and efficacious dose. If the drug is deemed to be safe and effective, it is submitted for approval to a national agency such as the FDA (Food and Drug Administration) or EMA (European Medicines Agency). Phase IV is the term used to describe the study of a drug's efficacy and safety profile post-approval. Clinical trials can far outweigh the cost of preclinical research and development (Cook, Brown *et al.* 2014). As a result, late stage failures can be hugely damaging to a pharmaceutical company.

The work presented in this thesis focuses on preclinical drug discovery, in particular the use of organic synthesis in hit-to-lead development. To better understand the context of hit-to-lead development in the preclinical drug discovery process I will now explore each of the five pre-clinical steps in more detail.

1.2.1 Target Validation

Target validation is the process of identifying a biomolecule which can be modulated to affect a disease. The majority of drug targets are proteins or enzymes but can also be DNA, RNA or

antibodies. Initially targets are identified using analysis of phenotypic assays, proteomics or gene expression (Swinney and Anthony 2011). Once a target is linked to a disease it can be further interrogated in cellular or *in vivo* models to understands its role in disease progression. Protein levels can be altered using mRNA (messenger RNA), RNAi (interference RNA) or gene knockouts. The downstream processes and other biological pathways associated with the target are ascertained to determine the safety of the gene as a therapeutic target. Identifying a target's natural substrate(s), isoforms, phylogenetic family and druggability are also important in assessing its suitability as a drug target (Figure 1.2). For example, protein-protein interactions (PPIs) are often considered challenging targets because the proteins lack a distinct lipophilic pocket into which a small molecule can bind.



Figure 1.2: Biological understanding assembled from target validation. An understanding of a target's role in disease is not the sole requirement. It is also important to have an understanding of a target's family tree, its natural substrates and its druggability before attempting hit discovery.

Both phenotypic assays and target-based approaches have had great success in identifying bioactive molecules which can be optimised into drugs (Swinney and Anthony 2011). However

identifying and characterising the target or targets can be a long and drawn out process. In part this is because the biological methods used to assess a target's link to disease do not always capture the dynamic effect of a drug. For example, a gene targeted by a knockout experiment is never expressed by the organism throughout its lifetime. This is very different from disease contraction and progression, which are often caused by biological changes (such as mutations or overexpression) after an organism is fully developed. Natural substrates can be used to help understand the effects of modulating a target, but ultimately a potent and selective chemical probe is needed (Bunnage, Chekler *et al.* 2013). Chemical probes are drug-like molecules which can be used in a dose dependent manner in cellular or *in vivo* experiments to validate targets. Chemical probes do not need to meet the same requirements as a drug candidate so can be developed at a fraction of the time or cost. The development of chemical probes is discussed in more depth later in this chapter.

1.2.2 Hit Discovery

A hit molecule is defined as a compound or biostructure which shows activity against a desired target *in vitro,* in cellular assays or *in vivo*. Hit discovery methods have changed dramatically in the past 40 years, turning drug discovery from a haphazard process into an ever more rational practice. Traditionally, potential drugs were developed following clinical observations, presentations and posters at scientific meetings, published reports in scientific journals or collaboration with academic investigators (Lipinski, Lombardo *et al.* 2001). Hits stemmed from natural products or the side-effects of approved drugs (Kubinyi 1995). Examples of natural product hit discovery include aspirin (approved in 1899 and derived from the active ingredient in willow tree bark), and Taxol (approved in 1992). The latter was first isolated from Pacific yew trees

in 1967 and its development and eventual approval 25 years later is characteristic of the *ad hoc* nature of drug discovery during that era.

The first high throughput screen (HTS) in 1989 allowed drug discovery to become more rational (Moos, Green *et al.* 1993). HTS identifies hits from libraries of millions of compounds using low-cost *in vitro* assays, and were widely adopted in the 1990s with the introduction of automation (Figure 1.3). The low-cost to performing large scale structure-activity relationship (SAR) studies meant HTS libraries became dominated by synthetically accessible compounds in favour of complex natural products (Ortholand and Ganesan 2004). An analysis of approved drugs by Lipinski *et al.* in 1997 gave one of the first descriptions of drug-like chemical space which further shaped the design of HTS compound libraries (Lipinski, Lombardo *et al.* 1997). The now famous Lipinski Rule of 5 (Ro5) of molecular weight ≤ 500, clogP ≤ 5, number of hydrogen bond donor ≤ 5 and number of hydrogen bond acceptors ≤ 10, was widely adopted by big pharma and still forms the core of lead optimisation projects today. The rules were proposed as a guideline for developing orally bioavailable drugs, but a number of drugs classes were found to fall outside the Ro5, including antibiotics, antifungals, vitamins and cardiac glycosides.

Hits obtained using HTS need validation as true positives. A post-screening triage removes assay artefacts and promiscuous compounds. An orthogonal assay is usually used to remove false positives. However special care must be taken to remove pan-assay interference compounds (PAINS) which display non-specific bioactivity and/or interfere with assay readouts across multiple targets (Baell and Holloway 2010; Dahlin, Nissink *et al.* 2015). The source of the behaviour can include aggregation, metal chelation, singlet oxygen production, compound fluorescence, redox activity, impure samples, membrane disruption and cysteine oxidation. Failure to remove these

compounds can result in wasted time and money (Devine, Mulcair *et al.* 2015), and can also result in misleading publications polluting the literature (Baell and Walters 2014; Baell 2015).

An alternative to HTS using smaller, fragment-like compounds was proposed in 1996 (Shuker, Hajduk *et al.* 1996). So-called Fragment Based Drug Discovery (FBDD) offered a number of advantages to HTS, primarily better coverage of chemical space with smaller compound libraries (Hann, Leach *et al.* 2001) and flexibility in growing hits towards Lipinski's Ro5 chemical space during hit-to-lead development (Congreve, Carr *et al.* 2003). FBDD is discussed in more detail later in this chapter.

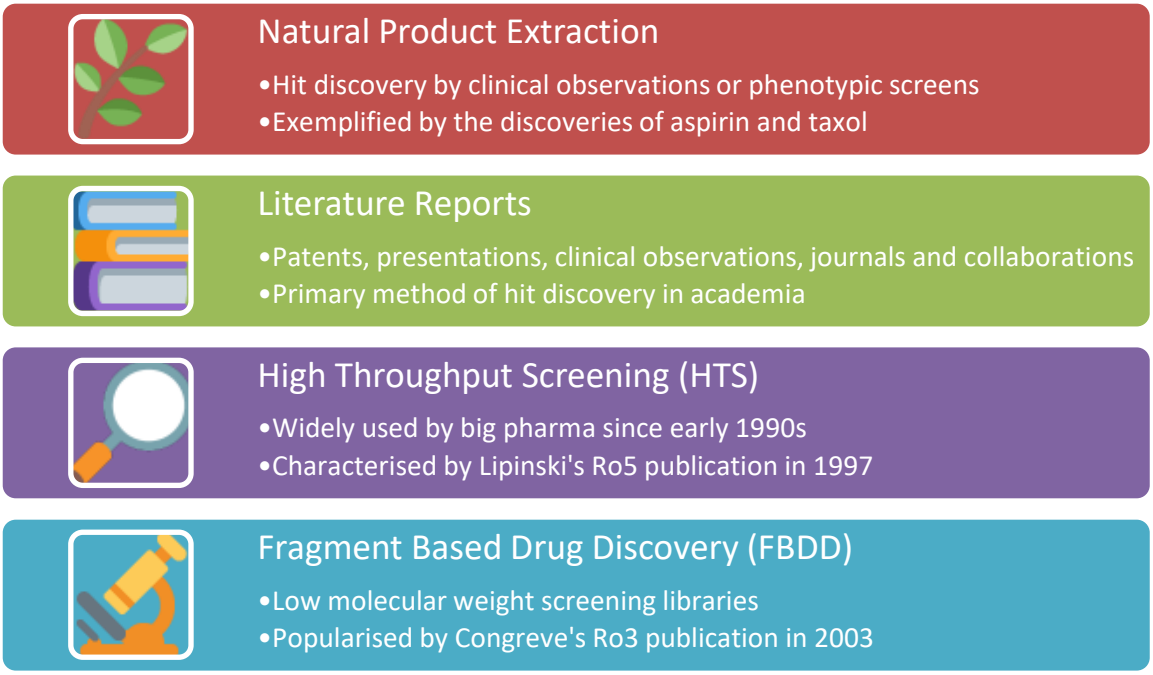


Figure 1.3: Methods used in hit discovery. Natural product extraction was superseded by HTS in big pharma during the 1990s. In comparison, hit discovery in academia is still driven by the sharing of clinical observations, scientific literature and patents.

The high cost of building and maintaining a HTS library, as well as the extensive infrastructure required to perform HTS, means the technique is rarely used in academia. The majority of academic hit discovery projects still rely on clinical observations, reports in the literature and serendipitous findings (Figure 1.3).

1.2.3 Hit-to-Lead Development

Once a hit against a target or group of targets has been identified, the molecule must be developed into a lead. A lead compound or lead series is a potent and selective molecule or series of molecules. The level of potency and selectivity varies between projects depending on the nature of the target and the predicted requirements for effective disease treatment. Generally the development of a hit into a lead utilises synthetic organic chemistry to design and synthesise structural analogues and test for desired biological effects (Kshirsagar and Editor 2008). The results of a given round of synthesis are then used to drive the next round of compound design in a cyclic process (Figure 1.4). This iterative process is often dependent on the knowledge of medicinal chemists who combine expertise in synthesising drug-like compounds with experience in how certain functional groups can affect biological activity. Decisions are often pragmatic, with compound design driven by a desire to access as many analogues as possible from a common starting point using robust chemical reactions. The rise of combinatorial chemistry (combichem) in the 1990s, driven by improvements in robotics and automation, allowed the rapid synthesis of large libraries containing thousands of compounds for hit-to-lead development (Ortholand and Ganesan 2004). However, the limited success of combichem projects is believed to be the result of a small number of chemical reactions giving poor coverage of chemical space (Kshirsagar and Editor 2008). Efforts have been made to utilise activity-directed synthesis (Karageorgis, Warriner *et al.* 2014). The method uses catalysed reactions with alternative possible outcomes, performed in parallel in arrays of different conditions. Products are isolated and tested for biological activity, and the results used to design the next reaction array. The method has been shown to generate submicromolar leads with novel scaffolds, but is not yet widely employed.

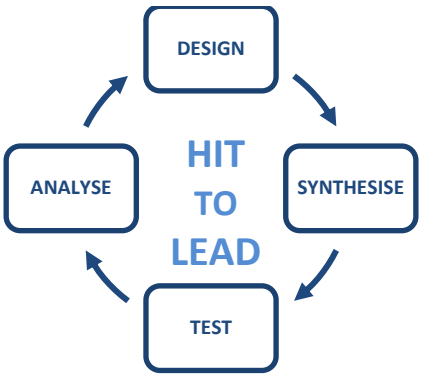


Figure 1.4: The iterative process used in hit-to-lead development. Medicinal chemists design and synthesise hit analogues and test the biological effect of the new compounds using in vitro solution based assays. The results are then analysed for structure-activity relationships (SAR).

The biological effect of a compound on a target can be measured using *in vitro* solution based assays. The assays use recombinant proteins which are expressed in organisms and purified in solution buffer. Depending on the nature of the target, full length or single domain proteins are used. Single domains have less biological relevance than full length protein, but are often easier to purify, characterise and crystallise. A wide variety of assays exist to assess the biological activity of a small molecule, and the assay selection is dependent on the biological system being targeted, the ability to reproducibly purify protein and the disease state under investigation. Common assays measure non-specific protein binding (e.g. thermal shift, SPR), inhibition of a specific binding event (e.g. FRET, NMR), disruption of protein function (e.g. proteomics-based LCMS) or effects on cellular systems (e.g. immunoassays). Dose dependent or time dependent outcomes of an assay can be measured to obtain IC_{50}, EC_{50} or K_d values. An IC_{50} is the compound concentration required to inhibit a biological process by 50%. Similarly, an EC_{50} is the compound concentration at which half of the compound's maximal effect is observed. The K_d (dissociation constant) is a measure of the reversible dissociation from a protein-ligand complex into the protein and free ligand. Improvements in IC_{50}, EC_{50} and K_d drive hit-to-lead development, and are commonly referred to as compound activity, affinity or potency. Iterations of synthesis and *in vitro* testing build an understanding of the SAR for ligand-protein systems. SAR studies assume improvements

in activity are the result of structural changes to the compound, and changes to different parts of the compound have independent and cumulative effects. The use of SAR observations drives the design of new compounds until the desired potency is achieved. Commonly, *in vitro* activity values in the nM range are considered a prerequisite for drug candidates (Hann 2011). If evidence of *in vivo* pharmacology has been obtained, it is common for researchers to file a patent at this stage to ensure a competitive advantage over rivials.

1.2.4 Lead Optimisation and Preclinical Development

Once a lead compound or lead series has been identified, the best candidate compounds are optimised ready for *in vivo* animal and human trials. As well as the potency and selectivity of a compound, the pharmacokinetic (PK), pharmacodynamic (PD) and toxicology profiles need to be optimised (Figure 1.5). Traditionally these five parameters were optimised independently, with a heavy bias towards improving the potency in the early stages of a hit-to-lead campaign (Lipinski, Lombardo *et al.* 2001). The result was a large failure rate for projects despite extensive and expensive chemistry efforts. Consequently multi-parameter optimisation of leads is now performed from an early stage after hit discovery, bringing about a big drop in project attrition rates caused by drug PK and PD (Hann 2011). Furthermore, medicinal chemists have developed tricks to improve PK and PD profiles of potent leads with small structural changes, such as the addition of fluorine or deuterium atoms, which can have minor affects on activity but a significant impact on drug metabolism and toxicology (Mullard 2016).

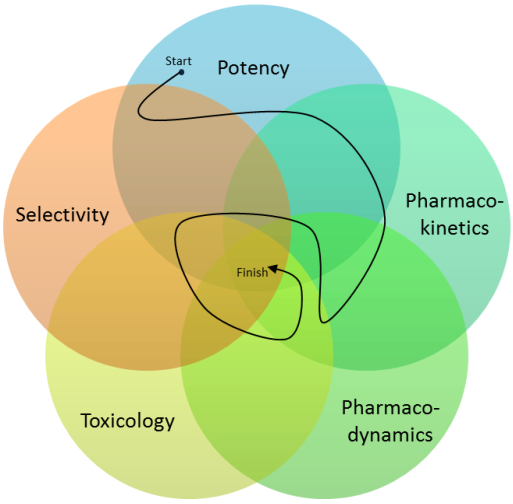


Figure 1.5: The challenges of multi-parameter lead optimisation. The black line represents a typical path taken during hit optimisation. All five parameters must be considered if the lead is to become a drug candidate. Adapted from Hann 2011.

The selectivity of a drug candidate is an important consideration to reduce off-target effects. Inhibition of key enzymes such as the hERG ion channel and the Cytochrome p450 (CYP) enzymes should be avoided (Sanguinetti and Tristani-Firouzi 2006; Guengerich 2008). *In vitro* activity assays are commonly used to assess selectivity and predict off-target effects. For popular disease targets, private companies offer screening panels against a diverse array of proteins within a family, such as the DiscoveRx platforms for Kinase, GPCR and Bromodomain profiling.[1] A minority of drugs affect multiple targets in order to achieve biological modulation. This so-called polypharmacology is more likely to show a therapeutic benefit, but can make lead optimisation challenging using standard medicinal chemistry and SAR studies (Hann 2011; Besnard, Ruda *et al.* 2012).

The PK profile of a drug candidate describes how *in vivo* administration affects the compound. The most common method employed to analyse PK properties, called the ADME scheme, is a series of mono-compartmental models, each representing one of four properties: absorption, distribution, metabolism and excretion. Each property can have a significant impact on a drug's

[1] www.discoverx.com, accessed 3rd October 2016

efficacy; a highly potent and selective compound could never become a clinical drug if it is quickly metabolised or not distributed to the biologically relevant target areas.

In contrast, the PD profile of a drug candidate describes how an organism reacts when the compound is administered. The primary focus of PD profiling is the dose-response relationship of a drug from which therapeutic windows and dose intervals can be established. The method of administration is an important consideration when carrying out dose-response studies. A compound administered intravenously is likely to have a much higher volume of distribution peak than an orally administered compound, which can take longer to be absorbed but also have longer clearance times.

Toxicology is the study of how a drug candidate can adversely affect an organism. Traditionally this has been done with animal experimentation. A number of variables need to be considered, including age, gender, long-term exposure and interactions with other drugs. Modern protocols use *in vitro* tests using cells or tissue before *in vivo* testing in order to minimise the number of experiments performed on animals (Hann 2011). Consequently a lead may be dropped after years of medicinal chemistry optimisation when an adverse toxicology profile is not identified until tests in higher species or Phase I clinical trials. Alongside drug efficacy, toxicology is one of the biggest causes of drug attrition.

A number of models have been developed to aid in lead optimisation. An early example is the Topliss method (Topliss 1977), in which a decision tree is used to suggest new substituents on aryl rings. While the method has been successful in improving potency, it does not take into consideration the other lead optimisation parameters. Structure Based Drug Discovery (SBDD) has had great success in aiding lead optimisation. Target-guided SBDD uses the 3D-structure of the target protein to guide rational SAR studies. The 3D-structures, generated using NMR or X-ray

crystallography, can enable rapid improvements in potency and selectivity, and are discussed in more detail later in this chapter. In contrast, ligand-guided SBDD can be used when no protein structure is available. Using databases of SAR data, computational models can be built to design new compounds with improved potency and/or chemical properties (Matlock and Swamidass 2014; Stumpfe, Hu *et al.* 2014). The technique is of particular use in preclinical development when optimising PK and PD profiles after a potent and selective lead series has been identified.

In contrast to the logical process presented by researchers in medicinal chemistry journals, lead optimisation without protein structures or SAR data can be *ad hoc* and subjective (Hann 2011). It is tempting for medicinal chemists to add large lipophilic substituents to bring about large improvements in potency while failing to consider the negative impact this may have on PK and toxicology profiles. Furthermore, it can be synthetically challenging to improve potency using polar functional groups (Lipinski, Lombardo *et al.* 2001). On the other hand, hits with low potency but high-quality ligand-protein interactions may be ignored. To combat these risks, researchers have designed metrics to aid hit-to-lead optimisation. Examples include ligand efficiency (LE) (Hopkins, Groom *et al.* 2004) and lipophilic efficiency (LipE) (Leeson and Springthorpe 2007). LE is a measure of the Gibbs free energy (ΔG) of binding per heavy atom in the ligand (N) (Figure 1.6). By maintaining the LE of a small, weak ligand, any improvements from additional molecular weight can be attributed to high quality interactions. Typically an optimised lead ready for testing in clinical trials has a LE around 0.3 to 0.4. Similarly, LipE is a measure of binding per unit of lipophilicity (logP), and maintaining a high LipE (5 to 7) increases the probability of a lead having a favourable PK profile. LipE is defined as the $p(IC_{50})$ minus the lipophilicity (Figure 1.6). Typically the calculated lipophilicity, clogP, is used.

$$LE = \frac{\Delta G}{N} \approx -\frac{1.4\ log\ IC_{50}}{N}$$

$$LipE = -log\ IC_{50} - c\ log\ P$$

Figure 1.6: Formulae used to calculate ligand efficiency (LE) and lipophilic efficiency (LipE). LE is a measure of the Gibbs free energy (ΔG) of binding per heavy atom (N) (Hopkins, Groom et al. 2004). LipE is predicted by subtracting the calculated partition coefficient (clogP) from the pIC_{50} (Leeson and Springthorpe 2007).

Many pharmaceutical companies have developed internal guidelines for lead optimisation (Hann 2011). For example, the Pfizer 3/75 rule takes into consideration the topological polar surface area (TPSA) of compounds (Hughes, Blagg *et al.* 2008), while the GSK 4/400 rule applies more restrictive limits on the Lipinski Ro5 (Gleeson 2008). Others have developed target specific guidelines, such as the Pfizer method for the multi-parameter optimisation of CNS drugs which monitors six physicochemical properties to increase the probability of a lead compound being suitable for targeting disease behind the blood-brain barrier (Wager, Hou *et al.* 2010).

1.2.5 A Faltering Pipeline

The pipeline outlined thus far has had huge success in bringing novel small molecule medicines to market. However, in recent years the pharmaceutical sector has struggled to develop enough blockbuster drugs to cover the rising cost of bringing a new medicine to market (Paul, Mytelka *et al.* 2010). A blockbuster drug is a drug which achieves $1 billion in annual sales for one or more years. For perspective, it is now estimated it takes around $2.8 billion to bring a drug to market (DiMasi, Grabowski *et al.* 2016). And even when a blockbuster drug does make it to market, the competitive advantage can be short lived thanks to 'Me Too' drugs developed by rival companies (Figure 1.7). 'Me Too' drugs are structurally similar small molecules with similar efficacy to the original drug but developed at a fraction of the cost.

Furthermore, many of the novel medicines brought to market target rare diseases with small patient populations. The percentage of approvals for orphan-designated drugs was 41% and 47%

for 2014 and 2015 respectively. An orphan disease is described as a condition which affects less than 200,000 people in the USA (i.e. less than 1 in every 1000 of the population) and includes many rare cancers as well as cystic fibrosis and Huntington's disease. The small patient size of orphan diseases results in pharmaceutical companies charging high prices for treatments to recoup the costs of their development, which in turn can damage a company's reputation and limit access for poorer patients in both developed and developing countries (Kessel 2014).

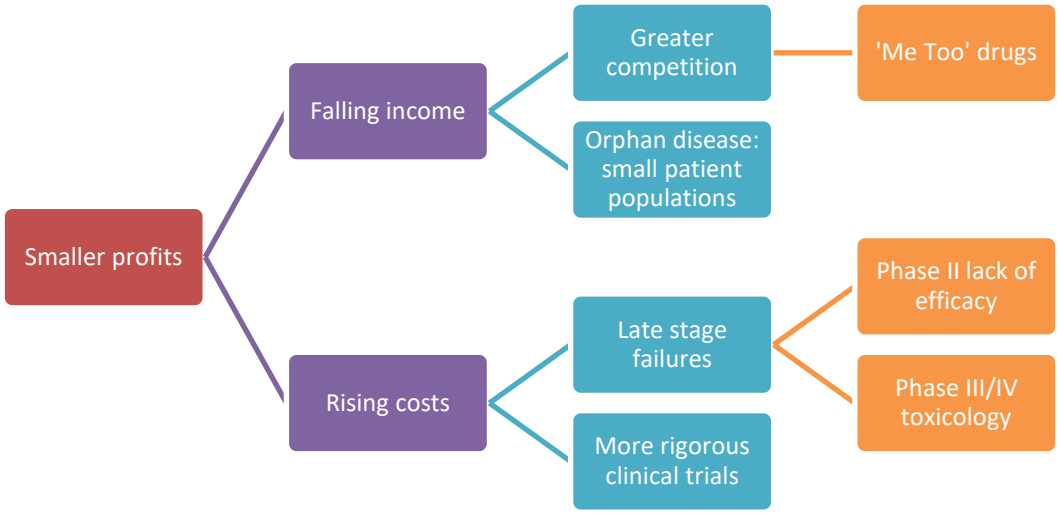


Figure 1.7: Significant factors affecting the profits of many companies in the pharmaceutical industry.

The rising cost of drug discovery can be attributed to two factors (Figure 1.7) (DiMasi, Grabowski *et al.* 2016). Firstly, the rising cost of clinical trials as national regulators increase the number of requirements before approval. While the Accelerated Approval Program introduced by the FDA in 2012 has accelerated drug-approval and helped many breakthrough medicines reach market, the overall approval costs are unlikely to decrease significantly in the future (Avorn 2015).

The second factor causing rising costs is the high rate of project attrition during clinical trials (Hann 2011). Often Phase III or IV trials identify dangerous side-effects in subsets of the population which were not spotted in the smaller test groups used in Phases I and II, and which force the medicine to be withdrawn from market. It is difficult to predict this type of failure, but it is hoped improved

understanding of early stage PK and PD and better *in vivo* models can help reduce the probability of these late stage failures. However, the biggest cause of drug candidate attrition is lack of efficacy in Phase II clinical trials (Cook, Brown *et al.* 2014). Too often the *in vitro* disease models used to develop lead compounds do not translate into clinical efficacy. The primary reason for this is poor preclinical target validation: a target's role in disease is not fully elucidated or the *in vivo* druggability of a target is overestimated. Many pharmaceutical companies now have more rigorous expectations from target validation. For example, Pfizer has published their Four Pillars of target validation: exposure at site of action, target engagement, functional pharmacology and relevant phenotypes (Bunnage, Chekler *et al.* 2013).

Increasingly researchers use chemical probes as an integral part of target validation (Workman and Collins 2010; Bunnage, Chekler *et al.* 2013; Arrowsmith, Audia *et al.* 2015). The demand for high quality chemical probes against a range of targets sets the context of this thesis. The SGC Chemical Probes project aims to develop potent and selective chemical tools for the elucidation of epigenetic protein function and their role in disease. The next two sections describe the utility of chemical probes and how they have been used to date in understanding the role of epigenetic proteins.

1.3 Chemical Probes

The sequencing of the human genome in the early 2000s revealed a large number of uncharacterised potential drug targets (Workman and Collins 2010; Arrowsmith, Audia *et al.* 2015). Chemical biology is playing an important part in elucidating the role of novel targets in human biology and in the treatment of disease, and the use of chemical probes is integral to the research. A chemical probe is a selective small molecule capable of modulating the function of a protein, allowing researchers to gain mechanistic insights and investigate phenotypic effects.

Chemical probes can be used in tandem with RNAi and CRISPR technologies for target validation. Unlike genetic approaches, small molecule probes offer the opportunity to inhibit specific protein functions (rather than removing the whole protein), act rapidly and reversibly in almost any cell type, and offer insight into the kinetics of inhibition. Furthermore, studies have shown the publication of a chemical probe stimulates further research around the protein or proteins it targets (Fedorov, Mueller *et al.* 2010; Edwards, Isserlin *et al.* 2011).

It is important to note that a good chemical probe would not necessarily make a good drug (Arrowsmith, Audia *et al.* 2015). A chemical probe does not require the same pharmacokinetic properties as a drug, such as bioavailability or metabolic stability. Conversely, a probe must have a high level of selectivity for a known target whereas a drug may have an undefined mechanism of action or activity against a number of targets (polypharmacology). Furthermore, a drug is protected by IP restrictions. If the scientific community is to make best use of a chemical probe, it should be freely available for purchase and use, along with data detailing its characterisation and bioactivity.

The utility of a chemical probe is defined by its quality. The literature is littered with examples of bad chemical probes or the misuse of poorly characterised probes (Workman and Collins 2010; Arrowsmith, Audia *et al.* 2015). In general, a compound with poor selectivity constitutes a bad chemical probe. Even more damaging are pan-assay interference structures (PAINS). The prevalence of PAINS in the academic literature is caused by a failure to recognise false positives during hit discovery, usually as a result of inadequate or non-existent orthogonal screening techniques, or failure to recognise common structural alerts (Baell and Walters 2014; Baell 2015). Analysis of the NIH MLI (National Institutes of Health Molecular Libraries Initiative) pilot project from 2004 to 2008 classified 25% of the 64 chemical probes generated as “dubious”, (Workman

and Collins 2010). The study concluded failure to sufficiently triage hits using properties regularly assessed in industry (e.g. solubility and chemical reactivity) as the primary contributor to the development of substandard probes. Misuse of a chemical probe in a poorly designed experiment can also result in misleading conclusions about a protein's biological function (Bunnage, Chekler *et al.* 2013). For example, using high compound concentrations may yield positive assay readouts but are likely to be the result of non-specific behaviour.

A range of properties are required for a chemical probe to be fit for purpose. Review articles have attempted to outline the properties of a good chemical probe (Baell and Holloway 2010; Workman and Collins 2010; Arrowsmith, Audia *et al.* 2015). A summary is depicted in Figure 1.8. The chemical structure should be well characterised (NMR, mass spectroscopy, synthetic route) and be stable in aqueous conditions. The compound's solubility and cell permeability are also important considerations. The expected potency of a chemical probe in biochemical and cellular assays is likely to be dependent upon the biological system under investigation. The general consensus is to target <100 nM in biochemical assays and 1-10 μM in cellular assays. Ideally the PK profile of the chemical probe would allow the compound to be tested *in vivo* and to target relevant tissue. The selectivity of a compound should be at least 10-100 fold selective over related targets, and profiled against the protein class (e.g. kinases) as well as broader pharmacology. Evidence of rational SAR studies and activity cliffs indicate potency is target specific. Further evidence of on-target effects can be determined using an inactive analogue, such as an enantiomer, as a control compound. Likewise, the use of two or more chemical probes from different structural classes can help rule out off-target pharmacology. Finally, the chemical probe should be free of IP restrictions and be freely available for use from a vendor, the original researchers or *via* a published synthetic route.

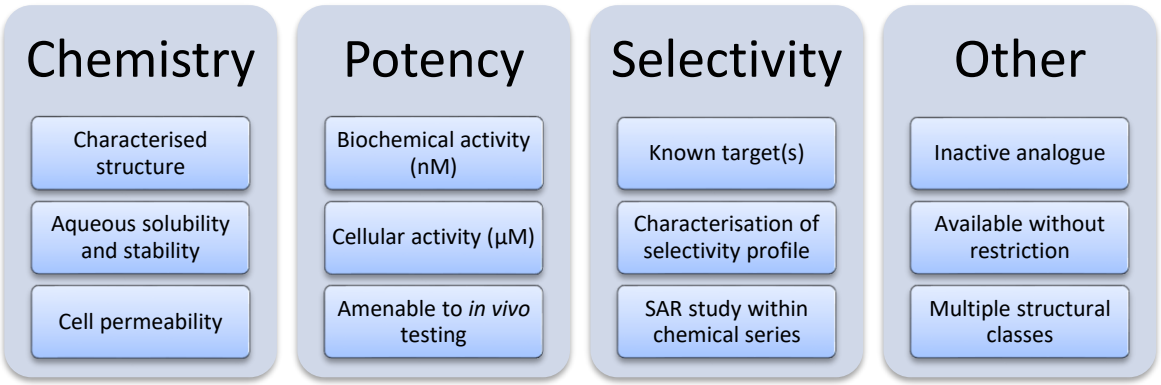


Figure 1.8: Property requirements for a chemical probe. Not all properties are necessarily required, but should be considered in the development of a chemical probe (Workman and Collins 2010).

The properties outlined in Figure 1.8 describe an ideal chemical probe. However, to expect this for every chemical probe is both unrealistic and, often, unnecessary. A number of high impact chemical probes which do not fulfil the ideal criteria have been reported for kinases, HSP90, LXR and bromodomains (Collins, Fivush *et al.* 2002; Workman, Burrows *et al.* 2007; Workman, Clarke *et al.* 2010; Bunnage, Chekler *et al.* 2013). In 2015 the Chemical Probes Portal[1] was founded to build a wiki-style record of published chemical probes (Arrowsmith, Audia *et al.* 2015). The resource aims to collate all available data about published chemical probes, including limitations and off-target effects. The hope is that researchers can use the information to select the best probe(s) for their target and how best to design an experiment to test their hypothesis.

Effective use of chemical probes in industry has significantly improved target validation, although it is too early to know the impact this will have on clinical trial attrition (Bunnage, Chekler *et al.* 2013). The comparatively low cost of developing a chemical probe offers academic labs the opportunity to contribute to discovering new drug targets (Edwards 2016). A leading example is the Chemical Probe Project at the Structural Genomics Consortium (SGC).[2] Here, a collaboration between academia and industry designs and characterises chemical probes for kinases and

[1] www.chemicalprobes.org, accessed 3rd October 2016

[2] www.thesgc.org, accessed 3rd October 2016

epigenetic proteins. Before a probe is released it is assessed using the following guidelines: *in vitro* potency <100 nM, >30-fold selectivity over sequence-related proteins, 'industry standard' profiling for off-target pharmacology, and on-target cellular activity < 1 μM (Arrowsmith, Audia *et al.* 2015). The project has had widespread success in producing high quality chemical probes for epigenetic targets such as bromodomains and lysine transferases.

1.4 Epigenetic Proteins

The experimental work described in this thesis centres on hit discovery for epigenetic proteins with the eventual aim of designing potent and selective chemical probes. While genetics is the study of genes encoded by DNA, epigenetics is the study of how the expression of genes is regulated (Bheda and Schneider 2014). Epigenetic regulation is believed to be non-hereditary, and underpins why the expression of a single DNA sequence can result in a wide variety of cell types, each with different sizes, structures, life expectancies and rates of mitosis. Each cellular phenotype is expressed by a complex regulatory network of covalent modifications to DNA and to histone proteins, which is affected by childhood development, disease, environmental chemicals, ageing and diet.

Epigenetic regulation can be split into two broad categories. First, methylation of DNA, specifically by the conversion of cytosine to 5-methylcytosine, is used to stabilise the DNA double-strand helix. As a result, areas of the genome containing high levels of methylation have been found to be less transcriptionally active.

The second category of gene regulation is post-translational modifications of histones. Histone octamers complex together to form nucleosomes, around which DNA is wrapped (Pande 2016). Nucleosomes organise to form chromatin, which is structured to allow long strands of DNA to be stored in cells just 10-100 μM across. Reversible modifications of histone tails protruding from the

nucleosomes regulate the winding and unwinding of the DNA coils. Unwound DNA is more accessible to transcription factors meaning genes in these regions are better expressed. Modifications occur on lysine and arginine residues among others, and include covalent tags such as methylation, acetylation, phosphorylation and hydroxylation. The proteins regulating and utilising the modifications can be split into three categories: readers, writers and erasers (Figure 1.9).

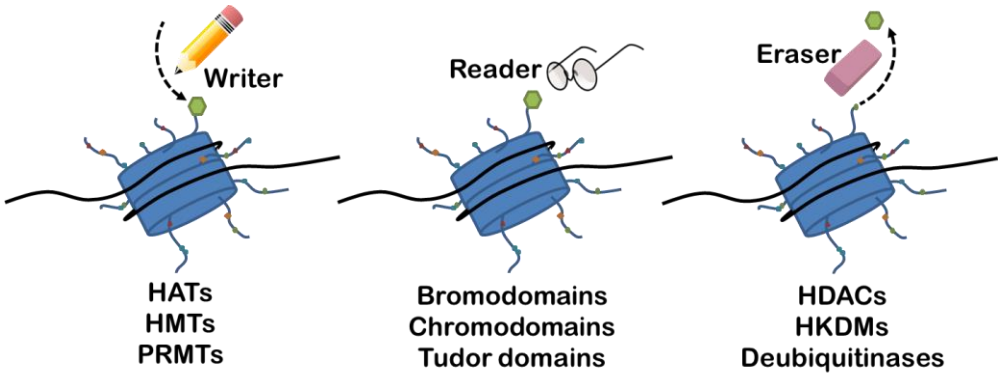


Figure 1.9: The three epigenetic protein classes, with examples, which regulate the post-translational modification of histone tails: writers, readers and erasers. DNA is depicted as a black line. The histone octamer around which DNA is wrapped is depicted as a blue cylinder.

Epigenetic proteins are relevant in the progression of diseases including cancers, metabolic disorders, inflammation, central nervous system diseases and viral infections (Pande 2016). Furthermore, modulation of epigenetic mechanisms has been shown to have a therapeutic effect. A number of protein classes have been shown to be druggable including HDACs, HMTs and bromodomains with small molecule inhibitors in the clinic or in clinical trials. The exploration of epigenetic target space is in its infancy and there appears to be substantial potential for drug discovery (Pande 2016).

This thesis reports hit discovery and hit-to-lead development against epigenetic proteins. The targets originate from two protein families: bromodomains and lysine demethylases.

1.4.1 Bromodomains

Bromodomains are acetyl-lysine (KAc) reader domains associated with the modulation of gene expression (Filippakopoulos, Picaud *et al.* 2012). The druggability of bromodomains have made them attractive targets for the treatment of diseases such as cancer and inflammation, leading to the development of a range of chemical probes for the investigation of bromodomain (Brd) biology (Brand, Measures *et al.* 2015). In 2010, JQ-1 and I-BET were reported as potent inhibitors of the BET bromodomains (subfamily II) (Filippakopoulos, Qi *et al.* 2010; Nicodeme, Jeffrey *et al.* 2010) and subsequent academic and industrial research has focused on targeting the BET bromodomains (Fish, Filippakopoulos *et al.* 2012; Hewings, Fedorov *et al.* 2013; Lucas, Wohlwend *et al.* 2013; Gosmini, Nguyen *et al.* 2014). However, recent publications indicated other subfamilies of the Brd family tree (Figure 1.10) could be targeted by small molecule inhibitors (Ferguson, Fedorov *et al.* 2013; Fedorov, Lingard *et al.* 2014), including CBP/p300 (subfamily III) (Hay, Fedorov *et al.* 2014), BRD7/9 (IV) (Clark, Vieira *et al.* 2015; Theodoulou, Bamborough *et al.* 2016), ATAD2 (IV) (Bamborough, Chung *et al.* 2016), BAZ2A/B (V) (Drouin, McGrath *et al.* 2015; Chen, Chaikuad *et al.* 2016), BRPF1/2/3 (IV) (Bamborough, Barnett *et al.* 2016), CECR2 (I), TAF1 (VII) and SMARCA2/4/PB1 (VIII).[1] A pan-Brd inhibitor, bromosporine, has also been released as a tool for the validation of functional assays and to elucidate further biological roles of Brds.

[1] www.thesgc.org/chemical-probes/epigenetics, accessed 3rd October 2016

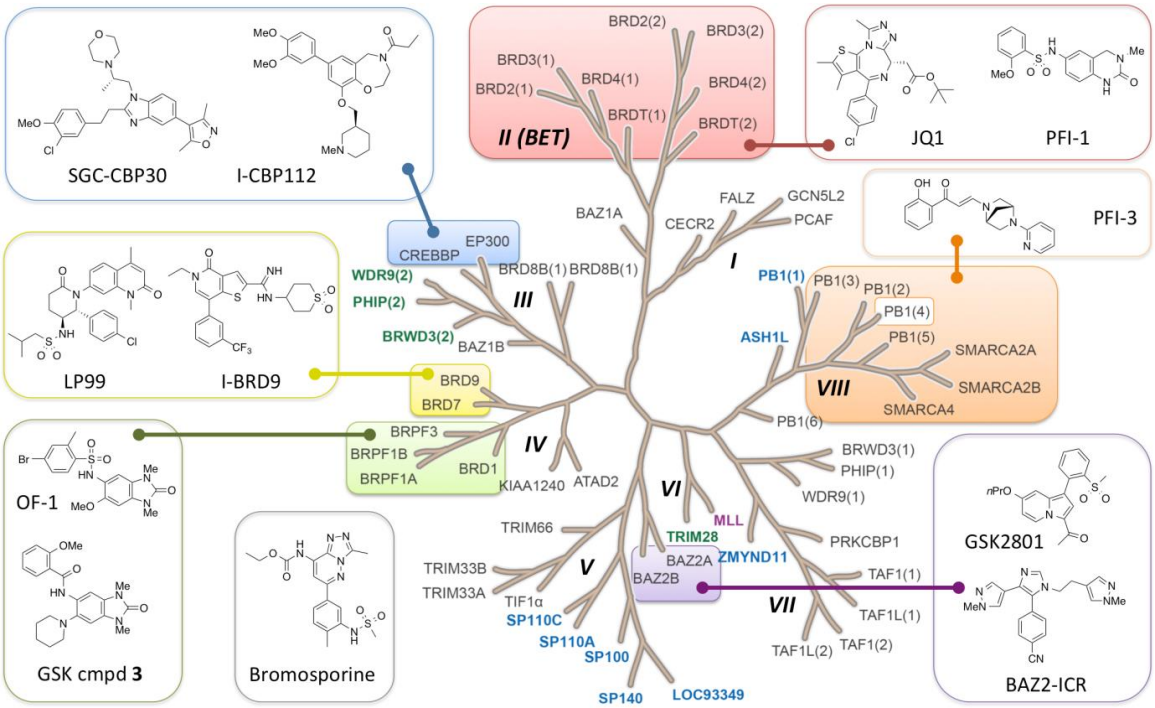


Figure 1.10: Bromodomain family tree with sub-families (I–VIII) as defined by Filippakopoulos et al (2012). Brds for which SGC Chemical Probes have been released before 2016 are highlighted and the structures depicted. Atypical bromodomains are highlighted for which the conserved asparagine residue in the KAc binding site is replaced by tyrosine (blue), threonine (green) or aspartic acid (purple). Figure reprinted from Cox et al. 2016 with permission from the authors.

All known bromodomain structures form an alpha-helical bundle comprised of four alpha helices, A, B, C and Z (Figure 1.11 A) (Filippakopoulos, Picaud *et al.* 2012). The A-Z and B-C loops form the KAc binding site. Most Brd inhibitors are anchored by a highly conserved asparagine (Asn) residue and a network of four water molecules in the KAc binding pocket. Together the Asn and first water in the chain of four donate two H-bonds to a pair of H-bond acceptors on the bound peptide or ligand. The importance of the conserved Asn residue in ligand binding reflects its critical role in KAc recognition of the natural ligand (Figure 1.11 B). However, only forty-eight of the sixty-one known human Brds (79%) have an Asn residue in the KAc binding pocket. The remaining thirteen Brds (21%) have a tyrosine, threonine or aspartic acid in place of the asparagine (Figure 1.11 C). To date, no high potency binders of these atypical, non-asparagine bromodomains have been reported.

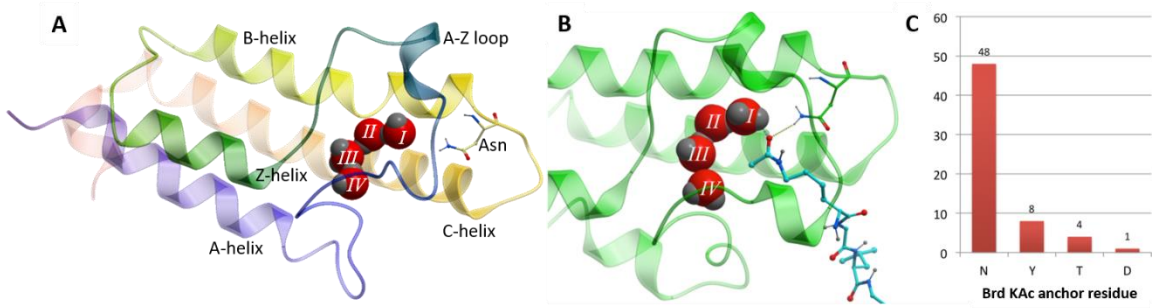


Figure 1.11: (A) The typical fold of a human Brd. BRD1 (rainbow ribbon) is composed of four alpha helices. The A-Z and B-C loops form the KAc binding site, characterised by a conserved Asn residue (yellow sticks) and a network of four waters (red CPK) numbered I-IV (PDB ID 5FG6). (B) Canonical water network and KAc binding in typical bromodomains (Brds). The H4K8AcK12ac peptide (cyan sticks) in complex with Brd4(1) (green ribbon) is anchored by N140 (green sticks) and a network of four waters (red CPK) numbered I-IV. The KAc makes H-bonds to N140 and water I (dotted lines) (PDB ID 3UW9). (C) Prevalence of typical and atypical Brds. The histogram shows the number of Brds with various residue types in the key KAc anchor position across the 61 known human bromodomains. Asparagine (N) is the most widely conserved, but tyrosine (Y), threonine (T) and aspartic acid (D) are also observed. Figure reprinted from Cox et al. 2016 with permission from the authors.

1.4.2 Lysine Demethylases

Lysine demethylases (KDMs) are epigenetic proteins that remove methyl groups from lysine residues on histone tails. The KDM family is split into eight sub-families. The first, KDM1, is structurally distinct from the KDM2-7 sub-families. The latter consists of 22 human proteins containing the Jumonji *C*-terminal domain (JmjC) (Johansson, Tumber *et al.* 2014). JmjCs catalyse the demethylation of mono-, di- and tri- methylated lysines by employing a catalytically active iron atom and 2-oxo-glutarate (2-OG) (Figure 1.12) (Walport, Hopkinson *et al.* 2012). Many proteins in the JmjC family have been linked to cancer and inflammation.

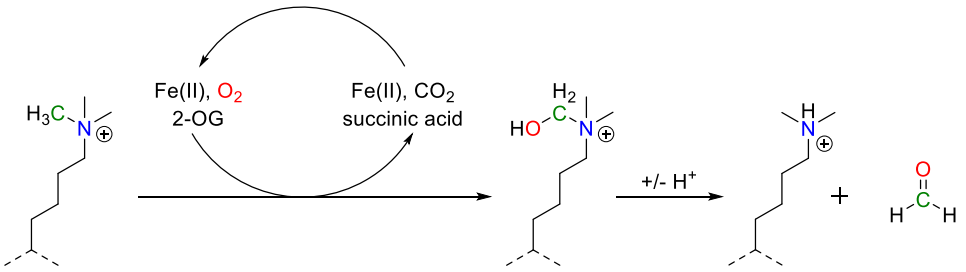


Figure 1.12: JmjC domain containing demethylation mechanism. Catalytically active iron (II) oxidises trimethyl lysine to hydroxymethyl-dimethyl lysine by converting oxygen and 2-oxo-glutarate (2-OG) into carbon dioxide and succinic acid. The final step is loss of formaldehyde to give dimethyl lysine.

The JmjC-containing KDM domains are based upon a Rossmann fold composed of a beta barrel surrounded by alpha helices (Figure 1.13 A). The catalytic binding site contains an iron centre to which 2-OG and two histidine residues are co-ordinated (Figure 1.13 B). The trimethyl lysine substrate chelates to the iron centre by displacement of a water molecule. There is low sequence conservation around the binding site offering the opportunity to develop selective chemical probes (Thinnes, England *et al.* 2014). However, only a handful of small-molecule inhibitors have been published for JmjC KDMs, the majority of which are 2OG-competitive and coordinate to the iron catalytic site (Bavetsias, Lanigan *et al.* 2016). For example, the hydroxamic acid containing compound, methylstat, has been shown to inhibit JmjC-containing KDMs in cellular assays *via* iron chelation (Luo, Liu *et al.* 2011). More recently, the chemical probe GSK-J1 was shown to inhibit the KDM5 and 6 sub-families by iron chelation *via* a 2-pyridylpyrimidine motif and a carboxylic acid (Kruidenier, Chung *et al.* 2012; Heinemann, Nielsen *et al.* 2014). An ethyl ester pro-drug, GSK-J4, is also available for use in cellular assays.

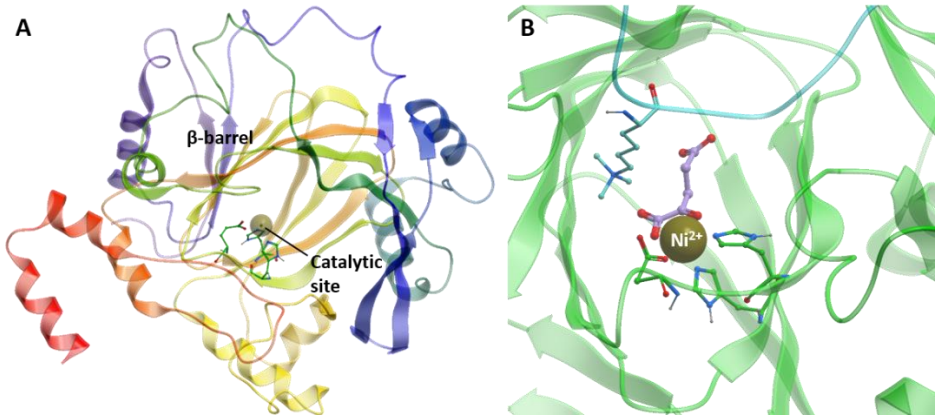


Figure 1.13: (A) The typical fold of a JmjC-containing KDM. JMJD2D (rainbow ribbon) is composed of a β*-barrel surrounded by α-helices. Fe^{2+} occupies the catalytic site and is coordinated by H192, E194 and H280 (green sticks). (B) The JMJD2D trimethyl lysine catalytic binding site. The TARKme3STGGKA peptide (cyan sticks and ribbon) is in complex with the JMJD2D (green ribbon). Ni^{2+} (brown sphere) occupies the Fe^{2+} catalytic site and is coordinated by H192, E194, H280 (green sticks), and 2-OG (mauve sticks). Protein synonym: KDM4D (PDB ID 4HON, Krishnan 2013).*

KDM2-selective compounds have been published, but a carboxylic head group makes them inactive in cell assays (England, Tumber *et al.* 2014). The pan-JmjC chemical probe, IOX1, also

contains a metal chelating carboxylic acid (Hopkinson, Tumber *et al.* 2013). Selectivity for KDM5 has been demonstrated using multiple distinct chemotypes (Johansson, Velupillai *et al.* 2016). However, it appears the requirement for a metal chelating motif, which restricts cellular and *in vivo* activity, is limiting further development of JmjC inhibitors (Pande 2016).

1.5 Fragment Based Drug Discovery

High throughput screening (HTS) revolutionised pharmaceutical hit discovery in the early 1990s (Lipinski, Lombardo *et al.* 2001). However the limitations became apparent when big increases in research and development spending during the 2000's did not yield the expected increase in drug approvals (Paul, Mytelka *et al.* 2010; DiMasi, Grabowski *et al.* 2016). Compound libraries proved to be expensive to maintain, hit discovery was plagued by high false positive rates and attrition rates of lead compounds were high (Cook, Brown *et al.* 2014). Many attempts were made to lower attrition and to spot leads likely to fail earlier (Lipinski, Lombardo *et al.* 2001; Hann 2011). However, lead optimisation using SAR often adds molecular weight, lipophilicity and polar functional groups, making optimisation within drug-like space difficult.

An analysis of chemical space demonstrates the poor coverage of potential drug-like compounds by HTS libraries (Hall, Mortenson *et al.* 2014). The number of possible compounds with MW in the range 400 to 500 is incomprehensively large, with each additional heavy atom increasing the number of potential molecules eight-fold. A HTS library, which usually contain millions of compounds, only covers 10^{-22}% of the estimated 10^{30} possible compounds with MW < 500. In comparison, a library of just 1000 compounds can cover 0.0001% of the estimated 10^{9} compounds with MW < 250. Because of the much better coverage of available chemical space, the use of screening libraries containing smaller, fragment-sized compounds (MW < 300) became a

technique in its own right. Fragment based drug discovery (FBDD) is now a technique used to complement HTS and has had success in the development of new drugs.

1.5.1 History of FBDD

In 1996, Shuker *et al*. demonstrated how small organic molecules can be used to identify different parts of a binding site and the weak-binding ligands subsequently linked to give an inhibitor with nM activity against BCL2 (Shuker, Hajduk *et al.* 1996). This is widely cited as the first example of what went on to become known as FBDD (Figure 1.14). A study by Hann *et al*. in 2001 used computational and statistical models to demonstrate how smaller, more hydrophilic molecules can make better screening compounds as well as better leads (Hann, Leach *et al.* 2001). It was one of the key findings to generate interest in using smaller compounds for hit discovery. By 2003, Congreve *et al.* had proposed a Rule of Three for fragment libraries: MW < 300, clogP < 3, hydrogen bond donors < 3, hydrogen bond acceptors < 3 and number of rotatable bonds < 3 (Congreve, Carr *et al.* 2003).

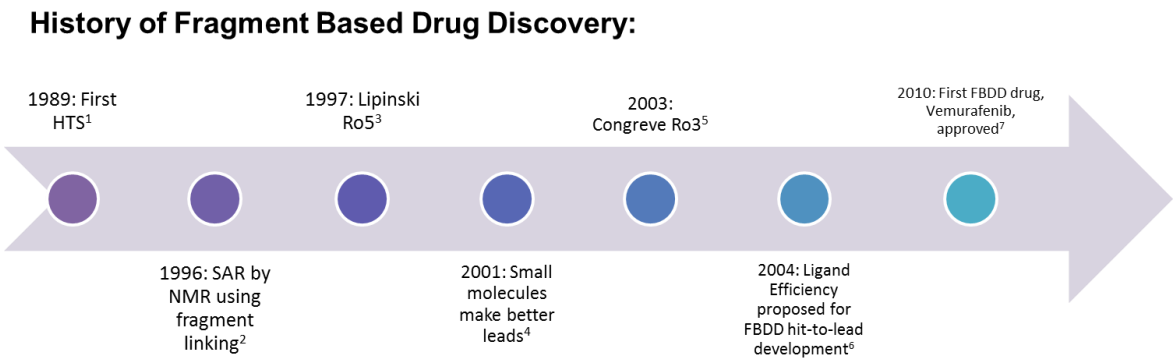


Figure 1.14: History of FBDD. [1]Moos et al. 1993. [2]Shuker et al. 1996. [3]Lipinski et al. 1997. [4]Hann et al. 2001. [5]Congreve et al. 2003. [6]Hopkins et al. 2004. [7]Baker 2013.

The smaller library size required for FBDD meant hit discovery could be performed at a fraction of the cost of HTS, and as a result the technique saw widespread success in academia as well as in industry (Hajduk and Greer 2007). In 2005 a drug developed using FBDD entered Phase II clinical

trials for the first time (LY-517717, Lily/Protherics) and in 2011 the first FBDD drug to reach the clinic was approved (vemurafenib, Plexxikon) (Baker 2013). In 2016, a total of 29 compounds have entered clinical trials, with two having reached Phase III and a further two having been approved by the FDA (Figure 1.15) (Efremov and Erlanson 2016).

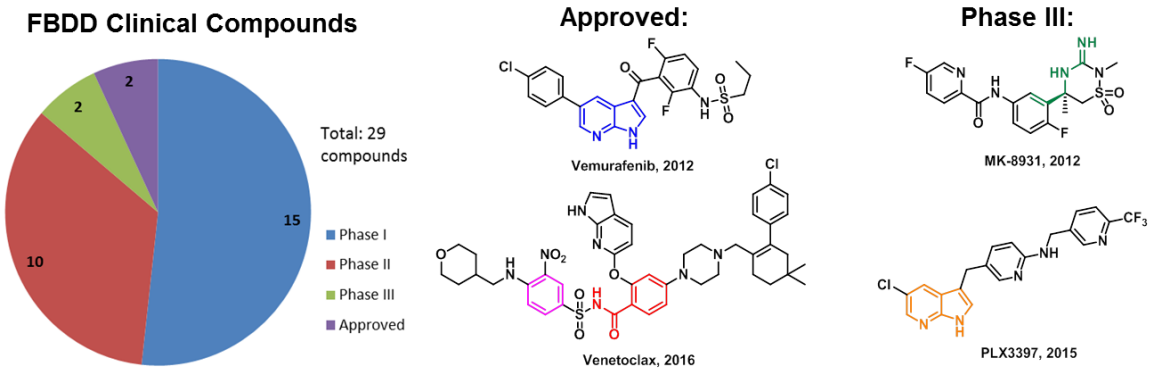


Figure 1.15: Success of clinical compounds developed by FBDD. The structural features present in the original fragments and maintained in the drug are highlighted in colour.

1.5.2 Hit Discovery in FBDD

One of the greatest obstacles in FBDD is detecting the weak interactions between the protein and fragment. A fragment hit is likely to only have one or two high quality protein-ligand interactions. High quality interactions are associated with large enthalpic terms driving the free energy of binding, and include H-bonds, salt bridges, aromatic π-stacking and hydrophobic interactions. As a result, binding affinity is often low and techniques for detecting binding have to be sensitive in the µM to mM scale rather than nM and sub-nM. In 1996, Shuker *et al*. demonstrated the use of NMR to observe interactions between proteins and small compounds, giving significantly greater sensitivity than the screening methods available at the time (Shuker, Hajduk *et al.* 1996). Early fragment screens were limited to NMR (usually ^{19}F or ^{15}N) and X-ray crystallography, both of which give exquisite structural information but limited quantitative information. Both techniques are also difficult to perform in a high-throughput process. More recently, a handful of techniques have become more widely used for fragment screens including thermal shift assays (TSA), mass

spectrometry (MS), surface plasmon resonance (SPR), high-throughput virtual screening, and a wide range of biochemical assays (Table 1.1) (Scott, Coyne *et al.* 2012).

Technique	Maximum Sensitivity	Description	Strengths	Weaknesses
Thermal Shift Assays (TSA)	1 mM	Fluorescence based detection of protein unfolding	High throughput Low cost High sensitivity	False positives False negatives Not all proteins are amenable No binding site information
Mass Spectrometry (MS)	0.1 mM	Soft ESI to detect protein-ligand complexes	High throughput Amenable to cocktailing	Low sensitivity No binding site information
Surface Plasmon Resonance (SPR)	10 mM	Changes in the reflective index of polarised light caused by fragment binding to immobilised protein	Calculations of binding affinity and kinetics Low protein consumption Low false positive rate	Sensitivity limited by compound solubility Medium-throughput No binding site information
Protein-observed NMR	10 mM	Chemical shift of isotope labelled protein	Low false positive and false negative rates High sensitivity Binding site information	Large amount of protein required Limited to small proteins (<50kDa)
Ligand-observed NMR	1 mM	waterLOGSY; STD; ^{19}F T2 relaxation (see main text)	High throughput Applicable to most systems	No binding site information False positives
X-ray Crystallography	100 mM	Co-crystallisation or fragment soaking	High sensitivity Low false positive rate High concentration Binding site information	Low throughput False negatives Requires crystallisable protein
Virtual Screening	N/A	Molecular docking	High throughput Binding site information	Poor correlation with experimental data for fragment-sized compounds

Table 1.1: Hit discovery techniques commonly used in FBDD.

Recent advances in robust assay design has transformed TSA into a high throughput hit detection method with good reproducibility (Erlanson, Fesik *et al.* 2016). The most common form of the technique, differential scanning fluorimetry (DSF), detects protein unfolding (also referred to as melting or denaturing) using a fluorescent dye that binds to the hydrophobic patches of the protein as it unravels. The presence of a ligand binding to the protein stabilises the structure resulting in a shift in the unfolding temperature. Often, the extent of thermal stabilisation correlates well with the ligand affinity. However, when the thermodynamics of protein unfolding dominate the system the technique is less useful, and false positives and false negatives are

common. Furthermore, protein binding does not necessarily correlate with inhibition of the protein's biological role, particularly if binding is well away from the active site.

Native MS has become a complementary hit discovery technique in FBDD (Scott, Coyne *et al.* 2012). Many methodologies exist, but in its simplest form soft electrospray ionisation (ESI) is used to ionise protein-fragment mixtures, and mass changes corresponding to the protein plus fragment detect binding. In reality the sensitivity required to detect non-covalent binding results in low resolution data, and other methods have been developed to detect binding indirectly.

SPR is the most commonly used FBDD hit discovery method (Erlanson, Fesik *et al.* 2016). Detection is achieved by immobilising protein on a sensor chip, usually using biotinylated protein and a streptavidin-coated chip. Changes in mass caused by compound binding are detected by changes to the refractive index of the sensor surface near the protein. Once a protein system has been optimised, the technique requires a very small amount of protein, and one protein chip can be used to screen hundreds of compounds. Dose-response curves can be generated to yield a K_d and the kinetics of binding, although this is dependent upon compound solubility as aggregation can interfere with the assay and yield false positives. Like TSA, the method does not give any indication of active site inhibition, although active-site mutants have been used to try to differentiate between specific and nonspecific ligands.

Protein-observed NMR is considered by many to be the gold standard of FBDD hit detection (Erlanson, Fesik *et al.* 2016). Fragment binding is observed directly by detecting chemical shift changes in an isotope-labelled protein. K_d values can be calculated from dose-response curves, and information on the binding site can be gathered by assigning resonance peaks. However, the method requires large amounts of isotope-labelled protein, which can be expensive and time

consuming to produce. Furthermore, deconvolution becomes increasingly difficult with increasing protein size, and is rarely performed on proteins with molecular weight > 50 kDa.

In comparison, ligand-observed NMR offers a faster method of hit detection against most targets, although with lower sensitivity and no information on the protein binding site or ligand binding mode (Scott, Coyne *et al.* 2012). Water-ligand observed via gradient spectroscopy (waterLOGSY) and saturation transfer difference (STD) uses irradiation pulses at the resonance frequency of bulk water (waterLOGSY) or the target biomolecule (STD), which results in a transfer of magnetization to nearby compounds. Both methods give a clear distinction between binding and non-binding fragments, and higher throughput can be achieved using compound cocktailing. ^{19}F-based NMR experiments offer a more sensitive and robust assay, but requires a fluorine-containing screening library or a known ligand with specific protein interactions.

Until recently, the low throughput nature of X-ray crystallography meant it was rarely used as a primary screening method. More regularly it was (and still is) used as a validation technique, with many organisations requiring a crystal structure before embarking upon medicinal chemistry efforts (Erlanson, Fesik *et al.* 2016). The technique is discussed in more detail in the next section, however it benefits from low false positive rates and high sensitivity. The ability to soak compounds at high concentration allows hits to be detected despite being inactive in traditional solution phase screens. Nevertheless, the technique appears to have a non-trivial false negative rate and requires a robust protein crystallisation protocol which tolerates high concentration fragment soaking. As a result, crystallography is often used as a secondary or tertiary screen by the majority of researchers. Even so, a handful of companies founded upon fragment based hit discovery using X-ray crystallography, such as Astex Pharmaceuticals, have had good success in hit-to-lead-to-clinic development of drug candidates.

Finally, increasing computational power has made virtual screening a viable hit discovery technique. Molecular docking can be used to triage large screening libraries to enrich hit rates during experimental screening efforts. However, the utility of docking and scoring algorithms is currently limited to large ligands in the MW range 300-600 Da that are able to make multiple protein-ligand interactions. As such, the technique is better suited to HTS campaigns or to later lead optimisation efforts (Erlanson, Fesik *et al.* 2016).

There are inherent risks when screening a fragment library at high concentration in a search for weak interactions. The varying false positives and false negative rates of each technique mean many researchers use two or more assays in order to validate hits. However, the overlap between screening methods has been shown to vary significantly and could result in promising hits being overlooked (Wielens, Headey *et al.* 2013; Kutchukian, Wassermann *et al.* 2015; Schiebel, Radeva *et al.* 2015). It has been suggested a more robust protocol would be to use one technique known to work well for the target class on multiple compound series until affinity has improved sufficiently to allow detection using orthogonal assays (Erlanson, Fesik *et al.* 2016).

1.5.3 Hit-to-Lead Optimisation in FBDD

Once a fragment, or fragments, has been detected, the hit must be optimised into a lead compound. Three main methodologies have been described: fragment growing, fragment linking and fragment merging (Figure 1.16). The method best associated with FBDD is fragment linking, and was demonstrated using SAR by NMR in 1996 in the discovery of BCL2 inhibitors (Shuker, Hajduk *et al.* 1996). In theory, linking fragments which bind in neighbouring binding sites gives an affinity boost greater than the sum of the initial fragments. In practice, fragment linking is extremely challenging. Linker length and geometry is hard to optimise and can have a detrimental effect of compound properties such as MW and cLogP.

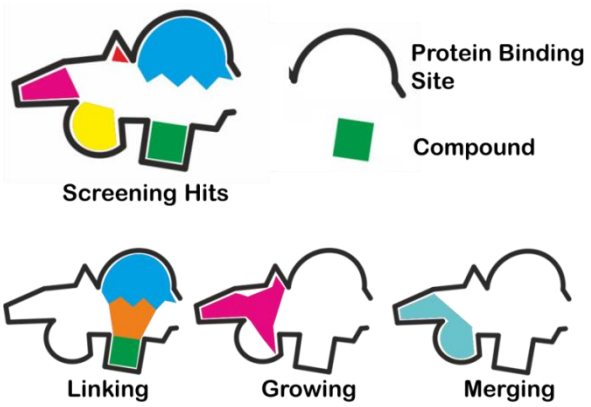


Figure 1.16: Hit-to-lead optimisation in FBDD: fragment linking, growing and merging. All three methods have been successfully used in drug development. Fragment linking is the textbook example of FBDD. In reality, most campaigns utilise a combination of fragment growing and merging.

More commonly, a novel binding site is identified by fragment screening, and the growth of a neighbouring fragment using structure based drug design (SBDD, see next section) allows exploitation of the new site. The technique often involves fragment merging in order to capture the best interactions from different fragments, and in practice most campaigns end up mixing fragment merging, linking and growing during hit-to-lead development.

FBDD success stories published to date use robust fragment binding models to aid hit-to-lead optimisation (Erlanson, Fesik *et al.* 2016). Follow-up chemistry plays a crucial role in early optimisation to answer specific questions about binding modes, key interactions and potential growth vectors. Synthesis can be a bottleneck at this step, and the synthetic tractability of hits play a critical role in the progress of a project (Murray and Rees 2016).

1.6 X-ray Crystallography and Structure Based Drug Design

Structural information can be a significant boost in hit-to-lead development and lead optimisation by rationalising binding affinities and SAR data. Three-dimensional models of protein-ligand complexes can be obtained using NMR spectroscopy, electron microscopy or X-ray crystallography. The Protein Data Bank (PDB) is the recognised repository for all biomolecular

structures, and currently contains >100,000 openly available models.[1] X-ray crystallography is the most frequently used method, with more than ten times the number of depositions in the PDB than NMR, the next most frequently used method.

Protein X-ray crystallography uses the diffraction patterns generated when crystallised protein is subjected to irradiation from a beam of X-rays. Diffraction patterns are formed as a result of photons interacting with the electrons found in each atom of a protein and are observed using a detector placed behind the sample. The diffraction pattern is a Fourier transform of the biomolecule's electron density. The position of the atoms dictates the phase of the X-rays, whereas the size of the atoms determines the amplitude. Detectors can only gather information of X-ray amplitude, so the phase information has to be estimated using multi-wavelength anomalous dispersion (MAD) or single-wavelength anomalous dispersion (SAD). The final model is determined using molecular replacement in which an estimate of the atom positions is made and improved iteratively until there is satisfactory agreement with the experimental data (Abergel 2013). The resolution of the resulting 3D-model is determined by the quality of the dataset, and is reported in angstroms. In general, a resolution of 2.5 Å or less is required to accurately interpret the binding mode of a ligand. However, if the ligand has low occupancy (<80-90%) it can be hard to unambiguously model the binding mode.

The other parameters used to determine the quality of a model include the R-factor, the free R-factor and B-factors. The R-factor measures the agreement between the model and the dataset. An R-factor within ten-fold of the resolution is generally considered acceptable (i.e. a 2.5 Å dataset with an R-factor ~20%). The free R-factor was introduced to reduce R-factor overfitting, and is an R-factor taken from a small subset (5%) of the unmodelled diffraction data (Brunger 1997). A B-

[1] www.rcsb.org, accessed 3rd October 2016

factor value is given to each atom to indicate the amount by which the atom deviates from its mean coordinates within each unit cell or across the crystal lattice. Higher B-factors imply less ordered parts of a protein or ligand and the values are used to identify disordered loops or flexible sidechains.

Building high-resolution datasets with high occupancy ligands (>90%) is a time-consuming process. A robust crystallisation protocol for the protein needs to be developed and optimised for co-crystallisation or soaking experiments. Co-crystallisation involves adding the ligand to the crystallisation buffer before crystallisation. The presence of the ligand may alter the rate of crystallisation or the conditions required, so may need further optimisation. Soaking ligands by adding compound solutions to the crystallisation plates once the protein crystals have been grown is a more robust method for screening compound libraries assuming the binding pocket is solvent exposed (i.e. not involved in crystal contacts) in the crystal lattice. However, addition of a ligand at high concentration can result in crystal death if the compound induces a change in protein conformation or disrupts crystal contacts. Collecting datasets is also a labour-intensive and time consuming process, limiting the number of compounds which can be screened. Until recently, crystallography was used for hit validation, but recent advances in crystallisation protocols and beamline data collection has allowed the technique to become a viable primary assay in hit discovery (Chayen 2009; Winter and McAuley 2011; Badger 2012; Hall, Mortenson *et al.* 2014; Patel, Bauman *et al.* 2014).

Once a 3D-structure of a protein-ligand complex has been obtained, hit-to-lead development can be significantly accelerated using Structure Based Drug Design (SBDD). In several cases SBDD has resulted in fast-tracked drug discovery. For example, vemurafenib was approved by the FDA just six years after the project was initiated (Erlanson, Fesik *et al.* 2016). SBDD can be split into two

steps: the identification of novel hits using an experimentally determined protein structure alongside virtual screening or the intuition of a medicinal chemist, and the iterative modification of a hit using *de novo* design tools to guide development (Anderson 2003). I will focus on the use of the latter in the development of hits obtained by X-ray crystallography, as depicted in Figure 1.17.

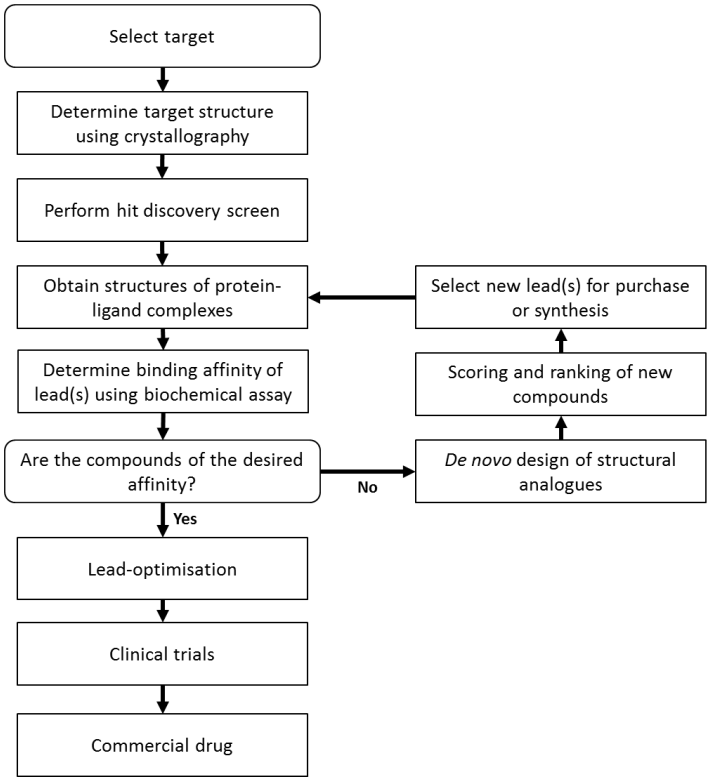


Figure 1.17: The iterative process of SBDD. Adapted from Anderson 2003.

Once a structure of a protein-ligand complex has been obtained, the data can be used to design new compounds. First, a medicinal chemist or a computational algorithm is used to design compounds *de novo*. A second algorithm then scores the compounds based upon structural and/or affinity data. Finally, compounds are selected for purchase or synthesis to be tested in experimental assays. Often *de novo* design generates synthetically challenging, unstable or insoluble compounds meaning the final step has to be performed manually by a chemist. The final selection process has to be pragmatic, but can also be subjective based upon the chemist's own

experiences within medicinal chemistry (Hartenfeller, Zettl *et al.* 2012). Algorithms have been built to design screening compounds using known, robust chemical reactions such as DOGS and SCUBIDOO (Hartenfeller, Zettl *et al.* 2012; Chevillard and Kolb 2015). Other algorithms, such as SPROUT and OOMMPPAA, circumvent the problem by mining databases to suggest small structural changes to a lead compound in order to optimise binding affinity (Gillet, Johnson *et al.* 1993; Bradley, Wall *et al.* 2014).

1.7 Medicinal Chemistry

The iterative process used in hit-to-lead development is driven by medicinal chemistry. Medicinal chemistry brings together expertise from synthetic organic chemistry, computational chemistry, chemical biology and structural biology to invent novel therapeutic molecules. As a corollary, the field can contribute to research in pharmacology and other areas of biology.

The development of a novel drug takes years of design cycles and the synthesis of hundreds of unique compounds. Medicinal chemists have established a number of tools and techniques to keep timelines and expenditure to a minimum. Chemical biology is used to select a target and predict how best to produce a therapeutic effect. Computational chemistry and structural biology can be used to prioritise compounds for synthesis. Parallel synthesis enables a chemist to synthesise multiple compounds in a single experiment. A series of structural analogues in which one part of the molecule is systematically varied enables better SAR understanding. Whenever possible a synthetic route is designed in which the scaffold of the lead compound is built and a reactive handle used in the final step to perform parallel synthesis (Figure 1.18). This late-stage derivatisation increases the success rate of synthesis, but in practice can be highly inefficient, requiring multiple protection and deprotection steps and often requiring significant optimisation by process chemists when scaled up.

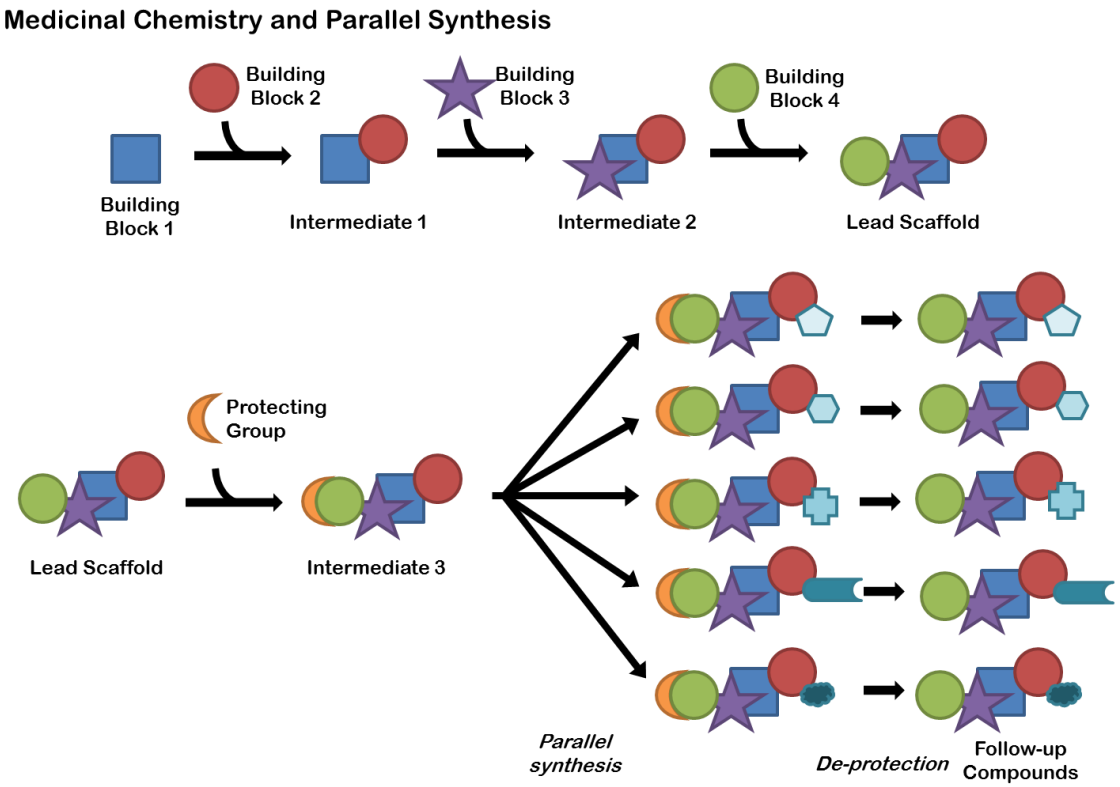


Figure 1.18: Typical synthetic route employed by medicinal chemists. A scaffold is built up using commercially available building blocks before a reactive handle is used to react with a library of building blocks using parallel synthesis. Often a protecting group is required to ensure the regioselectivity of reactions.

In order to maximise success rates, the reactions commonly used in parallel synthesis are well-established, robust reactions amenable to a range of substrates. A recent analysis of reactions used in medicinal chemistry revealed how reliant the field is upon a handful of reactions, such as amide couplings, Suzuki-type cross-couplings and *N*-alkylations, which are suitable for parallelisation, as well as the extensive use of protection and deprotection reactions (Roughley and Jordan 2011).

The work presented in this thesis utilises parallel synthesis to synthesise libraries of amides and thioureas. Amide bonds are widespread in biological and synthetic systems. The bond links amino acid building blocks together into peptide chains, which in turn fold to give proteins. Furthermore, the amide bond can be found in a wide range of synthetic molecules including many commercial drugs such as paracetamol, Lipitor and Tamiflu. Formation of an amide bond is most frequently

the result of condensation between a carboxylic acid and an amine. However to perform the reaction directly requires high temperatures (>200 °C) (Valeur and Bradley 2009). To avoid such harsh reaction conditions, the –OH of the carboxylic acid is converted to a good leaving group before subsequent treatment with the amine (Figure 1.19).

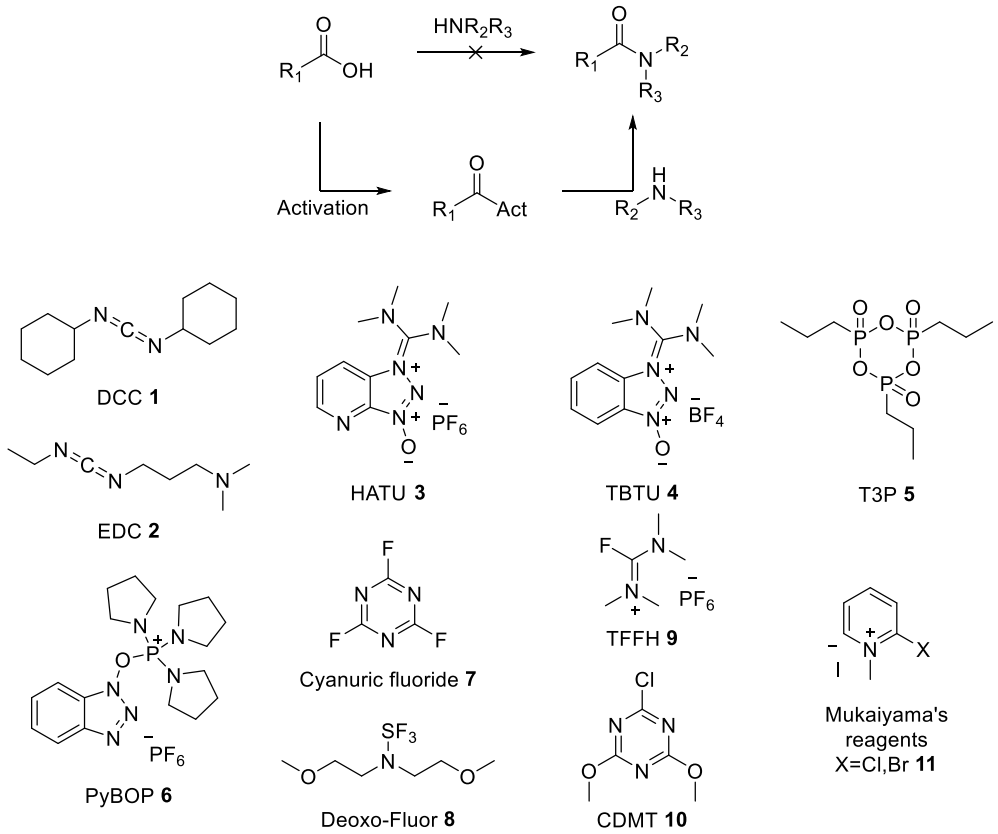


Figure 1.19: Amide coupling schematic and selected coupling reagents. The condensation of a carboxylic acid and an amine requires high temperatures. In medicinal chemistry, the acid is activated using coupling reagents, such as the examples shown.

A vast array of reagents designed to activate carboxylic acids, called coupling reagents, have been published or are sold by vendors (Figure 1.19). Choosing the most appropriate coupling reagent is critical, particularly when performing parallel synthesis on a library of compounds with varying reactivity (e.g. anilines, secondary amines, bulky substrates). Common reagents include carbodiimides (e.g. DCC **1**, EDC **2**), 1*H*-benzotriazoles and derivatives (e.g. HATU **3**, TBTU **4**), phosphorus containing reagents (e.g. T3P **5**, PyBOP **6**) and mixed anhydrides (Valeur and Bradley

2009). Polymer-supported versions of the same reagents are designed to aid purification (Convers, Tye *et al.* 2004; Valeur and Bradley 2007). In situ generation of acid halides is also possible using fluoride sources (e.g. cyanuric fluoride **7**, Deoxo-Fluor **8**, THHF **9**) or chloride sources (e.g. CDMT **10**, Mukaiyama's reagents **11**). More commonly, the acid chloride can be purchased and reacted directly. The mild reaction conditions and the high levels of specific reactivity of acid chlorides make them ideal for amide bond formation and parallel synthesis, but they have limited commercial availability and are not amenable to multi-step synthesis.

Thioureas are found in fungicides, herbicides and rodenticides, and have been shown to inhibit biological enzymes (Maddani and Prabhu 2010). The primary methods for thiourea synthesis involve condensation of primary and secondary amines with isothiocyanates, thiophosgene or carbon disulfide in the presence of mercury acetate (Figure 1.20).

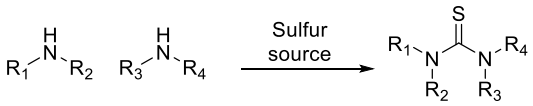


Figure 1.20: Synthesis of thioureas from primary and secondary amines using isothiocyanates, thiophosgene or carbon disulfide as a source of sulfur.

Isothiocyanates offer the simplest method of thiourea synthesis, but the limited availability of isothiocyanates from vendors means they often need preparing from the amine. The sequential addition of two amines to thiophosgene is more amenable to parallel synthesis. The toxicity of thiophosgene has led to the creation of novel reagents, such as thiocarbonyldiimidazole (TCDI) which offers a safe and efficient alternative method (Venkatachalam and Uckun 2007). Di-substituted thioureas can be synthesised by reacting TCDI with an amine at low temperature, followed by addition of a second amine at room temperature after 30-60 minutes.

1.8 Project Aims

Recent developments at beamline I04-1 at Diamond Light Source (DLS) around soaking, harvesting and data analysis has enabled a 1000-compound fragment screen by crystallography to be completed within a week. As a consequence, X-ray crystallography is now a viable primary screen for academic and industrial researchers. In particular, this makes it realistic to screen by soaking single compounds per crystal, rather than cocktails, so that concentrations >100 mM can be achieved, one hundred to one thousand times higher than possible in solution-based techniques, with accordingly increased sensitivity and hit rates. The very weak hits identified by X-ray screening pose a practical problem, in that they are in general too weak to be verified by orthogonal solution assays; yet such information is a common requirement in medicinal chemistry operations, especially if the crystallographic evidence is ambiguous. Optimising weak compounds by traditional SAR is equally undermined without reliable assay readout in the range of the compounds' binding affinity. In any case, progressing weak compounds to potency in general requires changes not only to the periphery, where conventional SAR operates, but also to the core of the compound (Erlanson, Fesik *et al.* 2016).

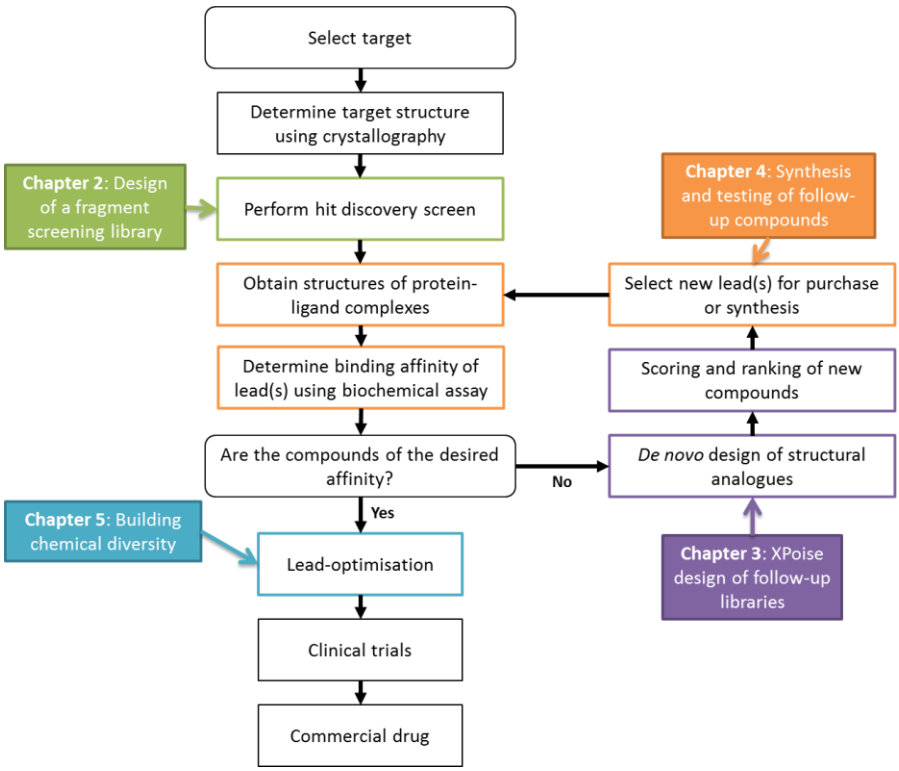


Figure 1.21: Project overview within the context of SBDD.

To overcome the problem of optimising weak compounds discovered by high concentration fragment soaking, I report the design and use of a poised fragment screening library to identify hits (Chapter 2) and a computational tool for the design of follow-up compounds using the poised paradigm (Chapter 3) (Figure 1.21). I demonstrate the utility of the method with the parallel synthesis of analogues against a range of epigenetic targets to deliver inhibitors (Chapter 4). Finally, I explore how modern synthetic chemistry methods can be used to complement poised drug discovery (Chapter 5).

2 Design of a Poised Fragment Library

The design of the fragment library described in this chapter has been published by Cox *et al*. in Chemical Science (Cox, Krojer *et al.* 2016). The Diamond-SGC Poised Library (DSPL) is in use at DLS and SGC Oxford and details of the library are available online.[1]

2.1 Overview

In this chapter I designed a fragment library using the principles of FBDD and poised chemistry. First, a list of poised functional groups was selected to ensure the synthetic tractability of fragments. Next, the ZINC Fragment-Like catalogue was downloaded and was analysed for poised functional groups. The commercial availability of each fragment as well as the starting materials required to synthesise the fragment was checked and a diverse subset of 1,000 compounds was selected for purchase using the USRCAT chemical fingerprint algorithm. Finally, 776 compounds were purchased and assembled to make the DSPL.

2.2 Background

2.2.1 Principles of Fragment Library Design

In 2003, Congreve analysed fragment-like compounds and proposed a Rule of Three (Ro3) for FBDD (Congreve, Carr *et al.* 2003). His guidelines of molecular weight ≤300 Da, cLogP ≤3, number of hydrogen bond donors ≤3 and number of hydrogen bond acceptors ≤3 have been widely used in the FBDD community ever since (Keseru, Erlanson *et al.* 2016). However many other physicochemical properties are often considered when designing fragment libraries including

[1] www.diamond.ac.uk/Beamlines/Mx/Fragment-Screening/Resources.html, accessed 3rd October 2016

heavy atom count, rotatable bond count, total polar surface area, number of chiral centres, and solubility.

Beyond chemical properties, a number of other factors influence the design of a fragment library. These factors can be broadly broken into three categories: chemical diversity, the demands of a specific experimental technique, and synthetic tractability (Figure 2.1).

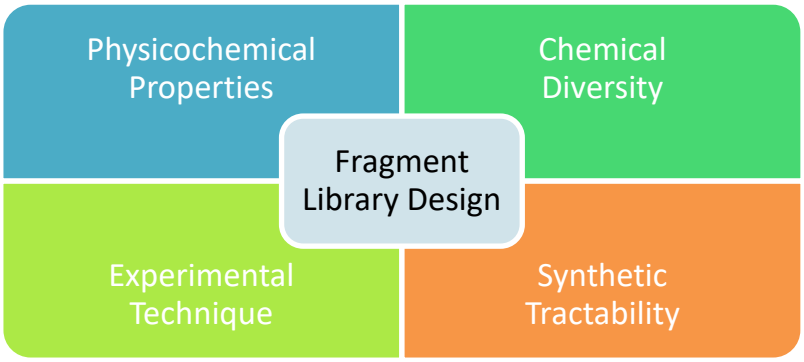


Figure 2.1: The principles of Fragment Library Design. A fragment library is designed by balancing four requirements: physicochemical properties, such as the Ro3; sufficient coverage of chemical space to bolster hit rates; the requirements of the experimental technique with which the library will be screened; and the ease with which the compounds can be investigated using chemical synthesis after being identified as a hit.

High coverage of chemical space increases the hit rate of a screening library. A number of diversity and clustering algorithms have been developed to aid library design. Some libraries utilise complexity (such as 3D-scaffolds) to increase chemical space coverage (Morley, Pugliese *et al.* 2013), while others use pharmacophore diversity from a library of small, simple starting points (Murray and Rees 2016).

Some experimental techniques restrict which compounds can be used in screening campaigns. For example, NMR-screening is often performed using ^{19}F-containing compounds (Vulpetti and Dalvit 2013) and the sensitivity of X-ray crystallography screens have been improved by using bromine-containing compounds (Koester, Craan *et al.* 2011).

To validate low potency fragment hits, screening of structural analogues is required. A synthetically tractable hit ensures follow-up compounds can be synthesised without the need for

time consuming and expensive chemistry efforts. Reactive handles, such as amines and carboxylic acids are often utilised, allowing fragments to be elaborated using robust and reliable chemistry.

2.2.2 Literature Review of Fragment Libraries

Designing and building a fragment library can be a time-consuming and expensive process. Academic groups often utilise commercial fragment libraries if they do not have the experience or resources to build bespoke in-house libraries (Wielens, Headey *et al.* 2013; Blaazer, Orrling *et al.* 2015). A number of vendors sell fragment libraries containing pre-selected compounds. The size and composition of the libraries vary between companies and the method used to design the library is often kept secret. However, most libraries claim to promote a particular property within its design, such as solubility or chemical diversity. A selection of commercial libraries are summarised in Table 2.1, highlighting the range of design principles and library sizes used by vendors.

The majority of bespoke fragment libraries reported in the literature are based upon Congreve's Ro3. Variations in molecular weight limits (e.g. ≤250 Da and/or ≥100 Da), cLogP filters (≤2 is sometimes used to improve solubility) or additional property filters (e.g. number of rotatable bonds, polar surface area or aqueous solubility) are common. Some researchers have found libraries containing Ro3 violations improve hit rates (Koester, Craan *et al.* 2011) while others have had success using stringent filters (Riccardi Sirtori, Caronni *et al.* 2015). Fundamentally it is important to bear in mind that Congreve's Ro3 were designed as guidelines, not as a set of hard and fast rules.

Vendor	Website	Library Description	No. of Compounds
ACB Blocks	www.acbblocks.com	^{19}F NMR, solubility	1280
Analyticon	www.ac-discovery.com	Fragments from nature	213
ChemDiv	www.chemdiv.com	3D designed library	4283
Enamine	www.enamine.net	Golden library (diversity)	1794
IOTA	www.iotapharma.com	Diversity	1500
Integrex	www.integrexresearch.com	Shape and chemical structure diversity	1500
Key Organics	www.keyorganics.net	Aqueous solubility	1166
		From nature, high fsp^3	183
		CNS fragment library	700
		Fluorine	1950
		Bromine	1656
Maybridge	www.maybridge.com	Diversity, DMSO solubility	2500
Otava	www.otavachemical.com	Fluorine	1217
		Metal chelators	1023
		Halogen-enriched	618
Prestwick Chemical	www.prestwickchemicals.com	Drug fragments	910
Selcia	www.selcia.com	Solubility, purity	1400
TimTec	www.timtec.net	Structural diversity, solubility	3200
Zenobia	www.zenobiafragments.com	Multiple design paradigms	968

Table 2.1: Selected list of commercially available fragment libraries. A range of design principles and library sizes are employed by vendors. Adapted from Keseru et al. 2016.

In 2013, Wilde and Link noted the scarcity of heteroatomic functionalities in fragment libraries, particularly heavy elements such as bromine and sulfur, and suggested using a heavy atom count instead of a molecular weight filter (Wilde and Link 2013). A number of researchers have used experimentally calculated solubility (logS) or computationally predicted solubility (clogS) in library design, perhaps reflecting why many commercial fragment libraries are sold with solubility guarantees (Lepre 2001; Na and Hu 2011; Schulz, Landstroem *et al.* 2011; Chen, Zhou *et al.* 2015). However, as discussed in the previous chapter, the importance of compound solubility appears to be technique dependent (Jhoti, Williams *et al.* 2013).

Libraries designed for a specific experimental technique are common. Ligand-based ^{19}F-NMR is a highly sensitive fragment screening technique which detects changes in fluorine spin relaxation upon ligand-protein binding, and bespoke fluorine containing fragment libraries have been designed and successfully screened against various targets (Vulpetti and Dalvit 2013; Garavis,

Lopez-Mendez *et al.* 2014). Libraries designed to improve hit rates for X-ray crystallography enrich for electron rich elements such as bromine and sulfur to improve electron density (Burley, Hirst *et al.* 2010; Koester, Craan *et al.* 2011). Fragment cocktailing, used to improve the throughput of X-ray crystallography, can be more effective if compounds are pooled based upon shape diversity (Chilingaryan, Yin *et al.* 2012; Joseph-McCarthy, Campbell *et al.* 2014).

Novel fragment screening techniques often require bespoke libraries, such as DNA-encoded small molecule screens (Daguer, Zambaldo *et al.* 2015; Wichert, Krall *et al.* 2015). The technique uses fragments covalently linked to DNA or RNA tag which becomes enriched if the compound interacts with a target protein and amplified and detected using PCR.

With increasing computational power, virtual screening has become a popular primary fragment screening method (Joseph-McCarthy, Campbell *et al.* 2014). Large virtual fragment libraries have been developed which try to encompass commercially available fragment space (Irwin and Shoichet 2005) and synthetically accessible fragment space (Blum and Reymond 2009; Chevillard and Kolb 2015).

The optimisation of a fragment hit is a key step in FBDD. Purchase of compounds within a certain similarity cut off provides a fast and low cost method to validate hits. Libraries designed specifically for the technique, known as SAR by catalogue, use compounds which have a minimum number of structural neighbours available from vendors (Schulz, Landstroem *et al.* 2011; Doak, Morton *et al.* 2013). However, commercial libraries are of limited value for fragment growth. Synthetic handles can be used to ensure a compound can be elaborated using standard chemistry procedures. Some researchers select synthesis-ready compounds by eye (Lau, Withka *et al.* 2011) while others have used computational tools to select desired functional groups (Baurin, Aboul-Ela *et al.* 2004). In 2011, Na and Hu published a list of useful reactive handles which included esters,

amines, amides and aromatic halides (Na and Hu 2011). However many popular reactive handles form key interactions with the protein due to their polarity and ability to form hydrogen bonds which limits a chemists ability to design follow-up compounds (Keseru, Erlanson *et al.* 2016). Astex have overcome the problem by ensuring each library member has at least three reactive handles at different points around the compound to increase the chances of successful fragment elaboration (Burley, Hirst *et al.* 2010).

2.2.3 Chemical Diversity Metrics

Chemical diversity is difficult to define within the context of screening libraries. High chemical diversity within a library should result in higher hit rates against a wide range of protein classes. Astex use pharmacophores to measure chemical diversity (Murray and Rees 2016) while others have used 3D-characteristics such as fraction of sp^3 centres (fsp^3) or principal moments of inertia (Hung, Ramek *et al.* 2011; Morley, Pugliese *et al.* 2013). Chemically enriched libraries using deconstructed natural products (Over, Wetzel *et al.* 2013), deconstructed drug-candidates (Chen, Zhou *et al.* 2015) or so-called privileged structures (Chen and Hubbard 2009; Kutchukian, Wassermann *et al.* 2015) have also been developed with the aim of improving hit rates. For well-studied proteins or protein-classes a focussed library containing structural motifs known to commonly bind to the protein-class can be utilised. Focussed libraries have been designed and successfully screened against metallo-enzymes (zinc and iron chelators), kinases (ATP mimetics) and bromodomains (KAc mimetics) (Agrawal, Johnson *et al.* 2010; Bamborough, Brown *et al.* 2011; Chung, Dean *et al.* 2012).

A number of diversity metrics exist for selecting either similar or diverse compounds but there is no consensus on the “best”. Common measures involve generating compound fingerprints and taking the Tanimoto distance between them to identify similar structures. For example, Molecular

Access System Structural (MACCS) keys use chemical substructures defined by an arbitrary list of SMARTS patterns to build a bit vector (Durant, Leland *et al.* 2002). If the small molecule contains the substructure the bit will be "on", and "off" if it does not. The final bit vector, which can vary in length and in the substructures used, becomes the fingerprint.

Circular fingerprints, such as Morgan fingerprints, also generate bit vectors of substructures (Morgan 1965; Rogers and Hahn 2010). However, rather than using pre-defined substructures each atom is described using an integer based upon its extended connectivity (Figure 2.2). The integers are generated based upon each atom's neighbours (element, heavy atom neighbours, number of hydrogens) and any chemical features (donor, acceptor, aromatic, halogen, base, acid). The number of neighbours considered is user-defined and is usually in the range zero to three.

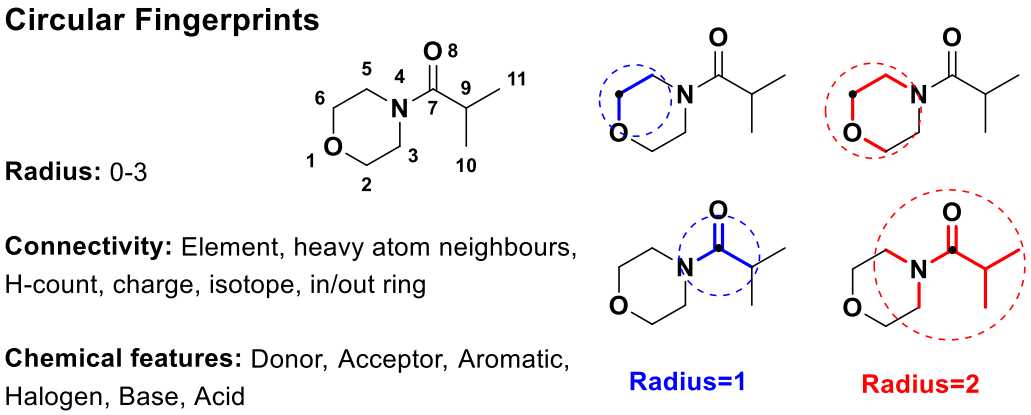


Figure 2.2: Circular fingerprint schematic. An integer is calculated for each atom in the compound. The value is based upon the user-defined radius, the atom connectivity, and its chemical features.

2.2.3.1 3D-similarity descriptors

Measuring similarity using 2D-fingerprints is often unhelpful in the context of small molecule drug discovery. Small changes which result in similar substructures, such as moving a bromide group from the 2-position to the 5-position of an aromatic ring, can have a significant impact upon a compound's size, shape and electrostatic potential. Fingerprints describing a molecule by linking 3D-shape with electrostatic properties better describe the interactions relevant to ligand-protein

binding. The ultra-fast non-superimpositional algorithms, USR (Ultrafast Shape Recognition), CSR (Chiral Shape Recognition) and ElectroShape, developed by Graham Richards and co-workers are examples of shape-based fingerprint generation (Armstrong, Morris *et al.* 2009; Armstrong, Morris *et al.* 2010; Armstrong, Finn *et al.* 2011). The methods also describe each compound using a moment-based fingerprint meaning they are computationally less expensive than alignment-based 3D similarity descriptors.

The methods developed by Richards and co-workers do not consider atom types. As a result, they cannot discriminate between similarly shaped molecules with different and distinct pharmacophores. For example, an aromatic ring substituted with a methoxy group may have a very similar shape to an equivalent ring substituted with a methylene alcohol at the same position, but the two ring systems will interact very differently with a protein. To overcome the problem, Schreyer and Blundell developed an extension called USRCAT (Ultrafast Shape Recognition with CREDO Atom Types) which combines USR with CREDO atom types (Schreyer and Blundell 2012). CREDO[1] is a database of protein-ligand interactions observed in experimentally-determined structures deposited in the Protein Data Bank (PDB) (Schreyer and Blundell 2009; Schreyer and Blundell 2013). In total 13 interactions types are described including hydrogen bonds, halogen bonds and carbonyl interactions.

The USRCAT algorithm uses a 12 element vector as calculated in the USR algorithm (Figure 2.3) (Armstrong, Finn *et al.* 2011). The vector is created by finding four centroids of the compound's 3D structure: the geometric centre, the closest atom to the first centroid, the furthest atom from the second centroid, and the furthest atom from the third centroid. A distribution of distances from each centroid to every other atom in the molecule is created and the mean, variance and

[1] marid.bioc.cam.ac.uk/credo, accessed 3rd October 2016

skewness of each distribution are calculated. The result is 12 descriptors: four centroids by three distribution moments. The distributions are then recalculated for four pharmacophore subsets using the original four centroids: hydrophobic, aromatic, hydrogen bond donor and hydrogen bond acceptor. The subsets are identified using the SMARTS patterns described in the CREDO database. The result is a 60 element vector (5x12).

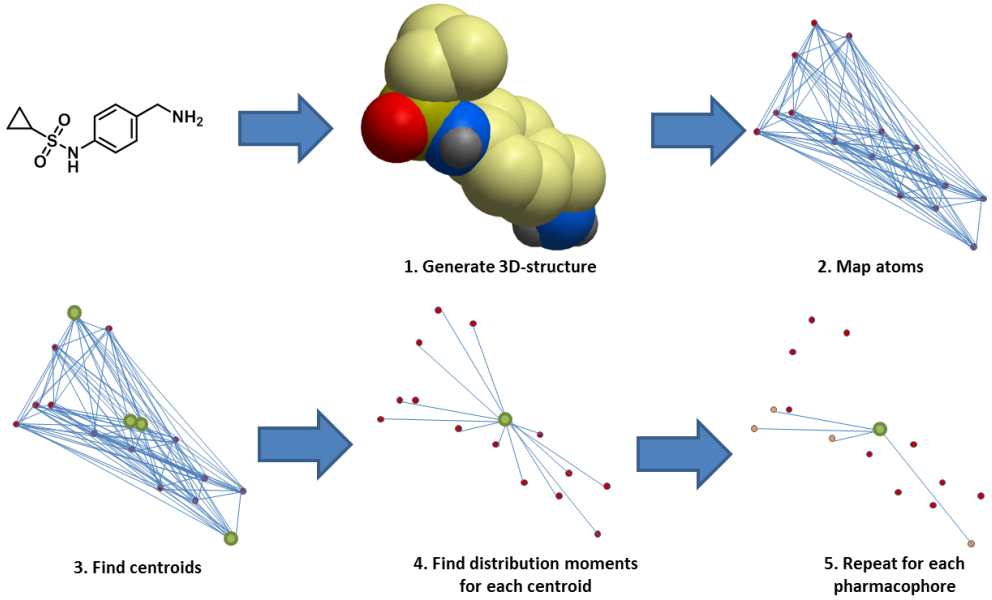


Figure 2.3: USRCAT fingerprint generation. A compound's structure is described using a distribution of vectors from four centroids to all other atoms in the molecule to give a 12 element vector. The process is repeated for each of four pharmacophore features: hydrogen bond donor, acceptor, aromatic and hydrophobic, giving a final 60 element vector (5x12) which becomes the compound fingerprint.

Once a vector has been generated for a library of compounds a suitable similarity metric must be used. Bit vectors commonly use the parameter-free Tanimoto method, although other methods are likely to be just as effective (Todeschini, Consonni *et al.* 2012). USRCAT fingerprints are compared using a variant of Manhattan distance in which each of the pharmacophore features can be weighted using pre-defined coefficients (Schreyer and Blundell 2012).

2.2.4 Project Aim

The XChem beamline facility at DLS (see Chapter 4, section 4.2.1) is able to perform high-throughput X-ray crystallography (>500 datasets per day).[1] My aim was to design a fragment library to be employed by DLS users in high concentration, single-compound soaking experiments against a wide range of target proteins.

Using the principles of fragment library design described in the previous section, I drew up a framework for the design of the Diamond-SGC Poised fragment library, DSPL (Figure 2.4). The library needed to be Ro3 compliant and to contain drug-like small molecules. Reactive and promiscuous compounds were to be avoided if possible. The size of the library was chosen to be 1000 compounds. The number reflects a compromise between affordability, library storage and upkeep, and sufficient coverage of chemical space, while also being a realistic number for screening by fragment soaking. The library was to include a diverse array of 3D-pharmacophore features. Compound solubility was not considered a critical factor in the library design because X-ray crystallography detects binding events even when saturated compound solutions are used (Patel, Bauman *et al.* 2014). Furthermore, fragment cocktailing or enriching for electron-rich elements was not considered a priority because of the high throughput and high resolution data collected at DLS.

[1] www.diamond.ac.uk/Beamlines/Mx/I04-1.html, accessed 3rd October 2016

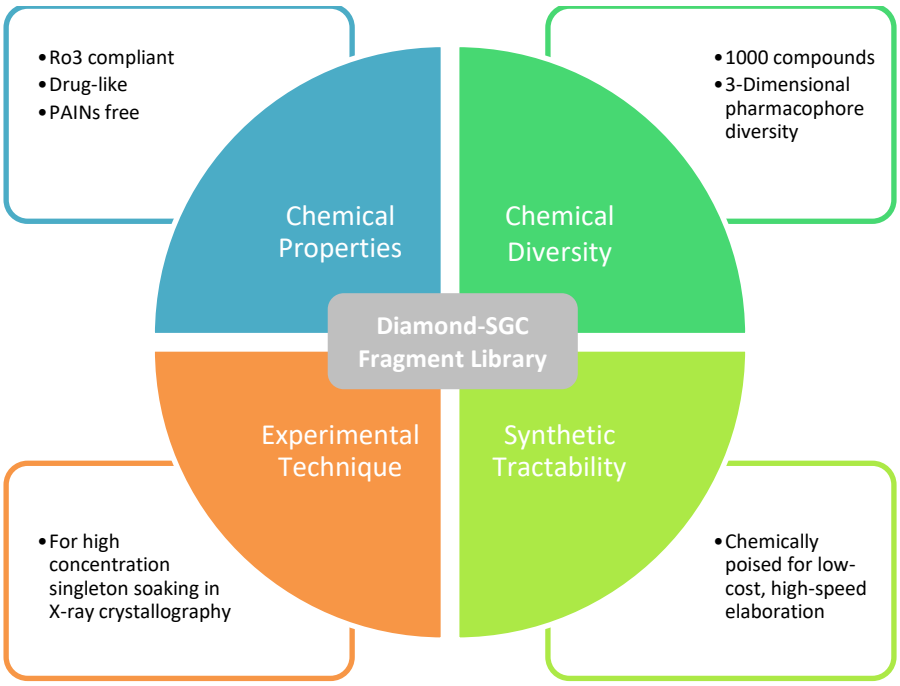


Figure 2.4: The framework used to design the Diamond-SGC Poised fragment library, DSPL.

Finally, the compounds in the fragment library needed to be synthetically tractable. Most academic users of the library do not have access to exhaustive chemistry resources so compounds were required to be amenable to low-cost, rapid elaboration. The concept of poised chemistry was proposed to address the problem. Poised fragments can be elaborated using robust and reliable chemistry. Poised fragment based drug discovery is discussed further in the next section.

2.2.4.1 KNIME and Schrödinger

In order to design and build a fragment library I needed to choose an appropriate computational program with which I could perform data manipulation and cheminformatics. The KNIME[1] Analytics Platform is a pipeline based software package containing pre-programmed modules ('nodes') as well as a range of commercial extensions which allows users to read, manipulate, transform, analyse and visualise a range of data types through a graphical user interface (Figure 2.5). The KNIME architecture is written in Java, although wrappers enable nodes to call other code

[1] www.knime.org, accessed 3rd October 2016

languages such as Python and Perl. A mechanism for extension nodes allows users to add plugins and novel functionality, through which a variety of cheminformatics tools have been added. Since 2006 pharmaceutical research has comprised the largest group of KNIME users (Tiwari and Sekhar 2007).



Figure 2.5: Schematic depicted the pipeline concept used in the KNIME Analytics Platform. Nodes in the data flow can be executed selectively or together, and allow for later inspection of results and models. Image adapted from the KNIME website.[1]

KNIME was selected as the platform with which to build DSPL. The library was designed using a workflow prepared with Schrödinger extensions for KNIME[2] (Berthold, Cebron *et al.* 2007; SchrodingerRelease2013-1 2013; Small-MoleculeDrugDiscoverySuite2013-1 2013) as well as the cheminformatics open source extensions by RDKit, Indigo, ChemAxon and OpenBabel (O'Boyle, Banck *et al.* 2011; SchrodingerRelease2013-1 2013; Small-MoleculeDrugDiscoverySuite2013-1 2013).

2.3 Drug-Like Commercial Fragments

2.3.1 ZINC Database

A variety of chemical databases have been developed to collate lists of known and available compounds. ChemSpider[3] is a chemical structure database containing over 55 million structures compiled from vendor catalogues and chemical publications. Similarly, eMolecules[4] provides an

[1] www.knime.org, accessed 3rd October 2016

[2] 2.6.4 Workbench version 2.5.1 v15901 from the Schrodinger 2012 Suite

[3] www.chemspider.com, accessed 3rd October 2016

[4] www.emolecules.com, accessed 3rd October 2016

online tool to search structures available from a list of trusted suppliers. ChEMBL[1] and PubChem[2] are databases containing biologically active small molecules containing 2D-structures and the associated bioactivity. All of these databases can be accessed online and searched by structure but cannot be downloaded or used offline.

ZINC (Zinc Is Not Commercial) is an online database of commercially available compounds for virtual screening (Irwin and Shoichet 2005).[3] The database can be downloaded in its entirety or in subsets built using predicted physicochemical property filters. Compounds can be downloaded as 3D-structures in sdf format or as SMILES strings in csv format. Chemical properties and commercial availability are provided with both formats.

The ZINC Fragment-Like database is a subset based upon Congreve's Ro3 (Congreve, Carr *et al.* 2003). Compounds are filtered by MW ≤250, cLogP ≤3.5 and NRB ≤5. Toxic or reactive functionalities are not removed. Initially the database was downloaded in sdf format but the file size made manipulation and filtering difficult. The ZINC Fragment-Like library dated 2013-01-26 was downloaded in csv format to overcome the problem. The library contains 701,228 compounds described using SMILES strings (Irwin and Shoichet 2005; Irwin, Sterling *et al.* 2012).

To ensure the commercial availability of compounds, forty catalogues from twenty seven vendors were downloaded from the ZINC website (Table 2.2). Vendors were selected based upon past experience of the company within the department. The most recent version of each catalogue was downloaded. No updated catalogues were downloaded during the course of the research as it was believed any changes in compound availability were unlikely to impact significantly on the

[1] www.ebi.ac.uk/chembl, accessed 3rd October 2016

[2] pubchem.ncbi.nlm.nih.gov, accessed 3rd October 2016

[3] www.zinc.docking.org, accessed 3rd October 2016

design and eventual purchase of DSPL. Of the 701,228 compounds in the ZINC Fragment-Like database, 54,292 (7.7%) were found to be available from at least one of the selected vendors.

Vendor	ZINC Catalogue date tag	Catalogue size
Aldrich CPR	2013-08-13	190,814
Alfa-Aesar	2013-08-07	33,089
Apollo Scientific	2012-09-13	42,187
Apollo Scientific Bioactives	2012-09-13	3,214
Asinex	2012-08-03	467,802
Asinex Building Blocks	2012-08-01	6,662
Bachem	2012-01-08	4,776
Cayman Chemical	2012-02-01	2,525
ChemBridge	2013-07-04	960,591
ChemBridge Building Blocks	2013-05-31	14,382
Combi-blocks	2011-12-06	3,320
Enamine	2013-05-17	1,698,842
Enamine Building Blocks	2013-05-17	70,929
Florida Heterocycle Compounds	2013-03-12	32,056
Fluorochem	2013-06-16	138,756
IBScreen	2013-03-04	453,762
IBScreen Bioactives	2013-02-04	752
IBScreen BuildingBlocks	2013-02-25	11,650
Key Organics	2013-08-12	41,645
Key Organics Building Blocks	2013-08-09	8,407
Life Chemicals	2013-05-30	361,641
Life Chemicals Building Blocks	2013-05-17	8,020
Matrix Scientific	2012-09-17	65,271
Maybridge	2012-05-09	55,544
Maybridge Building Blocks	2012-05-07	4,368
Maybridge Hit Finder	2012-05-08	14,400
Oakwood Chemicals	2013-02-22	72,418
Otava	2013-08-06	254,755
Peakdale	2011-12-06	15,339
Pharmeks	2011-12-29	259,523
Selleck Biochemicals	2012-12-14	563
Selleck Chemicals	2013-05-08	1,675
Sequoia Research Products	2013-02-19	2,266
SigmaAldrich	2013-08-13	98,352
Specs	2013-02-19	206,615
Specs Building Blocks	2013-02-16	8,505
Synthonix Building Blocks	2012-05-14	3,354
TCI	2012-03-01	20,310
Zelinsky Institute	2013-02-19	311,687
Zelinsky Institute Building Blocks	2012-06-19	10,732

Table 2.2: Vendor catalogues used to build the DSPL. Catalogues were downloaded from ZINC[1] between 23rd and 27th August 2013.

[1] www.zinc.docking.org, accessed 3rd October 2016

2.3.2 Rapid Elimination of Swill

The importance of using drug-like compounds in fragment screening campaigns is well documented (Walters, Stahl *et al.* 1998; Baell and Holloway 2010; Baell and Walters 2014; Baell 2015; Dahlin, Nissink *et al.* 2015). Yet there is no consensus of what makes a drug-like compound. Some researchers prefer to screen all compounds and keep hits making genuine interactions. This way scaffolds that would otherwise be missed are discovered, and any unwanted reactivity or toxicology can be subsequently designed out. However the method can result in wasted resources trying to prove if a hit is genuine. A more popular method is to computationally remove the most common toxic and reactive functionalities using Rapid Elimination of Swill (REOS) filters before beginning a screening campaign (Walters, Stahl *et al.* 1998). More recently Baell and co-workers began listing Pan Assay Interference Structures (PAINS) (Baell and Holloway 2010). The majority of PAINS would be removed using a REOS filter, but some, such as 2-aminothiazoles, are also classified as PAINS because of high false positive rates across most screening platforms (Dahlin, Nissink *et al.* 2015). However when using crystallography as a primary screen, the presence of PAINS are less of a concern (Keseru, Erlanson *et al.* 2016). False positive rates are low and it is possible to build specificity with just one or two rounds of synthesis before downstream validation using biochemical or biophysical assays (Burley, Hirst *et al.* 2010).

The Schrodinger REOS node was used to remove toxic and reactive functional groups. Unwanted substructures were identified using SMARTS strings. The list of functional groups and the SMARTS strings used are listed in Appendix A, and the KNIME workflow used is depicted in Appendix B. The resulting drug-like commercial fragment library contains 41,271 compounds.

2.4 Poised FBDD

2.4.1 Defining poised chemistry

Weak hits identified by X-ray screening pose a practical problem, in that they are in general too weak to be verified by orthogonal solution assays; yet such information is a common requirement in medicinal chemistry operations, especially if the crystallographic evidence is ambiguous. Optimising weak compounds by traditional SAR is equally undermined without reliable assay readout in the range of the compounds' binding affinities. In any case, progressing weak compounds to potency in general requires changes not only to the periphery, where conventional SAR operates, but also to the core of the compound. To overcome the problem of optimising weak compounds discovered by high concentration fragment soaking, the concept of poised chemistry was proposed.

I define a poised fragment as a fragment synthesised from a robust and general synthetic reaction such that rapid elaboration of the fragment hit into a library of analogues can be done using parallel chemistry. Identification of a poised bond (or bonds) in a fragment allows the compound to be deconstructed into at least two synthons (Figure 2.6). As a simple example, an amide bond can be deconstructed into an acid chloride and an amine. Upon discovery of a poised amide hit, purchase of similar synthons to those found in the hit allows the synthesis of a library of analogous amides, which can be soaked in crystals just days after the initial hit has been obtained.

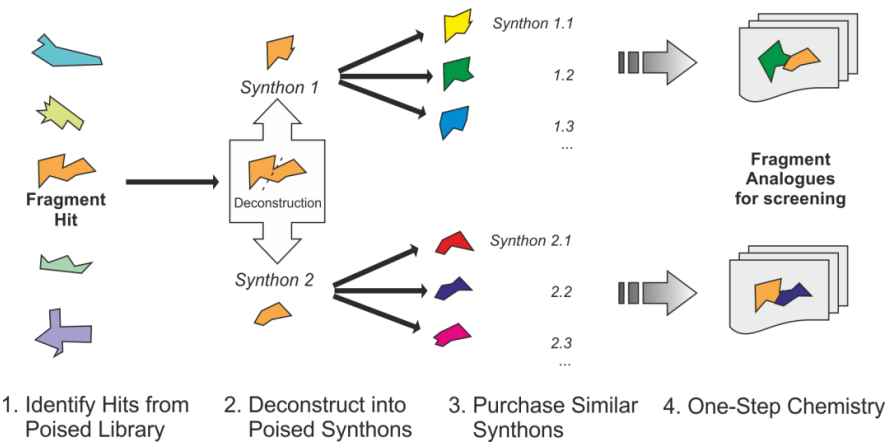


Figure 2.6: Schematic illustrating how a poised fragment library can be used to rapidly synthesise a follow-up library of analogues. If the orange fragment is identified as a hit, a library of similar compounds can be rapidly constructed using parallel synthesis with commercial analogues of synthons 1 and 2.

Poised chemistry requires reliable, robust reactions which tolerate a range of substrates and can be performed using a wide range of commercially available starting materials. The reaction products should be formed in high yield and contain drug-like functionalities with no known toxicological moieties.

Analyses of experimental data by Cooper *et al.* (Cooper, Campbell *et al.* 2010) and Roughley and Jordan (Roughley and Jordan 2011) found the majority of chemical reactions used in medicinal chemistry can be classified by a relatively small number of reaction types. The authors propose the reactions are regularly used because experienced medicinal chemists know they provide reliable methods to synthesise drug-like molecules. The top ten reactions identified by Roughley and Jordan, including amide coupling, reductive amination and Suzuki-type aryl–aryl coupling, can be performed using standardised procedures for a wide range of substrates. The starting materials for each reaction contain common functionalities which are well represented in commercial supplier space and are compatible with this diverse range of commercial compounds. Once protection/deprotection reactions were discarded, nine of the reactions were selected to create poised scaffolds with which to perform substructure searches: amide coupling, *N*-arylation, ether/ester synthesis, amine synthesis from S_N2 reactions, reductive aminiation, Sonogashira

coupling, sulfonamide coupling, Suzuki-type aryl-aryl coupling and (thio)urea synthesis. Recognising the importance of heterocycles in drug discovery (Pitt, Parry *et al.* 2009), poised fragment chemical space was further augmented with thirteen reactions used to build heterocycles: tetrahydro-isoquinolines, indoles, benzoxazoles, benzimidazoles, triazoles, tetrazoles, imidazoles, oxazoles and oxadiazoles (Hartenfeller, Eberle *et al.* 2011). The complete list of poised scaffolds is depicted in Figure 2.7.

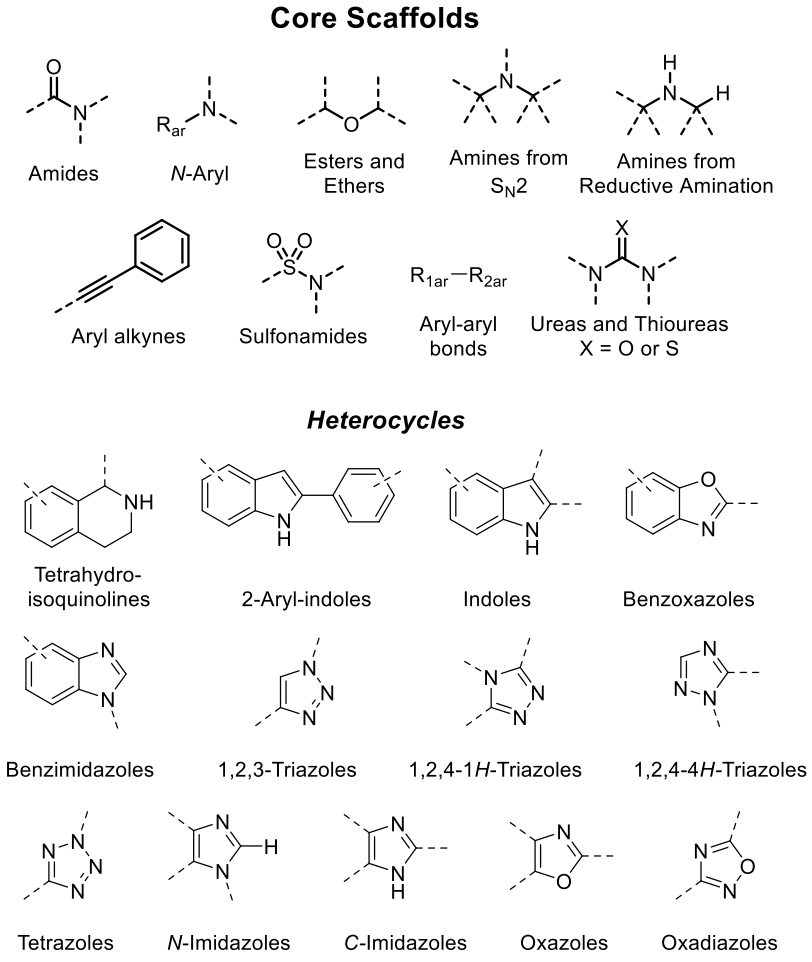


Figure 2.7: Scaffolds used to identify poised fragments. Core scaffolds are based upon the most commonly used medicinal chemistry reactions as described by Roughley and Jordan (Roughley and Jordan 2011). Heterocycle scaffolds are based on reactions described by Hartenfeller et al. (Hartenfeller, Eberle et al. 2011). The synthetic reactions used to build each core scaffold are described in Figure 2.9. Figure reprinted from Cox et al. 2016 with permission from the authors.

Previous attempts to codify these reactions (Hartenfeller, Eberle *et al.* 2012) led to the development of the *de novo* compound design databases DOGS and SCUBIDOO (Hartenfeller,

Zettl *et al.* 2012; Chevillard and Kolb 2015). However, poised chemistry differs from these methods by only considering compounds which can be synthesised in a single step.

2.4.2 Identifying Poised Fragment Space

The KNIME workflow used to design the DSPL is outlined in Figure 2.8, and is depicted in more detail in Appendix B. First, poised functional groups were identified using the Schrödinger Substructure Search node. Substructures were defined using the MarvinSketch node as SMILES strings. Secondly the compounds were converted to Indigo format and the two synthons needed to synthesise the molecule identified using the Indigo Combinatorial Reaction Enumeration node. Disconnections were based upon the synthetic steps depicted in Figure 2.9 and defined using the MarvinSketch node from ChemAxon in RXN (MDL Molfile Reaction) format. Next, all three compounds (original fragment, synthon one and synthon two) were converted to Unique SMILES using the OpenBabel node (O'Boyle, Banck *et al.* 2011).

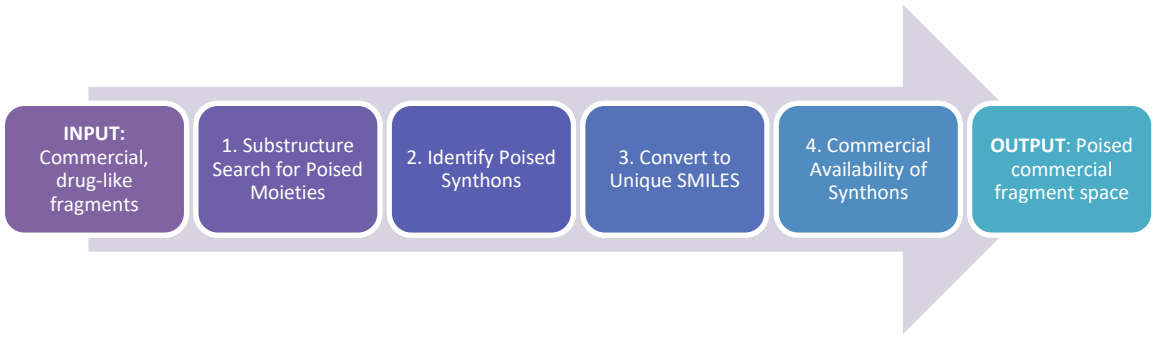


Figure 2.8: Schematic of KNIME workflow used to identify poised fragments. The workflow was designed using Schrodinger, Indigo, ChemAxon and OpenBabel extensions for KNIME.

The final step was to discard fragments if one or both of the synthons were not commercially available. Commercial availability was determined using an All Purchasable library built in Knime using the vendor catalogues downloaded from ZINC (Table 2.2) and a SigmaAldrich catalogue downloaded from MolPort, an online compound ordering service.[1]

[1] www.molport.com, accessed 3rd October 2016

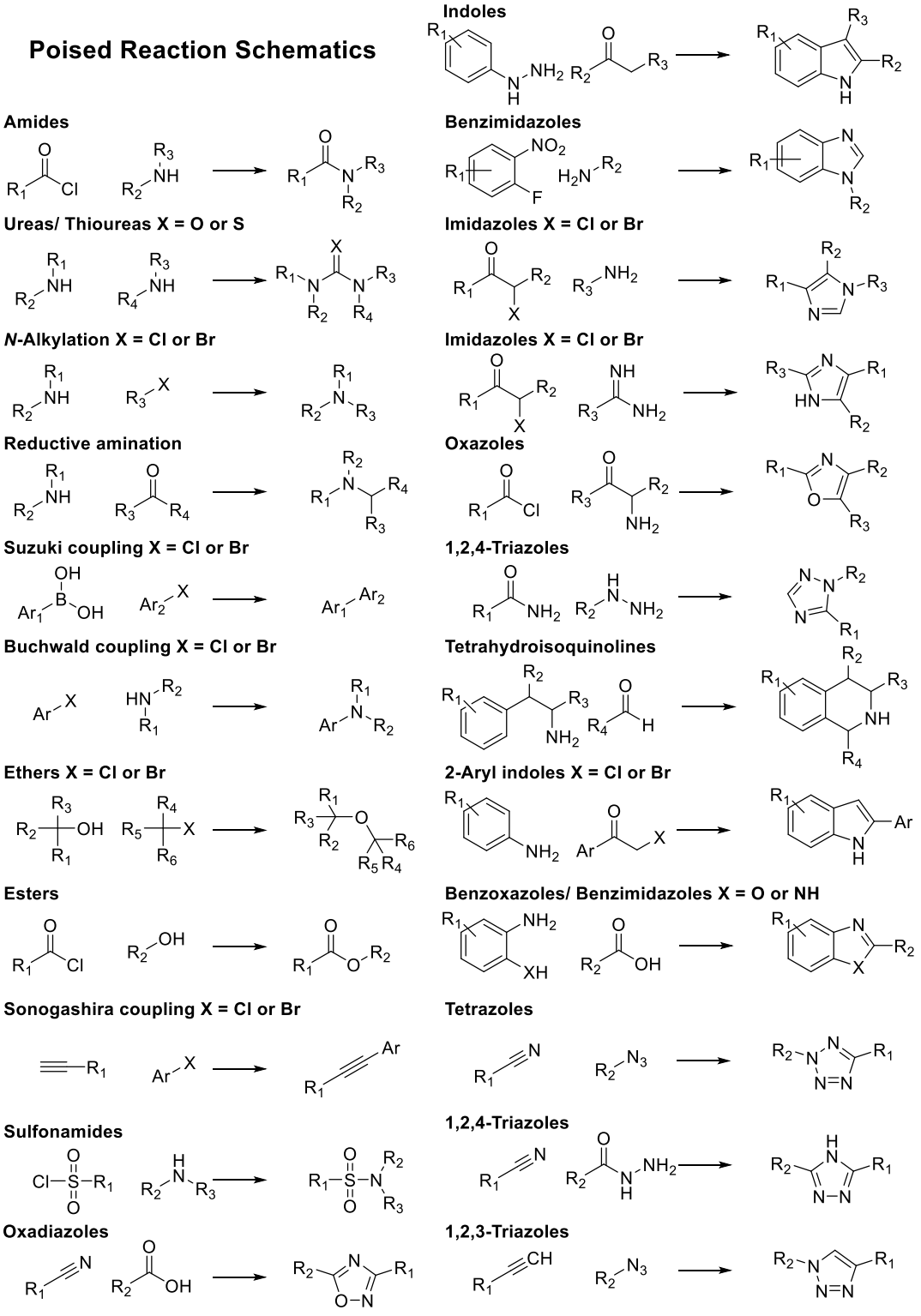


Figure 2.9: Synthetic reactions used to describe each of the selected poised reactions.

From the >41,000 fragments identified from the ZINC Fragment-Like database, 28,438 fragments (68.8%) were identified as containing at least one poised functionality. The fragments covered

42,997 reactions with a large number of fragments being found to be poised via multiple reaction types. While initially surprising that such a large number of commercial fragments could be synthesised by only twenty-two different chemical reactions, I believe this reflects the very limited coverage of potential synthetic compounds by commercial vendors. Of the >28,000 fragments, 11,875 poised fragments (42%) were believed to have two, well-represented commercially available starting synthons, covering 15,277 reactions.

Recognising the importance of ring systems in drug discovery (Pitt, Parry *et al.* 2009), further filters were applied to remove compounds without a ring system, along with quaternary nitrogens, imines and diacyl hydrazides. In total, 1,347 compounds were removed, leaving a library of 10,448 fragments. In theory, this library covers the whole of commercially accessible poised fragment space using the chosen reactions. The finding demonstrates how a good coverage of commercially available fragment space can be achieved using a relatively small number of reaction types. However, it also reveals the bias of commercially available compounds towards commonly used reaction types. The library is dominated by amides, with 55.1% of fragments containing a poised amide bond; ether bonds (18.2%) and reductive amination products (10.2%) are also heavily represented. The least represented chemistries are Sonogashira products (0.2%) and sulfonamides (2.0%) (Figure 2.10). Unfortunately, poised heterocyclic reactions were observed in just 2.2% of the compounds, despite thirteen reactions (only eight of which were found to be present) being used to identify such fragments and the importance of such scaffolds in drugs and chemical tools (Pitt, Parry *et al.* 2009).

2.5 Diversity Filter

The USRCAT algorithm was chosen to ensure chemical diversity was maintained upon selection of 1000 compounds for DSPL from the total of >10,000 poised fragments (Schreyer and Blundell

2012). USRCAT was selected because it was believed the additional conformation information makes it a superior measure of diversity for library design in comparison to traditional 2D-diversity measures.

The diversity filter was applied using a python script written by Dr Anthony Bradley (Bradley 2015). The USRCAT algorithm was implemented using the method outlined by Schreyer and Blundell (Schreyer and Blundell 2012). A conformation for each molecule was generated using a single universal force field (UFF) (Rappe, Casewit *et al.* 1992). Ten conformers were generated for each compound to ensure the compound's chemical shape was sufficiently sampled.

Selection of diverse compounds was performed using the MaxMin algorithm in the RDKit. The algorithm selects a user-determined number of compounds from a list of compound fingerprints. In this case 1000 compounds were selected. The number reflects a compromise between sufficient coverage of chemical space and affordability, and library storage and upkeep. Selection was performed by first picking a random compound and calculating its similarity to every other compound in the library. The least similar compound to the initial compound is then selected. The process is repeated for the new compound until 1000 compounds have been selected. The method is a computationally more efficient method of selecting diverse compounds than clustering because only 1000 similarity vectors are computed rather than a complete all versus all matrix.

Analysis of the 1000 compounds selected by USRCAT showed enrichment towards some of the less represented chemistry (Figure 2.10), perhaps reflecting that any single poised reaction is likely to generate similar 3D-pharmacophore fingerprints. The 1000 compounds, named DSPL_1000, were dominated by amides, but at 41.9% they represent thirteen percentage points less than poised fragment space (55.1%). In comparison, heterocycles (2.2% to 2.6%), Suzuki-type *C-C*

bonds (10.4% to 12.1%) and amines formed by *N*-alkylation (4.9% to 10.2%) are all better represented.

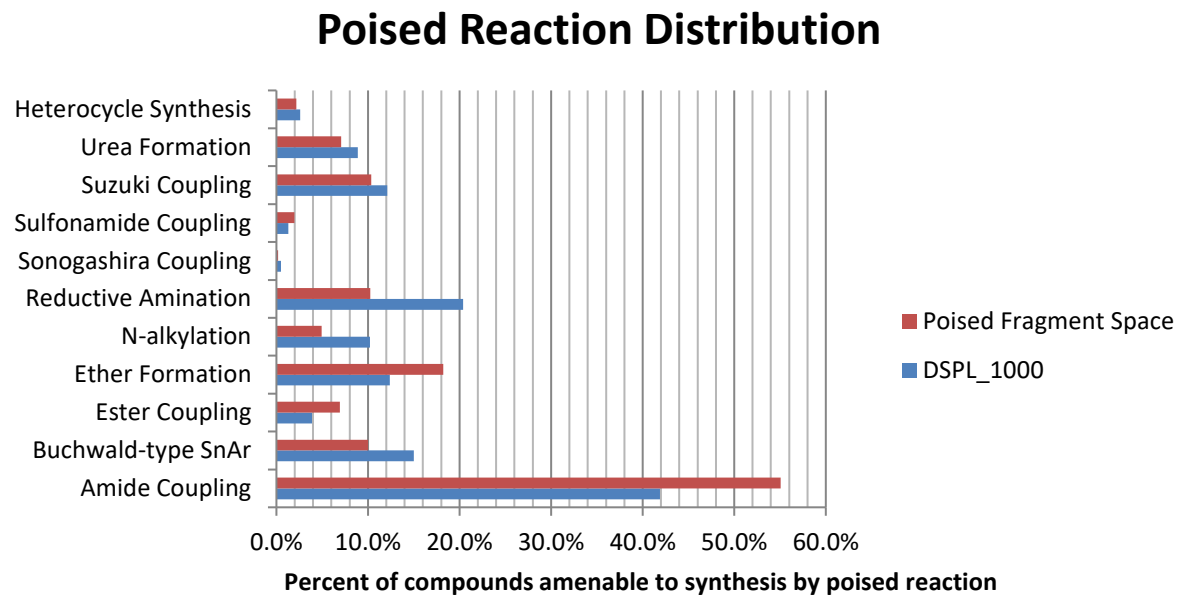


Figure 2.10: Distributions of Poised Reactions in Poised Fragment Space (10,448 compounds) and DSPL_1000 (1000 compounds selected using USRCAT diversity).

2.6 Purchase of DSPL

The DSPL_1000 was sent to vendors for purchasing quotes. Only quotes for compounds available in the catalogues downloaded from the ZINC database were requested from each vendor. Eighteen vendors responded with quotes for 10 mg of each compound. In total, 840 compounds were offered by one or more of the vendors contacted. No quotes for the remaining 160 compounds could be obtained because the compounds were unavailable.

A subset of 330 compounds were selected and purchased in March 2015. Compounds were selected based upon price and enriching for poorly represented poised reaction types. Compounds were dissolved to 500 mM in DMSO and transferred to LDV 1536-well plates with help from Mr Faruq Hossain, Dr Patrick Collins and Ms Alison Howarth.

The process was repeated in October 2015 with help from Ms Julia Holloway. Quotes were obtained for 543 of the remaining 670 compounds. Analysis of the quotes showed the 81 most expensive compounds were dominated by well-represented chemistries (amides 49% and ureas 33%) but contributed around 40% to the total cost. Therefore 462 compounds were ordered, of which sixteen were never delivered (either human error or drug-precursor restrictions). The new 446 compounds were dissolved to 500 mM in d_6-DMSO and added to the LDV 1536-well plates with help from Dr Patrick Collins, Dr Alice Douangamath and Mrs Elena Seraia. Therefore the final DSPL contains 776 compounds, and is now in regular use at DSPL for fragment soaking campaigns. The full library is listed as 2D structures in Appendix C.

2.6.1 Properties of DSPL

Many of the compounds in DSPL library have multiple poised disconnections, meaning the 776 compounds comprise 1003 reactions. Products of amide coupling (42.0%) and reductive amination (22.0%) dominate the library (Figure 2.11). The least represented poised chemistries are Sonogashira coupling (0.6%) and sulfonamide coupling (1.4%). The distribution is representative of poised fragment chemical space, and reflects the poor coverage of potential building blocks by commercial suppliers.

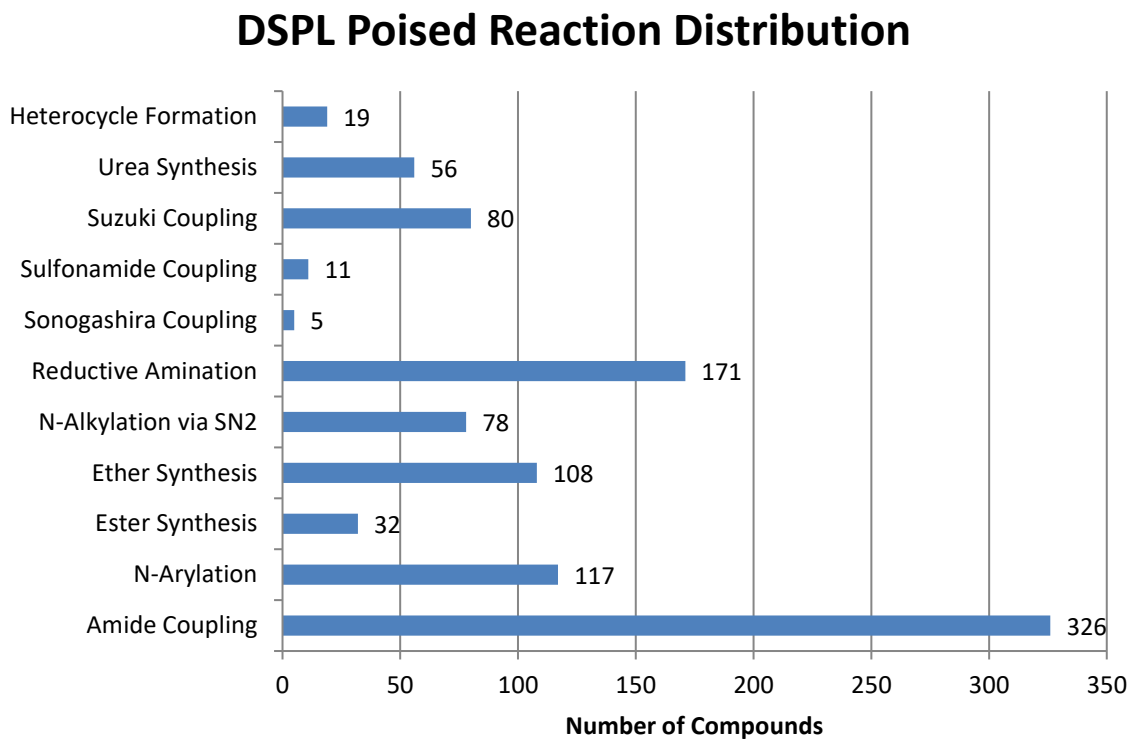


Figure 2.11: DSPL poised reaction distribution. The 776 compounds in DSPL cover 1003 reactions.

Selected physicochemical properties of DSPL are depicted in Figure 2.12. The properties were predicted using the RDKit Descriptor Calculator node in KNIME from the compound SMILES strings. The lipophilicity values generated by RDKit, SlogP, range from -1.10 to 3.87. Only six compounds fall outside Congreve's Ro3 (cLogP < 3.5) and is probably a result of the different computational methods used to calculate lipophilicity between the ZINC database and RDKit. The mean value is 1.93. The distribution of molecular weight is reflective of chemical space, in which the number of possible compounds increases exponentially with molecular weight. The maximum molecular weight is 250, and the mean value is 219.6. The pattern is repeated with heavy atom count, which ranges from 7 to 19 heavy atoms with a median value of 16.

The median number of hydrogen bond donors (1) is less than the median number of hydrogen bond acceptors (3). No filter for hydrogen bond acceptors/donors was implemented during the design of DSPL, which explains why a large number of compounds break Congreve's Ro3 (HBD ≤3,

HBA ≤3). Fifteen compounds possess exactly one hydrogen bond donor or acceptor (no compounds have zero hydrogen bond donors and acceptors). Three compounds have the highest number of hydrogen bond donors plus acceptors (nine).

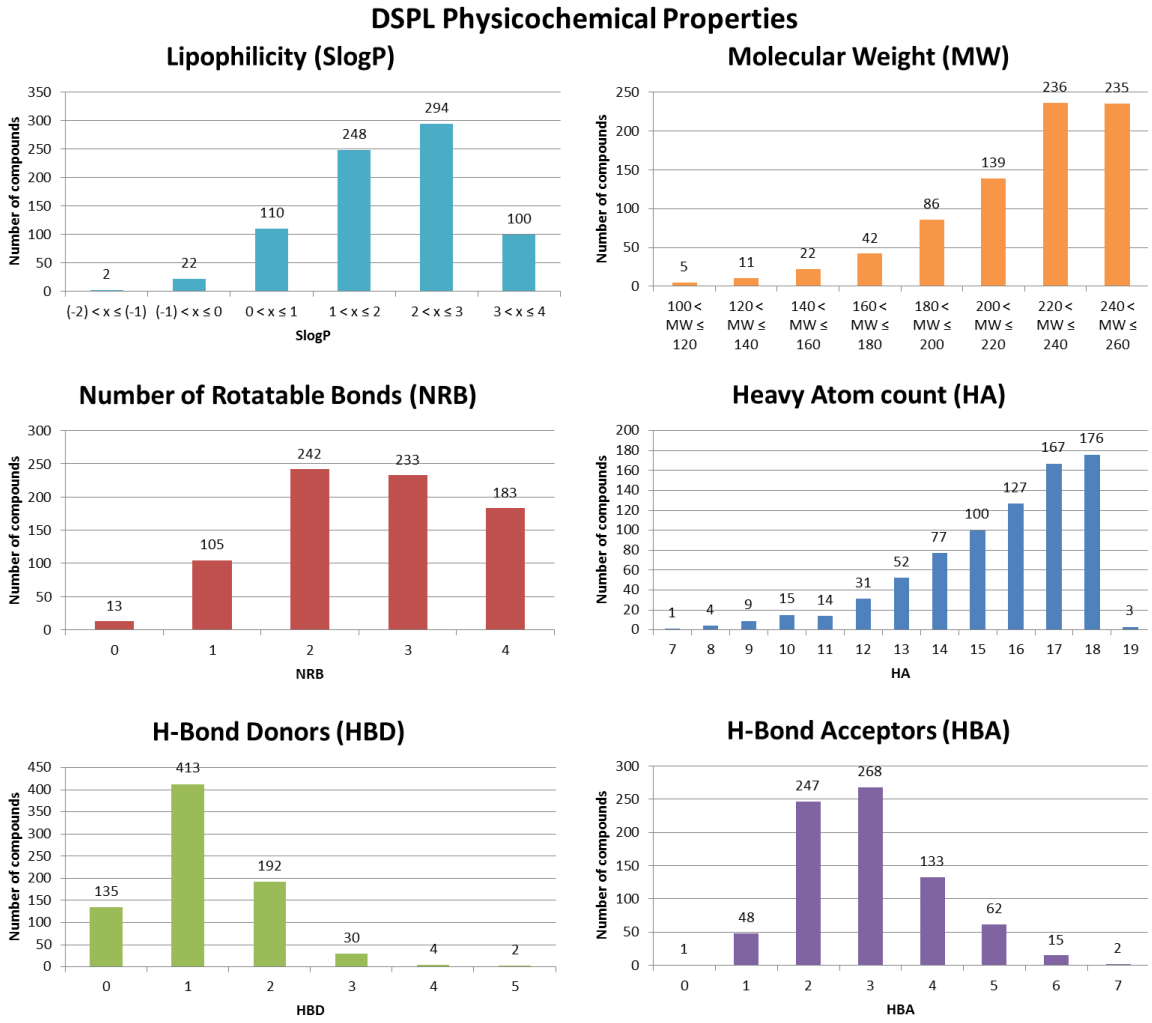


Figure 2.12: Physicochemical Properties of DSPL. Values for the 776 compounds were predicted using the RDKit Descriptor Calculator.

2.6.2 Comparison of DSPL to other fragment libraries

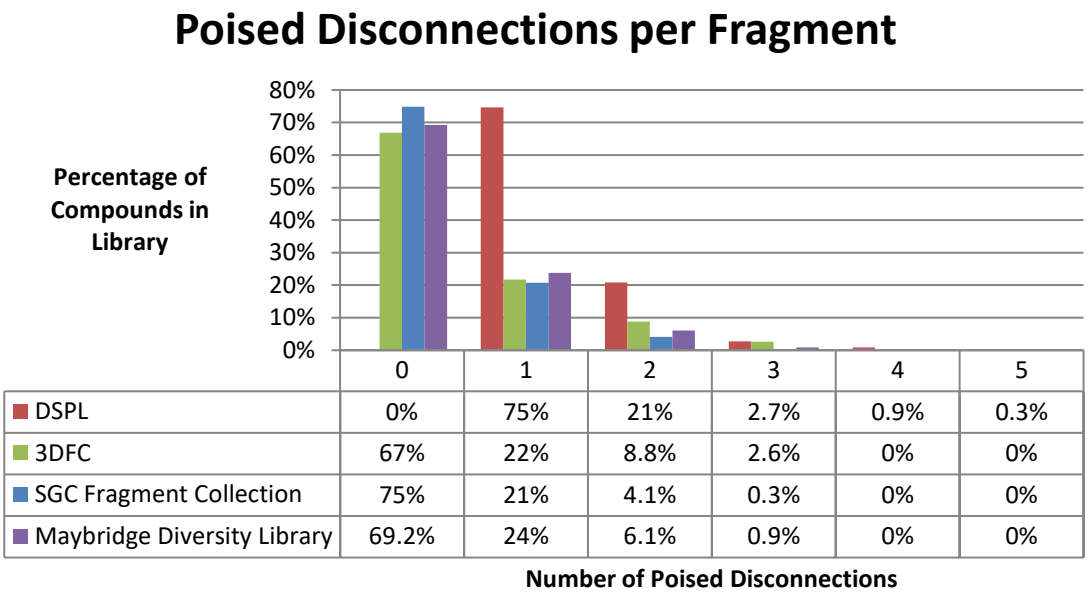
To demonstrate the utility of DSPL, a comparison to other fragment libraries was performed. Three compound sets were analysed: the fragment library compiled by the 3D fragment

consortium[1] (3DFC) (Morley, Pugliese *et al.* 2013), the SGC fragment collection assembled in-house and from collaborator collections, and the commercially available Maybridge Diversity library[2] of 1000 fragments. The 3DFC included a large portion of bespoke fragments synthesised by consortium collaborators, whereas the majority of the SGC Fragment Collection had been purchased from vendors. The Maybridge library is a subset of the SGC fragment collection and is regularly used as a part of internal hit discovery campaigns. The analysis was performed using the same KNIME protocol previously described.

2346 poised compounds (from 9319, 25.2%) were identified in the SGC fragment collection and 270 poised compounds (from 763, 35.4%) in the 3DFC. Likewise, the Maybridge library was found to contain 308 poised compounds (from 1000, 30.8%). The SGC fragment collection has a similar proportion of commercially represented poised fragments as the ZINC Fragment-Like database (10,448 from 41,271, 25.3%). In comparison the 3DFC has a much larger proportion of poised fragments, perhaps reflecting a heavy reliance on poised reactions in the synthesis of bespoke fragments. DSPL has a greater proportion of compounds with more than one poised disconnection than any of the other three libraries analysed demonstrating the library's synthetic tractability (Figure 2.13).

[1] Library dated 29th June 2016

[2] www.maybridge.com/portal/alias__Rainbow/lang__en/tabID__230/DesktopDefault.aspx, accessed 21st December 2016



	0	1	2	3	4	5
DSPL	0%	75%	21%	2.7%	0.9%	0.3%
3DFC	67%	22%	8.8%	2.6%	0%	0%
SGC Fragment Collection	75%	21%	4.1%	0.3%	0%	0%
Maybridge Diversity Library	69.2%	24%	6.1%	0.9%	0%	0%

Figure 2.13: Number of poised disconnections per fragment in the DSPL, the 3DFC, the SGC Fragment Collection and the Maybridge Diversity Library.

For the subset of poised compounds in the 3DFC and SGC Fragment Collection, the distribution of poised reaction types is similar to the DSPL (Figure 2.14). The 3DFC has a low number of heterocycles (0.7%), Suzuki-type *C-C* bonds (0.7%) and no Sonogashira-type bonds reflecting the library's design principle of enriching for 3D-characteristics. In comparison, the SGC Fragment Collection and the Maybridge Diversity Library have surprisingly low percentages of amide-containing compounds (27.5% and 17.9%) than the other two libraries. The former could be an artefact of the target classes the SGC has investigated in the past ten years, which include kinases and bromodomains, while the latter maybe a reflection of the diversity metric used by Maybridge to build the library.

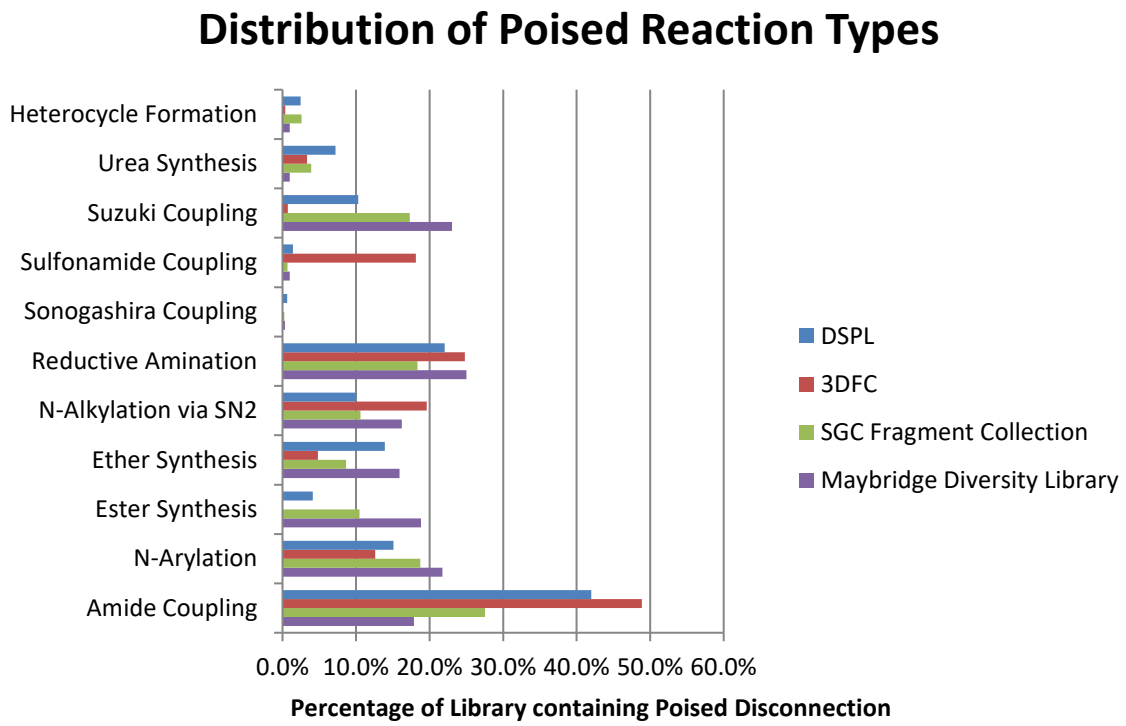


Figure 2.14: Comparison of the distribution of poised reaction types between four fragment libraries.

2.7 Ongoing Work and Conclusions

The DSPL has been designed and purchased for FBDD at DLS and SGC Oxford. The library was designed using the principles of fragment library design. The design process is summarised in Figure 2.15, and a full list of the fragments as 2D structures can be found in Appendix C or downloaded as SMILES strings from the DLS website.[1]

Congreve's rules were used as a guideline meaning the library is not completely Ro3 compliant. However all compounds have molecular weight less than 250 and cLogP less than 3.5. Toxic and reactive molecules were removed meaning the library is composed entirely of drug-like compounds. A library size of 1000 compounds was originally desired, but only 776 compounds were purchased. However it is believed the library has sufficient coverage of poised fragment

[1] www.diamond.ac.uk/Beamlines/Mx/Fragment-Screening/Fragment-Libraries.html, accessed 21st December 2016

space alongside a diverse range of poised reaction types. The library is synthetically tractable to elaboration: every fragment member can be synthesised using at least one of seventeen poised chemical reactions from commercially available starting points.

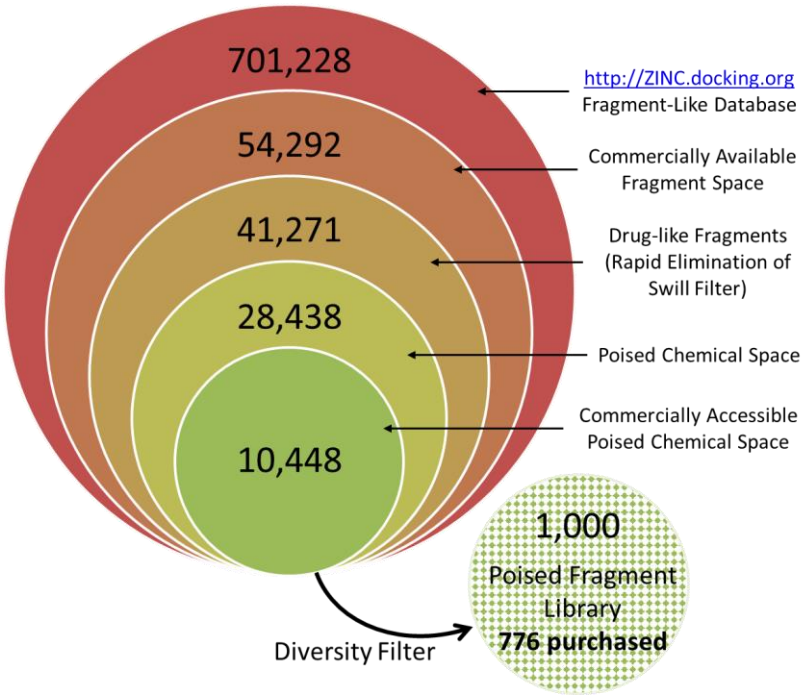


Figure 2.15: Overview of the DSPL design process. The 776 fragments in DSPL are listed in Appendix C.

The next two chapters explore how best to elaborate a poised fragment once it is identified as a hit. Chapter 3 describes the design of a computational tool for designing a library of follow-up compounds. Chapter 4 uses case studies of poised fragment hits for epigenetic targets to demonstrate the utility of using poised fragments.

3 XPoise - Fragment Elaboration Software

3.1 Overview

In this chapter I describe the design of a computational tool called XPoise, which is used to build follow-up libraries from poised fragments. The workflow, built using the KNIME pipeline tool (section 2.2.4.1), can be implemented on open-access software and used by researchers who do not have training in medicinal chemistry. The algorithm is split into three parts (Figure 3.1). First, the poised bonds of a hit obtained using X-ray crystallography are identified and the compound split into two or more synthons. Secondly, each synthon is computationally reacted with a library of commercially available building blocks compiled from vendor catalogues. Finally, the resulting list of compounds is filtered to contain no more than 3000 compounds ready to be analysed by LLOOMMPPAA, an algorithm which prioritises virtual compounds for experimental testing by assessing structural data and protein-ligand interaction fingerprints (PLIFs).

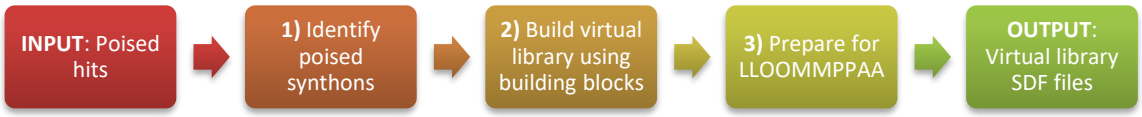


Figure 3.1: Overview of the workflow used by the XPoise algorithm to generate fragment follow-up libraries.

3.2 Background

3.2.1 Reaction Cheminformatics

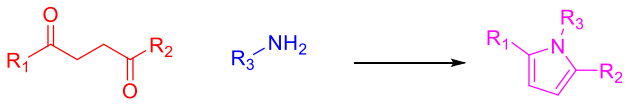
In order to build virtual libraries using Xpoise, I needed to choose a codeable chemical reaction language. Databases of encoded, computer-readable synthetic reactions are rare in the literature (Hartenfeller, Eberle *et al.* 2011). In 2003, Vinkers *et al.* used reaction SMARTS to build a dataset containing 75 reactions (Vinkers, De Jonge *et al.* 2003). They used the reactions to build a drug

design program, SYNOPSIS, which generates synthetically feasible molecules with desired chemical properties. More recently, Hartenfeller *et al.* published a dataset of 58 reaction schemes using reaction SMARTS (Hartenfeller, Eberle *et al.* 2011). The reactions did not take into consideration how reactivity could be affected by neighbouring-groups or regioselectivity, and consequently "construction schemes … should be understood as a source of inspiration rather than readily applicable synthesis schemes". Even so, Chevillard and Kolb used the dataset to enumerate a virtual library called SCUBIDOO containing 21 million compounds from 8000 building blocks (Chevillard and Kolb 2015). The dataset has also been converted to Reaction MQL format to develop the DOGS (Design Of Genuine Structures) program (Hartenfeller, Zettl *et al.* 2012). The software generates synthetically accessible compounds which can be made in four or less chemical steps. The subsequent library is compared to known bioactive ligands using structural and pharmacophore features to select compounds for synthesis.

MQL (molecular query language) is a cheminformatics search tool which incorporates spacial and physicochemical properties of atoms and bonds (Proschak, Wegner *et al.* 2007). Properties are split into three categories: settable, checkable and element symbol. Settable keywords allow users to define the value of a chemical property, and include "charge", "explicitHydrogens" and "aromatic". Checkable keywords are used to annotate functional groups neighbouring the query structure such as "ring", "valence", "explicitConnections" and "sp3". Element symbols include atom sets such as "Heavy" and "Halogen" as well as chemical symbols. A user can also add new property keywords to further enhance a query.

Reaction-MQL is an extension of MQL for describing molecular transformations (Reisen, Schneider *et al.* 2009). A reaction between reagents A and B to form product C is expressed in the format "A ++ B >> reactionIdentifier >> C" (Figure 3.2). The reactants are described by the substructures

directly involved in the reaction (the reaction centre) allowing for variation in substituents (R-groups). The direction of the reaction is described using ">>" (irreversible reaction) or "<<" (reversible reaction) followed by ">>". Between the two symbols an annotation is used to identify the transformation. Only the parts of the product which have been changed by the reaction are described. All other parts of the reactants are maintained. Solvents, catalysts and unchanged reagents are not described. Reaction-MQL uses the same property keywords as MQL, and is also extendable with user-defined keywords.



Reaction-MQL:
C-C1[!ring](=O)-C4[!aromatic & bound(-H)]-C5[!aromatic & bound(-H)]-C2(=O6)-C
++ C[!bound(=O)]-N3[allHydrogens=2 & charge=0] >> Paal-Knorr pyrrole >>
C1$1-N3-C2=C5-C4$=1

SMARTS:
[#6:1]-[#6!R:2](=[#8])-[Ch:3]-[Ch:4]-[#6:5](=[#8])-[#6:6].[#6:7](~[!O])-[NH2+0:8] >>
[#6:1]-[c1:2]:[c:3]:[c:4]:[c:5](-[#6:6]):[n1:8]-[#6:7]

Figure 3.2: Reaction query strings for the Paal-Knorr pyrrole synthesis using Reaction-MQL and SMARTS.

SMARTS (SMILES arbitrary target specification) is a cheminformatics query language developed by David Weininger and colleagues (Daylight Chemical Infomation Systems). A comprehensive description of the language can be found in the Daylight SMARTS theory manual.[1] Atoms are described using their atomic number (e.g. "[#7]" for nitrogen). Alternatively the chemical symbol can be used; capital letters represent aliphatic atoms, while lower case denotes aromaticity. Other symbols are used to describe chemical properties, including valence ("v"), wildcard atoms ("*"), hydrogen count ("H") and chirality ("@" and "@@"). Four logic operators allow the creation of complex queries: "&" (high priority and), ";" (low priority and), "," (or) and "!" (not). A reaction between reagents A and B to form product C is expressed in the format "A.B >> C" (Figure 3.2). Product and reactant substructures are encoded in separate SMARTS strings, and only the

[1] www.daylight.com/dayhtml/doc/theory/theory.smarts.html, accessed 3rd October 2016

reaction centre is defined. In contrast to Reaction-MQL, any part of the molecule described in the reactant but not in the product is removed. Undefined R-groups are maintained unmodified upon execution of the reaction. A comparison of Reaction-MQL and SMARTS reaction strings for the Paal-Knorr Pyrrole synthesis is depicted in Figure 3.2. SMARTS strings were selected to encode for reactions in the XPoise workflow because they are widely supported by non-proprietary software and editable using a simple text editor.

3.2.2 LLOOMMPPAA

LLOOMMPPAA is a tool designed to select diverse compounds from a virtual library for synthesis and screening when limited structural and SAR data is available to guide compound design (Bradley 2015). The LLOOMMPPAA method requires an input library composed of structural analogues to a crystallographic hit compound based upon a common substructure, which is analysed by means of four computational steps (Figure 3.3). First, the 3D conformation of each compound in the input library is generated using the conserved part of the ligand as a static core. This is computationally the most expensive step in the LLOOMMPPAA method. When executing LLOOMMPAA, the static core is defined using the Context SMILES. The Context SMILES can be deduced from the hit SMILES and knowledge of which reaction is being used to generate the follow-up library. In the second step, each compound is overlaid with the original crystal hit using the Context SMILES, and filtered based upon Van der Waal clashes with the protein crystal structure. PLIFs are calculated for the remaining compounds and finally a diverse subset of compound fingerprints is selected for testing.

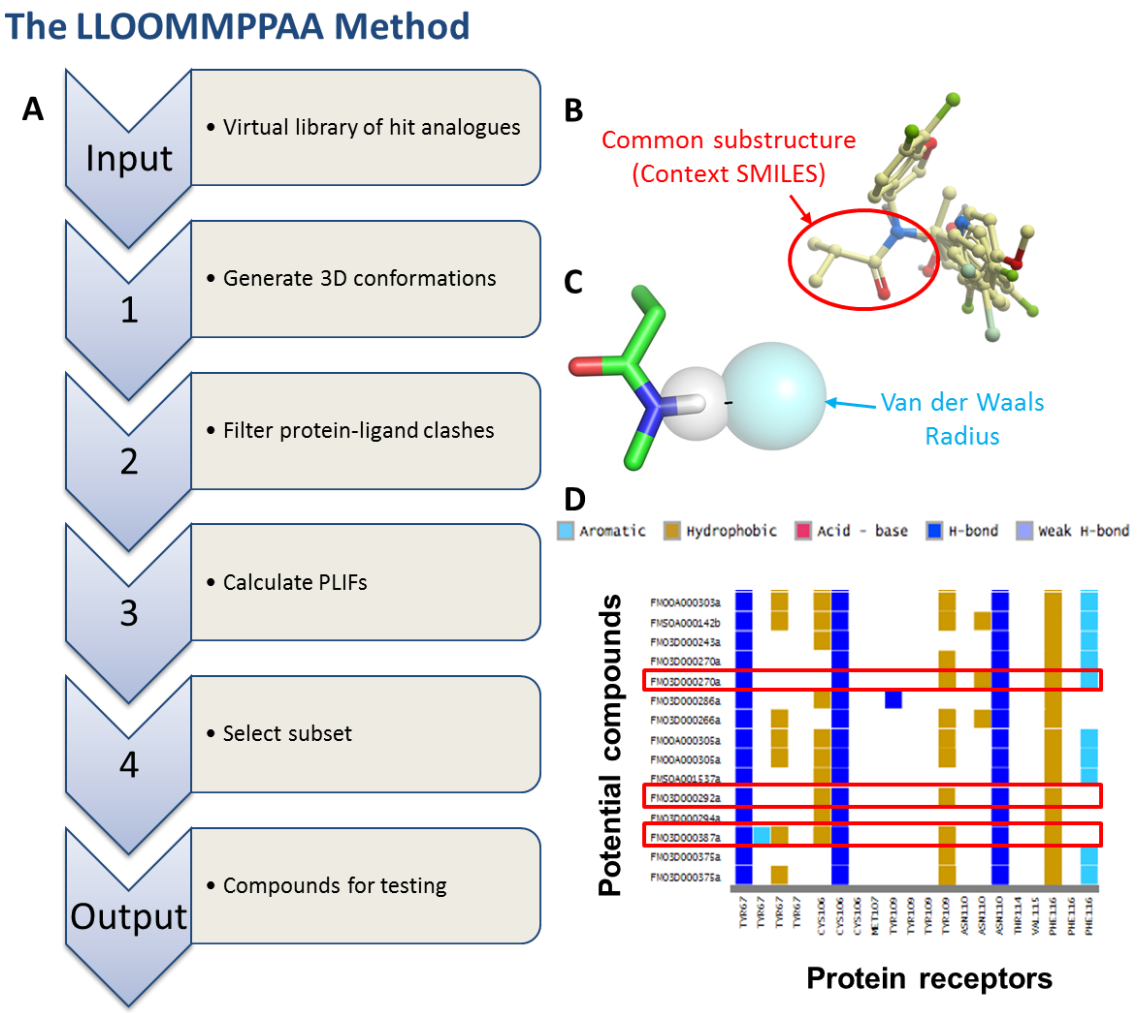


Figure 3.3: The LLOOMMPPAA method. (A) Overview of the four computational steps used by LLOOMMPPAA to select compounds for synthesis and screening. (B) A 3D-conformation is generated for each compound in a virtual library of structural analogues based upon a common substructure, defined by the Context SMILES. (C) The compounds are overlaid with the original crystal structure and filtered based upon Van der Waal clashes with the protein. (D) PLIFs are calculated for each remaining compound from which a diverse subset is selected. Image adapted from Bradley 2015.

PLIFs are bit vectors used to describe how a ligand interacts with a protein (Deng, Chuaqui *et al.* 2004; Bradley 2015). Each vector bit represents an interaction point on the protein defined by atom coordinates and a pharmacophore feature (aromatic, hydrophobic, acid-base, H-bond or weak H-bond). For a given ligand, each bit of the vector is set to one if the ligand forms an interaction to that part of the protein and a zero if not. The presence of an interaction is defined using distance and geometric constraints pre-calculated from other protein-ligand structures.

LLOOMMPPAA uses three methods for the selection of compounds. The first selects a diverse subset of PLIFs (PLIF_default). The second selects compounds with the maximum number of potential protein-ligand interactions (PLIF_max). The final method selects PLIFs with the highest similarity to the experimentally determined protein-ligand structures of the target (Focussed).

3.2.3 Project Aim

When a poised fragment hit is identified using X-ray crystallography, a library of structural analogues is compiled, synthesised and tested to better understand binding modes and SAR. The use of computational tools such as LLOOMMPPAA allows compounds from a large, diverse library to be selected based upon structural data, but selected compounds need to be commercially available or amenable to synthesis using parallel chemistry.

My aim was to develop a computational tool which could automatically build follow-up libraries using the principles of poised drug discovery outlined in the previous chapter (Figure 3.4). I also wanted the software to support the use of two other popular hit-to-lead methodologies: SAR by catalogue, and fragment growth using reactive handles.

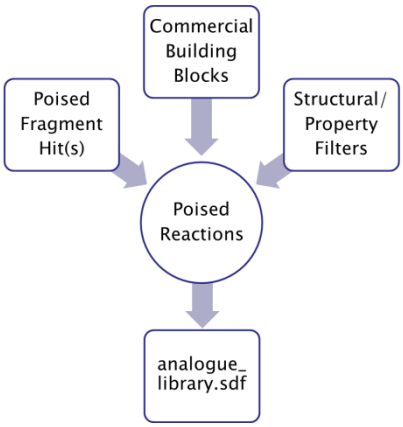


Figure 3.4: Proposed schematic for the design of follow-up libraries for poised crystal hits using the self-contained computational tool, XPoise.

3.3 Methods and Materials

3.3.1 Overview

The XPoise workflow was built using RDKit[1] for KNIME.[2] Reactions were encoded using SMARTS strings because they are widely supported by other software suites and editable using a simple text editor. Building block libraries were compiled from vendor catalogues and converted to a uniform format using a KNIME protocol. The KNIME workflows used by XPoise are depicted in Appendix B.

3.3.2 Building Blocks library

Initially the Aldrich CPR[3] catalogue was used as the building blocks library. Later versions were built using vendor catalogues provided upon request. Of the twelve vendors contacted, eight responded with catalogues containing price, weight and structural identifiers (such as SMILES or CAS number). The Chemical Identifier Resolver KNIME node, developed by the CADD group at NCI/NIH and provided by Talete,[4] was used to convert CAS numbers into SMILES strings where required. The final library is comprised of catalogues from Fluorochem,[5] Enamine,[6] Acros Organics,[7] Alfa Aesar,[8] Maybridge,[9] Chem-Impex International,[10] CombiBlocks,[11] and TCI.[12] If the

[1] www.rdkit.org,accessed 3rd October 2016
[2] Version 3.2 workbench 2.12.0
[3] www.sigmaaldrich.com/chemistry/chemistry-services/discovery-cpr.html, accessed 3rd October 2016
[4] www.talete.mi.it/, accessed 3rd October 2016
[5] www.fluorochem.co.uk, accessed 3rd October 2016
[6] www.enaminestore.com, accessed 3rd October 2016
[7] www.acros.com, accessed 3rd October 2016
[8] www.alfa.com, accessed 3rd October 2016
[9] www.maybridge.com, accessed 3rd October 2016
[10] www.chemimpex.com, accessed 3rd October 2016
[11] www.combi-blocks.com, accessed 3rd October 2016
[12] Tokyo Chemical Industry www.tcichemicals.com, accessed 3rd October 2016

same chemical was provided by two or more vendors, the cheapest supplier (calculated in GBP per gram) was kept. Compounds with twenty or more carbon atoms were discarded.

Analysis of the first version of the building blocks library by substructure indicated the distribution of price is independent of functional group (Figure 3.5). Furthermore, unpublished analysis of the Enamine catalogue by Prof Paul Brennan found no correlation between commonly calculated chemical properties (including HDB, HBA, MW, clogP, *N* count, *O* count, and fsp^3) or chemical diversity with price. Therefore a cut-off of 400 GBP/g, covering 61% of the building blocks library, was applied as a compromise between affordability and sufficient coverage of chemical space. The latest version of the library, which uses the same price cut-off, contains 192,667 entries.

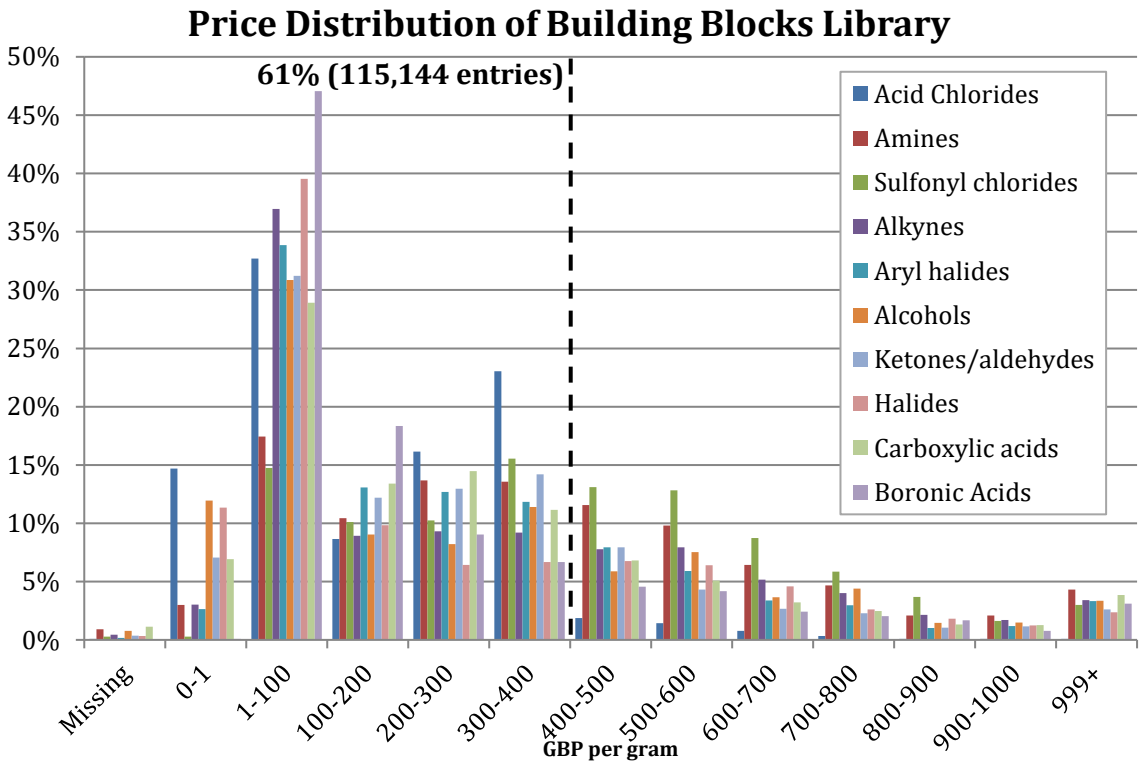


Figure 3.5: Price distribution of the first version of the building blocks library. The library was compiled from the catalogues of Fluorochem, Acros Organics, Enamine, Maybridge and Alfa Aesar. Compounds above 400 GBP/g or with no quoted price were removed, leaving 115,144 chemicals.

3.3.3 RDKit for KNIME

RDKit is an open source toolkit for cheminformatics.[1] Core functionalities are written using the C++ computer language, with a Python-based user interface. A full description can be found in the RDKit Documentation.[2] XPoise was built using the RDKit cheminformatics nodes for KNIME distributed through the KNIME community site.[3] The KNIME workflows used by XPoise are depicted in Appendix B.

3.3.3.1 Poised Reaction SMARTS

Reaction SMARTS were built using the MarvinSketch node by ChemAxon and adapted for execution by the RDKit One or Two Component Reaction nodes. The SMARTSviewer[4] was used to confirm the identity of reactant and product SMARTS. Further SMARTS strings are used to filter unwanted functional groups. An overview of a typical reaction workflow is depicted in Figure 3.6. The exact SMARTS strings used for each reaction are listed in Appendix D.

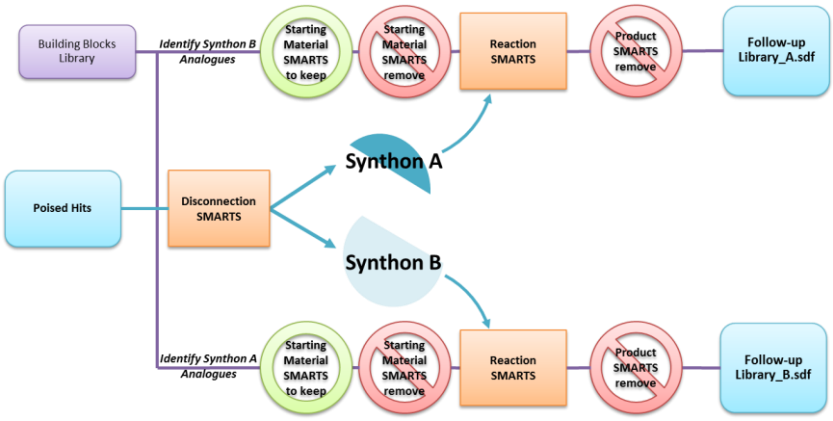


Figure 3.6: Use of SMARTS in XPoise. Reaction SMARTS (orange boxes) and substructure filters (green and red circles) were implemented using RDKit nodes for KNIME. The SMARTS strings used for each reaction are listedin Appendix D.

[1] www.rdkit.org, accessed 3rd October 2016

[2] www.rdkit.org/RDKit_Docs.current.pdf, accessed 3rd October 2016

[3] tech.knime.org/community/rdkit, accessed 3rd October 2016

[4] smartsview.zbh.uni-hamburg.de/smartsview/view, accessed 3rd October 2016

A REOS filter (described in the previous chapter) is used to ensure promiscuous and reactive groups are removed from the building blocks library. When a reactive handle is required for the reaction (e.g. acid chlorides for amide coupling or aldehydes for reductive amination), the REOS filter was reconfigured to keep the desired functional group or the filter is applied only to the reaction products.

Phosphorus is used as a dummy-atom in the reaction centre for three reactions (S_NAr, reductive amination and *N*-alkylation) to aid the removal of unwanted products. The undesirable products are the result of deactivated amide or sulfonamide nitrogen atoms taking part in the encoded reaction (Figure 3.7).

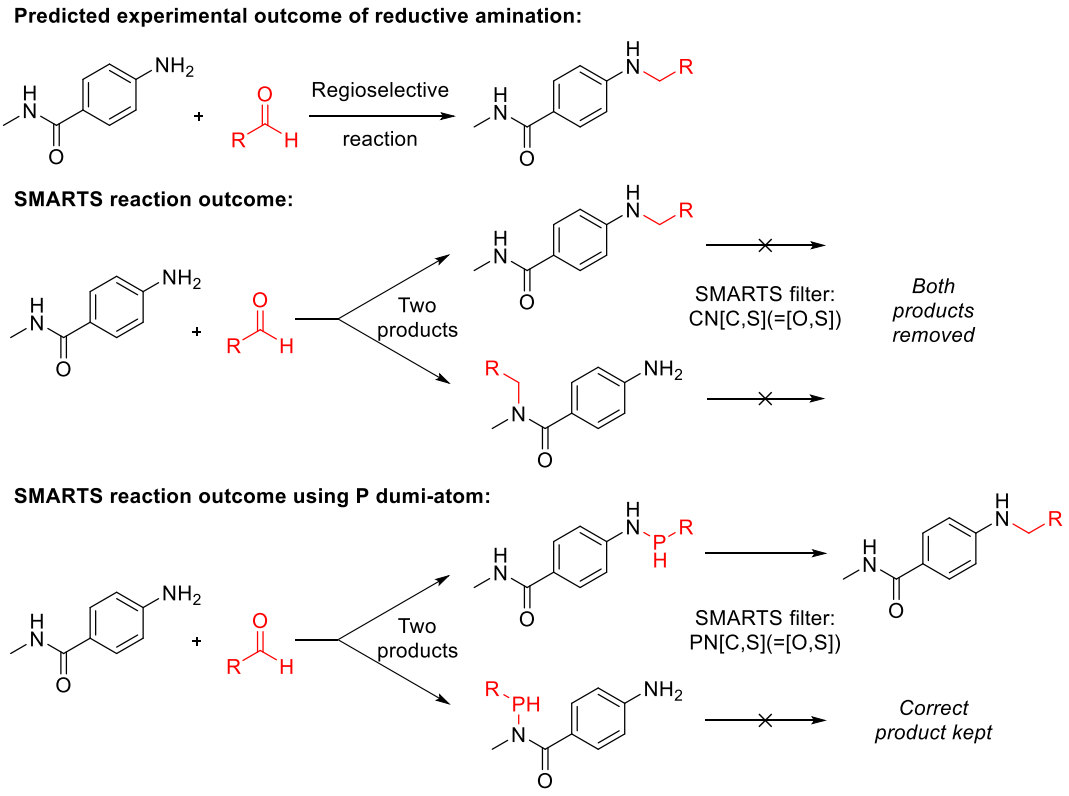


Figure 3.7: Use of a phosphorus dummy-atom in reaction SMARTS to aid the removal of unwanted products. A one-component SMARTS reaction is performed after the filter to convert [PH] into [CH_2].

Starting materials with multiple reactive handles (and therefore likely to face regioselectivity issues) are identified by counting the number of reaction products per starting material. For each reaction, building blocks yielding two or more products were removed (Figure 3.8).

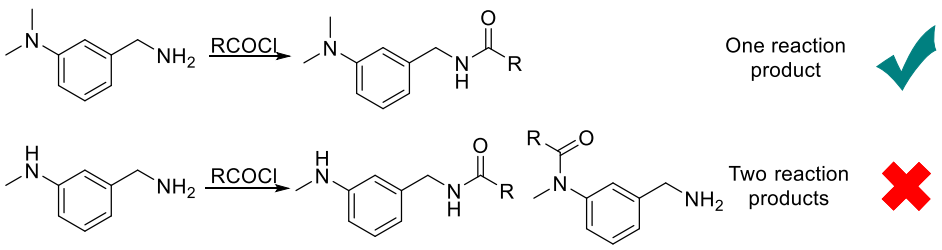


Figure 3.8: Reaction regioselectivity was ensured by removing building blocks which afforded two or more reaction products.

An alternative method was explored in which the number of each functional group on each building block were counted, and compounds with two or more of the same reactive handle were removed. However, this was found to be less reliable and less computationally efficient than counting reaction products.

3.4 Results

3.4.1 Overview

The XPoise workflow creates virtual libraries of follow-up compounds for poised fragment hits using commercially available building blocks. A library is generated for each poised synthon and each library saved as a separate SDF file. Chemical property filters are used to ensure each library contains no more than 3000 compounds, meaning the libraries are amenable to LLOOMMPPAA analysis without further manipulation. Screening libraries are also mined using Murcko scaffold substructure searching and Morgan fingerprint similarity, to aid the use of SAR by catalogue studies.

3.4.2 The XPoise Workflow

XPoise analyses a CSV file containing fragment SMILES and their IDs identified as hits in a crystallographic fragment screen. The fragments are subjected to the disconnection SMARTS to identify all possible poised synthons. The synthons are then reacted with the pre-filtered building blocks library using the relevant poised reaction SMARTS (Figure 3.9). Building blocks which are found to yield two or more products are removed. The resulting compounds are converted to SDF format and drug-likeness ensured by filtering compounds that do not contain a ring or have SlogP outside the range zero to four. The fragment ID, fragment SMILES, reaction ID and the context SMILES are all added to the data table, as well as the CAS number, price and supplier of each building block. Finally, the SDF files are saved in the same location as the input CSV file using a naming system which includes the hit ID, the reaction used, the number of compounds in the library and the Context SMILES.

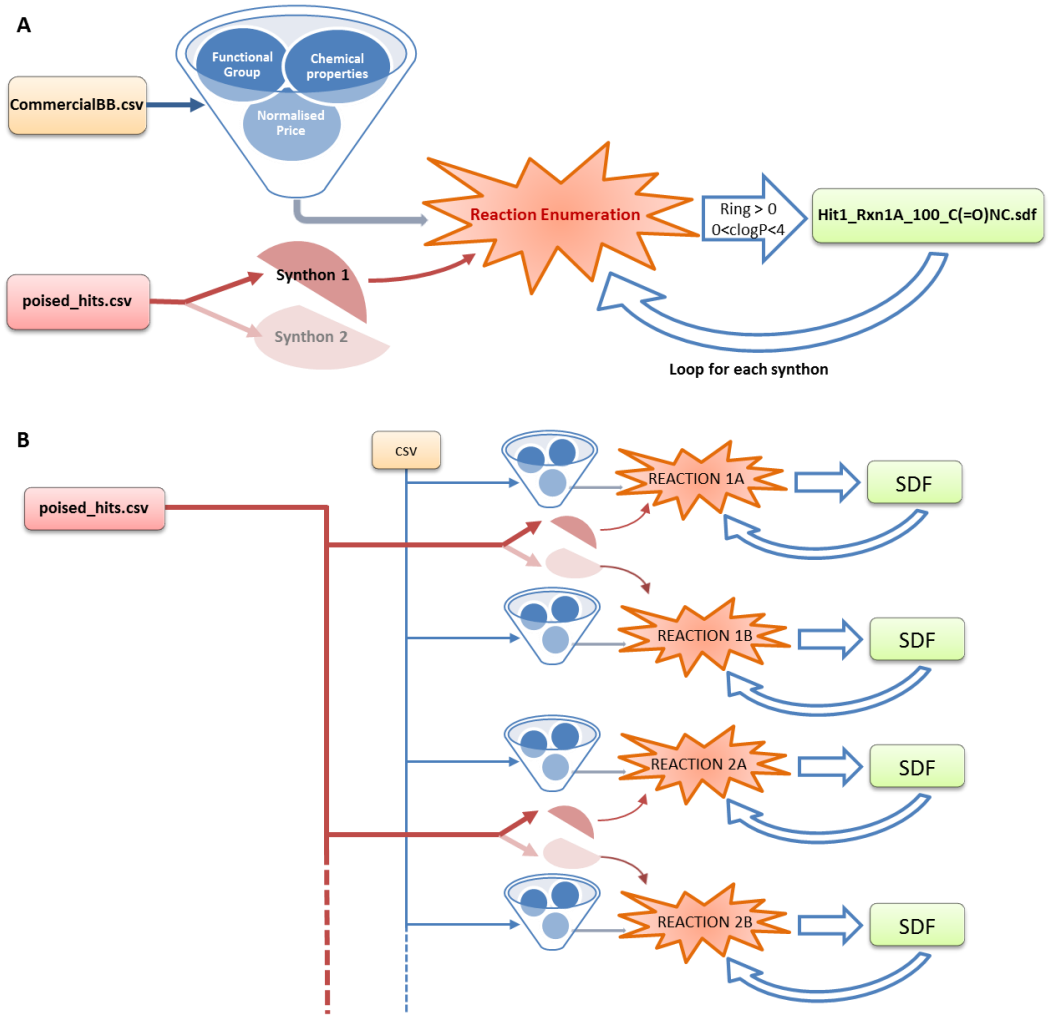


Figure 3.9: Schematic representation of the XPoise workflow. (A) For each reaction, the building blocks library (CommercialBB.csv) is filtered using substructure matching, chemical properties and price filters. Each synthon identified for the reaction is enumerated sequentially and the SDF files saved individually. (B) The analysis of each reaction is performed in parallel meaning multiple compounds can be analysed simultaneously.

In the following sections I detail the XPoise SMARTS filters, SMARTS reactions and property filters used to perform the amide coupling reaction, the benzimidazole formation reaction and to build SAR by catalogue follow-up libraries. Full tables containing SMARTS and filters for each poised reaction are listed in Appendix D.

3.4.2.1 Example 1: Amide coupling

In this section I detail the XPoise workflow used to combine an acid chloride synthon with amine building blocks to form a virtual library of amides. Visual representations of reaction SMARTS and filter SMARTS are depicted in Figure 3.10. Table 3.1 details all filters and SMARTS strings used in the workflow.

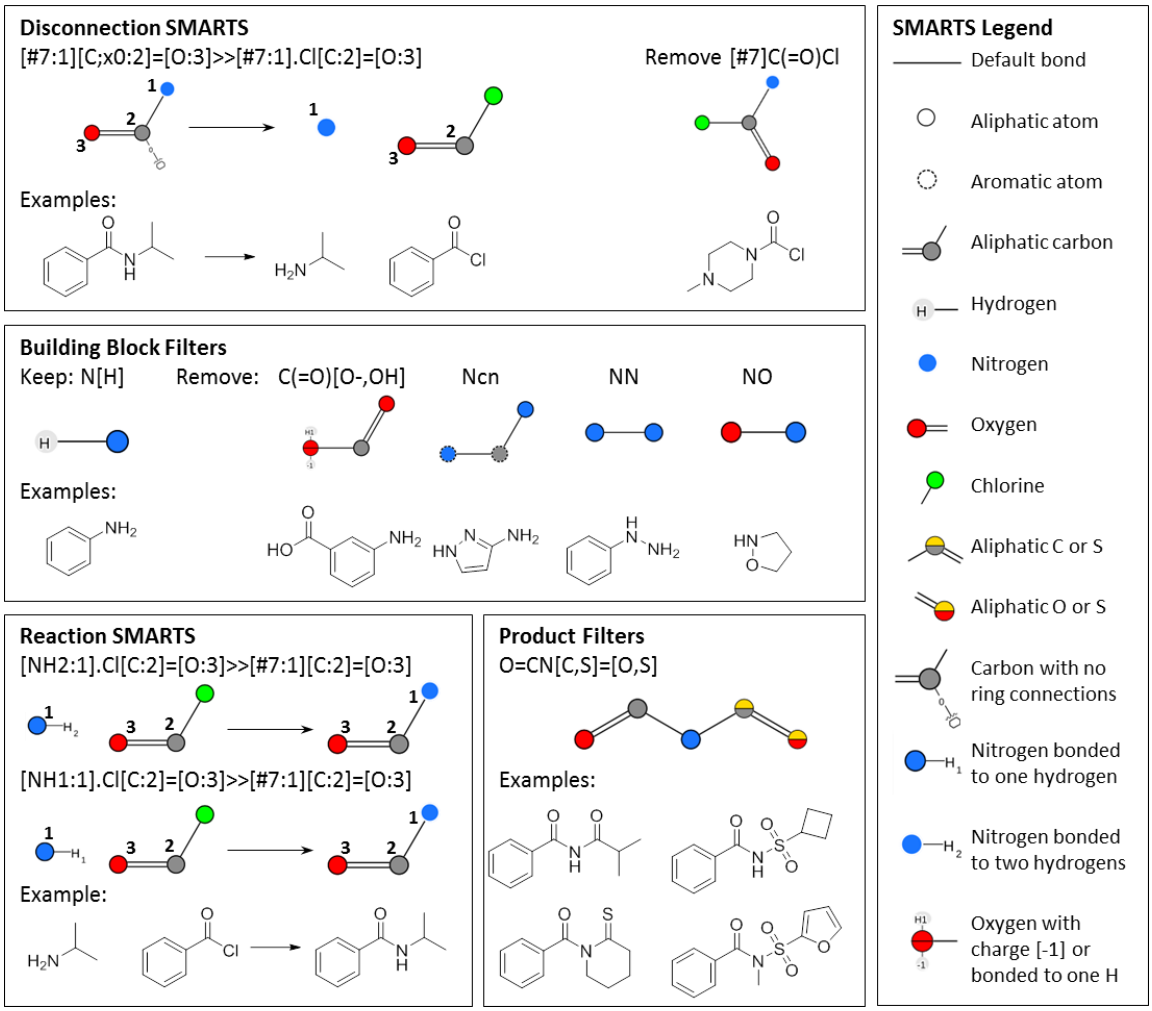


Figure 3.10: 2D representations of the SMARTS strings used generate libraries of amides, including example compounds. Images created using SMARTSviewer.[1] Copyright: ZBH Center for Bioinformatics Hamburg.

First, an amide bond is identified and disconnected if the carbonyl carbon is not part of a ring system. The amide nitrogen can be aromatic or aliphatic. The disconnection gives an amine and

[1] smartsview.zbh.uni-hamburg.de/smartsview/view, accessed 3rd October 2016

an acid chloride synthon. The acid chloride synthon is identified and carried forward for enumeration unless it is the product of urea disconnection.

XPoise Step	Description	Filter value	SMARTS
Disconnection SMARTS	Identify poised synthons		[#7:1][C;x0:2]=[O:3]>>[#7:1].Cl[C:2]=[O:3]
	Remove urea synthons	0	[#7]C(=O)Cl
Building block filters	Find amines	≥ 1	N[H]
	REOS filter		
	Remove carboxylic acids	0	C(=O)[O-,OH]
	Remove deactivated amines	0	Ncn
	Remove hydrazines and hydroxylamines	0	NN NO
	Rotatable bond filter	≤ 3	
	Heavy atom filter	≤ 14	
	Price filter	≤ 100 GBP/g	
Reaction SMARTS	For primary amines		[NH2:1].Cl[C:2]=[O:3]>>[#7:1][C:2]=[O:3]
	For secondary amines		[NH1:1].Cl[C:2]=[O:3]>>[#7:1][C:2]=[O:3]
Product filters	Reaction with primary (sulfon)amides	0	O=CN[C,S]=[O,S]
	Reaction with secondary (sulfon)amides	0	O=CN[C,S]=[O,S]
	Count reaction products	1	
	Number of rings	≥ 1	
	SlogP	0 ≤ SlogP ≤ 4	

Table 3.1: Filters and SMARTS reactions used to generate amide libraries from an acid chloride synthon and a library of amine building blocks.

Separately, the building blocks library is filtered to identify primary and secondary amines. A REOS filter is applied, and any carboxylic acids removed to avoid self-condensation of the building block in the amide coupling reaction conditions. Further undesirable amines are removed, including hydrazines, hydroxylamines and amines deactivated by the presence of a neighbouring heterocycle. Property filters (NRB, HA and price) are applied to the remaining amines before reaction with the acid chloride synthon using the reaction SMARTS. Primary and secondary amines are encoded for separately. Products generated as the result of amides or sulfonamides reacting with the acid chloride synthon are removed. If a building block generates multiple products they are also removed. Finally, the library is filtered to ensure each compound has one or more ring system and SlogP between zero and four.

3.4.2.2 Example 2: Benzimidazole formation

In this section I detail the XPoise workflow used to combine a 1,2-dianiline synthon with carboxylic acid building blocks to form a virtual library of benzimidazoles. Visual representations of reaction SMARTS and filter SMARTS are depicted in Figure 3.11. Table 3.2 outlines the filters and SMARTS strings used in the workflow.

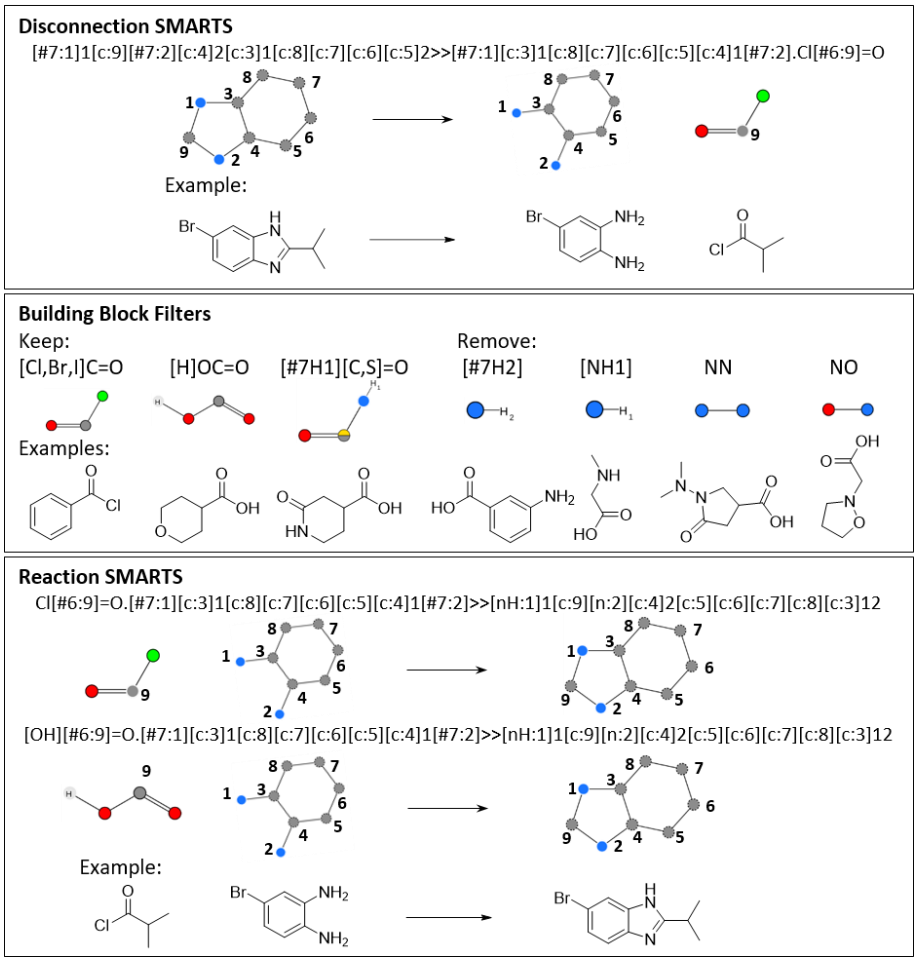


Figure 3.11: 2D representations of the SMARTS strings used generate libraries of benzimidazoles, including example compounds. For legend see Figure 3.10. Images created using SMARTSviewer.[1] Copyright: ZBH Center for Bioinformatics Hamburg.

First, a benzimidazole is identified and fragmented into dianiline and acid chloride synthons. Separately, the building blocks library is filtered to identify acid halides and carboxylic acids. A

[1] smartsview.zbh.uni-hamburg.de/smartsview/view, accessed 3rd October 2016

REOS filter is applied, and any additional amines are removed to avoid the formation of an amide bond in the reaction conditions, including hydrazines and hydroxylamines. Property filters (NRB, HA and price) are applied to the remaining acids before reaction with the dianiline synthon using the reaction SMARTS.

XPoise Step	Description	Filter value	SMARTS
Disconnection SMARTS	Identify poised synthons		[#7:1]1[c:9][#7:2][c:4]2[c:3]1[c:8][c:7][c:6][c:5]2>> [#7:1][c:3]1[c:8][c:7][c:6][c:5][c:4]1[#7:2].Cl[#6:9]=O
Building block filters	Find acid chlorides	≥ 1	[Cl]C=O
	Find carboxylic acids	≥ 1	[H]OC=O
	Remove primary amines	0	[#7H2]
	Remove secondary amines	0	[NH1]
	Keep secondary (sulfon)amides	≥ 1	[#7H1][C,S]=O
	REOS filter		
	Remove hydrazines and hydroxylamines	0	NN NO
	Rotatable bond filter	≤ 3	
	Heavy atom filter	≤ 14	
	Price filter	≤ 200 GBP/g	
Reaction SMARTS	For acid chlorides		Cl[#6:9]=O.[#7:1][c:3]1[c:8][c:7][c:6][c:5][c:4]1[#7:2]>> [nH:1]1[c:9][n:2][c:4]2[c:5][c:6][c:7][c:8][c:3]12
	For carboxylic acids		[OH][#6:9]=O.[#7:1][c:3]1[c:8][c:7][c:6][c:5][c:4]1[#7:2]>>[nH:1]1 [c:9][n:2][c:4]2[c:5][c:6][c:7][c:8][c:3]12
Product filters	Count reaction products	≤ 2	
	Number of rings	≥ 1	
	SlogP	0 ≤ SlogP ≤ 4	

Table 3.2: Filters and SMARTS reactions used to generate benzimidazole libraries from a 1,2-dianiline synthon and a library of acid chloride and carboxylic acid building blocks.

Benzimidazole formations using acid chloride and carboxylic acid building blocks are encoded for separately. The symmetry of the dianiline synthon means two identical products are produced per acid moiety (Figure 3.12). Therefore only if a building block generates four or more products is it removed. Finally, the library is filtered to ensure each compound has one or more ring system and SlogP between zero and four.

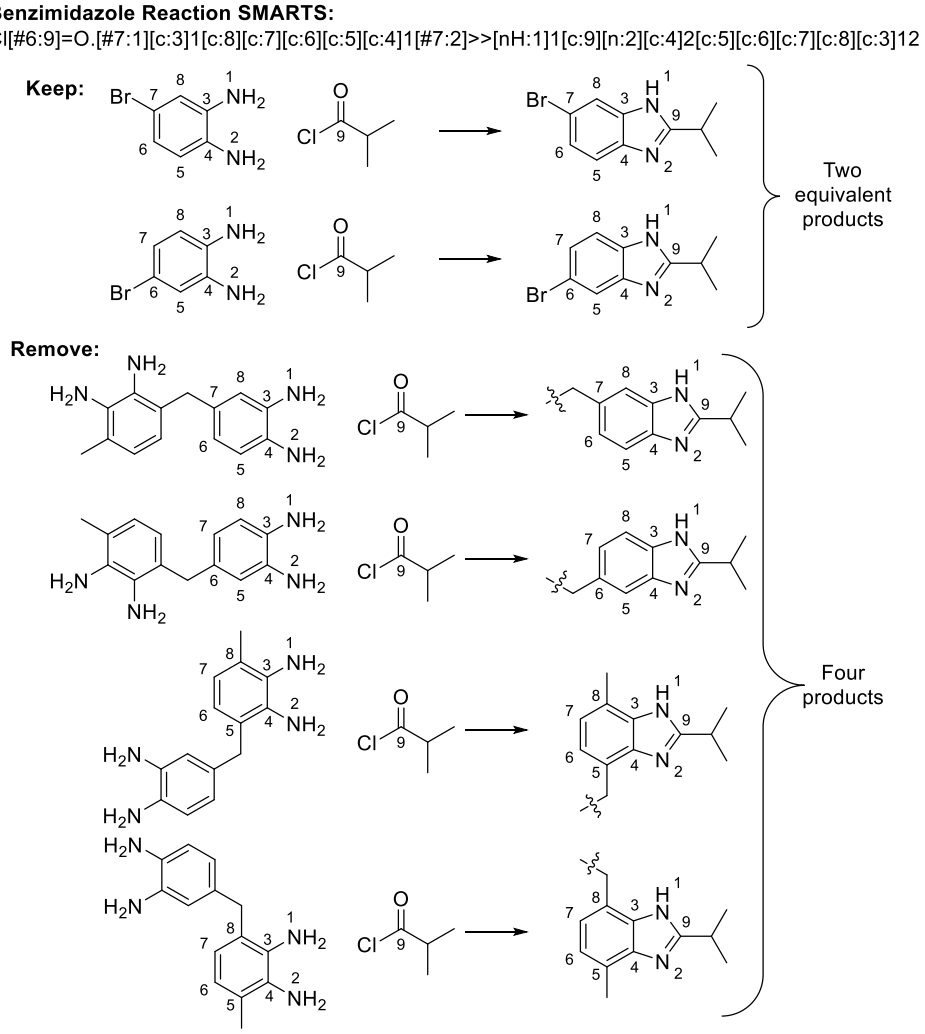


Figure 3.12: How atom numbering results in the formation of two equivalent benzimidazole products from one acid chloride building block. As a result, only the building blocks which generate four or more products are removed.

3.4.2.3 Example 3: SAR by catalogue

In this section I detail how the XPoise workflow enables SAR by catalogue by mining a screening library for structural analogues. SAR by catalogue is a rapid and low cost method used to explore the potential of a fragment for hit-to-lead development (Keseru, Erlanson *et al.* 2016). The purchase of catalogue compounds from commercial screening collections allows researchers to test compounds without incurring the expense or time delays associated with organic synthesis. Most commonly a similarity search is performed to identify structural analogues, and is normally

performed using online catalogues such as eMolecule[1] or Sigma Aldrich.[2] XPoise automatically searches a user-defined screening library for the twenty nearest neighbours to each fragment hit (Figure 3.13 A). Similarity is measured using Morgan fingerprints (radius = 2) as described in the previous chapter, and nearest neighbours identified using the Tanimoto distances between fingerprints.

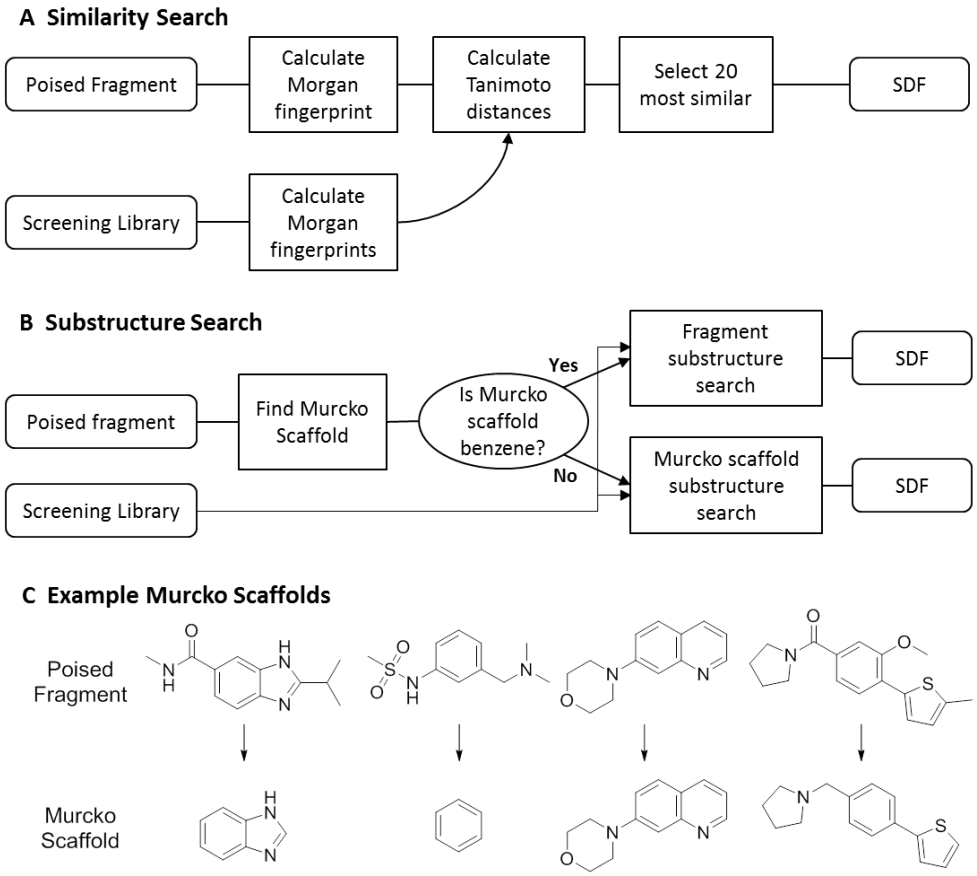


Figure 3.13: SAR by catalogue using XPoise. (A) Identification of structural analogues to a poised fragment in a commercial screening library using fingerprint similarity. (B) Identification of screening library compounds containing a common substructure to a poised fragment. The large number of commercially available compounds with a benzene Murcko scaffold means a substructure search of the whole compound is performed on fragments with a benzene Murcko scaffold. (C) Examples of Murcko scaffolds. Ring system and linker atoms are retained while side chain atoms are removed.

[1] www.emolecules.com, accessed 3rd October 2016

[2] www.sigmaaldrich.com, accessed 3rd October 2016

An alternative method of SAR by catalogue is substructure matching. Substructure searching allows motifs important for binding to be maintained and offer a structural template which can be used as the LLOOMMPPAA Context SMILES. XPoise performs a substructure search of a user-defined screening library using the Murcko scaffold of each fragment hit (Figure 3.13 B). The Murcko scaffold is a computationally derived framework of a molecule composed of the ring system atoms and linking atoms but with side chain atoms removed (Figure 3.13 C) (Bemis and Murcko 1996). If a fragment is found to have benzene as a Murcko scaffold, a substructure search is instead performed on the whole fragment. This is to avoid the large number of benzene-containing commercially available compounds that would be returned if a benzene substructure search was performed.

3.5 Ongoing Work and Conclusions

The XPoise workflow has been developed to automatically identify poised synthons from a list of fragment hits and to build a virtual library of follow-up compounds for each synthon. The workflow uses a building blocks library built from the catalogues of eight reliable commercial vendors, meaning each compound in the final library can be synthesised in one chemical step. Bespoke reactions SMARTS and functional group filters are implemented to improve the likelihood of successful follow-up synthesis. Each follow-up library is saved as a separate SDF file and contains data about the hit ID, the reaction used, and building block price, CAS number and suggested supplier. All SDF files produced contain no more than 3,000 compounds and are ready for analysis using LLOOMMPPAA. XPoise also suggests compounds to aid SAR by catalogue studies. Twenty structural analogues are found from a commercial screening library using Morgan similarity, and a separate library of compounds containing the same Murcko scaffold substructure as the original hit are identified.

XPoise is optimised to build virtual libraries for LLOOMMPPAA analysis by varying one synthon at a time. However, the flexibility of the KNIME software means users can readily adapt property filters and reaction protocols to their own requirements. For example, by skipping the Fragment Disconnection step, a user may input a compound with a reactive handle to build a follow-up library of elaborated compounds. Further changes to the input library of building blocks allow researchers to explore different areas of chemical space with different virtual libraries. In fact, given sufficient computational power, a full exploration of poised chemical space can be performed with only minor tweaks to the algorithm. This would represent a full matrix expansion rather than the single vector expansion currently performed by XPoise (Figure 3.14). The resulting library could be used to test novel computational scoring methods, based on pharmacophore matching or docking, because each compound can be easily accessed in one chemical step from commercially available starting points.

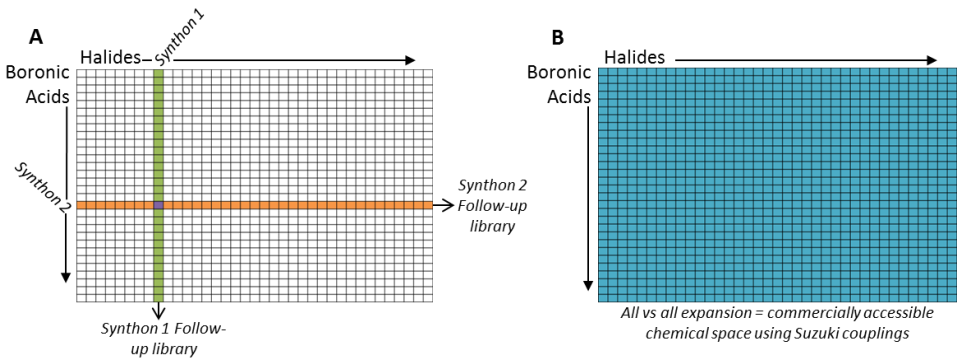


Figure 3.14: Poised expansion of Suzuki coupling building blocks using XPoise. (A) The encoded XPoise method yields single vector coverage of chemical space by only varying one synthon at a time. (B) Illustration of full matrix expansion to generate a complete library of Suzuki coupling products.

The XPoise workflow has been installed at beamline I04-1 at DLS and is used by researchers as a part of their hit discovery programs at the facility. The workflow allows users to identify all poised fragment hits for their crystal system regardless of the fragment library used to perform the initial screen and regardless of their expertise in organic chemistry. Follow-up compounds designed

using XPoise and selected using LLOOMMPPAA have been synthesised and tested against a number of epigenetic compounds. These campaigns are described in the next chapter.

4 Poised Fragments for Epigenetic Targets

The work described in this chapter was carried out in collaboration with Dr Anthony Bradley, Dr Patrick Collins, Dr Tobias Krojer, Dr Romain Talon, Mr Lin Mei, Ms Octovia Monteiro and Mr James Bennett. Dr Bradley helped with the use of LLOOMMPPAA for the selection of compounds for testing. Drs Collins, Krojer and Talon performed fragment soaking experiments and X-ray crystallography data collection and data analysis. I carried out the synthesis of follow-up compounds for BRD1 and PHIP(2). Mr Mei also synthesised a subset of thiourea compounds to target PHIP(2). The AlphaScreen assay was performed by myself (PHIP(2)), Ms Monteiro (BRD1 and PHIP(2)) and Mr Bennett (JMJD2D). The follow-up campaign against PHIP(2) has been published by Cox *et al.* in Chemical Science (Cox, Krojer *et al.* 2016).

4.1 Overview

In this chapter I describe the hit-to-lead development of poised fragments against four epigenetic proteins. Fragment soaking experiments were performed as a primary screen against each target. Poised hits were selected for further investigation based upon binding poses, previous knowledge of the target and binding affinity. The synthesis of follow-up compounds for two poised BRD1 hits yielded insight into the binding modes of both series, but the improvement in potency was not enough to be detected by an activity assay. In contrast, the elaboration of poised hits against PHIP(2) bore a ten-fold increase in activity as well as further structural data. A second phase thiourea library was designed and selected for synthesis using the XPoise/LLOOMMPPAA method. The resulting compounds showed further increases in potency, demonstrating the utility of the LLOOMMPPAA computational methodology in combination with poised fragments for structure-based hit optimisation.

The use of SAR by catalogue for two allosteric binders against the lysine demethylase target, JMJD2D, gave marginal improvements in potency and no further structural data. An analysis of chemical space coverage by SAR by catalogue compared to poised elaboration revealed the benefits of using poised synthesis in a follow-up campaign. Finally, a hit-to-lead campaign for the Tudor domain of SETDB1 using the XPoise/LLOOMMPPAA method demonstrated how outsourcing poised synthesis of a follow-up library to a CRO (contract research organisation) compares favourably with SAR by catalogue in terms of cost and delivery time.

4.2 Background

4.2.1 Fragment Soaking

Beamline I04-1 at DLS[1] has recently seen an order-of-magnitude speed-up in X-ray crystallography thanks to robotic sample handling, improved algorithms and detectors, and additional technical advances within the synchrotron facility. Further developments in crystal soaking, harvesting and data analysis have further reduced the effort to such an extent that a 1000- compound fragment screen by crystallography can be completed within a week (Figure 4.1). As a consequence, X-ray crystallography is now a viable primary screen for academic and industrial researchers. In particular, this makes it realistic to screen by soaking single compounds per crystal, rather than cocktails, so that concentrations >100 mM can be achieved, one hundred to one thousand times higher than possible in most solution-based techniques, with accordingly increased sensitivity and improved hit-rate.

[1] www.diamond.ac.uk/Beamlines/Mx/I04-1.html, accessed 3rd October 2016

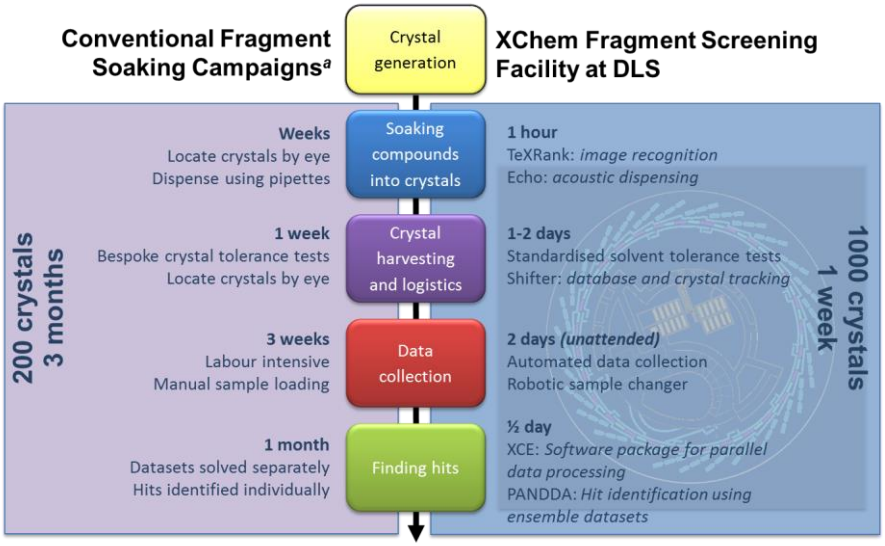


Figure 4.1: Comparison of conventional fragment screening with the DLS XChem facility. Developments in soaking, harvesting and data analysis at DLS have reduced experiment times to such an extent that a 1000-compound fragment screen by crystallography can be completed within a week. Figure adapted with permission from Dr Patrick Collins. More information can be found on the Diamond website.[1] *[a]Estimates based upon data published by Spurlino (2011).*

The fragment screening facility at DLS (XChem) is a four step workflow: compound soaking, crystal harvesting and pre-screening, data collection, and hit identification. Compound soaking has been improved with two key tools. Firstly, an in-house image recognition tool, called TeXRank, allows for more efficient identification and tagging of protein crystals in crystallisation plates (Ng, Dekker *et al.* 2014). Secondly, an Echo liquid handler[2] means fragment solutions can be rapidly and accurately transferred to target plate wells on the nL scale. As a result, 900 crystals can be soaked with unique fragments in just one hour, a process which would have taken at least a week using conventional methods (Spurlino 2011).

The DLS facility has developed a robust solvent tolerance test for protein crystals to ensure a crystal system is suitable for fragment screening. The whole process is recorded using a central database which is automatically updated during crystal harvesting using a shifter: a plate handler

[1] www.diamond.ac.uk/Beamlines/Mx/I04-1.html, accessed 3rd October 2016

[2] www.labcyte.com, accessed 3rd October 2016

designed and built in-house (Wright and Von Delft 2016). As a result, the preparation and pre-screening required for a fragment screen can be completed in one to two days, down from a week previously.

Data collection can be a labour intensive and time consuming process, as each crystal needs to be mounted, aligned and interrogated with human input. The use of robotics and alignment algorithms at DLS means data can be collected unattended in a continuous process. Human intervention is only required when all the crystals in the liquid nitrogen storage tank have been analysed and a new batch is added. As a result, 1000 crystals can be processed in just forty hours with minimal human input and largely unattended. The same number of crystals would previously have taken weeks of manual labour to interrogate.

Software tools developed by SGC researchers have made the processing of large datasets significantly easier. XCE (XChem Explorer) and PANDDA (Pan-Dataset Density Analysis) use the structural similarities between fragment screening datasets to solve structures and identify hits, even if the electron density is too weak for detection using conventional methods (Pearce, Bradley *et al.* submitted). As a consequence, finding hits can take less than a day of computational analysis versus the weeks required in the past (Spurlino 2011).

4.2.2 Fragment Screening of Epigenetic Proteins

Epigenetic proteins regulate the expression of DNA in cells (Section 1.4). One of the predominant regulatory pathways is *via* the covalent modification of histone tails. Histone modifications are effected by reader, writer and eraser proteins. In this chapter I detail hit-to-lead campaigns against reader and eraser proteins from the bromodomain, lysine demethylase and Tudor domain families. All three families are purported to be druggable and have been targeted by various drug discovery campaigns (Pande 2016). Fragment screening has been especially effective in

discovering Brd inhibitors using both focused and diverse libraries (Chung, Dean *et al.* 2012; Chaikuad, Petros *et al.* 2014; Harner, Chauder *et al.* 2014), and fragment hits have been optimized into potent and selective chemical probes (Fish, Filippakopoulos *et al.* 2012; Clark, Vieira *et al.* 2015).

Researchers at the SGC and DLS have performed high concentration fragment soaking experiments against a number of human epigenetic targets. This chapter focuses on follow-up campaigns against four epigenetic targets: the Brds BRD1 and PHIP(2), the KDM JMJD2D, and the Tudor domain of SETDB1. Fragment soaking campaigns, including protein crystallisation, data collection and structure solution, as well as follow-up compound soaking, was performed by Drs Krojer, Talon and Collins. Compounds were dissolved in DMSO or, in the case of Brds, in ethylene glycol. Ethylene glycol was selected as an alternate solvent to avoid DMSO competition, which has a relatively high affinity for Brds (Philpott, Yang *et al.* 2011). The nominal concentration varied between 100-500 mM, with the supernatant of compounds which did not fully dissolve used on the assumption it was a saturated solution. The solutions were soaked into protein crystals in crystal buffer in a 1:1 volume ratio to give an approximate final compound concentration of 50-250 mM. Full experimental conditions are given in Appendix E.

4.2.3 AlphaScreen

In vitro screening of compounds was carried out by myself, Ms Monteiro and Mr Bennett using an amplified luminescent proximity homogeneous assay (AlphaScreen) (Peppard, Glickman *et al.* 2003; Kawamura, Tumber *et al.* 2010; Philpott, Yang *et al.* 2011). The AlphaScreen assay is a fluorescence-based proximity assay; when phthalocyanine-containing donor beads are irradiated at 680 nm, ambient oxygen is excited to the singlet state. The half-life of singlet oxygen in aqueous solution is 4 μsec (Ullman, Kirakossian *et al.* 1996), and as a result the distance it can diffuse before

it decays is limited. The concentration of singlet oxygen decays exponentially with distance from the donor bead surface and at 200 nm is reduced by approximately 97% (Ullman, Kirakossian *et al.* 1994). If singlet oxygen is formed in close proximity to an appropriately derivatised acceptor bead and therefore within the 200 nm diffusion limit, energy is transferred to thioxene derivatives contained in the acceptor beads resulting in an emission of light at 520–620 nm (Figure 4.2) (Peppard, Glickman *et al.* 2003). The experimental conditions used for each protein are listed in Appendix E.

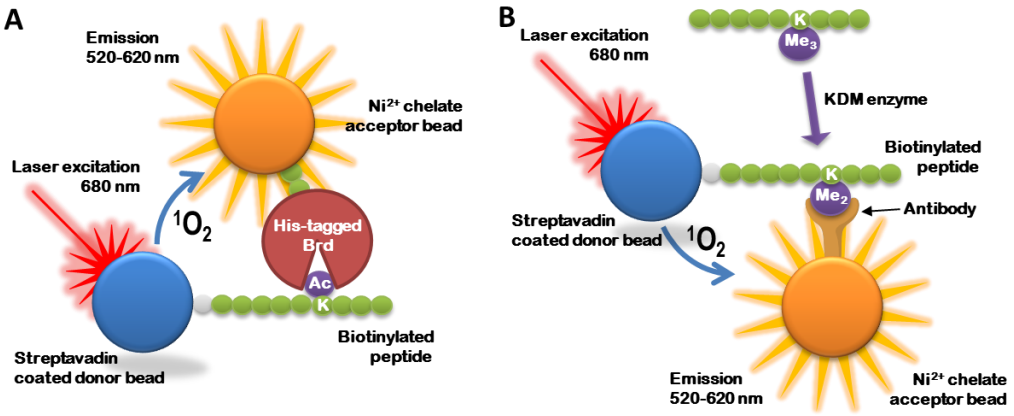


Figure 4.2: Schematic depicting the AlphaScreen peptide displacement assay. (A) His-tagged Brd binds to the KAc peptide, bringing the donor and acceptor beads into close proximity. Presence of an inhibitor disrupts KAc binding causing a reduction in light emission (Philpott, Yang et al. 2011). (B) The KDM enzyme converts Kme3 to Kme2. The Kme2 peptide is recognised by an antibody, bringing the donor and acceptor beads into close proximity. The presence of an inhibitor slows the production of the Kme2 peptide, and thereby reduces light emission (Kawamura, Tumber et al. 2010).

The Brd AlphaScreen uses the biotinylated natural KAc peptide of the relevant Brd (Philpott, Yang *et al.* 2011). Binding between the peptide and the His-tagged Brd brings the donor and acceptor beads into close proximity, causing an increase in light emission (Figure 4.2 A). The presence of an inhibitor disrupts Brd-KAc binding, resulting in a reduction of light emission.

In the KDM AlphaScreen assay, the Kme2 peptide product of the relevant KDM is detected by the acceptor bead (Figure 4.2 B) (Kawamura, Tumber *et al.* 2010). Biotinylated peptide substrate is used so that it can bind to donor beads through a biotin–streptavidin interaction. An antibody

with selectivity for the Kme2 peptide is used to couple the substrate to acceptor beads. The donor bead is brought into close proximity with acceptor beads only when the methyl mark of interest is present so that demethylation of the substrate results in a gain of fluorescence and is modulated by the presence of inhibitors.

4.2.4 Project Aim

To demonstrate the utility of poised fragment based drug discovery I wanted to perform hit development campaigns against a range of target proteins with different binding sites and using distinct chemotypes. I aimed to design and synthesise follow-up compounds using poised chemistry in order to gain insight into the binding site of each protein, and where possible improve the activity of the series. I hoped to identify one or more lead series for each target which could be taken forward with a full medicinal chemistry program (Figure 4.3). The proteins selected for the project were the highly druggable Brd, BRD1; the atypical and previously untargeted Brd, PHIP(2); the KDM, JMJD2D, for which allosteric fragment binders had been identified; and the Tudor domain of SETDB1.

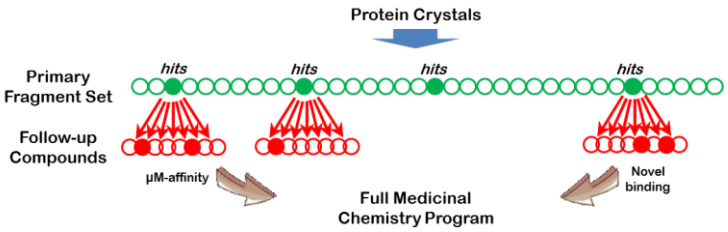


Figure 4.3: Hit-to-lead development of poised fragment hits. Multiple fragment hits obtained from a crystal soaking screen can be investigated by the synthesis and testing of follow-up compounds. The best series can then be taken forward for lead optimisation using a full medicinal chemistry program. Figure adapted with permission from Prof Frank von Delft.

Hits were obtained by SGC researchers using the fragment soaking facility offered at DLS. Each protein was screened using in-house fragment libraries and analysed using the WONKA evaluation tool (Bradley, Wall *et al.* 2015). Hits for follow-up campaigns were selected based upon poised

functionality and interesting binding features. In house AlphaScreen assays were used to track changes in activity against the relevant proteins.

4.3 Brd BRD1

4.3.1 BRD1 Fragment Soaking Campaign

BRD1, also known as BRPF2, is in subfamily *IV* of the Brd phylogenetic tree and is structurally similar to BRPF1A, BRPF1B and BRPF3 (Figure 1.10, page 24). The BRD1 gene has been associated to schizophrenia and bipolar affective disorder (Bjarkam, Corydon *et al.* 2009; Nyegaard, Severinsen *et al.* 2010). The protein structure is characteristic of the Brd family with a conserved asparagine residue acting as the KAc binding anchor and four structurally conserved water molecules at the back of the binding site (Figure 4.4). The BRPF-selective chemical probes, OF-1 and NI-57, both inhibit BRD1 *in vitro* with IC_{50} values of 500 nM and 108 nM respectively.[1] Recently a TAF1 chemical probe, BAY-299, was released which also inhibits BRD1 (IC_{50} 6 nM) although neither the structure of the probe nor its binding mode have been disclosed.

The apo-bromodomain structure of human BRD1 was deposited in 2011 (PDB ID 3RCW) (Filippakopoulos, Picaud *et al.* 2012). Subsequently the ligand-protein complex of OF-1 in complex to BRD1 was deposited in 2016 (PDB ID 5FG6, Figure 4.4). Ligand binding is anchored by H-bonds from the carbonyl head group of OF-1 to N642 and water *I*, characteristic of KAc mimetic binding. Potency is further improved by the formation of a π-stacking interaction between F648 and the OF-1 benzimidazolone core.

[1] www.thesgc.org/chemical-probes, accessed 3rd October 2016

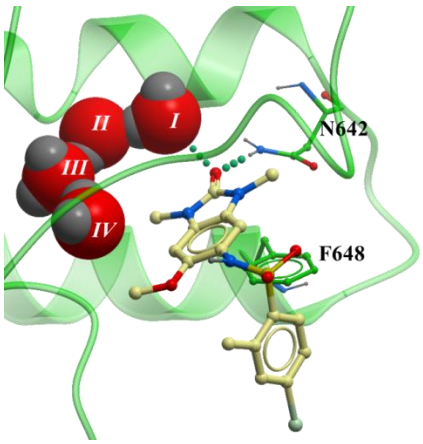


Figure 4.4: BRD1 (green ribbon) in complex to OF-1 (beige sticks). The carbonyl oxygen acts as the KAc mimetic by H-bonding (dotted lines) to N642 (green sticks) and water I. The N-methyl lies in the pocket formed by the canonical network of four water molecules numbered I-IV (red CPK). F648 (green sticks) forms a π-stacking interaction with OF-1 (PDB ID 5FG6).

Researchers at the SGC screened three internal fragment libraries against BRD1 using the XChem fragment soaking protocol outlined in the previous section. The first library was a 12-member Brd-focussed library built using fragments containing commonly observed KAc mimetics including cyclic and acyclic amides. The second library was a bespoke 60-member library synthesised by a collaborator, Dr John Spencer, containing a diverse set of heterocycles. The final library was composed of over 1000 poised compounds from the 3D-fragment consortium (Section 2.6.2, Morley, Pugliese *et al.* 2013) and a commercially available diverse Ro3 library purchased from Maybridge.[1] Poised compounds were identified using the KNIME protocol outlined in Chapter 2. Twenty-five hits (2.1% hit rate) were identified binding to the KAc binding site (unpublished results, Figure 4.5). The structures of all fragment-protein complexes are available online (Bradley, Wall *et al.* 2015).[2]

[1] www.maybridge.com/portal/alias__Rainbow/lang_en/tabID_230/DesktopDefault.aspx, accessed 4th October 2016

[2] http://wonka.sgc.ox.ac.uk/WONKA/2/Summarise/, accessed 3rd October 2016

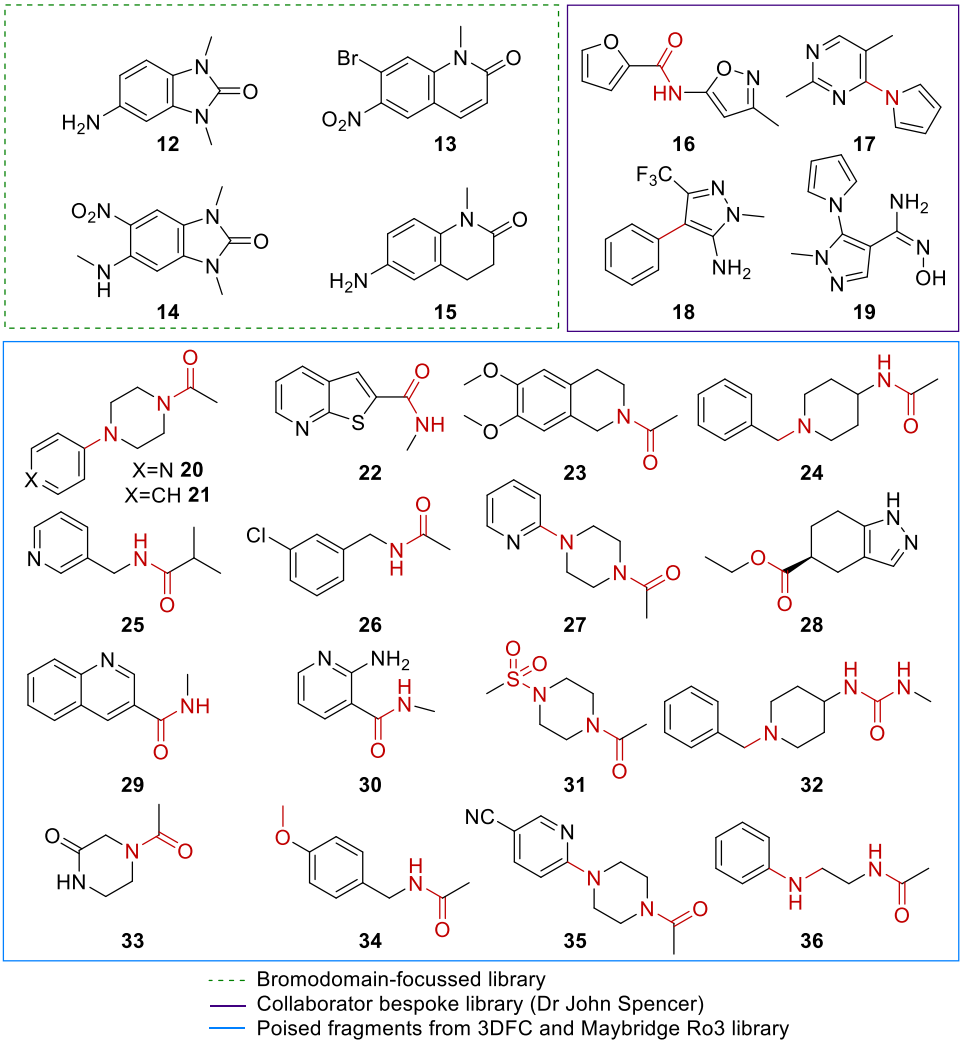


Figure 4.5: Fragment hits of BRD1 identified by fragment soaking experiments performed at DLS (unpublished results). Poised bonds are highlighted in red. Hit compounds are grouped by screening library.

The majority of the hits are poised thanks to the presence of an amide bond. Other poised bonds include S_NAr and reductive amination products, ureas and sulfonamides. Eight of the hits contain two poised bonds, offering more scope in potential follow-up campaigns.

Over 80% of the BRD1 fragment hits contain an amide KAc mimetic, which form H-bonds with N642. Three compounds, **17**, **18** and **19**, have *N*-containing heterocycles acting as the KAc mimetic. Only one compound, **28**, does not form a H-bond to N642. Instead the pyrazole ring displaces the first of the conserved water molecules to form a H-bond to Y599. Compound **25** also

displaces conserved water molecules. The pyridine ring displaces waters *III* and *IV*, allowing the amide group to act as the KAc mimetic.

4.3.2 BRD1 Follow-up Campaigns

Poised hits **17** and **25** were selected for follow-up campaigns because of their binding poses (Figure 4.6). Both fragments have unique binding features: pyrimidine **17** occupies an area of the binding site not occupied by any other screening hit; while amide **25** displaces two of the structural waters at the back of the binding site.

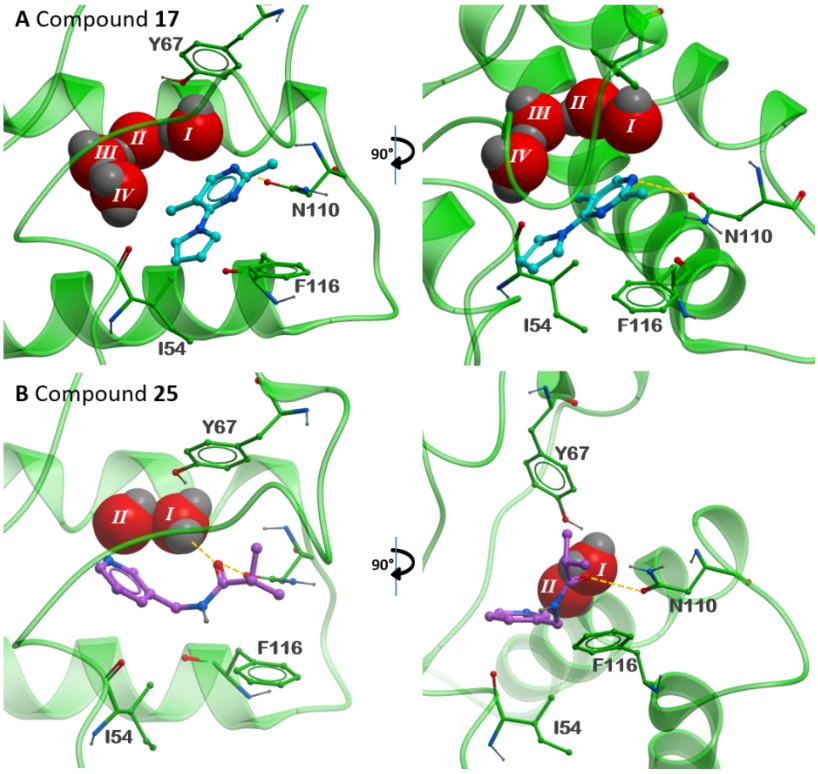


*Figure 4.6: Fragment hits of BRD1 selected for follow-up campaigns. (A) Pyrimidine **17** (cyan sticks) binding to BRD1 (green ribbon and sticks). Pyrimidine nitrogen forms a H-bond to N110 (orange dashes), and the pyrrole occupies a lipophilic pocket formed by I54. The 5-methyl group occupies the hydrophobic pocket formed by the four canonical water molecules (red CPK). (B) Amide **25** (magenta sticks) binding to BRD1 (green ribbon and sticks). The amide oxygen acts as a KAc mimetic, forming a H-bond to N110 (orange dashes). The pyridine ring displaces water III and IV, leaving only two of the Brd waters (red CPK).*

Compound **17** acts as a KAc mimetic by forming a H-bond to N110 *via* a pyrimidine nitrogen (Figure 4.6 A). The pyrrole group occupies a lipophilic shelf formed by I54 in an area of the binding site

under-explored by the other fragment hits. Compound **17** is the product of an S_NAr reaction between 2,5-dimethyl-4-chloropyrimidine **37** and pyrrole **38** (Figure 4.7).

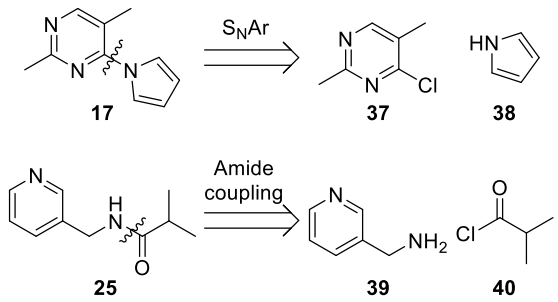
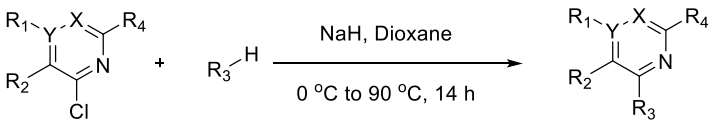


*Figure 4.7: Retrosynthesis of BRD1 hits **17** and **25**. Both fragments can be synthesised in one step using a poised reaction.*

Amide **25** acts as a KAc mimetic by forming a H-bond to N110 through the amide oxygen (Figure 4.6 B). The 3-pyridine group displaces waters *III* and *IV* of the Brd water network, and appears to form a H-bond to water *II*. The fragment is the product of an amide coupling reaction between pyridin-3-ylmethanamine **39** and isobutyryl chloride **40** (Figure 4.7).

4.3.2.1 Pyrimidine KAc mimetic

A follow-up library containing fifteen analogues of compound **17** was proposed and synthesis attempted using parallel chemistry. Six compounds contained changes to the pyrimidine head group, and the other nine compounds were designed to find alternatives to the pyrrole substituent. In total, seven compounds were synthesised (47% success rate) with an average yield of 26% (Table 4.1). The low success rate may be attributed to the harsh reaction conditions used to activate aromatic amines as nucleophiles causing degradation of the pyrimidine starting material (e.g. compounds **44** and **46**) or causing non-specific reactivity of aliphatic amines (e.g. compounds **52**, **53** and **54**). The original hit, **17**, was also resynthesised in 46% yield.



17, 41-55

Compound	R_1	R_2	R_3	R_4	X	Y	Yield[a]
17	H	CH_3	Pyrrol-1-yl	CH_3	N	C	46%
41	-	CH_3	Pyrrol-1-yl	CH_3	CH	N	27%[b]
42	CH_3	H	Pyrrol-1-yl	H	N	C	NI
43	Benzo		Pyrrol-1-yl	H	N	C	37%
44	H	Cl	Pyrrol-1-yl	Cl	N	C	NI
45	H	CH_3	Pyrrol-1-yl	Cl	N	C	43%
46	H	Br	Pyrrol-1-yl	Cl	N	C	NI
47	H	CH_3	Imidazol-1-yl	CH_3	N	C	NI
48	H	CH_3	2,4-Dimethylpyrrol-1-yl	CH_3	N	C	9%
49	H	CH_3	2-Cyanopyrrol-1-yl	CH_3	N	C	6%
50	H	CH_3	Indol-1-yl	CH_3	N	C	43%
51	H	CH_3	Phenylamino	CH_3	N	C	18%
52	H	CH_3	Benzylamino	CH_3	N	C	NI
53	H	CH_3	Pyrrolidin-1-yl	CH_3	N	C	NI
54	H	CH_3	4-methylpiperazin-1-yl	CH_3	N	C	NI
55	H	CH_3	Morpholino	CH_3	N	C	NI

*Table 4.1: Synthesis of structural analogues to BRD1 hit **17**. Reactions performed in parallel using S_NAr conditions. [a]NI: Product not isolated. [b]Reagents and conditions: Cs_2CO_3, DMF, 100 °C, 14 h.*

The seven pyrimidine analogues, plus the original hits, were tested using single shot AlphaScreen experiments at a compound concentration of 100 µM and 200 µM. No BRD1 inhibition was observed for any of the compounds at either concentration. Fragment soaking yielded one new dataset. Pyrimidine **51** is bound to the BRD1 KAc recognition pocket, but with a different binding pose to pyrimidine **17**. The heterocycle core of compound **51** is flipped 180° in comparison to the original hit, losing an H-bond to N110 but gaining one to water *I* (Figure 4.8). However, the amine linker allows the phenyl ring to occupy the same lipophilic pocket as the pyrrole ring present in compound **17**.

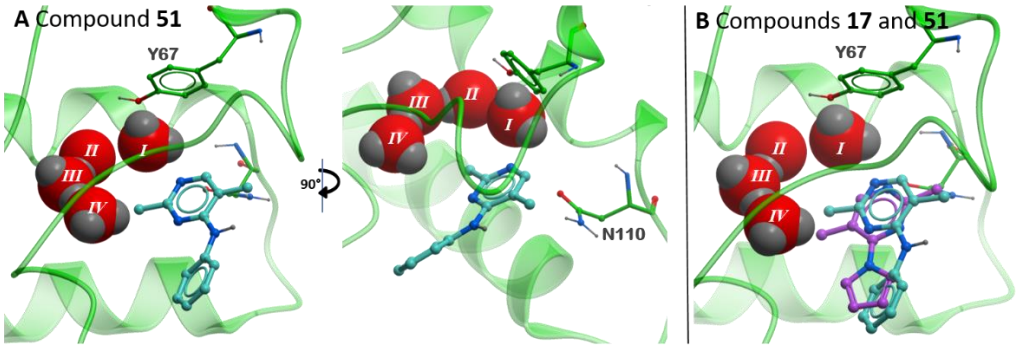
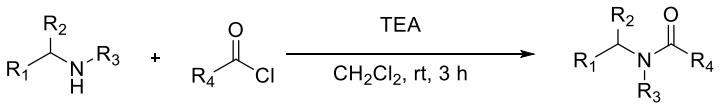


*Figure 4.8: (A) Binding mode of follow-up **51** (cyan sticks) to BRD1 (green ribbon and sticks). The 2-methyl substituent of the pyrimidine ring occupies the pocket made by the structural waters (red CPK), leading to the formation of an H-bond between the pyrimidine nitrogen and water I. (B) An overlay of pyrimidine hit **17** (magenta sticks) and follow-up **51** (cyan sticks). The 4-pyrrolyl and 4-phenylamino groups occupy the same lipophilic pocket made by I54.*

4.3.2.2 Amide KAc mimetic

A follow-up library containing nineteen analogues to amide **25** was proposed and synthesis attempted using parallel chemistry. Ten compounds explored alternatives to the pyridine head group, including possible bioisosteres such as tetrahydropyran (THP) and pyrazine. The remaining nine compounds used a range of acid chlorides to explore the potential formation of new interactions in place of the isopropyl group. In total sixteen analogues (84% success rate) were synthesised (Table 4.2). Synthesis of two compounds was not attempted because the starting materials were not delivered in time (pyridine **61** and oxazole **62**). Synthesis of alcohol **69** failed as a result of reaction regioselectivity. The original hit, **25**, was also resynthesised in 82% yield.



25, 56-74

Compound	R_1	R_2	R_3	R_4	Yield[a]
25	Pyridin-3-yl	H	H	^{i}Pr	82%
56	Pyridin-2-yl	H	H	^{i}Pr	ND
57	Pyridin-3-yl	CH_3	H	^{i}Pr	ND
58	Pyridin-4-yl	H	H	^{i}Pr	ND
59	Pyridin-3-yl	H	CH_3	^{i}Pr	ND
60	Pyrazin-2-yl	H	H	^{i}Pr	ND
61	4-Methylpyridin-3-yl	H	H	^{i}Pr	NA
62	Oxazol-4-yl	H	H	^{i}Pr	NA
63	THP-3-yl	H	H	^{i}Pr	ND
64	Phenyl	H	H	^{i}Pr	ND
65	THP-4-yl	H	H	^{i}Pr	ND
66	Pyridin-3-yl	H	H	Phenyl	82%
67	Pyridin-3-yl	H	H	Dimethylamino	72%
68	Pyridin-3-yl	H	H	3-Methoxybenzyl	76%
69	Pyridin-3-yl	H	H	Hydroxymethyl	NI
70	Pyridin-3-yl	H	H	Furan-2-yl	95%
71	Pyridin-3-yl	H	H	^{c}Pr	86%
72	Pyridin-3-yl	H	H	Benzyl	77%
73	Pyridin-3-yl	H	H	^{c}Hexyl	93%
74	Pyridin-3-yl	H	H	Morpholino	81%

*Table 4.2: Synthesis of structural analogues to BRD1 hit **25**. Reactions performed in parallel using standard amide coupling conditions between amines and acid chlorides. [a]ND: Product isolated, but yield not determined. NA: Synthesis not attempted. NI: Product not isolated.*

All sixteen follow-up compounds, plus the original hit, were tested using single shot AlphaScreen experiments at compound concentrations of 100 µM and 200 µM. No BRD1 inhibition was observed for any of the compounds at either concentration. Fragment soaking yielded two new datasets. Amides **57** and **71** were both found to bind to BRD1 in an alternative binding pose to hit **25**. The pyridinylmethanamide is flipped, making the pyridin-3-yl group solvent exposed (Figure 4.9). The smaller *iso*-propyl and *cyclo*-propyl groups do not displace waters *III* and *IV*, instead occupying the pocket formed by the four Brd water molecules.

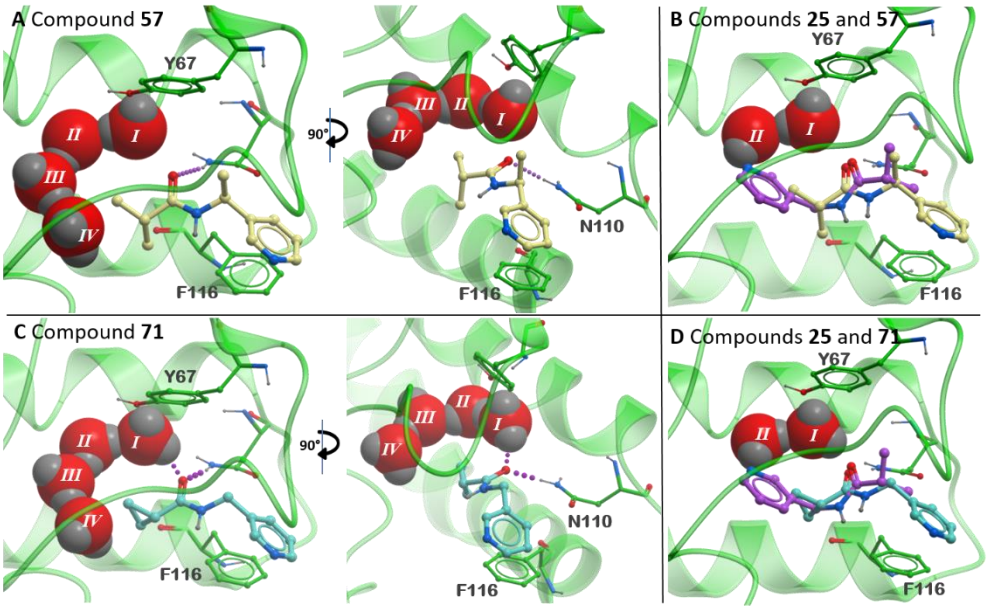


*Figure 4.9: Crystal complexes of BRD1 (green ribbon and sticks) follow-up compounds. (A) Amide **57** (beige sticks) forms a H-bond to N110 (purple dots). (B) An overlay of follow-up compound **57** and fragment hit **25** (purple sticks) illustrates the flipped binding pose of **57**. (C) Amide **71** (cyan sticks) forms H-bonds to N110 and water I (purple dots). (D) An overlay of follow-up compound **71** and fragment hit **25** (purple sticks) illustrates the flipped binding pose of **71**.*

Changes in fragment binding poses, like those observed between pyrimidines **17** and **51**, and amides **25**, **57** and **71**, have been documented before (Keseru, Erlanson *et al.* 2016). A shift in fragment binding could be due to new, stronger interactions being made by the modified compounds. Alternatively it may be that the original fragment had multiple binding modes which were disguised by ambiguous electron density. However, many successful hit-to-lead campaigns report the conservation of fragment binding modes (Murray, Verdonk *et al.* 2012; Erlanson, Fesik *et al.* 2016). The results offer an opportunity to investigate structure-binding relationships, although the research would be further enhanced with the addition of binding affinities, enabling structure-activity-binding relationship studies.

4.3.3 Re-Synthesis of BRD1 Fragment Hits

Poised fragment hits of BRD1 from the 3D fragment consortium were resynthesised in order to measure full IC_{50}'s using the AlphaScreen assay. The synthesis of fifteen amides was attempted,

of which twelve were isolated (Table 4.3). Failed synthesis was the result of by-product formation (*bis*-substituted urea analogue of compound **32**), apparent product volatility (piperizinone **33**) or problems with reaction regioselectivity (aniline **36**).

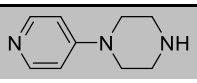
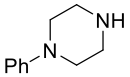
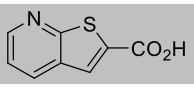
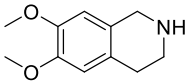
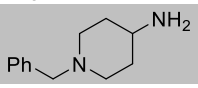
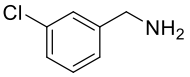
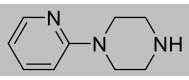
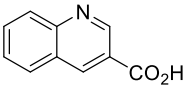
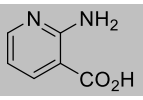
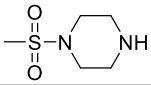
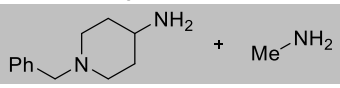
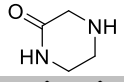
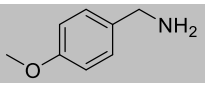
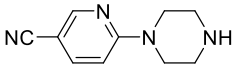
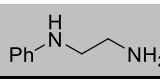
Compound	Starting Amine	Starting Acid	Reagents and Conditions[a]	Yield[b]
20			A	8%[c]
21			A	13%[c]
22			B	43%
23			A	6%[c]
24			A	9%[c]
26			A	8%[c]
27			A	20%[c]
29			B	73%
30			B	35%
31			A	35%
32		-	C	NI
33			A	NI
34			A	13%[c]
35			A	82%
36			A	NI

Table 4.3: Resynthesis of BRD1 fragment hits containing amide bonds. [a]Reagents and conditions: (A) Triethylamine, CH_2Cl_2, rt. (B) HATU, N-methyl morpholine, H_3CCN, rt. (C) CDI, DCM, rt. [b]NI: Product not isolated. [c]Final product isolated from 25 mg of crude material using HPLC.

The twelve resynthesised fragment hits were tested for BRD1 inhibition by AlphaScreen, as well as the Brds CECR2 and FALZ from subfamily *I*. Similar to pyrimidine hit **17** and amide hit **25**, ten of the fragments were found to be inactive at the top concentration of 200 μM. However, thienopyridine **22** and quinoline **29** were both found to have pIC_{50} values of 4.9 (Table 4.4), and showed no activity against the Brds CECR2 and FALZ. Interestingly, dihydroisoquinoline **23** was found to have an pIC_{50} value of 4.0 against CECR2 despite showing no inhibition of BRD1 or FALZ at 200 μM.

Compound	BRD1 pIC_{50}	HA	LE
OF-1	6.3	26	0.34
NI-57	6.9	25	0.39
PFI-4	6.1	28	0.31
22	4.9 ± 0.01 (2)	13	0.53
29	4.9 ± 0.01 (2)	14	0.48

*Table 4.4: Comparison of fragment hits **22** and **29** to the chemical probes OF-1, NI-57 and PFI-4. Despite the fragments' lower affinity, they have much higher ligand efficiency (LE). LE calculated as described in section 1.2.4 by dividing log(IC_{50}) by heavy atom count (HA).*

The SGC pan-BRPF chemical probes OF-1 and NI-57, and the BRPF1B specific chemical probe, PFI-4, all inhibit BRD1 on the nM scale.[1] The ligand efficiency of the probes is in the range of typical drug candidates, between 0.3 and 0.4 (Table 4.4). In contrast, the ligand efficiencies of fragments **22** and **29** are around 0.5 despite a 50-fold decrease in affinity. The two compounds would make excellent starting points for hit-to-lead development. However, the existence of at least three BRD1 chemical probes led to the decision to terminate investigation into BRD1 follow-ups in favour of targeting the previously intractable proteins PHIP(2), JMJD2D and SETDB1.

[1] www.thesgc.org/chemical-probes/epigenetics, accessed 3rd October 2016

4.4 Brd PHIP(2)

4.4.1 PHIP(2), an Atypical Bromodomain

Human pleckstrin homology domain interacting protein (PHIP) is believed to mediate the activity of insulin receptor proteins (De Semir, Nosrati *et al.* 2012), and has been identified as a possible marker in melanoma prognosis (Haqq, Nosrati *et al.* 2005). More recently, PHIP has been observed to be the most overexpressed protein in metastatic melanomas (De Semir, Nosrati *et al.* 2012). The PHIP protein contains two Brds, PHIP(1) and PHIP(2). The first Brd, from family *VII* of the Brd phylogenetic tree, has no reported crystal or NMR structures. However, in 2010 the crystal structure of the second PHIP Brd was deposited in the PDB (PDB ID 3MB3, Figure 4.10 A), bound to NMP but in a flipped conformation compared to typical Brds, including BRD1 (PDB ID 3RCW, Figure 4.10 B) (Filippakopoulos, Picaud *et al.* 2012).

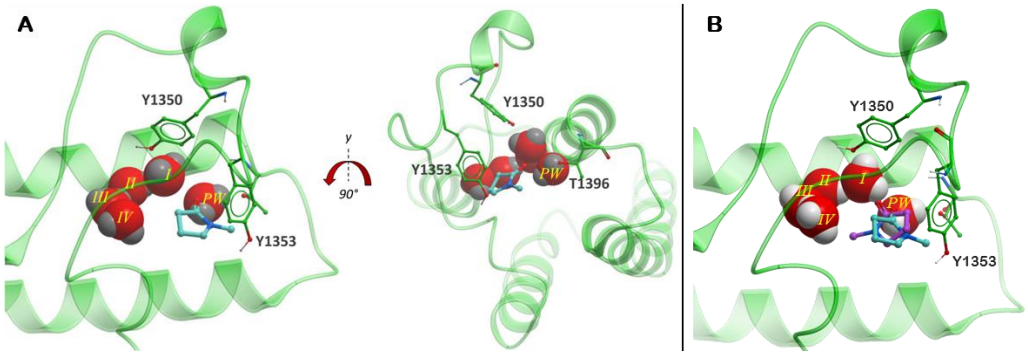


Figure 4.10: PHIP(2) (green sticks and ribbons) in complex with NMP (cyan sticks). (A) The substitution of an Asn for the smaller Thr allows an extra water molecule, the PHIP water (red CPK, PW) to fill the space between T1396 and the usual four Brd waters (red CPK I–IV) (PDB ID 3MB3, Filippakopoulos et al. 2012). (B) Overlay of BRD1 bound NMP (magenta sticks) with the NMP-bound PHIP(2) crystal structure (PDB ID 3RCW). NMP binds to PHIP(2) in a flipped conformation compared to typical Brds such as BRD1. Figure reprinted from Cox et al. 2016 with permission from the authors.

PHIP(2) belongs to Brd subfamily *III* alongside CREBBP (Figure 1.10, page 24). The atypical binding site of PHIP(2) is characterised by a threonine (Thr) residue (T1396) in place of the conserved Asn (Figure 4.10). Because the Thr side chain is one atom shorter than Asn, there is space for an additional water molecule, dubbed the 'PHIP water', to mediate an interaction between T1396

and the typical water network at the back of the binding pocket. The water-network is otherwise comparable to the network observed in most Brd binding sites, consisting of four water molecules and anchored by a tyrosine residue (Y1353). Despite being identified as a druggable Brd (Vidler, Brown *et al.* 2012), to date no small molecule inhibitors of PHIP(2) have been reported. An unpublished solution-phase screening campaign by the SGC did not yield any hits from a library of approximately seven thousand Brd-focused compounds.

4.4.2 PHIP(2) Fragment Soaking Campaign

Researchers at the SGC performed a fragment screen against PHIP(2) using the X-ray crystallography protocol outlined at the beginning of this chapter. Nine hits (0.9% hit rate) were found to bind to the KAc binding site (Figure 4.11). The fragment-protein structures are available online (Bradley, Wall *et al.* 2015).[1]

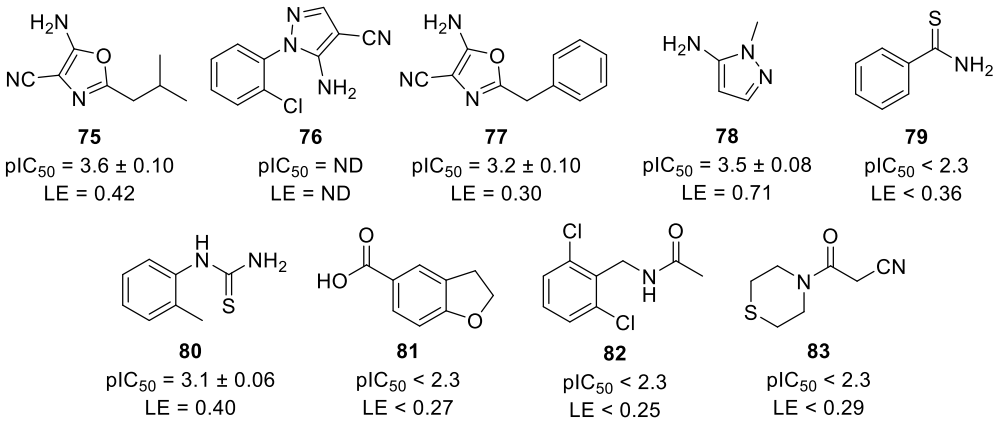


*Figure 4.11: Fragment hits of PHIP(2) identified by fragment soaking experiments performed at DLS. Compounds **75**, **77**, **80** and **82** were reported by Cox et al. (2016). The remaining hits have not been published. pIC_{50} values measured by AlphaScreen peptide displacement assay. All experiments performed in duplicate. ND = not determined.*

Compounds **75**, **76** and **77** form direct H-bonds to T1396 through primary amines, which allows the lipophilic head groups (*iso*-butyl, 2-chlorophenyl and benzyl respectively) to occupy the pocket

[1] http://wonka.sgc.ox.ac.uk/WONKA/8/Summarise/, accessed 3rd October 2016

made by the four canonical water molecules (Figure 4.12 A). The primary amine of pyrazole **78** forms a direct H-bond to Y1353 by displacing water *I* (Figure 4.12 B). Compounds **79**, **80** and **81** also displace one or more of the structural water molecules, indicating the molecules are easier to displace than in typical Brds (Figure 4.12 C). Only compound **82** binds like a classical KAc mimetic, forming a H-bond to T1396 *via* the PHIP water in an analogous binding mode to NMP (Figure 4.13 B). The final compound, amide **83**, binds at the edge of the binding pocket and does not appear to form any meaningful interactions.

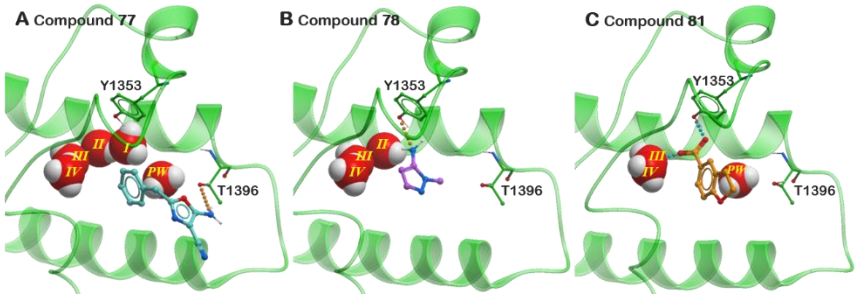


*Figure 4.12: Crystal structures of selected fragment hits bound to PHIP(2) (green ribbon and sticks). (A) Oxazole **77** (cyan sticks) forms an H-bond to T1396 (orange dots), allowing the benzyl group to occupy the pocket made by the four structural waters. Heterocycles **75** and **76** have analogous binding poses. (B) Pyrazole **78** (magenta sticks) displaces water I and the PW. The 3-amino group forms an H-bond to Y1353. (C) Carboxylic acid **81** (orange sticks) displaces waters I and II, allowing the acid head group to form H-bonds to Y1353 and water III (blue dots).*

Each fragment was tested using AlphaScreen displacement assay to select compounds for follow-up synthesis. The IC_{50} values of four of the fragments were found to be outside the sensitivity of the assay. The most potent compound, pyrazole **78**, has high ligand efficiency (LE) and a primary amine which could be used as a reactive handle for fragment elaboration. However closer inspection of the fragment's binding mode reveals the amine to be forming an important H-bond interaction to Y1350 at the back of the binding pocket, diminishing any hopes of using it for fragment growth. Similarly, the reactive handles of the primary thioamide **79** and carboxylic acid **81** make key interactions to the protein by displacing structural water molecules.

The oxazoles **75** and **77** show good activity. As a result, a poised oxazole-forming reaction was utilised by Dr Paul Brennan and co-workers to perform a follow-up campaign, leading to modest gains in potency but no new crystal structures (Cox, Krojer *et al.* 2016).

The poised thiourea **80** shows modest activity, good LE and a novel binding mode. The thiourea sulfur of compound **80** displaces water *I* at the back of the binding pocket allowing the formation of hydrogen bonds with Y1353 and water *II* (Figure 4.13 A). The secondary thiourea nitrogen forms an H-bond to T1396 via the structural PHIP water (*PW*). The primary thiourea nitrogen lies in the hydrophilic pocket made by waters *II*, *III* and *IV*. The phenyl ring of compound **80** makes an edge-face aromatic interaction with Y1350, which is shifted significantly towards the ligand compared to the NMP-bound conformation (Figure 4.10, page 114). The *ortho*-methyl substituent on the phenyl ring makes additional interactions with the hydrophobic residues I1403 and P1340.

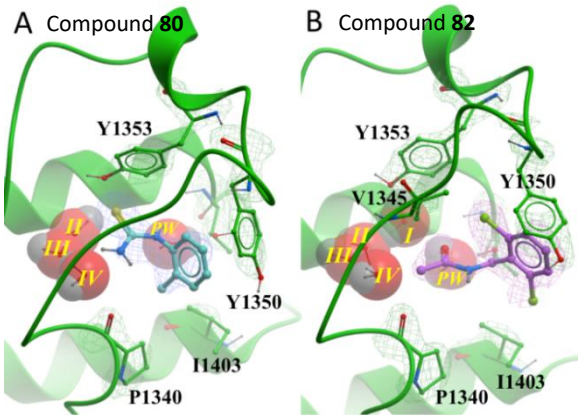


*Figure 4.13: PHIP(2) (green sticks, ribbons and mesh (2Fo-Fc)) in complex with fragment hits. (A) Thiourea **80** (cyan sticks and mesh (2Fo-Fc)) displaces waters I (PDB ID 5ENB). (B) N-Benzylacetamide **82** (purple sticks and mesh (2Fo-Fc)) binds with all water molecules intact and forms an edge–face interaction with Y1350 (PDB ID 5ENC). Figure reprinted from Cox et al. 2016 with permission from the authors.*

Compound **82** binds like a classical acetyl-lysine mimetic (Figure 4.13 B). The acetyl group lies deep in the binding pocket allowing the methyl group to sit in the hydrophobic pocket created by the four structural waters in a manner analogous to inhibitors of typical Brds and acetyl lysine. The carbonyl oxygen interacts with T1396 through a hydrogen bond mediated by the PHIP water. Like

compound **80**, the phenyl ring of compound **82** appears to make an edge-face aromatic interaction with the shifted Y1350 side chain, allowing the 2,6-dichloro substituents to occupy the hydrophobic pockets formed by V1345 and I1403 above and below the ring.

A distinct series for each of the poised hits **80** and **82** were rapidly populated with analogues via parallel, solution-phase synthesis. Compounds were selected for synthesis based upon structural similarity and manual inspection of the PHIP(2) crystal complexes.

4.4.3 Thioureas as PHIP(2) Inhibitors

The thioureas were synthesised by the sequential addition of the relevant amines to 1,1'-thiocarbonyldiimidazole (TCDI) (Figure 4.14 A). Analysis of the crystal data obtained for compound **80** indicated a lipophilic pocket adjacent to the 3-position of the benzene ring so emphasis was placed on exploring this position. Substitution of the terminal thiourea nitrogen was also proposed in an attempt to displace more of the structural waters at the back of the binding site. In total thirty-five compounds were proposed for synthesis and twenty-nine (83%) were synthesised and purified in moderate to good yield within two weeks (Table 4.5). The thiourea series was tested for inhibition of PHIP(2) ligand binding using an AlphaScreen assay and LE calculated using the relationship described in Chapter 1.

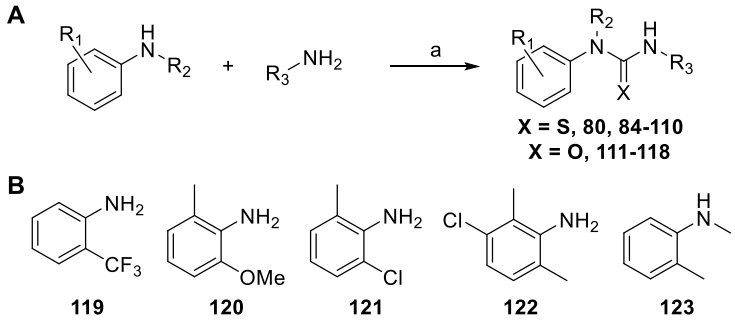
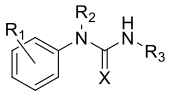


Figure 4.14: (A) Synthesis of thiourea follow-up library. Reagents and conditions: (a) Amine 1, TCDI, DCM, rt, 30 min, then amine 2, rt, 12 h. (B) Starting anilines from which one or more desired thioureas/ureas could not be isolated.

Failed synthesis can be attributed to the de-activated anilines used as starting materials (Figure 4.14 B). The presence of an electron-withdrawing neighbouring-group (trifluoromethyl **119**), 2,6-substitution patterns (anilines **120-122**) and *N*-methylation (secondary amine **123**) caused each amine to be unreactive towards TCDI, as well as potassium isocyanate in the case of anilines **121** and **122**.

The original fragment hit **80** had pIC_{50} of 3.1 in the AlphaScreen assay. Addition of a substituent at the 3-position of the phenyl ring to give compounds **84-87** increases potency. The methyl derivative **85** and the trifluoromethyl derivative **86** have improved pIC_{50} around 3.8. The addition of a methyl group at the 6-position of the phenyl ring to give compound **88** slightly increases potency to pIC_{50} 3.3 as does methylation of the primary thiourea nitrogen to give compound **89** (pIC50 3.4). The SAR improvements are not cumulative (compounds **91**–**109**), but replacement of 2-methylphenyl with 2-chlorophenyl on the methylated thiourea, giving compound **90**, improves PHIP(2) activity to pIC_{50} 3.9. Functional group conversion of the thiourea to a urea causes a complete loss of activity (compounds **111**–**118**). The LE of fragment hit **80** was calculated to be 0.40. The value increases to 0.45 for the most potent thioureas **85** and **90**.



Thiourea series

Compound Number	R_1	R_2	R_3	X	Yield[a]	Soaking Outcome[c]	PHIP(2) pIC_{50}[e]	LE
80	2-Me	H	H	S	66%	Success	3.1 ± 0.06 (2)	0.40
84	2-Me-3-OMe	H	H	S	71%	Error	3.0 ± 0.18 (2)	0.32
85	2,3-Me_2	H	H	S	12%	Success[d]	3.8 ± 0.06 (2)	0.45
86	2-Me-3-CF_3	H	H	S	33%	NL	3.9 ± 0.21 (2)	0.36
87	2-Me-3-Cl	H	H	S	26%	Success[d]	3.6 ± 0.03 (2)	0.42
88	2,6-Me_2	H	H	S	8%	NL	3.3 ± 0.08 (2)	0.39
89	2-Me	H	Me	S	14%	NL	3.4 ± 0.30 (2)	0.39
90	2-Cl	H	Me	S	19%	NL	3.9 ± 0.05 (2)	0.45
91	2-Me	H	Et	S	74%	NL	<2.3 (2)	<0.25
92	2-Me	H	*i*-Pr	S	78%	NL	2.7 ± 0.63 (2)	0.27
93	2-Me	H	$(CH_2)_2OH$	S	69%	Error	<2.3 (2)	<0.23
94	2-Me	H	Me_2	S	65%	NL	<2.3 (2)	<0.25
95	2-Cl	H	H	S	29%	NL	2.8 ± 0.14 (2)	0.35
96	2-OMe	H	H	S	19%	Error	<2.3 (2)	<0.27
97	2-CF_3	H	H	S	NI			
98	2,4-Me_2	H	H	S	32%	Error	2.9 ± 0.06 (2)	0.34
99	2,5-Me_2	H	H	S	44%	NL	3.1 ± 0.06 (2)	0.36
100	2-Me-3-F	H	H	S	45%	NL	2.5 ± 0.07 (2)	0.29
101	2-OMe-6-Me	H	H	S	NI			
102	2-Cl-6-Me	H	H	S	NI			
103	2,6-Me_2-3-Cl	H	H	S	36%	NL	2.7 ± 0.06 (2)	0.29
104	2,3-Me_2	H	Me	S	33%	NL	<2.3 (2)	<0.25
105	2-Me-3-Cl	H	Me	S	44%	NL	2.7 ± 0.14 (2)	0.29
106	2-OMe	H	Me	S	47%	NL	<2.3 (2)	<0.25
107	2-Cl-6-Me	H	Me	S	NI			
108	2-OMe-6-Me	H	Me	S	48%	NL	<2.3 (2)	<0.23
109	2,6-Me_2-3-Cl	H	Me	S	12%	NL	2.6 ± 0.11 (2)	0.26
110	2-Me	Me	H	S	NI			
111	2-Me	H	H	O	Purchased	NL	2.7 ± 0.71 (2)	0.35
112	2,3-Me_2	H	H	O	49%[b]	NL	2.7 ± 0.05 (2)	0.31
113	2-Me-3-Cl	H	H	O	52%[b]	NL	2.7 ± 0.75 (2)	0.32
114	2-Cl	H	H	O	42%[b]	NL	<2.3 (2)	<0.29
115	2-OMe	H	H	O	46%[b]	NL	2.5 ± 0.08 (2)	0.29
116	2-Cl-6-Me	H	H	O	NI[b]			
117	2-OMe-6-Me	H	H	O	19%[b]	NL	2.5 ± 1.04 (2)	0.27
118	2,6-Me_2-3-Cl	H	H	O	NI[b]			

Table 4.5: Synthesis and screening of thiourea follow-up series. [a]Synthesised using procedure described in Figure 4.14 A unless stated. NI: not isolated. [b]Reagents and conditions: KOCN, AcOH, H_2O, rt. [c]NL: No ligand found in model. Error: Experiment failure during soaking, at beamline or during data processing or model refinement. [d]Binding pose uncertain as a result of partial occupancy. [e]By AlphaScreen peptide displacement assay.

When the thiourea compounds were subjected to soaking experiments with PHIP(2) crystals, two of the most potent compounds, **85** and **87**, appeared to bind in the PHIP(2) KAc recognition site (Figure 4.15). The binding poses are uncertain because of partial occupancy, but the apparent poses are similar to compound **80**, displacing structural water *I* and *PW*.

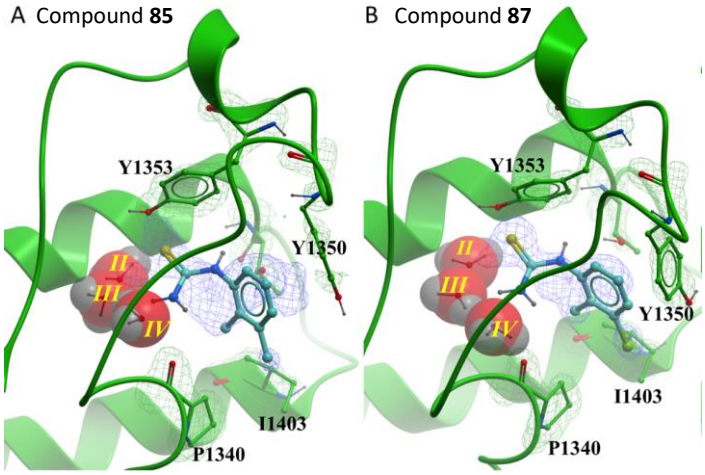


*Figure 4.15: Structures of thioureas (A) **85** and (B) **87** binding to the PHIP(2) KAc recognition pocket. Compounds: cyan sticks and blue mesh (2Fo-Fc). Protein: green sticks and green mesh (2Fo-Fc). Water: red CPK. Figure reprinted from Cox et al. 2016 with permission from the authors.*

4.4.4 Computational Design of Thiourea Analogues

A second library of follow-up compounds for PHIP(2) were designed using the XPoise algorithm described in the previous chapter. A library of amine building blocks was built using the ZINC commercial libraries (Section 3.3.2, page 80). Compounds were short-listed using the LLOOMMPPAA method (Section 3.2.2, page 77) as detailed by A. Bradley (2015). The PLIFS_default selection method was used to build a 20 compound library (Figure 4.16). Similarly, the PLIFS_max and Focussed methods were used to build 20 and 100 compound libraries respectively. Final compound selection was made by visual inspection with pragmatic decision making based upon knowledge of medicinal chemistry, taking into consideration the cost of starting materials, regioselectivity concerns and undesirable functional groups, such as carboxylic acids, leaving a library of 32 thioureas.

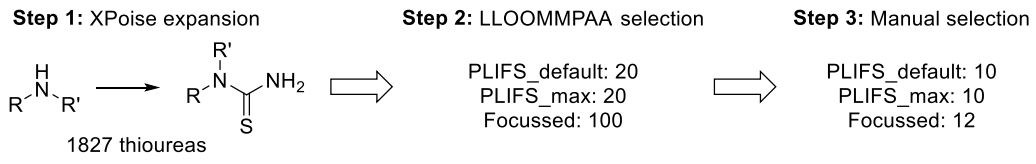


Figure 4.16: Design of a thiourea follow-up library using the XPoise/LLOOMMPPAA method. A final library of 32 compounds was selected for synthesis and testing against PHIP(2).

Alternative reaction conditions were sought as an alternative to the TCDI method used in the previous section, which had been low yielding and challenging to purify in the synthesis of terminal ureas. The direct reaction of aniline and benzylamine with ammonium isothiocyanate in acidic water was attempted using a literature procedure (Siddiqui and Ahsan 2010), but no product was detected using LCMS after stirring at 100°C for three days (Figure 4.17). However activation of the isothiocyanate using benzoyl chloride resulted in good conversion of aniline and benzylamine to the corresponding carbamothioyls before the loss of benzoic acid using sodium hydroxide to afford the terminal thiourea (Frank and Smith 1948). Although products **127** and **128** were not isolated, LCMS indicated greater than 95% conversion and a high level of crude purity in both cases.

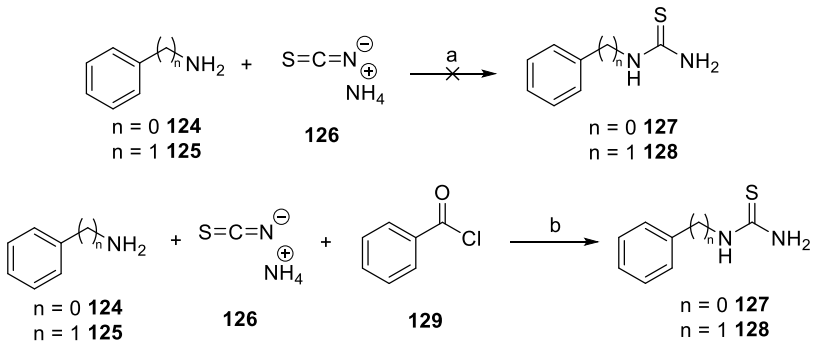


*Figure 4.17: Alternative reaction conditions for the synthesis of terminal thioureas. Reagents and conditions: (a) 0.5M HCl, H_2O, 100°C (Siddiqui and Ahsan 2010). (b) Reagents **126** and **129**, acetone, 60°C, 5 mins; then amine, 60°C, 1 h; then 5M NaOH, 100°C, 2 h (Frank and Smith 1948).*

The new reaction conditions were adapted to skip isothiocyanate activation by using the commercially available intermediate benzoyl isothiocyanate. The conditions were used by myself and Mr Lin Mei to synthesise the short-listed LLOOMMPPAA library (Figure 4.18). Of the thirty-two compounds proposed, only 16 were isolated (50% success rate). The most common reasons for failed synthesis were deactivated amines next to electron-withdrawing heterocycles (triazinone **156**, triazine **157** and pyrimidine **158**), reactive functional groups (pyrimidines **137** and **138**, thiadiazoles **140** and **141**, and sulfone **148**), or unstable heterocycles that decomposed in the

harsh reaction conditions (thiazole **132**, thiadiazole **146**, tetrazoles **150**, **151**, **154** and **155**, and oxazolidinone **161**). This is particularly evident for the Focussed subset, which contains a large proportion of heterocycles, and for which only four out of twelve products were isolated. However the products of some deactivated, reactive or unstable amines were still isolated, such as triazole **131**, acid **145** and pyrazole **152**.

Two reaction by-products were also isolated and characterised, formed by the reaction of the amine directly with the benzoyl isothiocyanate carbonyl, and resulting in the loss of hydrogen isothiocyanate (Figure 4.18).

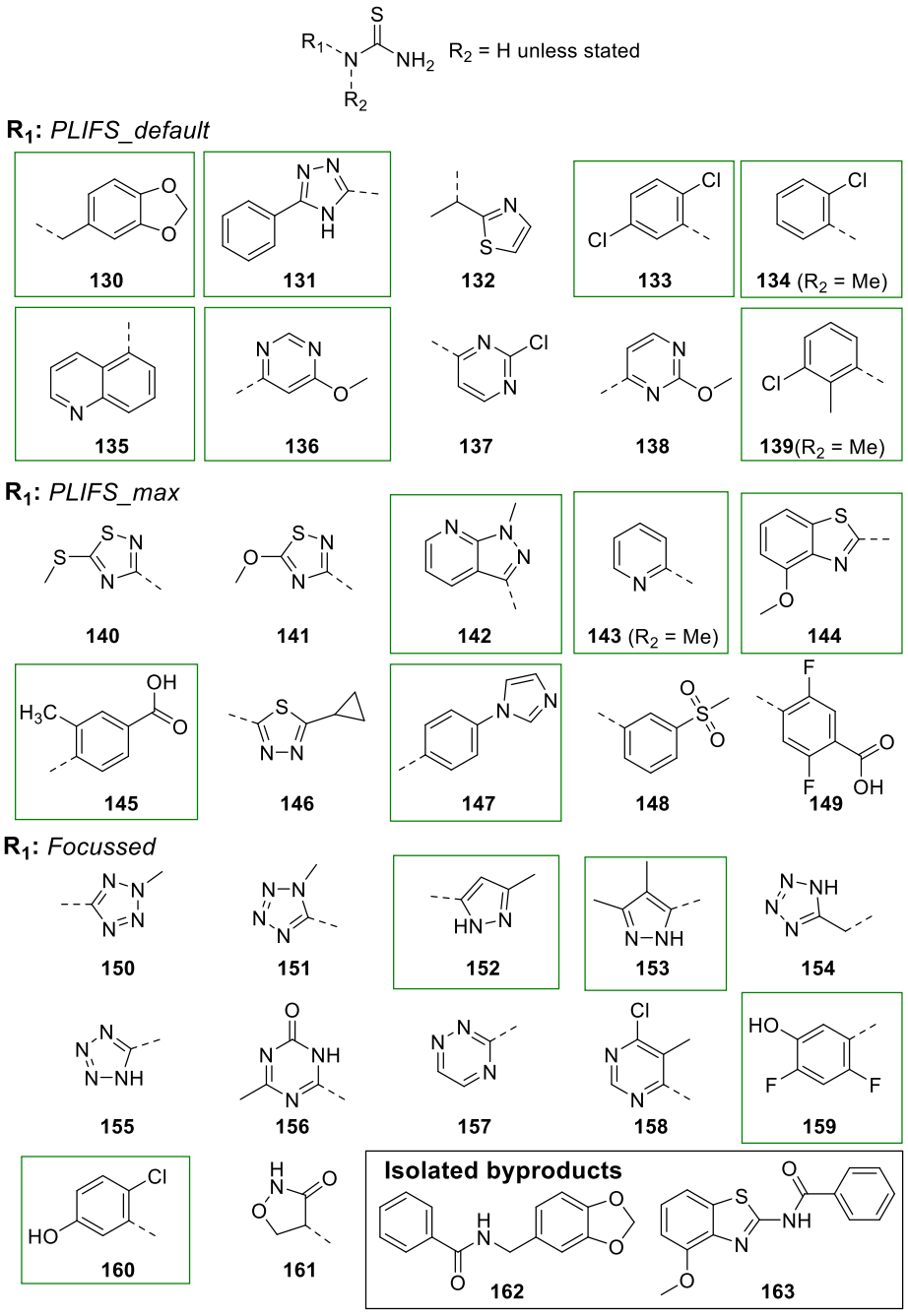


*Figure 4.18: Thiourea library designed using LLOOMMPPAA (Bradley 2015). Sixteen compounds were isolated (green boxes) plus two reaction by-products, isolated from the reaction mixtures of compounds **130** and **144**. Reagents and conditions: Benzoyl isothiocyanate, acetone or CH_2Cl_2, 60°C, 5 mins; then amine, 60°C, 1 h; then either 5 M NaOH or 3.6 M K_2CO_3, 100°C, 2 h.*

The sixteen thioureas plus the reaction by-products **162** and **163** were tested for PHIP(2) inhibition (Table 4.6). Only four of the compounds were found to be inactive in the AlphaScreen: quinoline **135**, pyrimidine **136**, pyridine **143** and imidazole **147**. A comparison of the phenylthioureas to the

first follow-up library reveals a number of SAR trends. The 2,5-dichloro substitution pattern of compound **133** shows comparable activity and LE to the 2,5-dimethyl substitution pattern of compound **99** (pIC_{50} 3.3 versus 3.1, and LE 0.38 versus 0.36). However, modification of compound **133** to the 2-chloro-5-hydroxy substituted compound **160** gives a small boost to activity (pIC_{50} 3.5) and improves LE to 0.41. The 2,4-difluoro-5-hydroxy substituted compound **159** also shows modest activity (pIC_{50} 3.0).

Addition of a carboxylic acid group *para*- to the thiourea head group (compound **145**) improves potency from pIC_{50} 3.1 (hit **80**) to pIC_{50} 3.5, likely as a result of H-bond formation between the acid and the hydroxyl-group of Y1350.

Compound Number	Yield[a]	Soaking Outcome[c]	PHIP(2) pIC_{50}[d]	LE
130	ND	NL	3.4 ± 0.05 (2)	0.34
131	ND	NL	3.3 ± 0.01 (2)	0.31
133	ND	NL	3.3 ± 0.06 (2)	0.38
134	ND	NL	2.9 ± 0.08 (2)	0.34
135	17.4%	NL	<2.3 (2)	<0.23
136	ND	NL	<2.3 (2)	<0.27
139	ND	Success	4.2 ± 0.07 (2)	0.45
142	ND	NL	3.1 ± 0.03 (2)	0.31
143	ND	NL	<2.3 (2)	<0.29
144	40.7%	NL	3.7 ± 0.05 (2)	0.35
145	2.9%	NL	3.5 ± 0.04 (2)	0.35
147	ND	NL	<2.3 (2)	<0.21
152	11.3%	NL	3.4 ± 0.02 (2)	0.47
153	ND	NL	3.6 ± 0.03 (2)	0.45
159	ND	NL	3.0 ± 0.03 (2)	0.32
160	ND	NL	3.5 ± 0.03 (2)	0.41
162	3.3%[b]	NL	3.9 ± 0.03 (2)	0.29
163	27.3%[b]	NL	3.5 ± 0.04 (2)	0.24

Table 4.6: Synthesis and screening of thiourea follow-up library suggested by LLOOMMPPAA. [a]Synthesised using procedure described in Figure 4.18. ND: Yield not determined. [b]Isolated reaction byproduct. [c]NL: No ligand found in model. [d]By AlphaScreen peptide displacement assay.

N-methylation of the secondary thiourea nitrogen increases the activity of compounds **134** (pIC_{50} 2.9) and **139** (pIC_{50} 4.2) in comparison to the unsubstituted analogues **95** (pIC_{50} 2.8) and **87** (pIC_{50} 3.6) while maintaining good LE. The pyrazole-containing compounds **152** and **153** show potency comparable to the benzene-containing series, but excellent LE (0.47 and 0.45 respectively) owed to the smaller ring size.

The 5,6-fused ring systems present in compounds **130**, **142** and **144** all show modest activity, but the increased MW results in a loss of LE to 0.35 or less. Similarly, 5-phenyltriazole **131** has pIC_{50} 3.3 but LE of just 0.31. The isolated by-products **162** and **163** yielded pIC_{50} values 3.9 and 3.5 respectively, but the presence of the bulky, lipophilic benzoyl moiety results in LE <0.3.

A comparison of pIC_{50} and LE reveals how the LLOOMMPPAA designed analogues are on average more active than the analogues designed by similarity and manual inspection (Figure 4.19 A). Ten LLOOMMPPAA compounds (56%) have higher pIC_{50} than the original hit in comparison to just six (17%) of the original follow-up library. In contrast, maintaining or improving LE of the original hit was less successful, with just seven (13%) of all thiourea analogues yielding a LE greater than hit **80**.

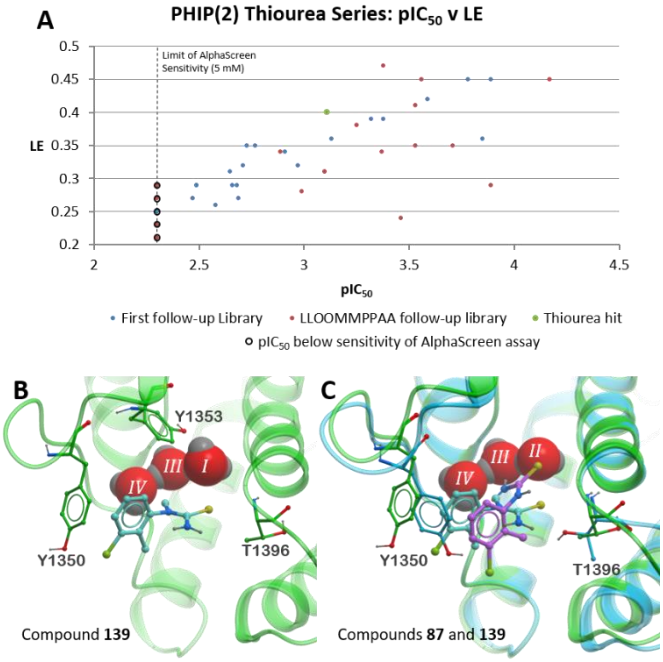


*Figure 4.19: (A) Comparing activity and LE of thiourea follow-up libraries and the thiourea hit **80**. Compounds which were found to be inactive in the AlphaScreen assay are shown at their maximum possible pIC_{50} and LE values. (B) Compound **139** (cyan sticks) in complex to PHIP(2) (green ribbon and sticks). The N-methyl occupies the hydrophobic pocket made by structural waters I-IV (red CPK; water II is eclipsed by water I). The thiourea sulfur occupies the PW pocket. (C) Overlay of compound **139** (cyan sticks) and analogue **87** (pink sticks) in complex to PHIP(2) (green ribbon and light blue ribbon). Y1350 moves to accommodate **139** which does not lie as deep in the binding pocket as **87**. The conformation of T1396 is also altered, forcing the hydroxyl group away from the ligand. Water I from the crystal structure of compound **139** not shown for clarity.*

The eighteen compounds from the LLOOMMPPAA follow-up library were subjected to soaking experiments. The most potent compound, thiourea **139**, was the only follow-up found to bind to the PHIP(2) KAc recognition site (Figure 4.19 B). The thiourea head group is flipped relative to hit **80**, allowing the *N*-methyl group to occupy the hydrophobic pocket formed by the structural waters, and moving the thiourea sulfur into the pocket normally occupied by the PHIP water. Consequently, all four Brd structural waters are present, but the phenyl group is forced further out of the binding site and Y1350 rearranges to maintain the edge-face aromatic interaction (Figure 4.19 C). Furthermore, the key Thr residue, T1396, is flipped into a new conformation in which the hydroxyl group is directed away from the binding pocket.

4.4.5 *N*-Benzyl Amides as PHIP(2) Inhibitors

The *N*-benzyl amides were synthesised using amine derivatisation with acyl chlorides (Figure 4.20). Substitution of the benzene ring was proposed as well as extension of the acetyl group with the aim of displacing structural waters. Displacement of water molecules often results in significant increases in activity thanks to the associated gain in entropy. In total twelve analogues were proposed for synthesis and eleven (92%) were synthesised in good to excellent yield within one week (Table 4.7).

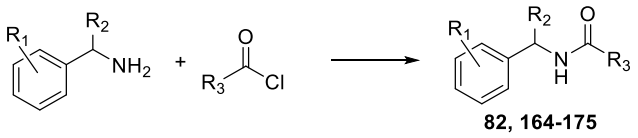
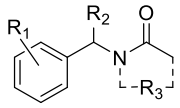


Figure 4.20: Synthesis of the N-benzylamide follow-up library. Reagents and conditions: Triethylamine, DCM, rt, 3 h.

The compounds were tested for inhibition of PHIP(2) ligand binding using the AlphaScreen assay. The original fragment hit **82** was inactive against PHIP(2) at the highest possible assay concentration, and would not have been detected as a hit if AlphaScreen had been used as the primary screen. Attempts to modify the acetyl amide group or the benzene ring gave no

improvements in potency. The exception was the 2,6-dimethoxy analogue **164** with a pIC_{50} of 3.7 and LE of 0.35.



Amide series

Compound Number	R_1	R_2	R_3[a]	Yield[b]	Soaking Outcome[g]	PHIP(2) pIC_{50}[i]	LE
82	2,6-Cl_2	H	*C*-Me	44%	Success	<2.3 (2)	<0.25
164	2,6-$(OMe)_2$	H	*C*-Me	79%	Success	3.7 ± 0.04 (2)	0.35
165	2,6-Cl_2	H	*C*-CH_2OH	15%[c]	Success	<2.3 (2)	<0.23
166	2,6-Cl_2	H	*N*-Me, *C*-Me	83%	Success	<2.3 (2)	<0.23
167	2,6-Cl_2	H	*C*-Et	82%	NL	<2.3 (2)	<0.23
168	2,6-Cl_2	H	*C*-^{i}Pr	82%	NL	<2.3 (2)	<0.21
169	2,6-Cl_2	H	*C*-CF_3	67%	NL	<2.3 (2)	<0.20
170	2-Cl	Me	*C*-Me	87%	NL	2.5 ± 5.68 (2)	0.26
171	2,6-Cl_2	H	*C*-$NHCH_3$	NI[d]			
172	2,4,6-$(OMe)_3$	H	*C*-Me	83%	NL	<2.3 (2)	<0.19
173	2-Cl-4-F	H	*C*-Me	82%	NL	<2.3 (2)	<0.25
174	2,4-Cl_2	H	*C*-Me	84%	NL	<2.3 (2)	<0.25
175	2-Cl-5-CF_3	H	*C*-Me	65%	NL	<2.3 (2)	<0.20
176	2,6-$(OMe)_2$	H	*C*-CH_2OH	23%[e]	Success[h]	<2.3 (2)	<0.20
177	2,6-$(OMe)_2$	H	5-lactam	6%[f]	NL	3.3 ± 0.04 (2)	0.27
178	2,6-$(OMe)_2$	H	6-lactam	18%[f]	NL	3.5 ± 0.04 (2)	0.27

Table 4.7: Synthesis and screening of amide follow-up series. [a]R_3 = N-H unless stated. [b]Synthesised using procedure described in Figure 4.20 unless stated. NI: not isolated. [c]Reagents and conditions: EDC, DMAP, triethylamine, DMF, rt. [d]CDI, triethylamine, DCM, rt. [e]Over three steps. See Appendix E for details. [f]Triethylamine, AcN, rt. [g]NL: No ligand found in model. [h]Binding pose uncertain as a result of partial occupancy. [i]By AlphaScreen peptide displacement assay.

The benzyl amide compounds were soaked into PHIP(2) crystals. Surprisingly the inactive compounds **165** and **166** were found to bind in the KAc binding site as well as the active compound **164**. Compound **164** has the same binding mode as fragment hit **82**, with the 2,6-dimethoxy substitution pattern on the benzene ring occupying the V1345 and I1403 lipophilic pockets above and below the ring (Figure 4.21 A).

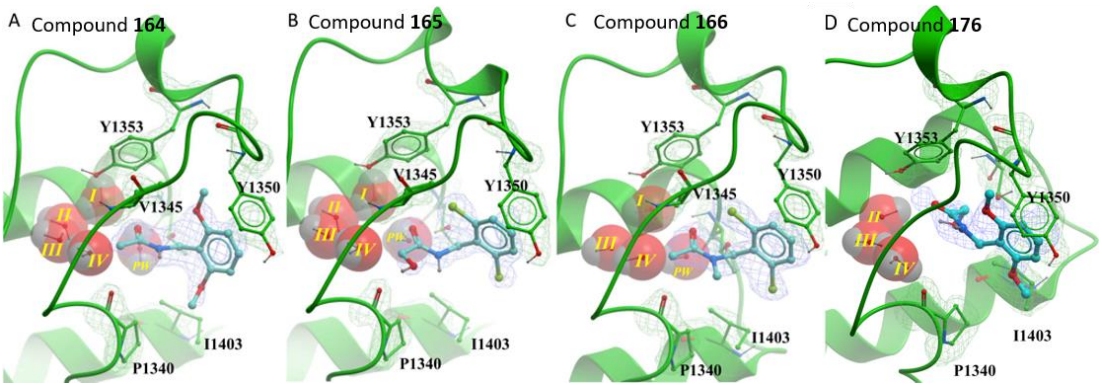


*Figure 4.21: PHIP(2) (green sticks and ribbons) in complex with N-benzylacetamides **164**–**166** and **176** (cyan sticks). Electron density shown as green and blue mesh (2Fo-Fc). (A) The amide **164** binds in the same pose as fragment hit **82**. PDB ID: 5ENH. (B) The hydroxyacetamide **165** forms an H-bond to P1340. PDB ID: 5ENI. (C) Tertiary amide **166** appears to displace structural water II. PDB ID: 5ENJ. (D) The hydroxyacetamide **176** appears to displace water I and PW, although the binding pose is uncertain. Figure reprinted from Cox et al. 2016 with permission from the authors.*

Alcohol **165** binds in the same position as fragment hit **82**, with the additional hydroxyl group making an H-bond interaction to the protein backbone at P1340 (Figure 4.21 B). In contrast, the *N*-methyl group of the tertiary amide **166** forms a hydrophobic interaction with the P1340 sidechain (Figure 4.21 C). Despite binding in a similar way to the other amide compounds, tertiary amide **166** appears to displace structural water *II*.

The activity and crystallographic data resulted in three new compounds being proposed for synthesis. All three were synthesized in low yield within two weeks (Table 4.7, compounds **176**–**178**). The 2,6-dimethoxy analogue of alcohol **165** (compound **176**) is inactive against PHIP(2) but appears to bind to the protein when subjected to soaking experiments, although the binding pose is uncertain as a result of low occupancy (Figure 4.21 D).

Overlay of fragment hit **82** with the previously reported NMP-bound structure (Figure 4.10, page 114) showed good overlap between the amide binding modes for each compound. A scaffold merging approach was adopted and the lactam containing compounds **177** and **178** proposed and synthesised. Both lactams were found to inhibit the activity of PHIP(2). pIC_{50}'s of 3.3 and 3.5 for compounds **177** and **178** respectively demonstrate good activity, and the lactam rings offer new

vectors for fragment growth. Unfortunately the LE (0.27) does not compare favourably with the best compound of the series, acetamide **164** (LE 0.35).

4.4.6 PHIP(2) Overview

Two hit-to-lead campaigns were performed on PHIP(2) fragment hits. Elaboration of thiourea **80** improved potency to sub-100 μM in two rounds of synthesis and generated three new crystal structures showing the primary thiourea head group acting as the KAc mimetic. A library of *N*-benzyl amides identified three sub-millimolar lead compounds from a hit which was inactive in the AlphaScreen assay. Furthermore, four novel crystal structures were generated which demonstrate how the amide bond acts as a KAc mimetic by H-bonding to T1396 *via* the *PW* molecule.

4.5 KDM JMJD2D

4.5.1 Background

JMJD2D, also known as KDM4D, is part of the KDM4 lysine demethylase sub-family. The KDM4 enzymes have been implicated in transcriptional regulation, cell-cycle progression, nuclear hormone signalling and embryonic stem cell self-renewal, while JMJD2D has been found to be overexpressed in prostate cancer and to promote cell proliferation in certain colon cancers (Krishnan and Trievel 2013). JMJD2D is distinct from other KDM4 enzymes in that it lacks PHD and Tudor domains, selectively demethylates H3K9me3/me2 over H3K36me3/me2, and can demethylate the di- and tri- methylated forms of H1.4K26 (Bavetsias, Lanigan *et al.* 2016).

A 2-OG competitive series of KDM4 inhibitors characterised by a pyrido-pyrimidinone head group and a pyrazole group directing iron chelation shows good activity against JMJD2D (Bavetsias, Lanigan *et al.* 2016). However the series shows no selectivity over KDM5 enzymes and low cell

activity. A crystal structure of JMJD2D complexed to an early compound from the series, KDOPP7, reveals the importance of forming a H-bond to K210, which forms a salt bridge to 2-OG in the active protein, and of forming a bidentate chelate to the catalytic iron centre (PDB ID 5F5C, Figure 4.22).

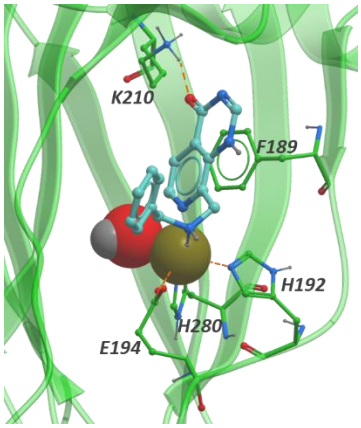


Figure 4.22: JMJD2D (green ribbon) in complex to KDOPP7 (cyan sticks). The pyrido-pyrimidinone head group is anchored by a π-stacking interaction with F189 and a H-bond to K210 (orange dashes). Bidentate chelation to Ni^{2+} is achieved through an aromatic nitrogen and secondary amine. The Ni^{2+} (brown sphere) occupies the Fe^{2+} catalytic site and is also coordinated by H192, E194, H280 (green sticks), and a water molecule (red CPK). Protein synonym: KDM4D (PDB ID 5F5C, Bavetsias et al. 2016).

4.5.2 JMJD2D Fragment Soaking Campaign

Researchers at the SGC performed a fragment screen against JMJD2D using the X-ray crystallography protocol outlined in the previous section. Each fragment was soaked twice: once with 2-OG present (1-3 µM) and once with it absent. Forty-four hits were found to bind to the protein at nineteen different sites, including six in the 2-OG binding pocket (unpublished results). The crystal structures are available online (Bradley, Wall *et al.* 2015).[1] Three fragments were found to bind in an allosteric site in the presence of 2-OG (Figure 4.23). An overlay of the fragment crystal structures with the trimethyl lysine bound complex reveals a conformation change to JMJD2D from an inactive to active state. The change in conformation dramatically alters the binding site

[1] http://wonka.sgc.ox.ac.uk/WONKA/12/Summarise/, accessed 3rd October 2016

of the fragments (Figure 4.23 B). It is thought the presence of the fragment could stabilise the inactive conformation of the protein and reduce trimethyl lysine binding, thereby acting as a non-competitive inhibitor.

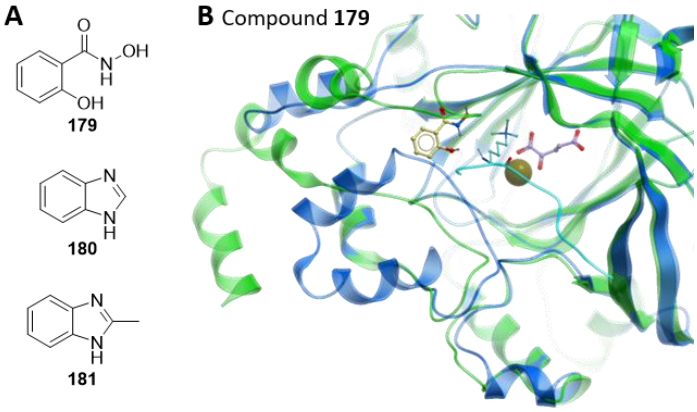


*Figure 4.23: (A) Three fragments identified to bind in a JMJD2D allosteric binding site in the presence of 2-OG. (B) An overlay of the JMJD2D-trimethyl lysine peptide complex (green ribbon, PDB ID 4HON, (Krishnan and Trievel 2013)) with the JMJD2D-compound **179** complex (blue ribbon). The conformation of JMJD2D is altered upon binding with the trimethyl lysine to accommodate the peptide (cyan ribbon and sticks), which causes a clash with the bound fragment **179** (light yellow sticks). Ni^{2+} (brown sphere) occupies Fe^{2+} catalytic site and is coordinated by 2-OG (mauve sticks) and the trimethyl lysine containing peptide, TARKme3STGGKA.*

Inspection of the hydroxamic acid **179** binding pose revealed an intramolecular H-bond from the acid N-H and the 2-hydroxyl oxygen. Binding is anchored by a H-bond from the 2-hydroxyl group to D195 and from the N176 amide to the acid carbonyl oxygen (Figure 4.24 A). As a result it was believed the hydroxamic acid was not needed for allosteric binding, but was observed to chelate to the active site nickel in the absence of 2-OG. Consequently, a follow-up campaign exploring secondary amide analogues in place of the hydroxamic acid was proposed.

Like hydroxamic acid **179**, the binding poses of benzimidazoles **180** and **181** are anchored by a H-bond between the imidazole nitrogen and D195 (Figure 4.24 B). The resolution of the crystal structure was not high enough to determine the relevance of the unsubstituted benzimidazole N-H, so a follow-up campaign was proposed using the 1- and 2- positions as growth vectors.

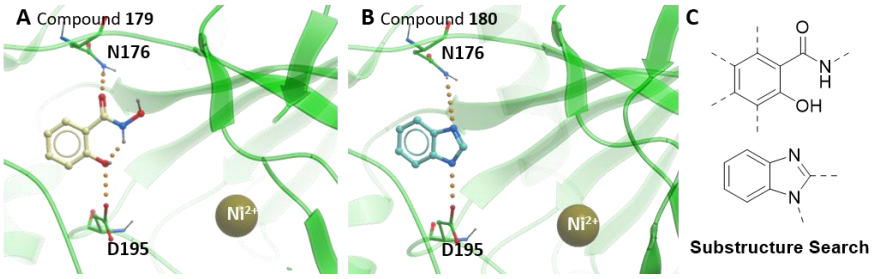


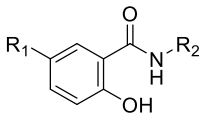
*Figure 4.24: Allosteric binders of JMJD2D (green ribbon). (A) Hydroxamic acid **179** occupies a lipophilic pocket adjacent to the Kme3 catalytic site, occupied by Ni^{2+} in the crystal structure. Binding is anchored by H-bonds to N176 and D195 (orange dashes). (B) Benzimidazole **180** occupies a lipophilic pocket adjacent to the Kme3 catalytic site. Binding is anchored by a H-bond to D195 (orange dashes). (C) Substructures used to build SAR by catalogue libraries.*

Both amide coupling and benzimidazole synthesis are poised reactions, but SAR by catalogue was chosen to speed-up initial hit-to-lead investigations. Two follow-up libraries for JMJD2D hits were built using systematic substructure searches on the eMolecules website[1] (Figure 4.24 C). From a library of 1156 amides and 614 benzimidazoles, sixty compounds (thirty from each library) were selected for testing using the LLOOMMPPAA PLIFS_default method (Section 3.2.2, page 77) as detailed by Bradley (2015).

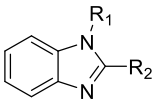
4.5.3 Using SAR by Catalogue to Study JMJD2D Hits

Seven amides and one hydroxamic acid were purchased and tested by AlphaScreen and X-ray crystallography (Table 4.8, series 1). Twenty-one compounds were not purchased because they were considered too expensive, to be sufficiently similar to purchased compounds or to contain reactive substructures. The eight amides were found to be inactive using a single shot AlphaScreen assay at 500 µM, except methoxybenzylamide **187** which was found to be an assay interference compound. Hydroxamic acids **179** and **182** were found to show good JMJD2D inhibition with pIC_{50}'s around 4.3. However, it is not clear if the inhibition is the result of allosteric binding or 2-OG competitive binding in the KDM active site.

[1] www.emolecule.com, accessed 3rd October 2016



Series 1: Amides
179, 182-189



Series 2: Benzimidazoles
180, 181, 190-218

Compound	Series	R_1	R_2	JMJD2D pIC_{50}[c]
Hit **179**	1	H	Hydroxy	4.3 ± 0.03 (2)
182	1	Methyl	Hydroxy	4.3 ± 0.03 (2)
183	1	H	H	<3.3 (1)
184	1	H	Methyl	<3.3 (1)
185	1	Methyl	Pyrazol-1-yl propyl	<3.3 (1)
186	1	H	Pyrazol-1-yl propyl	<3.3 (1)
187	1	H	4-Methoxybenzyl	Assay interference
188	1	H	2-(dimethylamino)pyridin-3-yl methyl	<3.3 (1)
189	1	H	*N*-Methylacetamid-2-yl	<3.3 (1)
Hit **180**	2	H	H	<3.3 (1)
Hit **181**	2	H	Methyl	<3.3 (1)
190	2	H	Cyanomethyl	<3.3 (1)
191	2	H	Hydroxymethyl	<3.3 (1)
192	2	H	Methylthio	<3.3 (1)
193	2	H	*tert*-Butyl	<3.3 (1)
194	2	H	Thiazol-4-yl	<3.3 (1)
195	2	H	Amino	<3.3 (1)
196[a]	2	H	Aminomethyl	<3.3 (1)
197	2	H	Benzyl	<3.3 (1)
198	2	H	Trifluoromethyl	<3.3 (1)
199	2	H	Furan-2-yl	4.2 ± 0.06 (2)
200	2	H	Bromo	<3.3 (1)
201	2	H	Hydroxy	<3.3 (1)
202	2	H	2-Aminoethyl	<3.3 (1)
203	2	H	4-Chlorobenzyl	<3.3 (1)
204[b]	2	H	(*R*)-1-Amino-3-methylbutan-1-yl	<3.3 (1)
205	2	H	*N*-(2-hydroxyethyl)amino	<3.3 (1)
206[b]	2	2-Aminoethyl	Methyl	<3.3 (1)
207	2	3-Aminopropyl	Methyl	~3.3 (1)
208	2	2-Hydroxyethyl	Methyl	<3.3 (1)
209	2	2-Cyanoethyl	Methyl	<3.3 (1)
210	2	Prop-2-yn-1-yl	Methyl	3.6 ± 0.03 (2)
211	2	Ethyl	Hydroxymethyl	<3.3 (1)
212	2	*neo*-Propyl	Hydroxymethyl	<3.3 (1)
213	2	Cyclopropyl	Ethyl	Assay interference
214	2	Methyl	Hydroxy	<3.3 (1)
215[a]	2	Ethyl	Aminomethyl	<3.3 (1)
216	2	Cyclopropyl	Aminomethyl	~3.3 (1)
217	2	Methyl	Thio	<3.3 (1)
218	2	*neo*-Propyl	Thio	<3.3 (1)

Table 4.8: SAR by catalogue JMJD2D follow-up library. [a]Di-hydrochloride salt. [b]Hydrochloride salt. [c]By AlphaScreen peptide displacement assay.

Twenty-nine benzimidazoles were purchased and screened against JMJD2D using AlphaScreen and X-ray crystallography (Table 4.8, series 2). Only three compounds were found to show >80% inhibition of JMJD2D at 500 μM. Full IC_{50} curves were calculated for furan **199** (pIC_{50} 4.2) and alkyne **210** (pIC_{50} 3.6), while cyclopropyl **213** was found to be an assay interference compound.

The activity of furan **199** is comparable to the hydroxamic acids **179** and **182**, and is likely result of metal chelation by the furan-2-yl heterocycle motif rather than allosteric binding.

A further eleven heterocyles were purchased to find alternatives to the benzimidazole core (Figure 4.25). The new scaffolds were based upon 5,6-fused *N*-containing heterocycles. All eleven compounds showed no JMJD2D inhibition at 500 µM except benzopyrazole **225** which was found to be an assay interference compound.

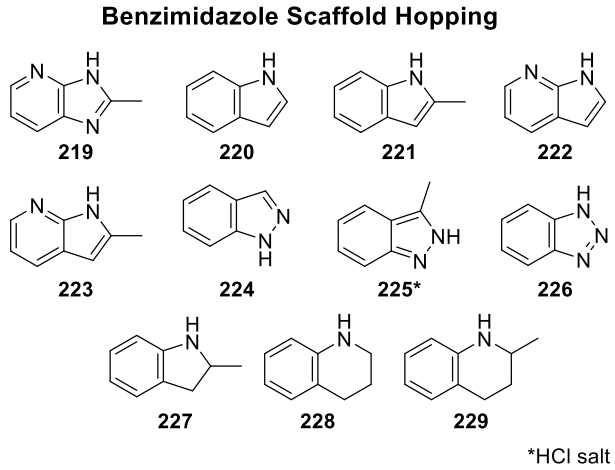


*Figure 4.25: JMJD2D follow-ups designed by scaffold hopping from benzimidazole hits **180** and **181**. All compounds were found to be inactive at 500 µM against JMJD2D by AlphaScreen peptide displacement assay, except benzopyrazole **225** which was found to be an assay interference compound.*

All compounds purchased as a part of the JMJD2D follow-up campaign were subject to fragment soaking experiments, however no crystal structures were identified. Further investigations are currently being undertaken in the Brennan group to find hydroxamic acid bioisosteres and identify other JMJD2D allosteric binders.

4.5.4 SAR by Catalogue vs Poised Elaboration

In order to assess the benefits of using poised elaboration to perform follow-up campaigns over SAR by catalogue, a comparison was made between the coverage of chemical space by each technique. First, poised follow-up libraries for JMJD2D hits **179** and **180** were designed using

XPoise. Phenol-2-amide analogues were built using an amide coupling reaction between 2-hydroxybenzoic acid and amine building blocks, giving a library of 1986 compounds. Benzimidazole analogues were built using the reaction of 1,2-dianiline and carboxylic acid and acid chloride building blocks, giving a library of 4624 compounds.

The two libraries were combined with their respective SAR by catalogue library built using substructure searching. Radial fingerprints were generated for each compound using MACCS keys (32-bit, N = 2). A pairwise distance matrix was generated using the Tanimoto distance between each fingerprint. The distances were then visualised as a 2D-scatter plot using a multi-dimensional scaling (MDS) algorithm (Figure 4.26) (Borg and Groenen 2005).

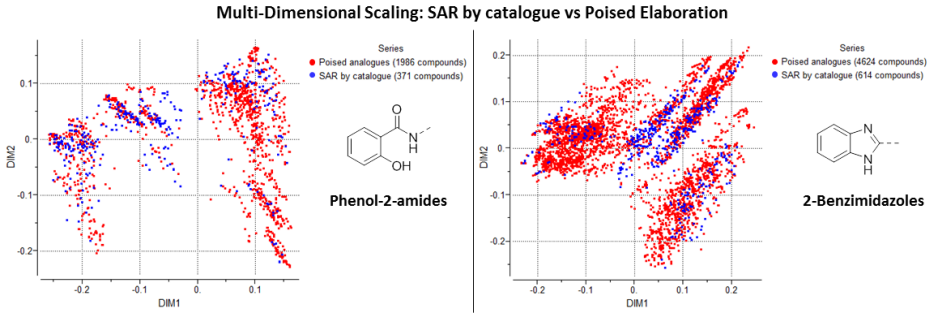


Figure 4.26: Comparison of SAR by Catalogue versus Poised Elaboration of JMJD2D fragment hits using the Tanimoto distance between radial fingerprints. Scatter plots were generated using MDS (N=2). SAR by catalogue libraries built using substructure searching on the eMolecule website. Poised analogues were designed using XPoise. Left: Comparison of follow-up libraries based upon hydroxamic acid ***179****. Right: Comparison of follow-up libraries based upon benzimidazoles* ***180*** *and* ***181****.*

The MDS scatter plots indicate SAR by catalogue and poised elaboration are able to explore similar areas of chemical space. However, the relatively high number of potential compounds accessible using poised elaboration gives much better coverage of the chemical space, and offers access to new areas of chemical space. This is particularly true of the poised analogues of the benzimidazole series, which cover large patches of chemical space not covered at all by the SAR by catalogue analogues.

The analysis reveals the benefits of using poised chemistry to design follow-up libraries. While SAR by catalogue offers a low cost and high-speed method to explore the chemical space around a hit, poised elaboration gives a much more rigorous assessment of potential growth vectors with only a small increase in time and financial commitment. In the next section I will demonstrate how a poised follow-up campaign can be used to generate hit analogues quickly and at relatively low cost using the Tudor domain of SETDB1 as a case study.

4.6 Tudor Domain of SETDB1

4.6.1 SETDB1 Fragment Soaking Campaign

The full length SETDB1 protein is a HMT with a catalytic SET domain and a triple Tudor domain containing a Kme binding site. Overexpression of SETDB1 has been observed to contribute to lung cancer and increases tumour invasiveness (Rodriguez-Paredes, Martinez de Paz *et al.* 2014). Meanwhile, depletion of SETDB1 using histone mutation reduces cancer growth in cell culture and nude mice models (Shah, Denton *et al.* 2014). To date, no HMT inhibitors of SETDB1 have been reported. Inhibition of non-catalytic domains of SETDB1 offers an alternative strategy to targeting the protein. In 2008, SGC Toronto solved and deposited the apo structure of the SETBD1 Tudor domain (PDB ID 3DLM), and recently performed a fragment screening campaign in collaboration with DLS to obtain crystal structures of compounds in complex to the Kme2 peptide binding site. Four fragment hits were identified from DSPL to bind in the peptide binding site of the SETDB1 Tudor domain, but outside the Kme2 recognition pocket (Figure 4.27 A and B).

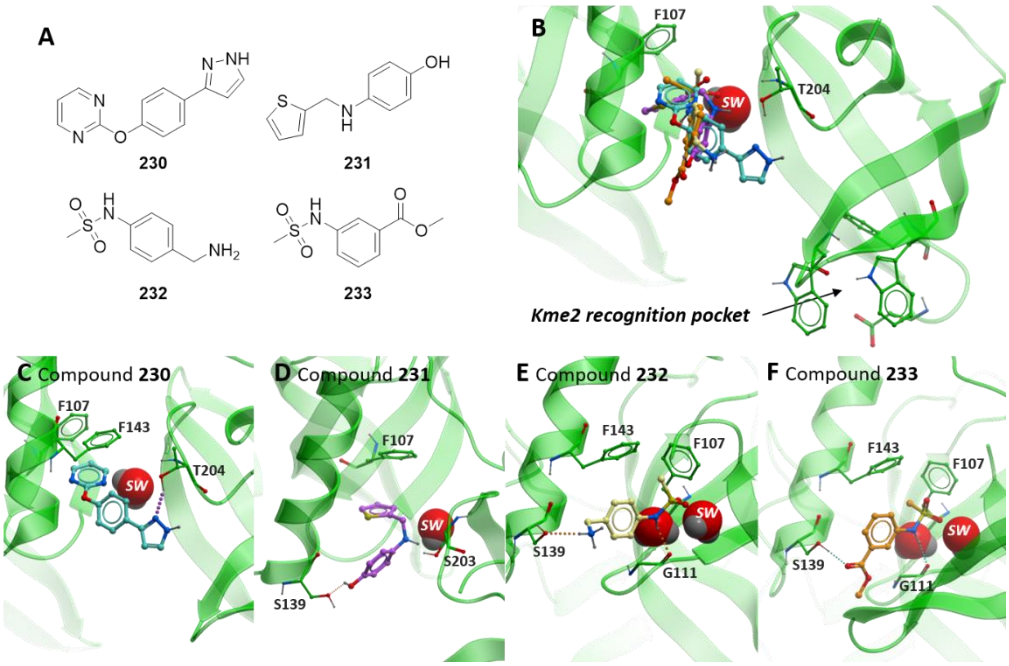


*Figure 4.27: Fragment hits in complex to the SETDB1 Tudor domain (green ribbon and sticks) identified by crystal soaking. SW: SETDB1 conserved water molecule (red CPK). (A) DSPL fragments **230-233**. (B) Overlay of fragments **230-233** in complex to SETDB1 and the nearby Kme2 recognition pocket. (C) The pyrimidine head group of hit **230** (cyan sticks) forms an edge-face interaction with F107 and a π-stacking interaction with F143. The pyrazole tail forms a H-bond to T204 (purple dots). (D) The thiophene head group of hit **231** (purple sticks) forms a π-stacking interaction with F107. The secondary amine forms a H-bond with either S203 or SW. The phenol tail forms a H-bond to S139 (orange dashes). (E) The methyl sulfonamide head group of amine **232** (light yellow sticks) is not bulky enough to displace a structural water observed in the apo structure. As a result, the sulfonamide NH is able to form a H-bond to the protein backbone, and the sulfonamide oxygens interact with SW (orange dashes). The primary amine forms a H-bond to S139. (F) Ester **233** (orange sticks) has a similar binding pose to **232**, with the sulfonamide head group interacting with SW and the ester carbonyl oxygen forming a H-bond with S139 (cyan dashes).*

The fragment binding site contains two water molecules in the apo protein form, both anchored by H-bonds to the protein backbone. One of the molecules, dubbed the SETDB water (*SW*) is present in all four crystal structures. In contrast the heteroaromatic head groups of hits **230** and **231** displace the other water molecule in order to occupy the aromatic cage formed by F107 and F143 (Figure 4.27 C and D). Binding of pyrimidine **230** is anchored by a H-bond between pyrazole and T204. Meanwhile, thiophene **231** forms a H-bond from the phenol OH to S139, and a second H-bond from the secondary amine NH to S203 and/or the *SW* molecule.

Sulfonamides **232** and **233** have similar binding poses (Figure 4.27 E and F). The sulfonamide oxygens form polar interactions with the *SW* molecule, and the sulfonamide NH forms a H-bond to the protein backbone at G111. The phenyl cores form edge-face π-bonds to F143, allowing the primary amine of **232** and the ester carbonyl of **233** to form H-bonds to S139.

4.6.2 SETDB1 Follow-up Libraries

Follow-up libraries for SETDB1 were designed using the XPoise algorithm described in the previous chapter in order to compare the price and lead times of performing poised parallel synthesis with performing SAR by catalogue. Eight libraries (two for each fragment hit, see Table 4.9) were built using the Aldrich CPR building blocks catalogue in XPoise.[1] LLOOMMPPAA was used to select 30 compounds from each library, which was then triaged by manual inspection to leave fifteen compounds per library. The follow-up series for compound **231** in which the aldehyde synthon was retained failed during LLOOMMPPAA analysis as a result of incorrect Context SMILES matching, leaving a final library of 105 compounds (7 series x 15 compounds) for synthesis. A quotation from a CRO (Contract Research Organisation) for the preparation of all 105 compounds was deemed prohibitively expensive. A subset of building blocks were contributing disproportionately to the cost of synthesis, with thirteen requiring multi-step synthesis. This resulted the decision to stop using the Aldrich CPR catalogue and to develop a commercial building blocks library containing prices, quantities and trusted vendors (Section 3.3.2, page 80).

[1] www.sigmaaldrich.com/chemistry/chemistry-services/discovery-cpr.html, accessed 3rd October 2016

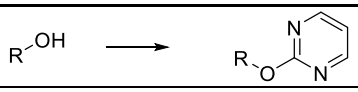
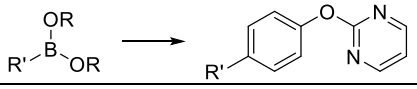
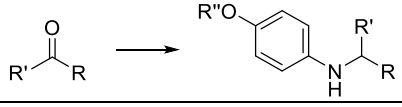
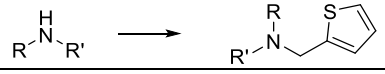
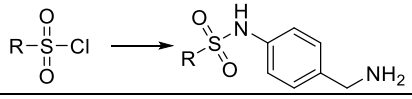
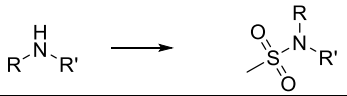
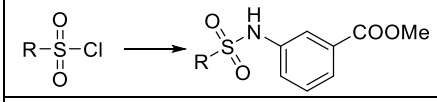
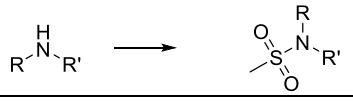
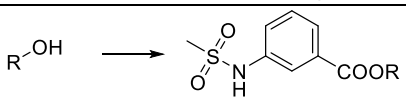
Hit	Poised Reaction	XPoise Reaction Scheme	BB Library[a]	XPoise Library Size
230	Ether coupling *via* S_NAr		A	1078
			B	1933
	Suzuki-type cross-coupling		A	924
			B	1684
231	Reductive amination		A (R''=H)	3183
			B (R''=Me)	1710
			A	12339[b]
			B	2190[b]
232	Sulfonamide coupling		A	453
			B	423[b]
			A	12334
			B	2184
233	Sulfonamide coupling		A	453
			B	423
			A	12334
	Esterification		B	1571

Table 4.9: Design of follow-up libraries for fragment hits of SETDB1.[a]Building block libraries. A: Aldrich CPR catalogue; B: Vendor BB catalogues. [b]Triaging using the LLOOMMPPAA failed as a result of incorrect Context SMILES matching.

The XPoise/LLOOMMPPAA method was performed a second time on eight series using the Vendor Building Blocks catalogue (Table 4.9). A series from each of the compounds **231** and **232** failed the LLOOMMPPAA method as a result of incorrect Context SMILES matching, meaning after manual triaging of the libraries, a final library containing 90 compounds (6 series x 15 compounds) was requested for synthesis by WuXi.[1] In total, 66 compounds were supplied (73%), including five compounds below 95% purity. Of the 24 compounds not delivered, six failed in synthesis, one failed in purification and the synthesis of eighteen compounds was not attempted because the starting materials could not be sourced, reflecting the difference in vendor catalogues between

[1] www.wuxiapptec.com, accessed 3rd October 2016

China and the UK. The time from placing the order to compound delivery was six weeks, and the final price worked out at around £100 per compound for 10-100 mg of material, including postage and packaging. This compares favourably with SAR by catalogue, for which delivery times range from 2-4 weeks, and compounds priced around £40-60 per compound for similar quantities, excluding postage and packaging. The cost of postage and packaging is difficult to estimate because the amount charged varies between suppliers and is often dependent on the number of compounds ordered.

The SETDB1 follow-up library is currently undergoing testing for SETDB1 inhibition and binding using AlphaScreen and X-ray Crystallography at SGC Toronto and DLS.

4.7 Ongoing Work and Conclusions

In this chapter I have used follow-up campaigns against epigenetic targets to demonstrate the benefit of using poised hit-to-lead development on low affinity fragments detected using X-ray crystallography. Initially, analogues for two fragment hits of BRD1 with unusual binding modes were synthesised using poised chemistry. The resulting library of 24 compounds did not show any improvement in potency, but demonstrated how low affinity fragments can have multiple binding poses.

Two poised fragment hits against PHIP(2) were selected for follow-up synthesis, resulting in two lead series for a previously intractable target with sub-mM activity, good LE and an array of structural data. A library of 35 thioureas and ureas improved *in vitro* activity of the initial hit alongside new structural data. A second thiourea library designed using the XPoise/LLOOMMPPAA method improved activity further, resulting in a 10-fold improvement on the initial hit while maintaining ligand efficiency. Furthermore, the second library showed enrichment for active compounds when compared to the first follow-up library designed by

manual inspection of the crystal structures. A library of fourteen *N*-benzylamides generated four new crystal datasets, including protein-ligand complexes for compounds found to be inactive in an AlphaScreen peptide displacement assay. Three compounds were found to have sub-mM activity despite not being able to measure the activity of the initial hit.

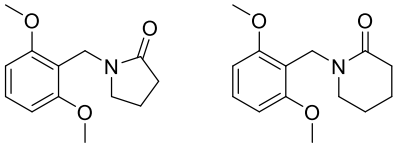
SAR by catalogue was utilised to investigate two poised hits for the KDM, JMJD2D. The two compounds had been found to bind to the protein in a novel allosteric site. However, a library of 48 analogues did not show any affinity for JMJD2D in an AlphaScreen assay and produced no new structural data. A comparison of chemical space coverage by SAR by catalogue versus poised elaboration revealed how the latter method offers a more rigorous exploration of chemical space around a hit compound.

Finally, the XPoise/LLOOMMPPAA method was used to build follow-up libraries for four hits of the SETDB1 Tudor domain. A library of 66 compounds (from a total of 90 requested) were synthesised and delivered by WuXi, a CRO, within six weeks and at a cost of around 100 GBP per compound. The time and price compare favourably with SAR by catalogue considering the improved coverage of chemical space. Analysis of compounds which were not delivered reveals the majority of failures to be a result of unavailable starting materials. It is hoped this can be reduced by working with WuXi to build an input library for XPoise containing Chinese vendor catalogues.

5 Synthesis of *N*-Dimethoxybenzyl Lactam Analogues as PHIP(2) Inhibitors

5.1 Overview

In this chapter I elaborate a lead series which had been optimised using poised chemistry (Chapter 4). The *N*-dimethoxybenzyl amide series was found to contain modest inhibitors of the PHIP(2) Brd. Two techniques were used to elaborate the series towards a potent and selective chemical probe. First, a late-stage functionalisation using microbial biotransformation was attempted in collaboration with Hypha Discovery to hydroxylate the lead compound **177** *via* C-H activation. Secondly, the lactams **177** and **178** (Figure 5.1) were decorated using a diastereoselective lactamisation cascade reaction to grow the fragment towards a previously unexplored part of the PHIP(2) binding site.



177
PHIP(2) pIC_{50} = 3.3
LE = 0.27

178
PHIP(2) pIC_{50} = 3.5
LE = 0.27

*Figure 5.1: PHIP(2) binders **177** and **178** used as starting points for lead optimisation.*

In the final part of the chapter I explore the use of one-step multi-component reactions (MCRs) to build bespoke, complex fragments which could be added to a screening library. The use of MCRs as poised reactions offers the opportunity to build complexity into fragments early in the hit-to-lead process and expand the scope and diversity of poised fragment space.

5.2 Microbial Biotransformation

Medicinal chemists use SAR data to direct lead optimisation. However, pragmatic decisions have to be made based upon viable synthetic routes and availability of commercial starting materials. Microbial biotransformation is a novel method for derivatisation of a lead compound to give synthetically challenging products. Hypha Discovery[1] specialises in using microbial panels to perform *C-H* activation on lead compounds to yield hydroxylated derivatives. The method has been used by pharmaceutical companies to reduce lead lipophilicity, to block metabolic hot spots and to incorporate new reactive chemical handles (Boer, Young-Sciame *et al.* 2016).[2]

The γ-lactam **177** was selected for testing using microbial biotransformation. At the time, the compound was the lead molecule in a series of *N*-dimethoxybenzyl amides which inhibit the PHIP(2) Brd. It was thought the aliphatic *C-H* bonds on the lactam ring could be hydroxylated to afford novel derivatives. Hypha Discovery used **177** to perform a biotransformation screen using a range of bacterial species known to metabolise small molecules and following the procedure described in Appendix F. Two species were identified as yielding metabolites of interest. The reaction with the Sp59 strain was scaled-up to afford around 5 mg of two hydroxylated compounds, **234** and **235**, each at >95% purity. The compounds were returned to SGC Oxford for characterisation and *in vitro* screening.

5.2.1 Characterisation of Hydroxylated Lactams

Compounds **234** and **235** were both found to be 251 Da (+16 Da) metabolites of the original lactam **177** (MW 235 Da), confirming a hydroxylation biotransformation. Both compounds were

[1] www.hyphadiscovery.co.uk, accessed 30th September 2016

[2] www.hyphadiscovery.co.uk/wp-content/uploads/2015/02/Biological-C-H-bond-activation-newsletter-Q1-2015.pdf, accessed 4th October 2016

subjected to NMR (nuclear magnetic resonance) analysis in deuterated chloroform using proton, APT (attached proton test), COSY (homonuclear correlation spectroscopy), HSQC (heteronuclear single quantum coherence) and HMBC (heteronuclear multiple-bond correlation) NMR spectroscopy. NMR is a commonly used spectroscopic technique for the identification and characterisation of organic molecules. The most commonly used methods are 1D ^{1}H and ^{13}C NMR, from which individual protons and carbon atoms within a molecule can be identified and potentially assigned based on chemical shifts, multiplicities and coupling constants. APT is a variation of ^{13}C NMR in which proton decoupling is performed to differentiate between primary, secondary, tertiary and quaternary carbons. An APT spectrum is the same as a ^{13}C spectrum except secondary and quaternary carbon centres give negative signals. COSY is a 2D-NMR experiment used to identify coupling between homonuclei, usually in the form of $^{3}J_{H,H}$ coupling. Similarly, HSQC is a 2D-NMR experiment used to identify coupling between covalently bonded heteronuclei, most commonly of the form $^{1}J_{H,C}$ coupling. Finally, the 2D-experiment, HMBC, measures long-range coupling between heteronuclei in the form of $^{3}J_{H,C}$ coupling.

The ^{1}H-NMR for compound **234** revealed a decrease in the number of aromatic protons compared to compound **177**, indicating hydroxylation of the aromatic ring (Figure 5.2 A). This hypothesis was further supported by the presence of a broad singlet with chemical shift 5.54, in the range normally associated with phenolic protons. Finally, COSY NMR confirmed the aromatic protons to be adjacent to one another (J = 8.8 Hz), allowing compound **234** to be identified as *N*-(3-hydroxy-2,6-dimethoxybenzyl)pyrrolidin-2-one (Figure 5.2 B).

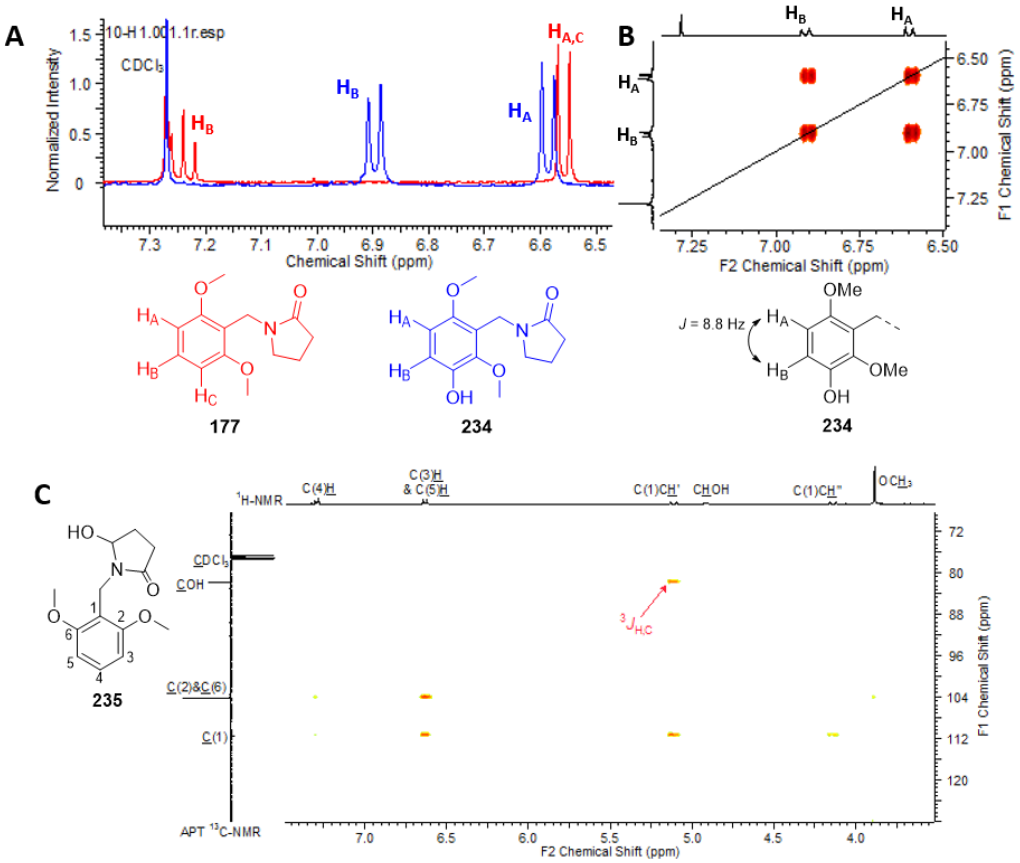


*Figure 5.2: NMR characterisation of hydroxylated lactams **234** and **235**. (A) Aromatic region from the proton NMR of compounds **177** (red) and **234** (blue). The distinctive triplet-doublet pattern observed for the 2,6-disubstituted benzene ring of compound **177** becomes a doublet-doublet pattern representative of a 2,3,6-trisubstituted benzene ring in compound **234**. (B) Aromatic region from the COSY NMR of compound **234** showing coupling between the aromatic protons (J = 8.8 Hz). (C) Selected region of the HMBC NMR spectra of compound **235** showing the $^{3}J_{H,C}$ coupling between one of the benzylic protons and the hydroxylated carbon, confirming the compound to be a 5-hydroxylactam derivative.*

The proton NMR for derivative **235** revealed one less proton in the aliphatic region when compared to starting compound **177**, indicating hydroxylation of the lactam ring or the benzylic carbon. The hydroxylated carbon is the only tertiary carbon centre in the molecule, and as a result was identified by APT. Observation of $^{3}J_{H,C}$ coupling between a benzylic proton and the hydroxylated carbon in the HMBC spectrum indicated the hydroxyl-group to be at the 5-position of the lactam ring as this is the only aliphatic position three bonds from the benzylic carbon (Figure 5.2 C). This was confirmed upon full proton and carbon assignment, allowing identification of compound **235** as *N*-(2,6-Dimethoxybenzyl)-5-hydroxypyrrolidin-2-one.

Unfortunately, screening of derivatives **234** and **235** at 2 mM against PHIP(2) in an AlphaScreen peptide displacement assay showed <60% inhibition. However the hydroxyl groups provide reactive handles from which future derivatisation could be undertaken. Potential reactions include oxidation, nucleophilic substitution or ether formation. Compounds **234** and **235** are particularly amenable to such derivatisations as the expected binding poses would direct the hydroxyl groups towards unexplored areas of the PHIP(2) binding site.

5.3 PHIP(2) Lead Optimisation

In Chapter 4 I described the identification of two lead series in the development of PHIP(2) inhibitors using poised chemistry. The *N*-benzyl amide series included the γ- and δ-lactams **177** and **178** which showed modest activity against PHIP(2) (pIC_{50} 3.3 and 3.5 respectively). Although crystal structures for these compounds were not obtained, an overlay of fragment hit **82** with the previously reported NMP-bound structure (Figure 4.10, page 114) indicated the lactam **177** could act as a KAc mimetic by binding in an equivalent pose. Furthermore, the location of the C4 and C5 positions of the lactam ring offered the opportunity for fragment growth towards a previously unexplored part of the PHIP(2) binding site containing polar residues including glutamine and glutamic acid. A diastereoselective cascade reaction was identified from the literature and used to decorate the lactams with incorporation of chiral carbon centres (Pelletier, Ray *et al.* 2011).

5.3.1 Chirality

The presence of three dimensional characteristics in a drug or chemical probe has been shown to improve potency and selectivity. The use of chiral centres is an effective way of increasing fsp^3 and incorporating molecular complexity (Morley, Pugliese *et al.* 2013). Chiral compounds are stereoisomers with non-superimposable mirror images. The rigid tetrahedral arrangement of a carbon centre with sp^3 hybridisation means any carbon with four unique substituents is a chiral

centre. The mirror images are called enantiomers and have the same chemical and physicochemical properties, but in chiral environments such as biological systems, opposing enantiomers can display differing properties. Amino acids are naturally occurring chiral compounds, with one enantiomer of each amino acid being found in nature (Figure 5.3). Some amino acids, such as isoleucine and threonine, are diastereomers as well as enantiomers. Diastereomers are stereoisomers with two or more chiral centres but which are not mirror images of one another. By extension, a compound with two chiral centres, which has four possible stereoisomers, has two pairs of enantiomers. However, the stereoisomers differing by just one chiral centre are diastereomers of one another and have different chemical and physicochemical properties as well as potentially differing biological properties (Figure 5.3).

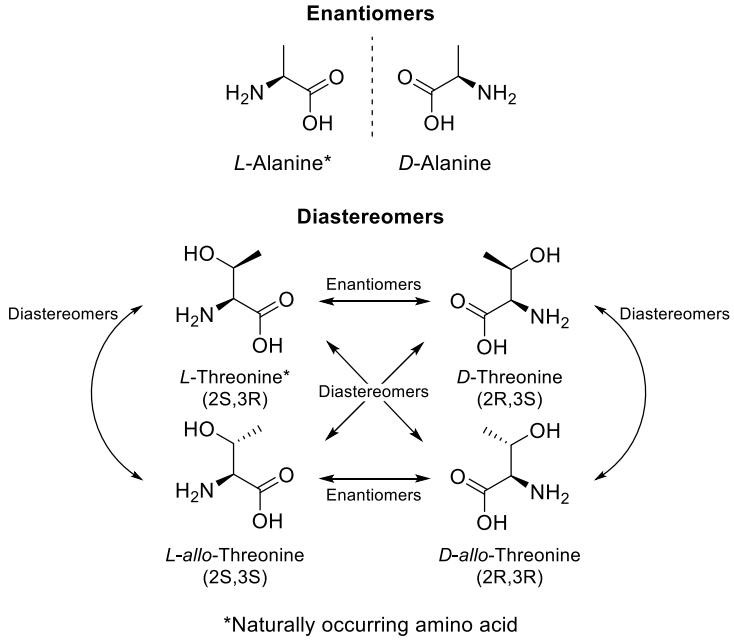


Figure 5.3: Chirality in amino acids. Most amino acids, such as alanine, have one chiral centre although only one enantiomer of each amino acid is found in nature. Threonine is an example of a diastereomer, meaning naturally occurring threonine has one enantiomer and two diastereomers.

A recent example of using chirality in the design of chemical probes is in the development of the BRD7/9 inhibitor, LP99 (Clark, Vieira *et al.* 2015). Starting from the planar, aromatic, Brd-promiscuous fragment hit, *N*-methylquinolone **236**, fragment growth at the 7-position to give δ-

lactam **237** returned good potency and selectivity for BRD9 (Figure 5.4). By utilising a nitro-Mannich lactamisation cascade, chiral lactam analogues such as diastereomer **238** could be rapidly synthesised for SAR studies. Diastereomerically enriched products were isolated and tested in racemic mixtures to identify a lead series. Once the desired activity and selectivity was achieved, preparative HPLC on chiral stationary phase was used to isolate the active enantiomer, compound (2*R*,3*S*)-**239**, the absolute configuration of which was later confirmed using protein X-ray crystallography. The inactive enantiomer, (2*S*,3*R*)-**239**, can be used as a negative control.

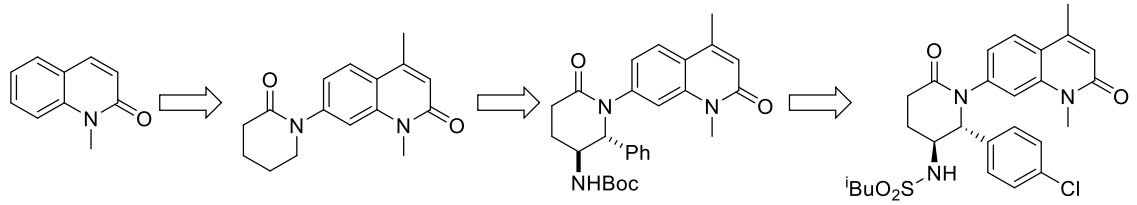


Figure 5.4: Development of the BRD7/9 chemical probe, LP99, using fragment growth followed by the incorporation of chiral centres (Clark, Vieira et al. 2015). All K_d values are for BRD9.

5.3.2 Nitro-Mannich Lactamisation Cascade in the Synthesis of PHIP(2) Inhibitors

In order to grow lactams **177** and **178** into a previously unexplored part of the PHIP(2) binding site, the strategy utilised in the development of LP99 was adopted. The nitro-Mannich Lactamisation cascade was utilised to decorate the 4- and 5-positions of compound **177**, and the 5- and 6-positions of compound **178**, in order to gather initial SAR data.

The treatment of methyl 3-nitropropanoate with an imine has been shown to form γ-lactams in high diastereomeric excess (Pelletier, Ray *et al.* 2011). Furthermore, *in situ* formation of the imine from an amine and aldehyde in the presence of a stoichiometric amount of acid yields the same product in good to excellent yield, and was utilised in the synthesis of lactam **177** derivatives (Figure 5.5). Similarly, the treatment of methyl 4-nitrobutanoate with an amine, aldehyde and

acid yields δ-lactams with high diastereomeric excess (Jakubec, Cockfield *et al.* 2012), and was used in the synthesis of lactam **178** derivatives (Figure 5.6).

Methyl 3-nitropropanoate **241** was formed from the commercially available acid **240** using a literature procedure and isolated in good yield (Silva, Costa *et al.* 2001). Treatment of ester **241** with 1.5 equivalents of 2,6-dimethoxybenzylamine, benzaldehyde and benzoic acid in toluene (0.25 M) at 70 °C afforded the γ-lactam **242** as a single diastereomer in 54% yield (Figure 5.5). The identity of the diastereomer was confirmed by NMR, which revealed the $^{3}J_{trans}$ (H-5/H-4) coupling constant to be 2.1 Hz, in the same range originally reported by Pelletier *et al.* (Pelletier, Ray *et al.* 2011). In contrast the diastereomer formed when isobutyraldehyde is used could not be immediately identified because the γ-lactam **243**, isolated in 63% yield, shows no $^{3}J_{trans}$ (H-5/H-4) coupling.

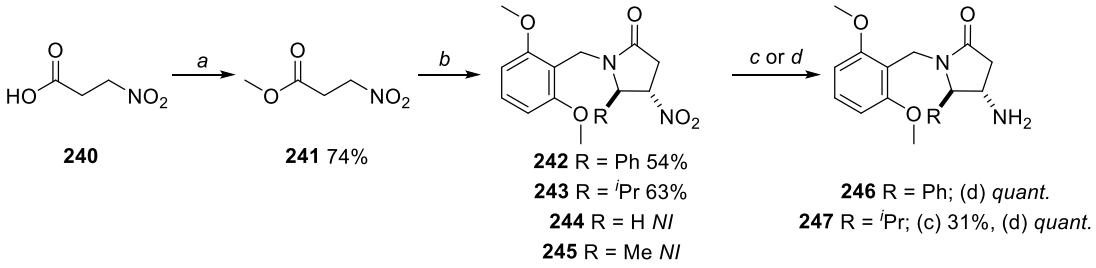
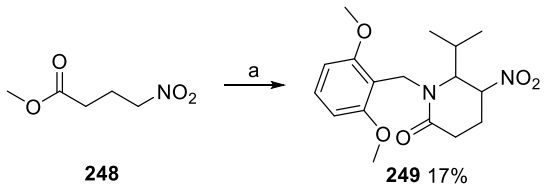


*Figure 5.5: Synthesis of amines **246** and **247** using a nitro-Mannich lactamisation cascade. Reagents and conditions: (a) H_2SO_4, MeOH, rt, 48 h. (b) 1,2-Dimethoxybenzylamine, RCHO, PhCOOH, toluene, 70 °C. (c) Zinc, HCl, EtOAc:MeOH (2:1). (d) Raney Nickel, H_2, MeOH, rt. NI: Not isolated.*

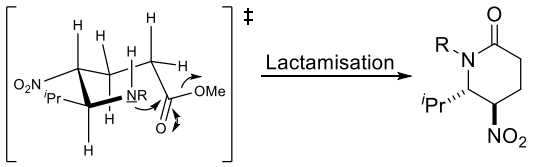
Efforts to synthesise lactams **244** and **245** from formaldehyde and acetaldehyde respectively were unsuccessful, likely as a result of the high reaction temperature vapourising the aldehydes slowing imine formation.

Reaction of methyl 4-nitrobutanoate **248** with 1.0 equivalents of 2,6-dimethoxybenzylamine and isobutyraldehyde, and 1.5 equivalents of benzoic acid in toluene at 70 °C afforded δ-lactam **249** in 17% yield as a single diastereomer (Figure 5.6). The identity of the diastereomer could not be

unambiguously assigned, although the *trans*-diastereomer is reported as the major reaction product in the literature (Jakubec, Cockfield *et al.* 2012). Two hypotheses are proposed to rationalise the preferential formation of the *trans*-diastereomer. Firstly, the nitro-Mannich reaction is fast and reversible, allowing preferential lactamisation *via* the thermodynamically lower energy transition state. Alternatively, lactamisation occurs to both diastereomeric products of the nitro-Mannich reaction at an equal rate, but post-reaction epimerisation of the nitro-bearing stereogenic centre leads to the preferential formation of the thermodynamically more stable product (Figure 5.6).



Hypothesis 1: Preferential cyclisation *via* the thermodynamically more stable transition state



Hypothesis 2: Epimerisation to the (assumed) thermodynamically more stable product

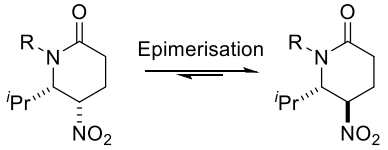


*Figure 5.6: Synthesis of δ-lactam **249**. The identity of the diastereomer could not be unambiguously assigned, although the trans-product has been reported as the major product of the reaction (Jakubec, Cockfield et al. 2012). The trans-diastereomer could be formed by preferential cyclisation through the lower energy transition state, or as a result of post-reaction epimerisation to the (assumed) thermodynamically more stable product. Reagents and conditions: (a) 1,2-Dimethoxybenzylamine, ^{i}PrCHO, PhCOOH, toluene, 70 °C.*

Reduction of the nitro-group of lactam **243** to the primary amine was attempted using zinc in acidic conditions, giving amine **247** in 31% yield (Figure 5.5). Less harsh reducing conditions were sought, and when lactams **242** and **243** were treated with an excess of Raney-nickel in MeOH under hydrogen, amines **246** and **247** were afforded in quantitative yield. The $^{3}J_{trans}$ (H-5/H-4) coupling constant for the phenyl analogue **246** was observed to be 3.6 Hz, indicating the reaction

had not affected the chiral centres. Again, no $^3J_{trans}$ (H-5/H-4) coupling was observed for the isopropyl analogue **247**. 3D-modelling of *trans*-**247** using ChemBio3D Ultra[1] indicated the amine and isopropyl groups to be axial with respect to the ring in the lowest energy conformation, driven by the steric bulk of the isopropyl group clashing with the dimethoxybenzyl group. Furthermore the conformation allows formation of an H-bond between the amine and the 2-methoxy-oxygen on the phenyl ring, which are only 3.1 Å apart (Figure 5.7). As a result, the H-4/H-5 dihedral angle is predicted to be 87.8°. The Karplus equation predicts coupling constants between adjacent carbons to be smallest when the dihedral angle is around 90° (Karplus 1959), explaining why no $^3J_{trans}$ (H-5/H-4) coupling is observed. Furthermore, the H-4/H-3_{AX} dihedral angle is 35.7° in comparison to an H-4/H-3_{EQ} dihedral angle of 88.5°. These predicted angles are in agreement with the observed $^3J_{trans}$ (H-4/H-3_{AX}) coupling constant of 7.2 Hz, together with no detectable $^3J_{trans}$ (H-4/H-3_{EQ}) coupling.

Comparable coupling constants are observed for the nitro-precursor **243**, indicating an analogous *trans*-axial conformation driven by the steric bulk of the isopropyl group. In contrast, the phenyl analogue **246** is likely to be in the favourable equatorial-equatorial conformation because of less steric clashes between the planar phenyl rings and the dimethoxybenzyl group.

[1] www.cambridgesoft.com, accessed 29th September 2016

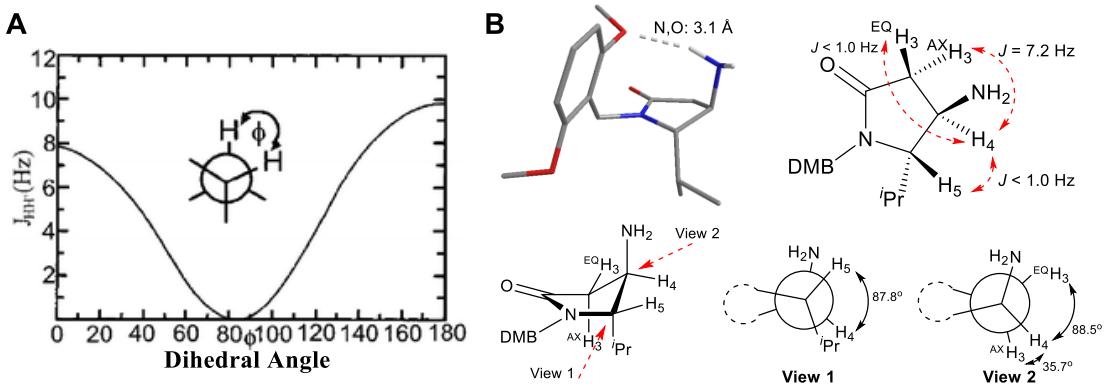


*Figure 5.7: Relative assignment of the isolated diastereomer, amine **247**, using NMR. (A) Typical example of the vicinal Karplus relationship between the dihedral angle (ϕ) and the coupling constant J. Reproduced with permission from The Royal Society of Chemistry (Benniston and Harriman 2006). (B) Predicted conformation of the single enantiomer (4R,5S)-**247**. The trans-amino and isopropyl groups are axial with respect to the lactam ring, allowing the formation of an intramolecular H-bond between the amine and 2-methoxy oxygen. The predicted dihedral angles are in agreement with the experimentally observed NMR coupling constants. DMB: dimethoxybenzyl.*

With amine **247** in hand, a library of amide and sulfonamide derivatives were sought for testing against PHIP(2). Seven acid chlorides and three sulfonyl chlorides were selected to include a range of chemotypes. All ten products were isolated in moderate to good yield and the same *trans*-axial-conformation was observed by NMR with respect to the isopropyl and amide/sulfonamide groups (Figure 5.8).

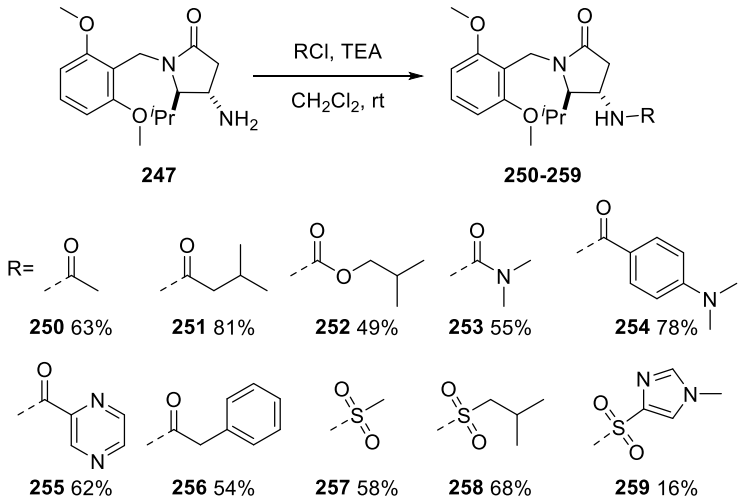
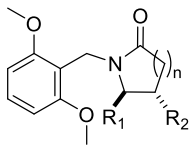


*Figure 5.8: Synthesis of amine **247** derivatives using amide and sulfonamide coupling reaction conditions.*

Nitro compounds **242**, **243** and **249**, amines **246** and **247**, amides **250-256**, and sulfonamides **124-259** were all tested in the AlphaScreen assay for activity against PHIP(2) (Table 5.1). Unfortunately, all amide and sulfonamide derivatives showed no activity in the AlphaScreen assay. In fact the only isopropyl analogues to show PHIP(2) activity were the nitro-containing γ- and δ-lactams **243** and **249** (pIC_{50} 2.9 and 3.3 respectively). In contrast, the phenyl analogues **242** and **246** showed good activity with pIC_{50}'s 3.5 and 3.9 respectively. The true pIC_{50} of amine **246** could be as high as 4.2 if only one of the enantiomers present has PHIP(2) activity. The results indicate the locked axial-axial conformation observed for the isopropyl derivatives is not favourable for binding in the PHIP(2) KAc binding site. Conversely, the equatorial-equatorial conformation adopted in the presence of a planar phenyl ring appears to be more favourable for binding, and inspection of the crystal structure indicates the binding site is likely to better accommodate this conformation.



Compound	R_1	R_2	n	PHIP(2) pIC_{50}[a]
177	H	H	1	3.3 ± 0.04 (2)
178	H	H	2	3.5 ± 0.04 (2)
242	Phenyl	Nitro	1	3.5 ± 0.11 (2)
243	Isopropyl	Nitro	1	2.9 ± 0.14 (2)
246	Phenyl	Amino	1	3.9 ± 0.07 (2)
247	Isopropyl	Amino	1	<2.3 (2)
249	Isopropyl	Nitro	2	3.3 ± 0.10 (2)
250	Isopropyl	Acetamide	1	<2.3 (2)
251	Isopropyl	3-Methylbutanamide	1	<2.3 (2)
252	Isopropyl	Isobutyl formamide	1	<2.3 (2)
253	Isopropyl	1,1-Dimethylurea	1	<2.3 (2)
254	Isopropyl	4-(Dimethylamino)benzamide	1	<2.3 (2)
255	Isopropyl	Pyrazine-2-carboxamide	1	<2.3 (2)
256	Isopropyl	2-Phenylacetamide	1	<2.3 (2)
257	Isopropyl	Methanesulfonamide	1	<2.3 (2)
258	Isopropyl	2-Methylpropane-1-sulfonamide	1	<2.3 (2)
259	Isopropyl	*N*-Methylimidazole-4-sulfonamide	1	<2.3 (2)

Table 5.1: PHIP(2) inhibition of the N-dimethoxybenzyl lactam series. [a]By AlphaScreen peptide displacement assay.

Efforts are ongoing to obtain crystal structures of the lactam compounds in complex to PHIP(2) and to isolate single enantiomers for biological testing. However, the SAR presented here provides

enough information to guide future lead optimisation. Further exploration at the 5-position of the γ-lactam is required, alongside further derivatisation of the 4-amino functionality using amide and sulfonamide coupling, reductive amination or Buchwald-type cross-coupling. Investigations into δ-lactam analogues would also offer the opportunity to consider 5,6-substituted lactams and examine the effect *trans*- versus *syn*- diastereomers have on biological activity.

As demonstrated here, utilisation of stereoselective chemistry allows medicinal chemists to incorporate complexity and 3D-characteristics into a lead series from a simple starting point. An alternative method would be to screen complex fragments which are amenable to poised chemistry to identify complex hits. In the next section I explore the use of multi-component reactions to build bespoke fragments using one-step, parallel chemistry.

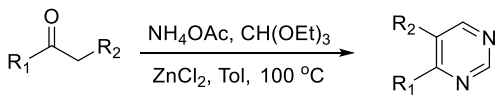
5.4 Multi-Component Reactions

In order to increase the diversity of DSPL, investigations were undertaken for poised reactions beyond the twenty-two shortlisted in Chapter 2. Multi-component reactions (MCRs) use three or more reactants to yield one product, and therefore offer a method for building complexity and diversity into a compound library. Domling *et al.* recently summarised the use of MCRs in drug discovery (Domling, Wang *et al.* 2012). Reactions based upon work undertaken by Ugi in the 1960s (Ugi 1962) offer the easiest route into biologically active chemical space (Barreto, Vercillo *et al.* 2011; Koes, Khoury *et al.* 2012). However a number of other reactions have been reported which access large libraries of biologically active compounds (Barrow, Stauffer *et al.* 2008; Sasada, Kobayashi *et al.* 2009; Gunawan, Petit *et al.* 2012; Frolova, Magedov *et al.* 2013; Maiti, Chanda *et al.* 2013). The 3D shape, drug-like properties and diverse chemical functionalities present in MCR products make them ideal poised fragments. Yet the pungent isonitriles used in Ugi-based MCRs

and the often temperamental nature of other MCRs have thus far made them an underexplored area in library design.

A survey of the MCR literature was completed to find new reactions which could be described as poised. Reactions were rejected if the chemistry was deemed too complex, if only a narrow selection of starting materials were commercially available or if the products contained reactive or promiscuous functional groups. In total, twenty eight reactions were shortlisted, including the Povarov reaction (Liu, Dagousset *et al.* 2009), the Pauson-Khand reaction (Khand, Knox *et al.* 1973), the Van Leusen reaction (Van Leusen, Wildeman *et al.* 1977) and Ugi-derived MCRs (Ugi 1962).

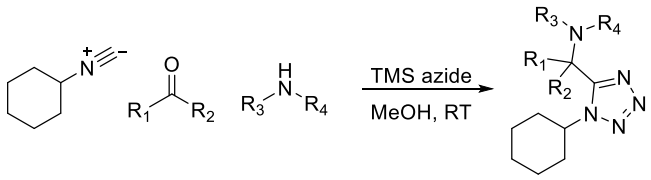
Due to the under-representation of heterocycle-forming reactions in poised space (Section 2.4.2), a pyrimidine forming reaction and an Ugi-type tetrazole synthesis were selected to undergo synthetic investigations. Five pyrimidines were synthesised from ketones using the MCR procedure outlined by Sasada *et al.* (Sasada, Kobayashi *et al.* 2009). A diverse range of ketones were used to generate the pyrimidines in low to moderate yield. Electron-rich ketones were found to react in lowest yield (compound **261**) or to not react at all (compound **262**). Bicyclic systems were also synthesised, but strained ring systems were isolated in low yield (e.g. compounds **263** and **264**).



Compound	R_1	R_2	Yield[a]
260	Phenyl	H	36%
261	2,4-dihydroxyphenyl	Phenyl	NI[b]
262	Pyrrol-2-yl	H	NI
263	Cyclooctanyl		12%
264	Cyclopentyl		5%
265	*N*-Boc piperidin-[4,3-*d*]-yl		38%

Table 5.2: Synthesis of 4,5-substituted pyrimidines using the procedure reported by Sasada et al. (2009) [a]NI: Not isolated. [b]Compound detected by crude NMR and LCMS, and isolated in low purity (<90%).

An adaption of Ugi's original four-component tetrazole synthesis was the second reaction chosen for investigation (Wang, Fang *et al.* 2013). Five tetrazole-containing compounds were synthesised in low to good yield (Table 5.3). Compound **267** demonstrates the use of a secondary amine (morpholine) as a starting material, while compound **268** indicates electron-deficient amines can be used but with a low yield. Tetrazoles **269** and **270** demonstrate the use of low molecular weight carbonyls (acetone and formaldehyde respectively).



Compound	R_1	R_2	R_3	R_4	Yield[a]
266	^{n}Propyl	H	3-Chlorophenyl	H	43%
267	Phenyl	Me	Morpholino		30%
268	^{i}Propyl	H	2-Pyridinyl	H	8%
269	Methyl	Methyl	4-Methoxyphenyl	H	73%
270	H	H	3-Chlorophenyl	H	66%

Table 5.3: Synthesis of tetrazoles using a variation of the Ugi-reaction (Wang et al. 2013). [a]Chiral compounds formed as a racemate.

The results of the preliminary synthetic investigations into MCRs demonstrate the utility and limitations of using such reactions to build bespoke fragments. The Ugi-tetrazole proved to be a robust reaction which can combine a wide range of chemotypes. However, the size of the products means many analogues do not fulfil the classical definition of a fragment. Of the five compounds synthesised in this study, the molecular weights range from 291 to 333 Da, on the upper limit of Congreve's Ro3 guidelines (MW ≤300) (Congreve, Carr *et al.* 2003). In contrast, the pyrimidine forming reaction produced small, drug-like compounds with molecular weights in the range 120 to 264. But the reaction proved to be low yielding and unable to tolerate certain chemotypes such as electron rich ketones. As a result, further investigations were halted in favour of the poised reactions defined in Chapter 2.

5.5 Ongoing Work and Concluding Remarks

Researchers in the field of FBDD are increasingly looking for simple compounds to populate fragment libraries, but with multiple growth vectors to enable fragment elaboration (Murray and Rees 2016). The use of poised chemistry enables progression of weak hits to potent leads by making changes to the periphery, where conventional SAR operates, as well as to the core of the compound. The ability to perform this hit optimisation quickly and effectively overcomes a common bottleneck in hit-to-lead development and saves time later in the process (Erlanson, Fesik *et al.* 2016). However, ensuring leads can be progressed from small, promiscuous hits often requires more complex chemistry, allowing the development of complex leads with a high degree of 3D-characteristics. The processes explored in this chapter – biotransformation, stereoselective reaction cascades and MCRs – are examples of reactions which increase the chemical diversity of a lead series. Many studies have demonstrated how medicinal chemistry is reliant upon a small number of reactions (Cooper, Campbell *et al.* 2010; Roughley and Jordan 2011), and calls have been made to identify and use novel reactions to improve the chemical diversity of lead compounds (Brown and Bostrom 2015; Murray and Rees 2016).

The lactam diastereomers synthesised in this chapter using a diastereoselective reaction cascade demonstrates how bespoke chemistry can be used to elaborate fragments. Although only modest gains in activity were obtained, the results can be used to direct future synthesis and SAR studies in order to reach sub-μM *in vitro* potency and enable investigations into cellular and *in vivo* activity.

6 Conclusions and Future Plans

The poised paradigm presented in this thesis aids the discovery and development of novel chemical probes by lowering the barriers of hit-to-lead development using organic synthesis. The method balances the pragmatism of SAR by catalogue, along with good exploration of chemical space. In this thesis I present a number of case studies which demonstrate the utility of poised FBDD against a range of epigenetic targets. The *N*-dimethoxybenzyl amide series of PHIP(2) inhibitors best illustrates how an inactive crystallographic hit can be developed into a lead series by utilising poised chemistry ready for lead optimisation using bespoke organic synthesis (Figure 6.1). First, a library of poised fragments is screened against a target using X-ray crystallography. The DSPL can be used to ensure the synthetic tractability of hits for follow-up campaigns even if the compound has low or immeasurable biochemical activity. Computational tools such as XPoise allows the design and selection of follow-up compounds for hit-to-lead development. The result is a lead series with modest activity and high-quality structural data ready for lead optimisation using traditional SAR studies and novel synthetic methodologies. Once the desired activity, selectivity and PK profiles have been achieved, lead compounds can be put forward as drug or chemical probe candidates.

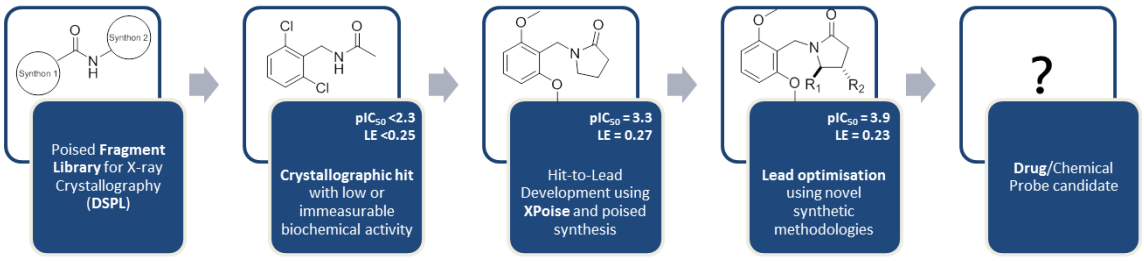


Figure 6.1: The poised paradigm in drug discovery. Hit discovery can be performed by screening DSPL using X-ray crystallography, enabling hit-to-lead development using XPoise. A lead compound can then be optimised using bespoke organic synthesis in order to identify a novel drug or chemical probe.

The future progress of the work presented in this thesis can be broadly split into three categories: the upkeep and evolution of DSPL, the development and distribution of the XPoise software, and the design, synthesis and characterisation of a potent and selective chemical probe to target the PHIP(2) Brd.

6.1 Evolution of DSPL

The DSPL has been designed and purchased for FBDD at DLS and SGC Oxford and is now in regular use by internal and external researchers at the XChem fragment soaking facility. Congreve's rules were used as a guideline meaning the library is not completely Ro3 compliant, although all compounds fulfil the criteria of molecular weight less than 250 and cLogP less than 3.5. Toxic and reactive molecules were removed meaning the library is composed entirely of drug-like compounds. Only 776 compounds of the desired 1000 compounds were purchased, yet it is believed the library has sufficient coverage of poised fragment space alongside a diverse range of poised reaction types. Furthermore, the library is synthetically tractable to elaboration: every fragment member can be synthesised using at least one of seventeen poised chemical reactions from commercially available starting points.

The upkeep of DSPL is not a trivial task. Full characterisation of the library using NMR as well as assessment of compound solubility and stability would aid quality control as well as future hit validation using biochemical and biophysical assays. Compounds may be removed from the library if they do not fulfil quality control criteria or if the synthesis of analogues proves to be more challenging than predicted. Likewise, some poised reactions may be removed from future versions of the library if they are found to be unsuitable. It is important to note the definition of poised chemistry described in Chapter 2 is not rigid, and new reactions may be added to the list if they

are found to be useful and reliable, such as the MCRs shortlisted in Chapter 5. As a result, it is expected the DSPL will continue to evolve and undergo a number of changes in the coming years.

6.2 Development of XPoise

The XPoise workflow described in Chapter 3 has been developed to automatically identify poised synthons from a list of fragment hits and to build a virtual library of follow-up compounds for each synthon. The workflow uses a library of commercially available building blocks meaning each compound in the final library can be synthesised in one chemical step.

XPoise is optimised to build virtual libraries for LLOOMMPPAA analysis by varying one synthon at a time. However, the flexibility of the KNIME software means users can readily adapt property filters and reaction protocols to their own requirements. Changes to the input library of building blocks would allow researchers to explore different areas of chemical space using different virtual libraries. Furthermore, a full exploration of poised chemical space can be performed given sufficient computational power and with only minor tweaks to the algorithm. This would represent a full matrix expansion rather than the single vector expansion currently performed by XPoise. The resulting library could be used to test novel computational scoring methods, based on pharmacophore matching or docking, because each compound can be easily accessed in one chemical step from commercially available starting points.

The XPoise workflow has been installed at beamline I04-1 at DLS and is used by researchers as a part of their hit discovery programs at the facility. The workflow allows users to identify all poised fragment hits for their crystal system regardless of the fragment library used to perform the initial screen and regardless of their expertise in organic chemistry. However, improvements to the workflow are required if the software is to be robust and flexible enough for widespread distribution to the wider scientific community.

6.3 PHIP(2) Chemical Probe

The fragment soaking screens and resulting follow-up campaigns against epigenetic targets described in Chapter 4 demonstrate the benefit of using poised hit-to-lead development on low affinity fragments. The two poised fragment hits against PHIP(2) selected for follow-up synthesis resulted in two lead series for a previously intractable target with sub-mM activity, good LE and an array of structural data. The *N*-benzylamide series generated four new crystal datasets, including protein-ligand complexes for compounds found to be inactive in an AlphaScreen peptide displacement assay.

Lead optimisation of the *N*-benzylamides using the diastereoselective reaction cascade described in Chapter 5 yielded preliminary SAR data to guide future synthetic efforts. Although only modest gains in activity were obtained, the results revealed promising growth vectors which can be explored in future rounds of synthesis. The eventual aim is to develop a potent and selective PHIP(2) KAc mimetic which fulfils the SGC Chemical Probe guidelines detailed in Chapter 1 and fuels future research into the PHIP gene.

6.4 Concluding Remarks

It is hoped that in time the use of DSPL and XPoise will enable academic and industrial users at Diamond to development novel, high impact chemical probes to further biological understanding of a target and accelerate drug discovery. Furthermore, the hit-to-lead development of BRD1, PHIP(2), JMJD2D and SETDB1 fragments could produce novel compounds with which to target epigenetic proteins and reveal new insight into the epigenome. Armed with the tools and projects I have described, the scientific community is poised to unearth new discoveries and build upon exisiting knowledge.

Bibliography

Abergel, C. (2013). Molecular replacement: tricks and treats. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **69**, 2167-2173.

Agrawal, A.; Johnson, S. L. *et al.* (2010). Chelator Fragment Libraries for Targeting Metalloproteinases. *ChemMedChem* **5**, 195-199.

Anderson, A. C. (2003). The Process of Structure-Based Drug Design. *Chem. Biol.* **10**, 787-797.

Armstrong, M. S.; Finn, P. W. *et al.* (2011). Improving the accuracy of ultrafast ligand-based screening: incorporating lipophilicity into ElectroShape as an extra dimension. *J. Comput.-Aided Mol. Des.* **25**, 785-790.

Armstrong, M. S.; Morris, G. M. *et al.* (2010). ElectroShape: fast molecular similarity calculations incorporating shape, chirality and electrostatics. *J. Comput.-Aided Mol. Des.* **24**, 789-801.

Armstrong, M. S.; Morris, G. M. *et al.* (2009). Molecular similarity including chirality. *J. Mol. Graphics Modell.* **28**, 368-370.

Arrowsmith, C. H.; Audia, J. E. *et al.* (2015). The promise and peril of chemical probes. *Nat. Chem. Biol.* **11**, 536-541.

Avorn, J. (2015). The $2.6 billion pill - methodologic and policy considerations. *N. Engl. J. Med.* **372**, 1877-1879.

Badger, J. (2012). Crystallographic fragment screening. *Methods Mol. Biol.* **841**, 161-177.

Baell, J. and Walters, M. A. (2014). Chemistry: Chemical con artists foil drug discovery. *Nature* **513**, 481-483.

Baell, J. B. (2015). Screening-Based Translation of Public Research Encounters Painful Problems. *ACS Med. Chem. Lett.* **6**, 229-234.

Baell, J. B. and Holloway, G. A. (2010). New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *J. Med. Chem.* **53**, 2719-2740.

Baker, M. (2013). Fragment-based lead discovery grows up. *Nat. Rev. Drug Discovery* **12**, 5-7.

Bamborough, P.; Barnett, H. A. *et al.* (2016). GSK6853, a Chemical Probe for Inhibition of the BRPF1 Bromodomain. *ACS Med. Chem. Lett.* **7**, 552-557.

Bamborough, P.; Brown, M. J. *et al.* (2011). Selectivity of Kinase Inhibitor Fragments. *J. Med. Chem.* **54**, 5131-5143.

Bamborough, P.; Chung, C.-W. *et al.* (2016). A Chemical Probe for the ATAD2 Bromodomain. *Angew. Chem.* **55**, 11382-11386.

Barreto, A. d. F. S.; Vercillo, O. E. *et al.* (2011). Fast and efficient microwave-assisted synthesis of functionalized peptoids via Ugi reactions. *Org. Biomol. Chem.* **9**, 5024-5027.

Barrow, J. C.; Stauffer, S. R. *et al.* (2008). Discovery and X-ray Crystallographic Analysis of a Spiropiperidine Iminohydantoin Inhibitor of β-Secretase. *J. Med. Chem.* **51**, 6259-6262.

Baurin, N.; Aboul-Ela, F. *et al.* (2004). Design and Characterization of Libraries of Molecular Fragments for Use in NMR Screening against Protein Targets. *J. Chem. Inf. Comput. Sci.* **44**, 2157-2166.

Bavetsias, V.; Lanigan, R. M. *et al.* (2016). 8-Substituted Pyrido[3,4-d]pyrimidin-4(3H)-one Derivatives As Potent, Cell Permeable, KDM4 (JMJD2) and KDM5 (JARID1) Histone Lysine Demethylase Inhibitors. *J. Med. Chem.* **59**, 1388-1409.

Bemis, G. W. and Murcko, M. A. (1996). The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **39**, 2887-2893.

Benniston, A. C. and Harriman, A. (2006). Charge on the move: how electron-transfer dynamics depend on molecular conformation. *Chem. Soc. Rev.* **35**, 169-179.

Berger, A. and Borgaes, E. E. (1968). Method of treating hypertension employing selected thioureas, Baxter Laboratories Inc. 2 pp.

Berthold, M. R.; Cebron, N. *et al.* (2007). KNIME: The Konstanz Information Miner. *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*, Springer.

Besnard, J.; Ruda, G. F. *et al.* (2012). Automated design of ligands to polypharmacological profiles. *Nature* **492**, 215-220.

Bheda, P. and Schneider, R. (2014). Epigenetics reloaded: the single-cell revolution. *Trends Cell Biol* **24**, 712-723.

Bjarkam, C. R.; Corydon, T. J. *et al.* (2009). Further immunohistochemical characterization of BRD1 a new susceptibility gene for schizophrenia and bipolar affective disorder. *Brain Struct. Funct.* **214**, 37-47.

Blaazer, A. R.; Orrling, K. M. *et al.* (2015). Fragment-based screening in tandem with phenotypic screening provides novel antiparasitic hits. *J. Biomol. Screening* **20**, 131-140.

Blum, L. C. and Reymond, J.-L. (2009). 970 Million Druglike Small Molecules for Virtual Screening in the Chemical Universe Database GDB-13. *J. Am. Chem. Soc.* **131**, 8732-8733.

Boer, J.; Young-Sciame, R. et al. (2016). Roles of UGT, P450, and Gut Microbiota in the Metabolism of Epacadostat in Humans. *Drug Metab. Dispos.* **44**, 1668-1674.

Borg, I. and Groenen, P. J. F. (2005). *Modern Multidimensional Scaling: theory and applications*, Springer.

Bradley, A. (2015). Development of Tools to Provide Prioritisation and Guidance in the Development of Chemical Probes and Small Molecule Leads. *Systems Approaches to Biomedical Sciences Industrial Doctoral Centre*, University of Oxford. **Doctor of Philosophy**.

Bradley, A. R.; Wall, I. D. *et al.* (2014). OOMMPPAA: A Tool To Aid Directed Synthesis by the Combined Analysis of Activity and Structural Data. *J. Chem. Inf. Model.* **54**, 2636-2646.

Bradley, A. R.; Wall, I. D. *et al.* (2015). WONKA: objective novel complex analysis for ensembles of protein-ligand structures. *J. Comput.-Aided Mol. Des.* **29**, 963-973.

Brand, M.; Measures, A. M. *et al.* (2015). Small Molecule Inhibitors of Bromodomain-Acetyl-lysine Interactions. *ACS Chem. Biol.* **10**, 22-39.

Brigas, A. F. (2004). Product class 30: tetrazoles. *Sci. Synth.* **13**, 861-915.

Brown, D. G. and Bostrom, J. (2015). Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone? *J. Med. Chem.* **59**, 4443-4458.

Brunger, A. T. (1997). Free R value: cross-validation in crystallography. *Methods Enzymol.* **277**, 366-396.

Bunnage, M. E.; Chekler, E. L. P. *et al.* (2013). Target validation using chemical probes. *Nat. Chem. Biol.* **9**, 195-199.

Burley, S. K.; Hirst, G. *et al.* (2010). Chapter 3: Fragment-based structure-guided drug discovery: stategy, process, and lessons from human protein kinases. *Drug Design: Structure- and Ligand-Based Approaches*. K. M. J. Merz, D. Ringe and C. H. Reynolds, Cambridge University Press.

Chaikuad, A.; Petros, A. M. *et al.* (2014). Structure-based approaches towards identification of fragments for the low-druggability ATAD2 bromodomain. *MedChemComm* **5**, 1843-1848.

Chayen, N. E. (2009). High-throughput protein crystallization. *Adv. Protein Chem. Struct. Biol.* **77**, 1-22.

Chen, H.; Zhou, X. *et al.* (2015). Evolutions in fragment-based drug design: the deconstruction-reconstruction approach. *Drug Discovery Today* **20**, 105-113.

Chen, I. J. and Hubbard, R. E. (2009). Lessons for fragment library design: analysis of output from multiple screening campaigns. *J. Comput.-Aided Mol. Des.* **23**, 603-620.

Chen, P.; Chaikuad, A. *et al.* (2016). Discovery and Characterization of GSK2801, a Selective Chemical Probe for the Bromodomains BAZ2A and BAZ2B. *J. Med. Chem.* **59**, 1410-1424.

Chen, V. B.; Arendall, W. B. *et al.* (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D-Biological Crystallography* **66**, 12-21.

Chevillard, F. and Kolb, P. (2015). SCUBIDOO: A Large yet Screenable and Easily Searchable Database of Computationally Created Chemical Compounds Optimized toward High Likelihood of Synthetic Tractability. *J. Chem. Inf. Model.* **55**, 1824-1835.

Chilingaryan, Z.; Yin, Z. *et al.* (2012). Fragment-based screening by protein crystallography: successes and pitfalls. *Int. J. Mol. Sci.* **13**, 12857-12879.

Chung, C.-w.; Dean, A. W. *et al.* (2012). Fragment-Based Discovery of Bromodomain Inhibitors Part 1: Inhibitor Binding Modes and Implications for Lead Discovery. *J. Med. Chem.* **55**, 576-586.

Clark, P. G. K.; Vieira, L. C. C. *et al.* (2015). LP99: Discovery and Synthesis of the First Selective BRD7/9 Bromodomain Inhibitor. *Angew. Chem.* **54**, 6217-6221.

Collins, J. L.; Fivush, A. M. *et al.* (2002). Identification of a Nonsteroidal Liver X Receptor Agonist through Parallel Array Synthesis of Tertiary Amines. *J. Med. Chem.* **45**, 1963-1966.

Congreve, M.; Carr, R. *et al.* (2003). A 'rule of three' for fragment-based lead discovery? *Drug Discovery Today* **8**, 876-877.

Convers, E.; Tye, H. *et al.* (2004). Preparation and evaluation of a polymer-supported Mukaiyama reagent. *Tetrahedron Lett.* **45**, 3401-3404.

Cook, D.; Brown, D. *et al.* (2014). Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat. Rev. Drug Discovery* **13**, 419-431.

Cooper, T. W. J.; Campbell, I. B. *et al.* (2010). Factors Determining the Selection of Organic Reactions by Medicinal Chemists and the Use of These Reactions in Arrays (Small Focused Libraries). *Angew. Chem.* **49**, 8082-8091.

Cox, O. B.; Krojer, T. *et al.* (2016). A poised fragment library enables rapid synthetic expansion yielding the first reported inhibitors of PHIP(2), an atypical bromodomain. *Chem. Sci.* **7**, 2322-2330.

Daguer, J. P.; Zambaldo, C. *et al.* (2015). DNA display of fragment pairs as a tool for the discovery of novel biologically active small molecules. *Chem. Sci.* **6**, 739-744.

Dahlin, J. L.; Nissink, J. W. M. *et al.* (2015). PAINS in the Assay: Chemical Mechanisms of Assay Interference and Promiscuous Enzymatic Inhibition Observed during a Sulfhydryl-Scavenging HTS. *J. Med. Chem.* **58**, 2091-2113.

Daylight Chemical Infomation Systems, L. N., CA, USA.

De Semir, D.; Nosrati, M. *et al.* (2012). Pleckstrin homology domain-interacting protein (PHIP) as a marker and mediator of melanoma metastasis. *Proc. Natl. Acad. Sci.* **109**, 7067-7072.

Deng, Z.; Chuaqui, C. *et al.* (2004). Structural Interaction Fingerprint (SIFt): A Novel Method for Analyzing Three-Dimensional Protein-Ligand Binding Interactions. *J. Med. Chem.* **47**, 337-344.

Devine, S. M.; Mulcair, M. D. *et al.* (2015). Promiscuous 2-Aminothiazoles (PrATs): A Frequent Hitting Scaffold. *J. Med. Chem.* **58**, 1205-1214.

DiMasi, J. A.; Grabowski, H. G. *et al.* (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. *J Health Econ* **47**, 20-33.

Doak, B. C.; Morton, C. J. *et al.* (2013). Design and Evaluation of the Performance of an NMR Screening Fragment Library. *Aust. J. Chem.* **66**, 1465-1472.

Domling, A.; Wang, W. *et al.* (2012). Chemistry and Biology Of Multicomponent Reactions. *Chem. Rev.* **112**, 3083-3135.

Doub, L.; Richardson, L. M. *et al.* (1958). Some phenylthiourea derivatives and their antituberculous activity. *J. Am. Chem. Soc.* **80**, 2205-2217.

Drouin, L.; McGrath, S. *et al.* (2015). Structure Enabled Design of BAZ2-ICR, A Chemical Probe Targeting the Bromodomains of BAZ2A and BAZ2B. *J. Med. Chem.* **58**, 2553-2559.

Durant, J. L.; Leland, B. A. *et al.* (2002). Reoptimization of MDL Keys for Use in Drug Discovery. *J. Chem. Inf. Comput. Sci.* **42**, 1273-1280.

Edwards, A. (2016). Reproducibility: Team up with industry. *Nature* **531**, 299-301.

Edwards, A. M.; Isserlin, R. *et al.* (2011). Too many roads not taken. *Nature* **470**, 163-165.

Efremov, I. V. and Erlanson, D. A. (2016). Chapter 6: Fragment-Based Lead Generation. *Lead Generation: Methods and Strategies*. J. Holenz, Wiley. **67,** 133-158.

Emsley, P.; Lohkamp, B. *et al.* (2010). Features and development of Coot. *Acta Crystallographica Section D-Biological Crystallography* **66**, 486-501.

England, K. S.; Tumber, A. *et al.* (2014). Optimisation of a triazolopyridine based histone demethylase inhibitor yields a potent and selective KDM2A (FBXL11) inhibitor. *MedChemComm* **5**, 1879-1886.

Erlanson, D. A.; Fesik, S. W. *et al.* (2016). Twenty years on: the impact of fragments on drug discovery. *Nat. Rev. Drug Discov*ery **15**, 605-619.

Fedorov, O.; Lingard, H. *et al.* (2014). [1,2,4]Triazolo[4,3-a]phthalazines: Inhibitors of Diverse Bromodomains. *J. Med. Chem.* **57**, 462-476.

Fedorov, O.; Mueller, S. *et al.* (2010). The (un)targeted cancer kinome. *Nat. Chem. Biol.* **6**, 166-169.

Ferguson, F. M.; Fedorov, O. *et al.* (2013). Targeting Low-Druggability Bromodomains: Fragment Based Screening and Inhibitor Design against the BAZ2B Bromodomain. *J. Med. Chem.* **56**, 10183-10187.

Filippakopoulos, P.; Picaud, S. *et al.* (2012). Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. *Cell* **149**, 214-231.

Filippakopoulos, P.; Qi, J. *et al.* (2010). Selective inhibition of BET bromodomains. *Nature* **468**, 1067-1073.

Fish, P. V.; Filippakopoulos, P. *et al.* (2012). Identification of a Chemical Probe for Bromo and Extra C-Terminal Bromodomain Inhibition through Optimization of a Fragment-Derived Hit. *J. Med. Chem.* **55**, 9831-9837.

Frank, R. L. and Smith, P. V. (1948). 1-Phenylthiourea. *Org. Synth.* **28**, 89-91.

Friedrich, F. and Pohloudek-Fabini, R. (1967). Organic thiocyanates. XXIX. Synthesis of some (thiocyanato-substituted)-aryl isothiocyanates and -aryl thioureas. *Pharmazie* **22**, 543-548.

Frolova, L. V.; Magedov, I. V. *et al.* (2013). Exploring Natural Product Chemistry and Biology with Multicomponent Reactions. 5. Discovery of a Novel Tubulin-Targeting Scaffold Derived from the Rigidin Family of Marine Alkaloids. *J. Med. Chem.* **56**, 6886-6900.

Garavis, M.; Lopez-Mendez, B. *et al.* (2014). Discovery of Selective Ligands for Telomeric RNA G-quadruplexes (TERRA) through 19F-NMR Based Fragment Screening. *ACS Chem. Biol.* **9**, 1559-1566.

Gillet, V.; Johnson, A. P. *et al.* (1993). SPROUT: A program for structure generation. *J. Comput.-Aided Mol. Des.* **7**, 127-153.

Gleeson, M. P. (2008). Generation of a Set of Simple, Interpretable ADMET Rules of Thumb. *J. Med. Chem.* **51**, 817-834.

Gold-Aubert, P. and Toribio, L. (1963). Synthesis and pharmacological activity of new disubstituted ureides. I. *Arch. Sci.* **16**, 405-410.

Gosmini, R.; Nguyen, V. L. *et al.* (2014). The Discovery of I-BET726 (GSK1324726A), a Potent Tetrahydroquinoline ApoA1 Up-Regulator and Selective BET Bromodomain Inhibitor. *J. Med. Chem.* **57**, 8111-8131.

Guengerich, F. P. (2008). Cytochrome P450 and Chemical Toxicology. *Chem. Res. Toxicol.* **21**, 70-83.

Gunawan, S.; Petit, J. *et al.* (2012). Concise One-Pot Preparation of Unique Bis-Pyrrolidinone Tetrazoles. *ACS Comb. Sci.* **14**, 160-163.

Hajduk, P. J. and Greer, J. (2007). A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discovery* **6**, 211-219.

Hall, R. J.; Mortenson, P. N. *et al.* (2014). Efficient exploration of chemical space by fragment-based screening. *Prog. Biophys. Mol. Biol.* **116**, 82-91.

Hann, M. M. (2011). Molecular obesity, potency and other addictions in drug discovery. *MedChemComm* **2**, 349-355.

Hann, M. M.; Leach, A. R. *et al.* (2001). Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *J. Chem. Inf. Comput. Sci.* **41**, 856-864.

Haqq, C.; Nosrati, M. *et al.* (2005). The gene expression signatures of melanoma progression. *Proc. Natl. Acad. Sci.* **102**, 6092-6097.

Harner, M. J.; Chauder, B. A. *et al.* (2014). Fragment-Based Screening of the Bromodomain of ATAD2. *J. Med. Chem.* **57**, 9687-9692.

Hartenfeller, M.; Eberle, M. *et al.* (2011). A Collection of Robust Organic Synthesis Reactions for In Silico Molecule Design. *J. Chem. Inf. Model.* **51**, 3093-3098.

Hartenfeller, M.; Eberle, M. *et al.* (2012). Probing the Bioactivity-Relevant Chemical Space of Robust Reactions and Common Molecular Building Blocks. *J. Chem. Inf. Model.* **52**, 1167-1178.

Hartenfeller, M.; Zettl, H. *et al.* (2012). DOGS: reaction-driven de novo design of bioactive compounds. *PLoS Comput. Biol.* **8**, e1002380.

Hay, D. A.; Fedorov, O. *et al.* (2014). Discovery and Optimization of Small-Molecule Ligands for the CBP/p300 Bromodomains. *J. Am. Chem. Soc.* **136**, 9308-9319.

Hay, M. P.; Turcotte, S. *et al.* (2010). 4-Pyridylanilinothiazoles That Selectively Target von Hippel-Lindau Deficient Renal Cell Carcinoma Cells by Inducing Autophagic Cell Death. *J. Med. Chem.* **53**, 787-797.

Heinemann, B.; Nielsen, J. M. *et al.* (2014). Inhibition of demethylases by GSK-J1/J4. *Nature* **514**, E1-E2.

Hewings, D. S.; Fedorov, O. *et al.* (2013). Optimization of 3,5-Dimethylisoxazole Derivatives as Potent Bromodomain Ligands. *J. Med. Chem.* **56**, 3217-3227.

Ho, P.-L.; Yang, H.-C. *et al.* (1960). Synthesis of trisubstituted thioureas. *Huaxue Xuebao* **26**, 1-6.

Hopkins, A. L.; Groom, C. R. *et al.* (2004). Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today* **9**, 430-431.

Hopkinson, R. J.; Tumber, A. *et al.* (2013). 5-Carboxy-8-hydroxyquinoline is a broad spectrum 2-oxoglutarate oxygenase inhibitor which causes iron translocation. *Chem. Sci.* **4**, 3110-3117.

Hughes, J. D.; Blagg, J. *et al.* (2008). Physiochemical drug properties associated with in vivo toxicological outcomes. *Bioorg. Med. Chem. Lett.* **18**, 4872-4875.

Hung, A. W.; Ramek, A. *et al.* (2011). Route to three-dimensional fragments using diversity-oriented synthesis. *Proc. Natl. Acad. Sci.* **108**, 6799-6804,.

Irwin, J. J. and Shoichet, B. K. (2005). ZINC--a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **45**, 177-182.

Irwin, J. J.; Sterling, T. *et al.* (2012). ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* **52**, 1757-1768.

Izvekov, G. I. and Skorikov, I. E. (1966). Automation of the alkylphenol washing in lubricant additive manufacture. *Neftepererabotka i Neftekhim*, 43-45.

Jakubec, P.; Cockfield, D. M. *et al.* (2012). Stereoselective, nitro-Mannich/lactamisation cascades for the direct synthesis of heavily decorated 5-nitropiperidin-2-ones and related heterocycles. *Beilstein J. Org. Chem.* **8**, 567-578,.

Jhoti, H.; Williams, G. *et al.* (2013). The 'rule of three' for fragment-based drug discovery: where are we now? *Nat. Rev. Drug Discovery* **12**, 644.

Johansson, C.; Tumber, A. *et al.* (2014). The roles of Jumonji-type oxygenases in human disease. *Epigenomics* **6**, 89-120.

Johansson, C.; Velupillai, S. *et al.* (2016). Structural analysis of human KDM5B guides histone demethylase inhibitor development. *Nat. Chem. Biol.* **12**, 539-545.

Joseph-McCarthy, D.; Campbell, A. J. *et al.* (2014). Fragment-Based Lead Discovery and Design. *J. Chem. Inf. Model.* **54**, 693-704.

Kang, I.-J.; Wang, L.-W. *et al.* (2009). Design, synthesis, and anti-HCV activity of thiourea compounds. *Bioorg. Med. Chem. Lett.* **19**, 1950-1955.

Karageorgis, G.; Warriner, S. *et al.* (2014). Efficient discovery of bioactive scaffolds by activity-directed synthesis. *Nat. Chem.* **6**, 872-876.

Karba, D. (1967). Syntheses of sulfonylthioureas. I. Reactions between sulfonamides and thioureas, thiocyanates, or isothiocyanates. *Arch. Pharm.* **300**, 844-848.

Karplus, M. (1959). Contact electron-spin coupling of nuclear magnetic moments. *J. Chem. Phys.* **30**, 11-15.

Kawamura, A.; Tumber, A. *et al.* (2010). Development of homogeneous luminescence assays for histone demethylase catalysis and binding. *Anal. Biochem.* **404**, 86-93.

Keseru, G. M.; Erlanson, D. A. *et al.* (2016). Design Principles for Fragment Libraries: Maximizing the Value of Learnings from Pharma Fragment-Based Drug Discovery (FBDD) Programs for Use in Academia. *J. Med. Chem.* **59**, 8189-8206.

Kessel, M. (2014). Restoring the pharmaceutical industry's reputation. *Nat. Biotechnol.* **32**, 983-990.

Khand, I. U.; Knox, G. R. *et al.* (1973). Organocobalt complexes. II. Reaction of acetylenehexacarbonyl dicobalt complexes, (RC2R1)Co2(CO)6, with norbornene and its derivatives. *J. Chem. Soc.* **1**, 977-981.

Koes, D.; Khoury, K. *et al.* (2012). Enabling large-scale design, synthesis and validation of small molecule protein-protein antagonists. *PLoS One* **7**, e32839.

Koester, H.; Craan, T. *et al.* (2011). A Small Nonrule of 3 Compatible Fragment Library Provides High Hit Rate of Endothiapepsin Crystal Structures with Various Fragment Chemotypes. *J. Med. Chem.* **54**, 7784-7796.

Krishnan, S. and Trievel, R. C. (2013). Structural and functional analysis of JMJD2D reveals molecular basis for site-specific demethylation among JMJD2 demethylases. *Structure* **21**, 98-108.

Kruidenier, L.; Chung, C.-w. *et al.* (2012). A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. *Nature* **488**, 404-408.

Kshirsagar, T. and Editor (2008). *High-Throughput Lead Optimization in Drug Discovery*, CRC Press.

Kubinyi, H. (1995). Strategies and recent technologies in drug discovery. *Pharmazie* **50**, 647-662.

Kutchukian, P. S.; Wassermann, A. M. *et al.* (2015). Large Scale Meta-Analysis of Fragment-Based Screening Campaigns: Privileged Fragments and Complementary Technologies. *J Biomol. Screen.* **20**, 588-596.

Laird, T. (2013). Introduction to Biological and Small Molecule Drug Research and Development: Theory and Case Studies. *Org. Process Res. Dev.* **17**, 1218.

Lau, W. F.; Withka, J. M. *et al.* (2011). Design of a multi-purpose fragment screening library using molecular complexity and orthogonal diversity metrics. *J. Comput.-Aided Mol. Des.* **25**, 621-636.

Leeson, P. D. and Springthorpe, B. (2007). The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discovery* **6**, 881-890.

Lepre, C. A. (2001). Library design for NMR-based screening. *Drug Discovery Today* **6**, 133-140.

Li, J.-T.; Meng, X.-T. *et al.* (2009). One-pot synthesis of benzylacetamide from oxime under ultrasound irradiation. *Ultrason. Sonochem.* **16**, 590-592.

Lipinski, C. A.; Lombardo, F. *et al.* (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **23**, 3-25.

Lipinski, C. A.; Lombardo, F. *et al.* (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **46**, 3-26.

Liu, H.; Dagousset, G. *et al.* (2009). Chiral Bronsted Acid-Catalyzed Enantioselective Three-Component Povarov Reaction. *J. Am. Chem. Soc.* **131**, 4598-4599.

Lucas, X.; Wohlwend, D. *et al.* (2013). 4-Acyl Pyrroles: Mimicking Acetylated Lysines in Histone Code Reading. *Angew. Chem.* **52**, 14055-14059.

Luo, X.-L.; Liu, Y.-X. *et al.* (2011). A Selective Inhibitor and Probe of the Cellular Functions of Jumonji C Domain-Containing Histone Demethylases. *J. Am. Chem. Soc.* **133**, 9451-9456.

Maddani, M. R. and Prabhu, K. R. (2010). A concise synthesis of substituted thiourea derivatives in aqueous medium. *J. Org. Chem.* **75**, 2327-2332.

Maiti, B.; Chanda, K. *et al.* (2013). Multicomponent Solvent-Free Synthesis Of Benzimidazolyl Imidazo[1,2-a]-pyridine Under Microwave Irradiation. *ACS Comb. Sci.* **15**, 291-297.

Matlock, M. and Swamidass, S. J. (2014). Sharing Chemical Relationships Does Not Reveal Structures. *J. Chem. Inf. Model.* **54**, 37-48.

Moos, W. H.; Green, G. D. *et al.* (1993). Recent advances in the generation of molecular diversity. *Annu. Rep. Med. Chem.* **28**, 315-324.

Morgan, H. L. (1965). Generation of a unique machine description for chemical structures--a technique developed at Chemical Abstracts Service. *J. Chem. Doc.* **5**, 107-113.

Morley, A. D.; Pugliese, A. *et al.* (2013). Fragment-based hit identification: thinking in 3D. *Drug Discovery Today* **18**, 1221–1227.

Mullard, A. (2015). 2014 FDA drug approvals. *Nat. Rev. Drug Discovery* **14**, 77-81.

Mullard, A. (2016). 2015 FDA drug approvals. *Nat. Rev. Drug Discovery* **15**, 73-76.

Mullard, A. (2016). Deuterated drugs draw heavier backing. *Nat. Rev. Drug Discovery* **15**, 219-221.

Murray, C. W. and Rees, D. C. (2016). Opportunity Knocks: Organic Chemistry for Fragment-Based Drug Discovery (FBDD). *Angew. Chem.* **55**, 488-492.

Murray, C. W.; Verdonk, M. L. *et al.* (2012). Experiences in fragment-based drug discovery. *Trends Pharmacol. Sci.* **33**, 224-232.

Murshudov, G. N.; Skubak, P. *et al.* (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D-Biological Crystallography* **67**, 355-367.

Na, J. and Hu, Q. (2011). Design of screening collections for successful fragment-based lead discovery. *Methods Mol. Biol.* **685**, 219-240.

Nair, G. V. (1965). Thiocarbamoylation of amines by bis(thiocarbamoyl) disulfides. *J. Indian Chem. Soc.* **42**, 359-366.

Ng, J. T.; Dekker, C. *et al.* (2014). Using textons to rank crystallization droplets by the likely presence of crystals. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **70**, 2702-2718.

Nicodeme, E.; Jeffrey, K. L. *et al.* (2010). Suppression of inflammation by a synthetic histone mimic. *Nature* **468**, 1119-1123.

Nyegaard, M.; Severinsen, J. E. *et al.* (2010). Support of association between BRD1 and both schizophrenia and bipolar affective disorder. *Am. J. Med. Genet., Part B* **153B**, 582-591.

O'Boyle, N. M.; Banck, M. *et al.* (2011). Open Babel: an open chemical toolbox. *J. Cheminf.* **3**, 33.

Ortholand, J.-Y. and Ganesan, A. (2004). Natural products and combinatorial chemistry: back to the future. *Curr. Opin. Chem. Biol.* **8**, 271-280.

Over, B.; Wetzel, S. *et al.* (2013). Natural-product-derived fragments for fragment-based ligand discovery. *Nat. Chem.* **5**, 21-28.

Pande, V. (2016). Understanding the Complexity of Epigenetic Target Space. *J. Med. Chem.* **59**, 1299-1307.

Patel, D.; Bauman, J. D. *et al.* (2014). Advantages of crystallographic fragment screening: Functional and mechanistic insights from a powerful platform for efficient drug discovery. *Prog. Biophys. Mol. Biol.* **116**, 92-100.

Paul, S. M.; Mytelka, D. S. *et al.* (2010). How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discovery* **9**, 203-214.

Pelletier, S. M. C.; Ray, P. C. *et al.* (2011). Diastereoselective synthesis of 1,3,5-trisubstituted 4-nitropyrrolidin-2-ones via a nitro-Mannich/lactamization cascade. *Org. Lett.* **13**, 6406-6409.

Peppard, J.; Glickman, F. *et al.* (2003). Development of a high-throughput screening assay for inhibitors of aggrecan cleavage using luminescent oxygen channeling (AlphaScreen). *J. Biomol. Screening* **8**, 149-156.

Philpott, M.; Yang, J. *et al.* (2011). Bromodomain-peptide displacement assays for interactome mapping and inhibitor discovery. *Mol. BioSyst.* **7**, 2899-2908.

Pitt, W. R.; Parry, D. M. *et al.* (2009). Heteroaromatic Rings of the Future. *J. Med. Chem.* **52**, 2952-2963.

Proschak, E.; Wegner, J. K. *et al.* (2007). Molecular Query Language (MQL) - a context-free grammar for substructure matching. *J. Chem. Inf. Model.* **47**, 295-301.

Rappe, A. K.; Casewit, C. J. *et al.* (1992). UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **114**, 10024-10035.

Reisen, F. H.; Schneider, G. *et al.* (2009). Reaction-MQL: line notation for functional transformation. *J. Chem. Inf. Model.* **49**, 6-12.

Riccardi Sirtori, F.; Caronni, D. *et al.* (2015). Establish an automated flow injection ESI-MS method for the screening of fragment based libraries: Application to Hsp90. *Eur. J. Pharm. Sci.* **76**, 83-94.

Rodriguez-Paredes, M.; Martinez de Paz, A. *et al.* (2014). Gene amplification of the histone methyltransferase SETDB1 contributes to human lung tumorigenesis. *Oncogene* **33**, 2807-2813.

Rogers, D. and Hahn, M. (2010). Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **50**, 742-754.

Roughley, S. D. and Jordan, A. M. (2011). The Medicinal Chemist's Toolbox: An Analysis of Reactions Used in the Pursuit of Drug Candidates. *J. Med. Chem.* **54**, 3451-3479.

Sanguinetti, M. C. and Tristani-Firouzi, M. (2006). hERG potassium channels and cardiac arrhythmia. *Nature* **440**, 463-469.

Sasada, T.; Kobayashi, F. *et al.* (2009). An Unprecedented Approach to 4,5-Disubstituted Pyrimidine Derivatives by a ZnCl2-Catalyzed Three-Component Coupling Reaction. *Org. Lett.* **11**, 2161-2164.

Schiebel, J.; Radeva, N. *et al.* (2015). One Question, Multiple Answers: Biochemical and Biophysical Screening Methods Retrieve Deviating Fragment Hit Lists. *ChemMedChem* **10**, 1511-1521.

Schreyer, A. and Blundell, T. (2009). CREDO: a protein-ligand interaction database for drug discovery. *Chem. Biol. Drug Des.* **73**, 157-167.

Schreyer, A. M. and Blundell, T. (2012). USRCAT: real-time Ultrafast Shape Recognition with pharmacophoric constraints. *J. Cheminf.* **4**, 27.

Schreyer, A. M. and Blundell, T. L. (2013). CREDO: a structural interactomics database for drug discovery. *Database*, bat049, 049 pp.

SchrodingerRelease2013-1 (2013). Maestro. *version 9.4* **Schrodinger**, LLC, New York.

Schulz, M. N.; Landstroem, J. *et al.* (2011). Design of a Fragment Library that maximally represents available chemical space. *J. Comput.-Aided Mol. Des.* **25**, 611-620.

Scott, D. E.; Coyne, A. G. *et al.* (2012). Fragment-Based Approaches in Drug Discovery and Chemical Biology. *Biochemistry* **51**, 4990-5003.

Shah, M. A.; Denton, E. L. *et al.* (2014). A global assessment of cancer genomic alterations in epigenetic mechanisms. *Epigenetics Chromatin* **7**, 29.

Shuker, S. B.; Hajduk, P. J. *et al.* (1996). Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274**, 1531-1534.

Siddiqui, N. and Ahsan, W. (2010). Triazole incorporated thiazoles as a new class of anticonvulsants: Design, synthesis and in vivo screening. *Eur. J. Med. Chem.* **45**, 1536-1543.

Silva, P. C.; Costa, J. S. *et al.* (2001). An expeditious synthesis of 3-nitropropionic acid and its ethyl and methyl esters. *Synth. Commun.* **31**, 595-600.

Small-MoleculeDrugDiscoverySuite2013-1 (2013). QikProp. *version 3.6* **Schrodinger**, LCC, New York.

Spurlino, J. C. (2011). Fragment screening purely with protein crystallography. *Methods Enzymol.* **493**, 321-356.

Stumpfe, D.; Hu, Y. *et al.* (2014). Recent Progress in Understanding Activity Cliffs and Their Utility in Medicinal Chemistry. *J. Med. Chem.* **57**, 18-28.

Swinney, D. C. and Anthony, J. (2011). How were new medicines discovered? *Nat. Rev. Drug Discovery* **10**, 507-519.

Theodoulou, N. H.; Bamborough, P. *et al.* (2016). Discovery of I-BRD9, a Selective Cell Active Chemical Probe for Bromodomain Containing Protein 9 Inhibition. *J. Med. Chem.* **59**, 1425–1439.

Thinnes, C. C.; England, K. S. *et al.* (2014). Targeting histone lysine demethylases - Progress, challenges, and the future. *Biochim. Biophys. Acta, Gene Regul. Mech.* **1839**, 1416-1432.

Tiwari, A. and Sekhar, A. K. T. (2007). Workflow based framework for life science informatics. *Comput. Biol. Chem.* **31**, 305-319.

Todeschini, R.; Consonni, V. *et al.* (2012). Similarity Coefficients for Binary Chemoinformatics Data: Overview and Extended Comparison Using Simulated and Real Data Sets. *J. Chem. Inf. Model.* **52**, 2884-2901.

Topliss, J. G. (1977). *A manual method for applying the hansch approach to drug design*, Elsevier.

Ubarhande, S. S.; Thakare, V. G. *et al.* (2010). Microwave induced improved synthesis of monoaryl thiocarbamides, 1,3-diarylthiocarbamides, 1,3-diarylcarbamides and monoaryl-2,4-dithiobiurets. *J. Indian Chem. Soc.* **87**, 1137-1141.

Ugi, I. (1962). The α-Addition of Immonium Ions and Anions to Isonitriles Accompanied by Secondary Reactions. *Angewandte Chemie International Edition in English* **1**, 8-21.

Ugi, I. and Steinbrucker, C. (1961). Isonitriles. II. Reaction of isonitriles with carbonyl compounds, amines, and hydrazoic acid. *Chem. Ber.* **94**, 734-742.

Ullman, E. F.; Kirakossian, H. *et al.* (1994). Luminescent oxygen channeling immunoassay: measurement of particle binding kinetics by chemiluminescence. *Proc. Natl. Acad. Sci.* **91**, 5426-5430.

Ullman, E. F.; Kirakossian, H. *et al.* (1996). Luminescent oxygen channeling assay (LOCI): sensitive, broadly applicable homogeneous immunoassay method. *Clin. Chem.* **42**, 1518-1526.

Valeur, E. and Bradley, M. (2007). PS-IIDQ: a supported coupling reagent for efficient and general amide bond formation. *Tetrahedron* **63**, 8855-8871.

Valeur, E. and Bradley, M. (2009). Amide bond formation: beyond the myth of coupling reagents. *Chem. Soc. Rev.* **38**, 606-631.

Van Leusen, A. M.; Wildeman, J. *et al.* (1977). Chemistry of sulfonylmethyl isocyanides. 12. Base-induced cycloaddition of sulfonylmethyl isocyanides to carbon,nitrogen double bonds. Synthesis of 1,5-disubstituted and 1,4,5-trisubstituted imidazoles from aldimines and imidoyl chlorides. *J. Org. Chem.* **42**, 1153-1159.

Vasilev, G. (1966). Some derivatives of phenylthiourea. I. *Farmatsiya* **16**, 20-24.

Venkatachalam, T. K. and Uckun, F. M. (2007). Synthesis of Substituted Heterocyclic Thiourea Compounds. *Synth. Commun.* **37**, 3667-3675.

Vidler, L. R.; Brown, N. *et al.* (2012). Druggability Analysis and Structural Classification of Bromodomain Acetyl-lysine Binding Sites. *J. Med. Chem.* **55**, 7346-7359.

Vinkers, H. M.; De Jonge, M. R. *et al.* (2003). SYNOPSIS: SYNthesize and OPtimize System in Silico. *J. Med. Chem.* **46**, 2765-2773.

Vulpetti, A. and Dalvit, C. (2013). Design and Generation of Highly Diverse Fluorinated Fragment Libraries and their Efficient Screening with Improved 19F NMR Methodology. *ChemMedChem* **8**, 1-14.

Wager, T. T.; Hou, X. *et al.* (2010). Moving beyond Rules: The Development of a Central Nervous System Multiparameter Optimization (CNS MPO) Approach To Enable Alignment of Druglike Properties. *ACS Chem. Neurosci.* **1**, 435-449.

Walport, L. J.; Hopkinson, R. J. *et al.* (2012). Mechanisms of human histone and nucleic acid demethylases. *Curr. Opin. Chem. Biol.* **16**, 525-534.

Walters, W. P.; Stahl, M. T. *et al.* (1998). Virtual screening - an overview. *Drug Discovery Today* **3**, 160-178.

Wang, S.-X.; Fang, Z. *et al.* (2013). Synthesis of tetrazole containing 1,2,3-thiadiazole derivatives via U-4CR and their anti-TMV activity. *Chin. Chem. Lett.* **24**, 889-892.

Wichert, M.; Krall, N. *et al.* (2015). Dual-display of small molecules enables the discovery of ligand pairs and facilitates affinity maturation. *Nat. Chem.* **7**, 241-249.

Wielens, J.; Headey, S. J. *et al.* (2013). Parallel screening of low molecular weight fragment libraries: do differences in methodology affect hit identification? *J. Biomol. Screen.* **18**, 147-159.

Wilde, F. and Link, A. (2013). Advances in the design of a multipurpose fragment screening library. *Expert Opin. Drug Discovery* **8**, 597-606.

Winn, M. D.; Ballard, C. C. *et al.* (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D-Biological Crystallography* **67**, 235-242.

Winter, G. (2010). xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* **43**, 186-190.

Winter, G. and McAuley, K. E. (2011). Automated data collection for macromolecular crystallography. *Methods* **55**, 81-93.

Workman, P.; Burrows, F. *et al.* (2007). Drugging the cancer chaperone HSP90: combinatorial therapeutic expolitation on oncogene addiction and tumor stress. *Ann. N. Y. Acad. Sci.* **1113**, 202-216.

Workman, P.; Clarke, P. A. *et al.* (2010). Drugging the PI3 Kinome: From Chemical Tools to Drugs in the Clinic. *Cancer Res.* **70**, 2146-2157.

Workman, P. and Collins, I. (2010). Probing the Probes: Fitness Factors For Small Molecule Tools. *Chem. Biol.* **17**, 561-577.

Wright, N. and Von Delft, F. (2016). Unpublished work.

Yahyazadeh, A.; Aliakbar, A. R. *et al.* (2010). Synthesis and characterization of N-[3-methyl-4-phenylthiazol-2(3H)-ylidene]benzenamine derivatives. *J. Pharm. Res.* **3**, 429-431.

Appendix A: REOS Filter used in Chapters 2 and 3

Substructure	Substructure SMARTS	Allowed Violations
Non-Fluorine Halogen	[Cl,Br,I]	2
Nitro-groups	[N;+0,+1;$(N(=O)~[O;H0;-0,-1])]	0
Multivalent Iodine	[I;$([IX2]),$([IX3])]	0
Long Aliphatic	[CH2][CH2][CH2][CH2][CH2][CH2][CH2][CH2]	0
Other Atom Types	[!#1;!C;!O;!N;!S;!P;!F;!Cl;!Br;!I;!c;!o;!n;!s;!p;!Na;!K;!Mg;!Ca;!Li]	0
Primary Alkyl Halides	[Cl,Br,I][CH2]	0
Esters	[#6][#6](=O)-&!@O-&!@[#6]	1
***α-β* Unsaturated Nitriles**	C=CC#N	0
Aldehydes	[HC]=O	0
Gem-dihalo Compounds	[Cl,Br,I]-C-[Cl,Br,I]	0
Acid Halides	[S,O]=C-[Cl,Br,I]	0
Sulfonyl Halides	S(=O)(=O)-[Cl,Br,I]	0
Perhalo Compounds i	c([F,Cl,Br,I])c([F,Cl,Br,I])c([F,Cl,Br,I])	0
Perhalo Compounds ii	c([F,Cl,Br,I])c([F,Cl,Br,I])c(*)c([F,Cl,Br,I])	0
Perhalo Compounds iii	c([F,Cl,Br,I])c(*)c([F,Cl,Br,I])c([F,Cl,Br,I])	0
Perhalo Compounds iv	c([F,Cl,Br,I])c([F,Cl,Br,I])cc([F,Cl,Br,I])	0
Perhalo Compounds v	c([F,Cl,Br,I])cc([F,Cl,Br,I])c([F,Cl,Br,I])	0
Anhydrides	O=[C,S,P]-O-[C,S,P]=O	0
Nitriles	C#N	1
Diimides, Isocyanates, Isothiocyanates	N=C=[N,O,S]	0
Halogen-Heterobonds	[Cl,Br,I]~[S,P,Si]	0
Fluorine	F	6
Acetals	O-&!@[CH2]-&!@O	0
Gem-diphenyl Phosphorous	c1c(cccc1)Pc2ccccc2	0
Sulfate	S(C)(~O)(~O)~O	0
Phosphate	P(~O)(~O)~O	1
Carboxylic Acids	C(=O)[O-,OH]	1
Phosphonic Acids	P(~O[#6])(~O)(~O)[!#8]	0
Sulphonic Acids	[#6]S(=O)(~O)O[#6]	0
***t*-Butyl**	C([CH3])([CH3])([CH3])	1
Disulfanes	[#16]-[#16]	0
Thioxophosphanes	S=P~*	0
Halogen-N/O/S	[F,Cl,Br,I]-[N,O,S]	0
Imides	ON(C=O)C=O	0
Halo-Ketones	[F,Cl,Br,I]-[CH]-C=O	0
Sulfurs	S	2
Mustards	[N,S][CH2][CH2][Cl,Br,I]	0
Diazo Compounds	N#N*	0

Substructure	Substructure SMARTS	Allowed Violations
Thioesters	*-S-C(=O)-*	0
Maleimides	O=C1NC(=O)C=C1	0
Quinones i	C1C(C=CC(C=1)=O)=O	0
Quinones ii	c21ccccc1C(=O)C=CC2=O	0
Quinones iii	C1(C=CC=CC1=O)=[C,N,O]	0
Quinones iv	C1=CC=CC(C1=O)=O	0
Quinones v	C1C(C:CC(C:1)=O)=O	0
Quinones vi	[O;$(O=C1c2ccccc2C(=[C,O,N])c3ccccc13),$(O=C1C=CC(=[O,C,N])C=C1)]	0
Quinones vii	C1=CC(C=CC1=&!@[C,N,O])=[!R][C,N,O]	0
Quinones viii	C1=CC(C(C(C1=O)=O)=*)=*	0
Quinones ix	C1:CC:CC(C1=O)=O	0
Thiocyanates	S-C#N	0
Three Membered Rings	*1~*~*~1	1
Peroxides	O-O	0
Triflates	FCS(=O)(=O)[F,Cl,Br,O]	0
Halo-Amines	N-[F,Cl,Br,I]	0
Diazo Methylketones	N=N=C-C=O	0
Chloramines	C-N-Cl	0
Unsaturated Ketones i	[O;$(O=[#6][#6]=&!@a),$(O=CC(=&!@C)C=O),$(O=C-C=&!@[CH])]	0
Unsaturated Ketones ii	[O;$(O=[#6][#6]=&!@A),$(O=CC(=&!@C)C=O),$(O=C-C=&!@[CH])]	0
Michael Acceptors	[O;$(O=CC(=C)C=O),$(O=C-C=&!@C)]	0
Allenes	C=C=C	0
Carbon attached to 4 N/O	C([N,O])([N,O])([N,O])[N,O]	0
Aromatic Hydroxyls	c[OH]	3
Terminal Vinyl	[CH2]=[CH]-&!@[N,O,S]	0
Long Aliphatic	[CH2]-&!@[CH2]-&!@[CH2]-&!@[CH2]-&!@[CH2]-&!@[CH2]-&!@[CH2]	0
Siloxanes	[Si]O	0
Azides	N=&!@N	0
Diazonium	N=&!@N=&!@*	0
Alkene Halogenide	C=&!@C[F,Cl,Br,I]	0
Conjugated Double Bonds i	C=C-&!@C=C-&!@C=O	3
Conjugated Double Bonds ii	C=C-&!@C=C-&!@C=C	3
Conjugated Double Bonds iii	C=C-&!@C=C-&!@C:C	3
Triphenyl	A(a)(a)a	0
4 Fused Rings i	*~&@*(~&@*~&@*(~&@*)~&@*~&@*)~&@*~&@*(~&@*)~&@*~&@*	0
4 Fused Rings ii	*(~&@*)~&@*(~&@*(~&@*~&@*)~&@*(~&@*)~&@*(~&@*~&@*)~&@*(~&@*~&@*)~&@*)~&@*~&@*	0
4 Fused Rings iii	*(~&@*)(~&@*(~&@*(~&@*(~&@*)~&@*)~&@*(~&@*)~&@*)~&@*(-&@*)~&@*)~&@*	0
Conjugated Acyclic Doublebonds	*=&!@*-&!@*=&!@*-&!@*=&!@*	0
Sulfoxides	[SX3](=O)	1

Substructure	Substructure SMARTS	Allowed Violations
Thiols	[SH]	0
Dipeptides	O=C-&!@C-&!@N-&!@C(=O)C-&!@N	0
Acyclic Ketals	C-&!@O-&!@C-&!@O-&!@C	0
Active Esters	c1(:c:c(:c:c:c:1OC(=O)[#6])[Cl,Br,I])[Cl,Br,I]	0
Halo-Amines	N-C-[F,Cl,Br,I]	0
Aromatic Bromide	cBr	1
N-Oxide	n-O	0
3 Carbonyls	C(C=O)(C=O)(C=O)	0
Thioguanadinea	NC(=&!@S)(-&!@S)	0
Hydroxyalkyne	C([OH])C#C	0
Alykne	C#C	1
Catechol	[NH2,OH]cc[NH2,OH]	0
Halogen *para*-to-Phenol	c1([Cl,Br,I])aac([OH])ca1	0
Large Rings	[r8,r9,r10,r11,r12,r13,r14,r15,r16,r17,r18,r19,r20]	0
Three Fused Aromatic	[a;$(aa[R3](a)a[R3](a)a),$(aa[R3](a)aa[R3](a)a)]	0
Epoxides	C1[O,N]C1	0
Adamantane	A1A2AA3AA1AA(A2)A3	0
Bromine	Br	1
Oxonium	[O+,o+]	0
Phosphine Oxides 1	P(~[!#8])(~[!#8])(~[!#8])=O	0
Phosphine Oxides ii	P(-[H])(~[!#8])(~[!#8])=O	0
Hydroxylamines	N-[OH]	1
Thioguanadine ii	S-C(=&!@N)-&!@N	0
Acyclic Olefins	C=&!@C	0
Two Carbonyls	C(=O)C([#6])C(=O)	0
Naphthalimide	O=C1NC(=O)c2cccc3cccc1c23	0
Sulphonate Esters	S(=O)(~O)O[#6]	0
Phosphinic Acids	[#6]P(=O)([OH])[#6]	0
Vinyl Halides	C=C-[Cl,Br,I]	0
Hexyl Chain	[CH2]-&!@[CH2]-&!@[CH2]-&!@[CH2]-&!@[CH2]-&!@[CH2]	0
Doublebond *α*-to-Carbonyl	C(=O)C=&!@[C,N]	0
Thioester	C(=S)O	0
Stilbene	cC=&!@Cc	0
Nitrosoamines	N-N=O	0
Sulfonyl Halides	S(=O)(=O)[F,Cl,Br,I]	0
Sulfonates	[#6]S(=O)(=O)O	0
Diketones i	C(=O)[CH,CH2]C(=O)	0
Diketones ii	[#6](=O)[#6](=O)	0
Phosphonate Esters	[#6]P(~O)(~O)O[#6]	0
Vinylagous Acid Chlorides	C(=O)C=CCl	0

Substructure	Substructure SMARTS	Allowed Violations
Imines	[#6]=[NH]	0
***p*-Phenols**	[OH]cccc[OH]	0
Vinyl	c-&!@C(=C)-&!@c	0

Table A.1: SMARTS strings used in REOS filter to remove unwanted functional groups from DSPL and XPoise virtual libraries.

Appendix B: KNIME Workflows used in Chapters 2 and 3

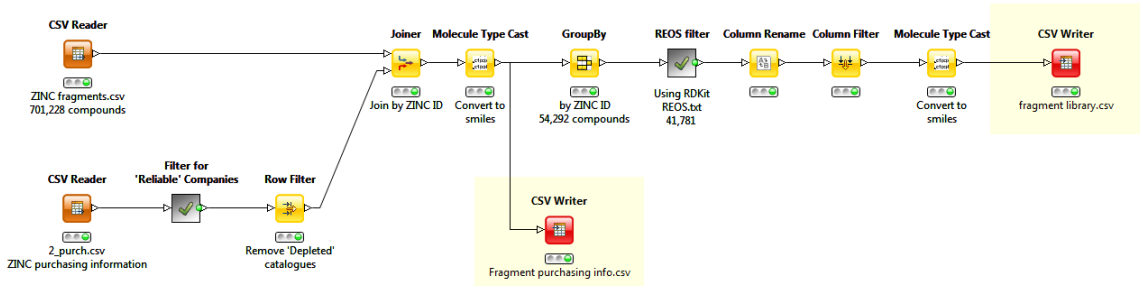


Figure B.1: DSPL design part 1 – identifying commercially available, drug-like fragments.

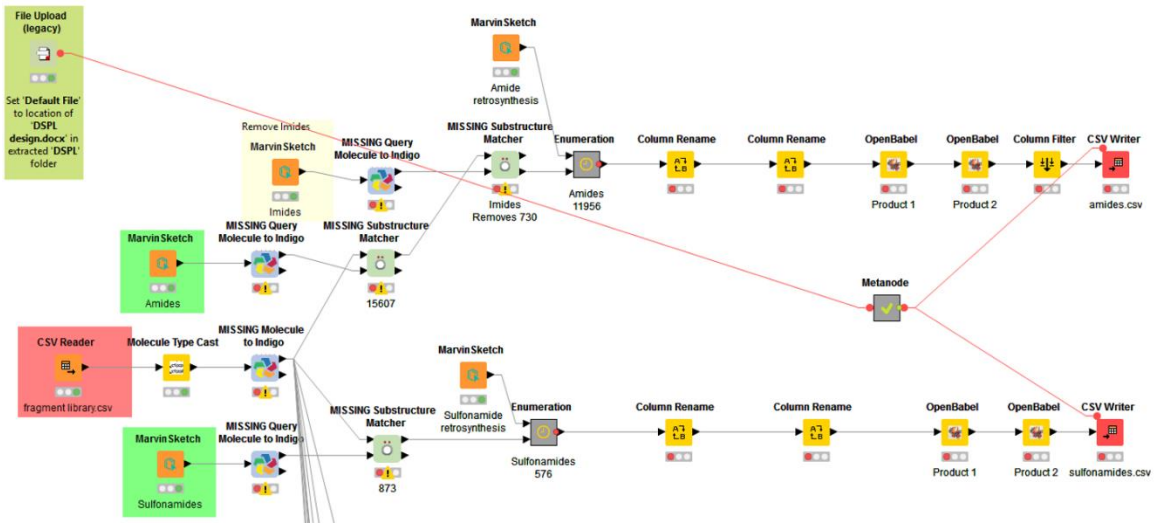


Figure B.2: DSPL design part 2 – identifying poised bonds. Only amide and sulfonamide depicted for clarity. 'Enumeration' subflow depicted in Figure B.3.

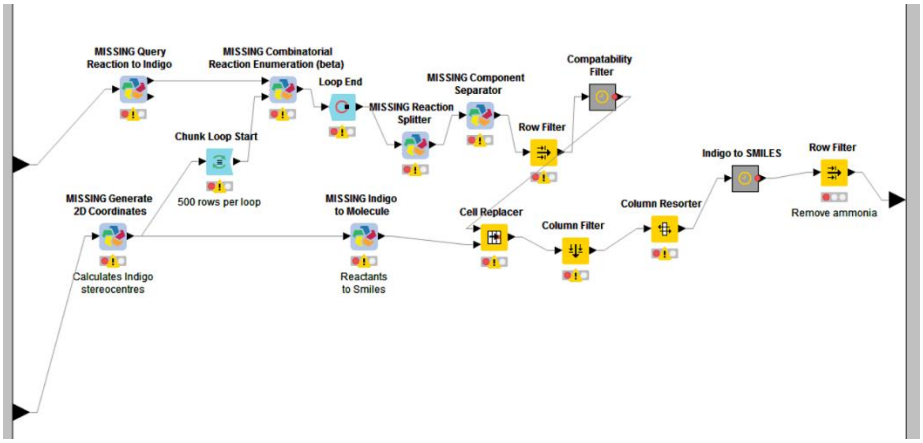


Figure B.3: DSPL design – subflow used to identify amide synthons. A similar workflow was used for the other poised reactions.

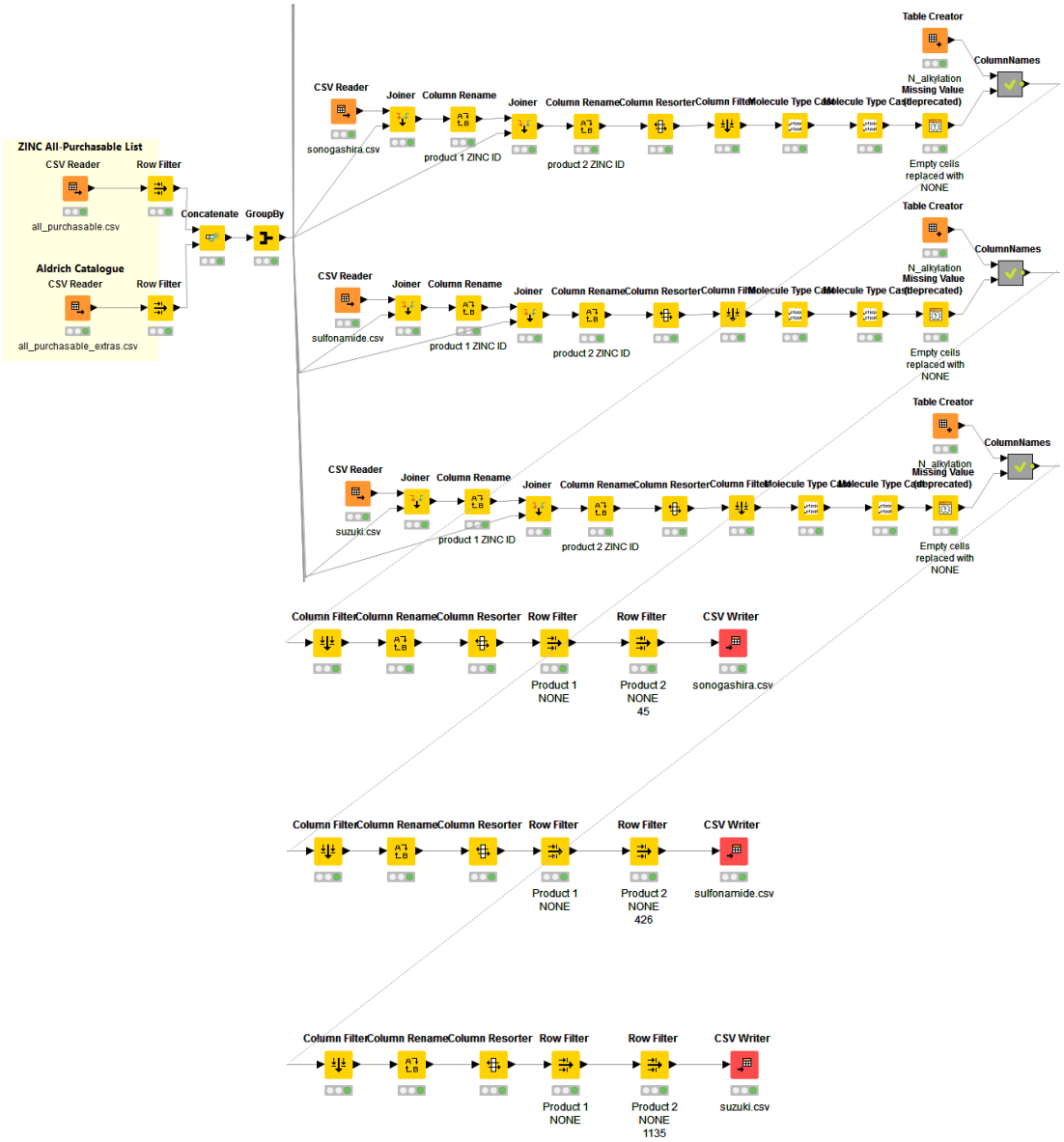


Figure B.4: DSPL design part 3 – identifying poised fragments with commercially available synthons. Only Sonogashira, sulfonamide and Suzuki-type fragment workflows shown for clarity.

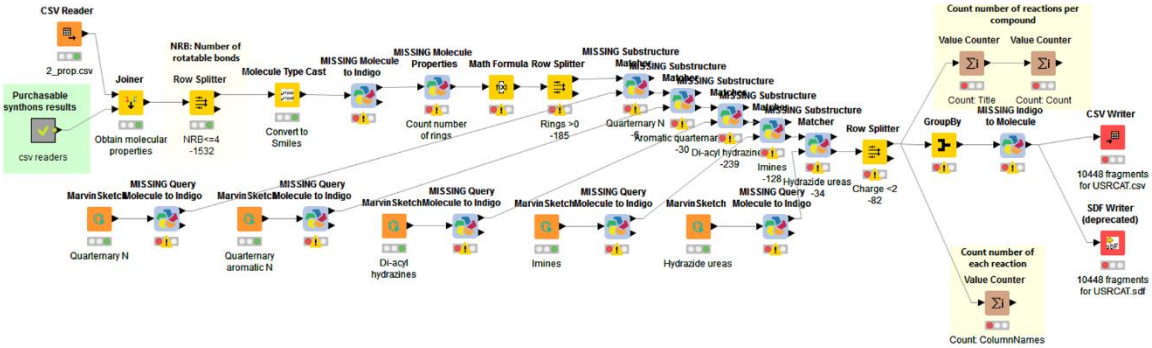


Figure B.5: DSPL design part 4 – final preparation of poised fragment library containing 10,448 compounds ready for diversity filtering by the USRCAT algorithm.

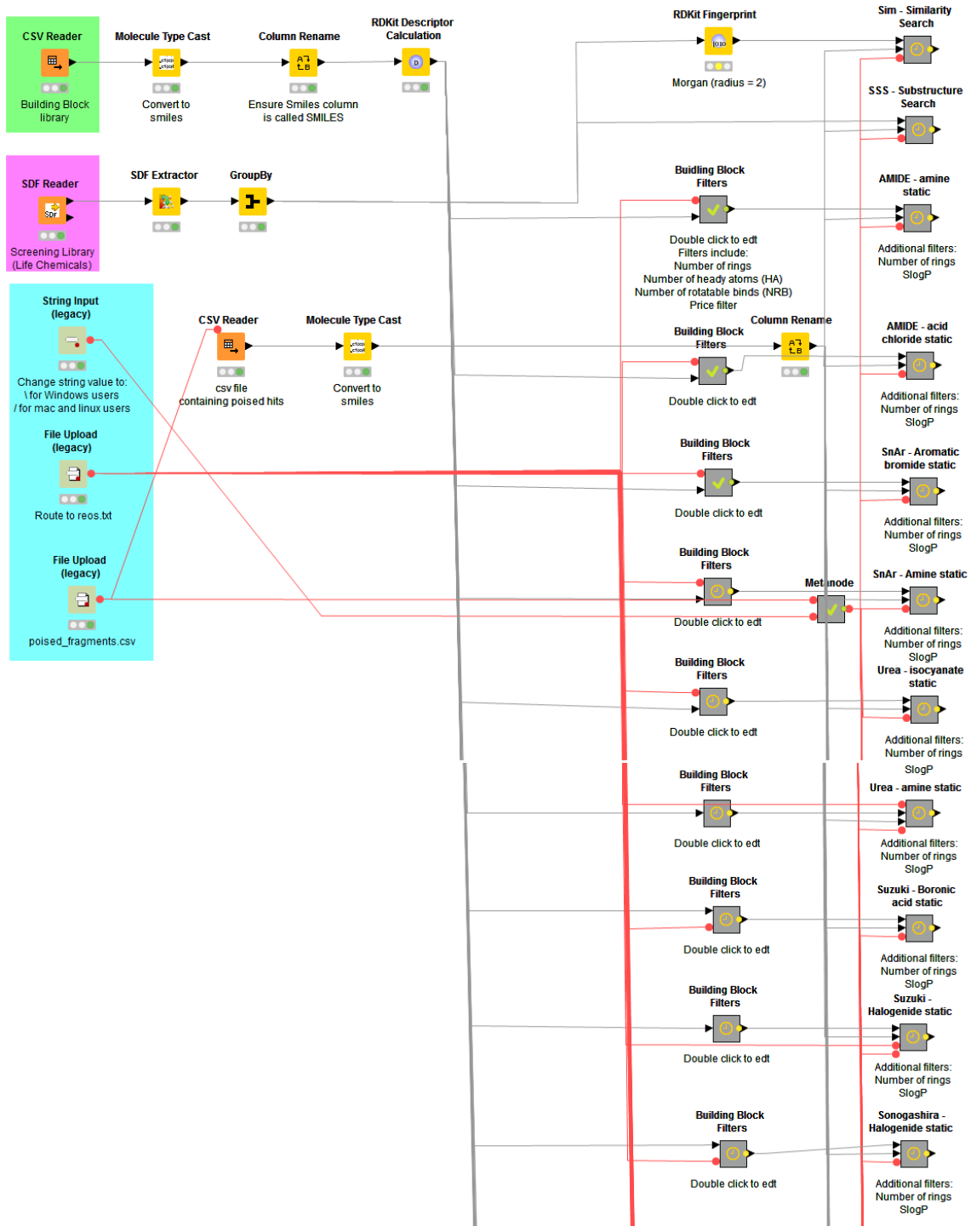


Figure B.5: XPoise workflow overview. Only five of the poised reactions are shown for clarity.

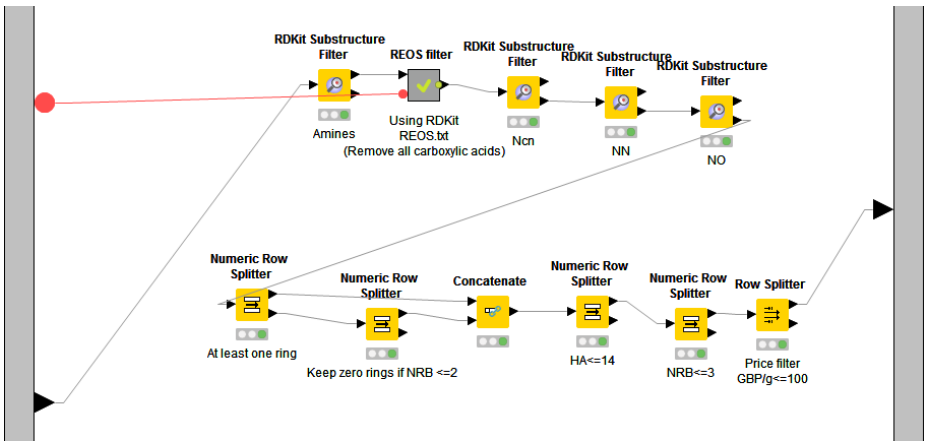


Figure B.6: Example 'Building Blocks Filters' subflow used in XPoise. Amines are identified and prepared for reaction with an acid chloride synthon to form an amide.

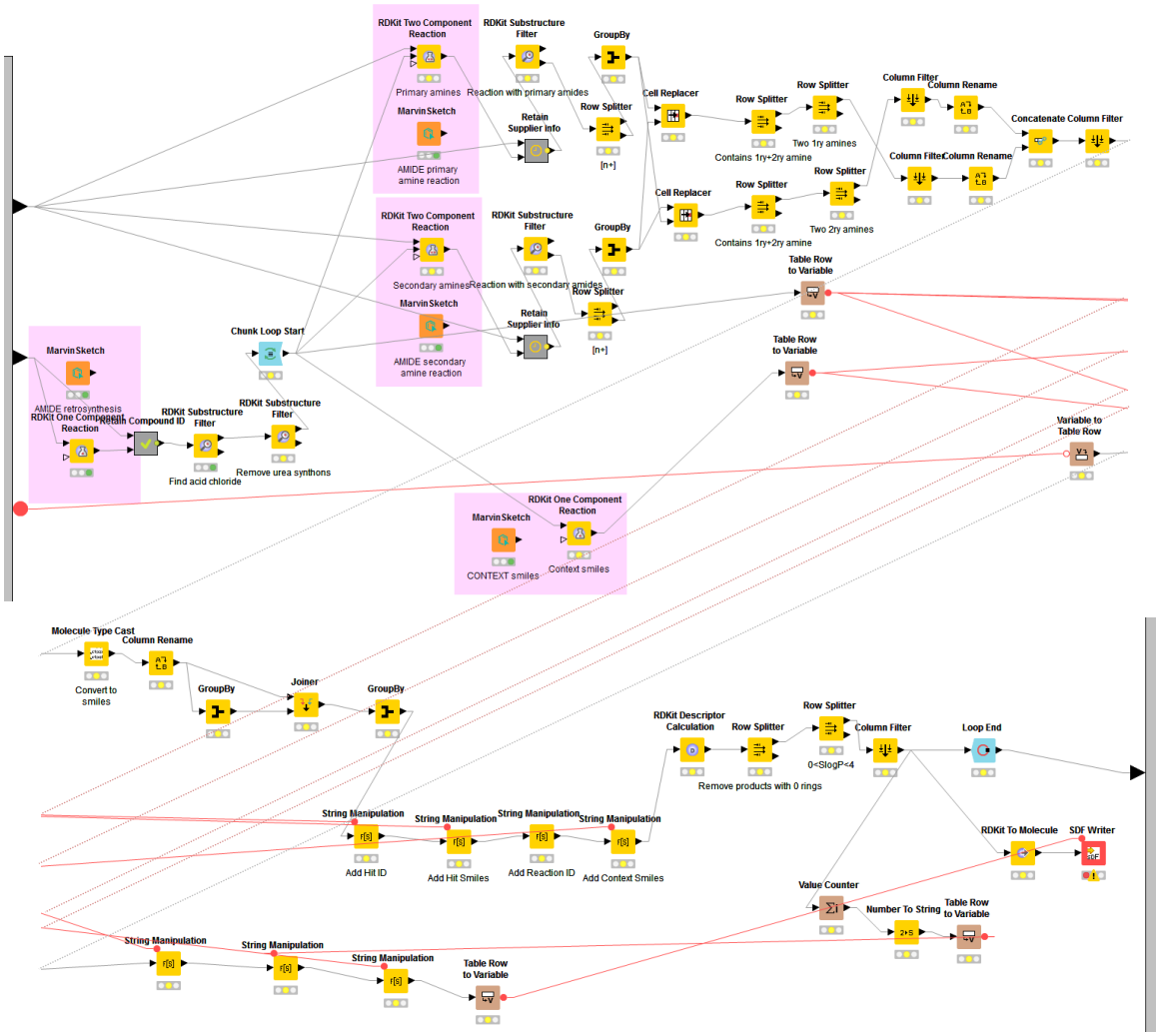


Figure B.7: Example subflow used to enumerate a library of follow-ups to a poised fragment in XPoise. An acid chloride synthon is identified and reacted with a library of amines.

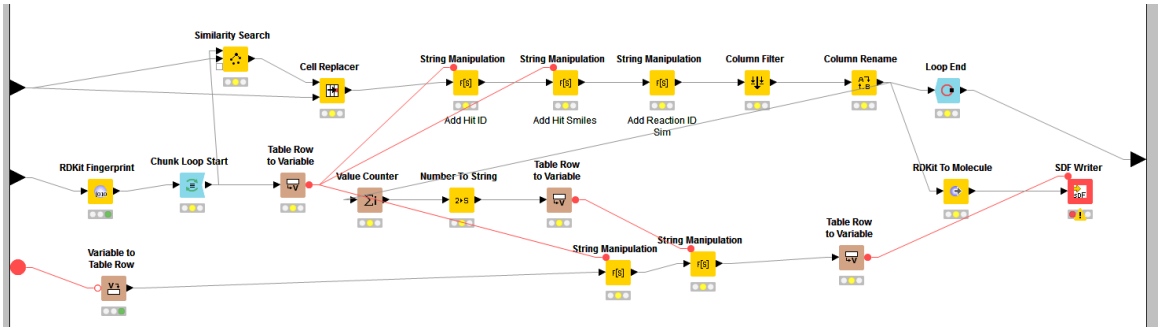


Figure B.8: Subflow used in XPoise to perform a similarity search to a poised fragment.

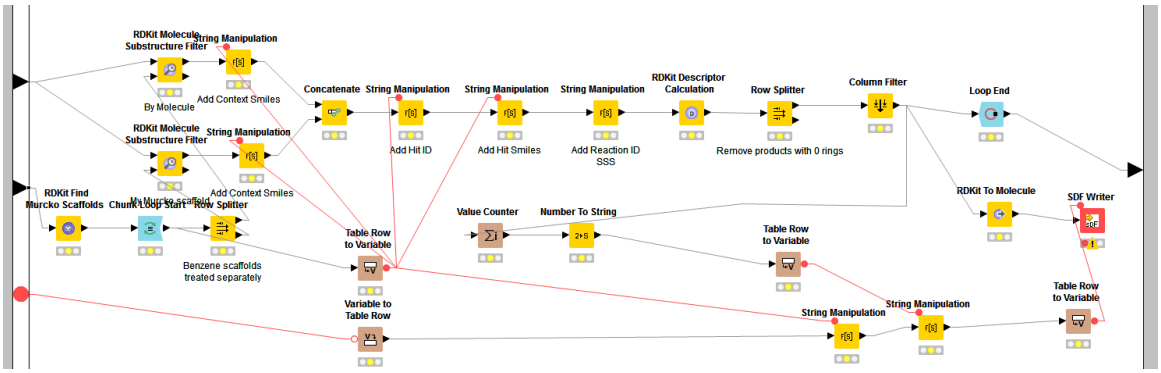
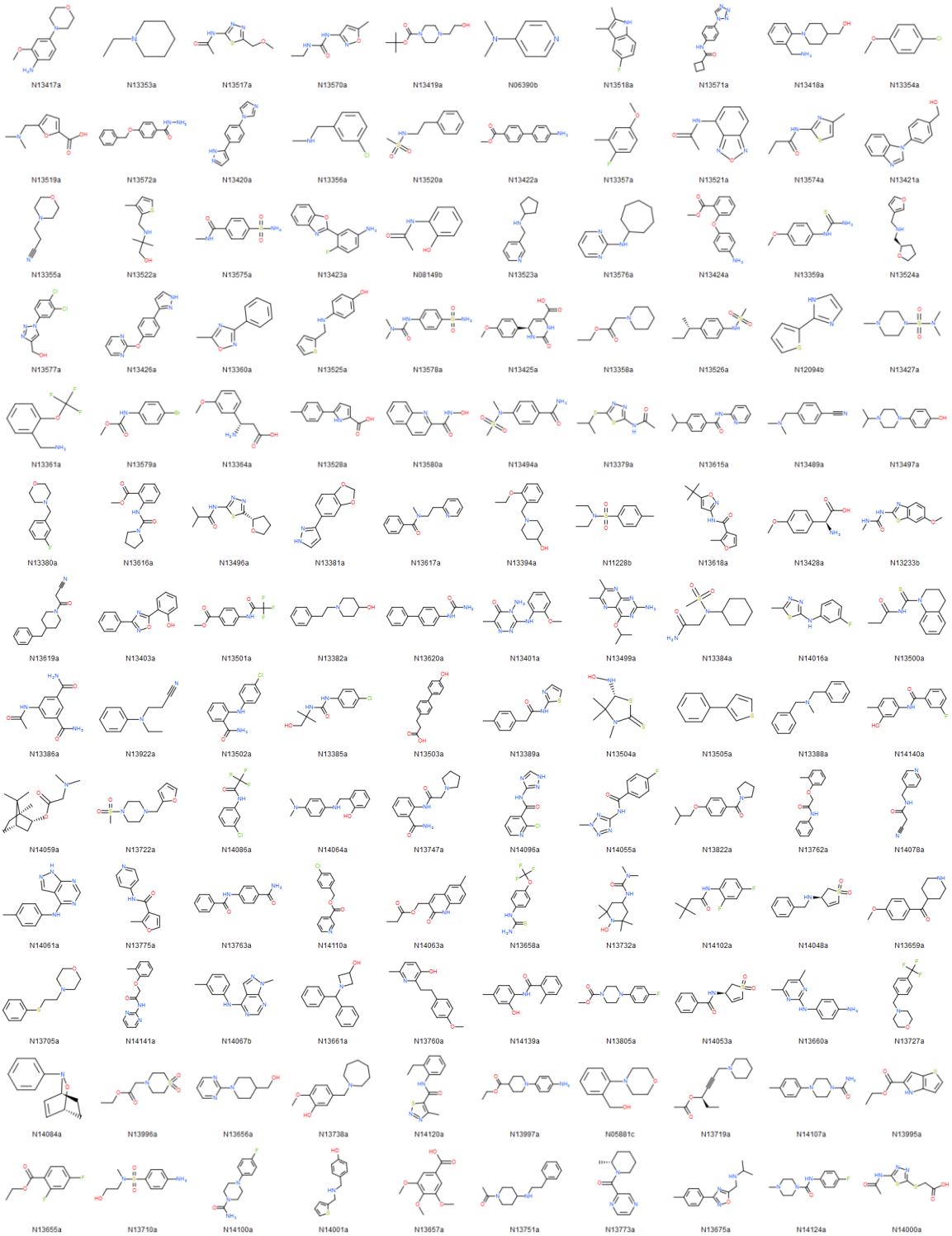
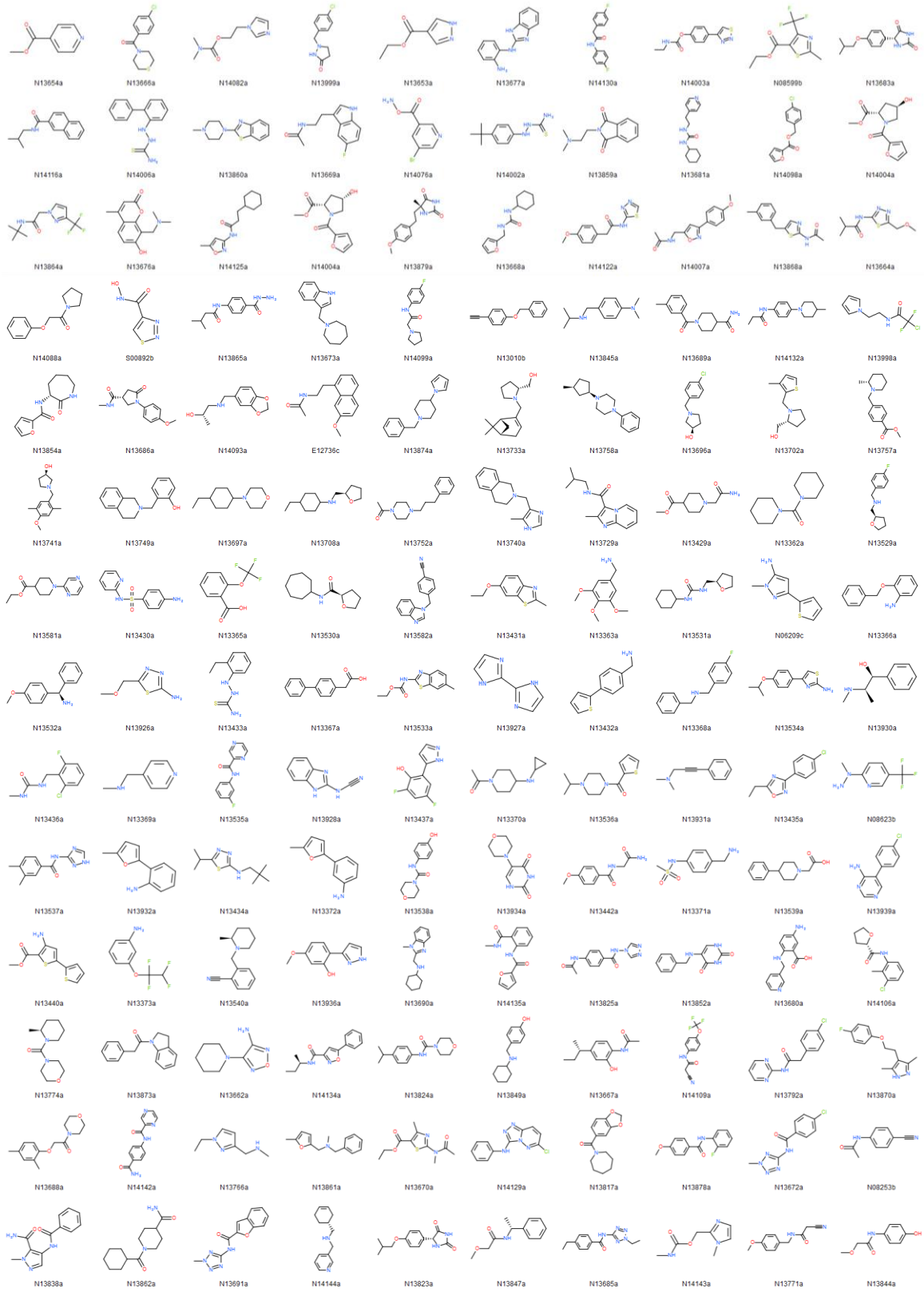
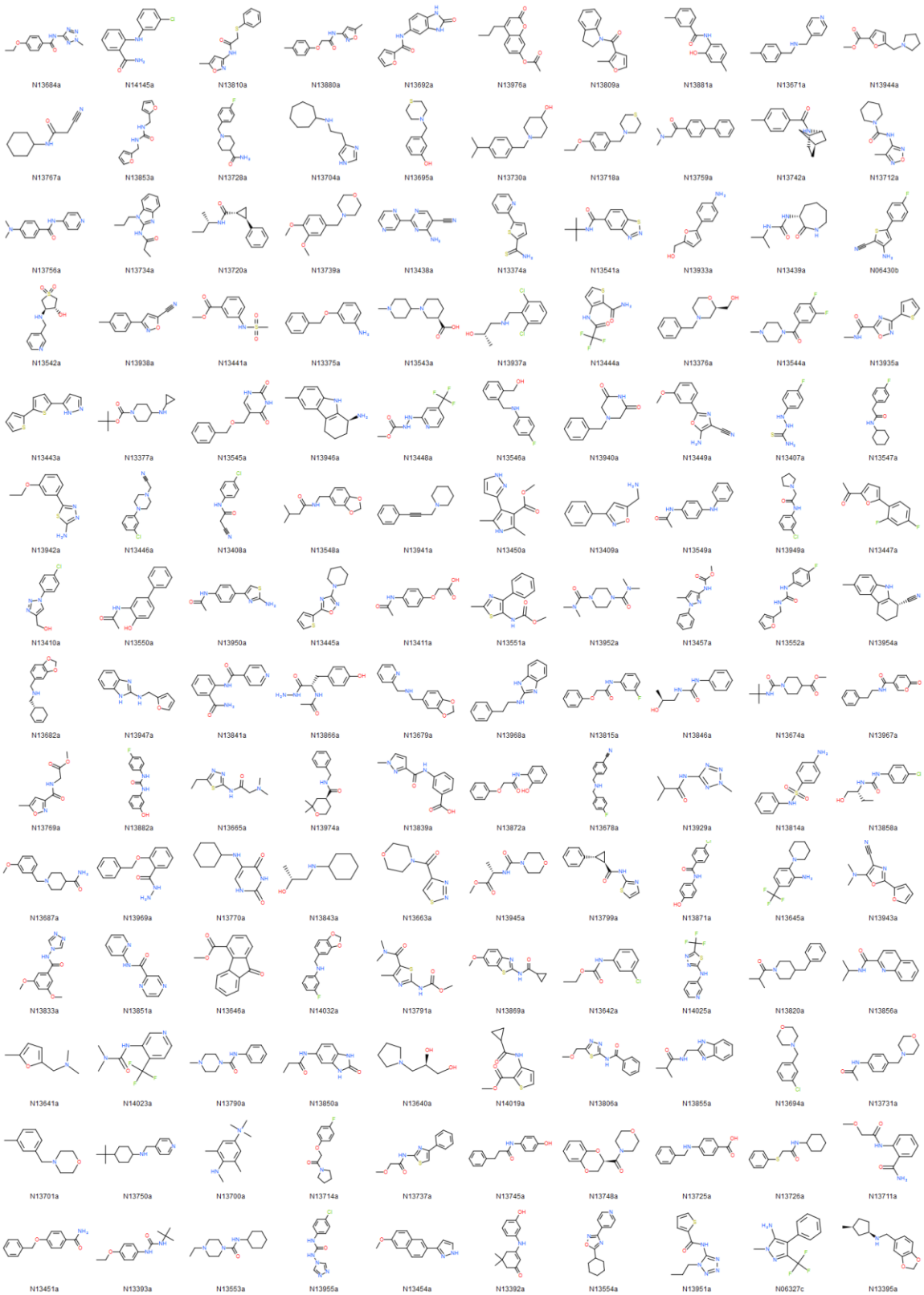


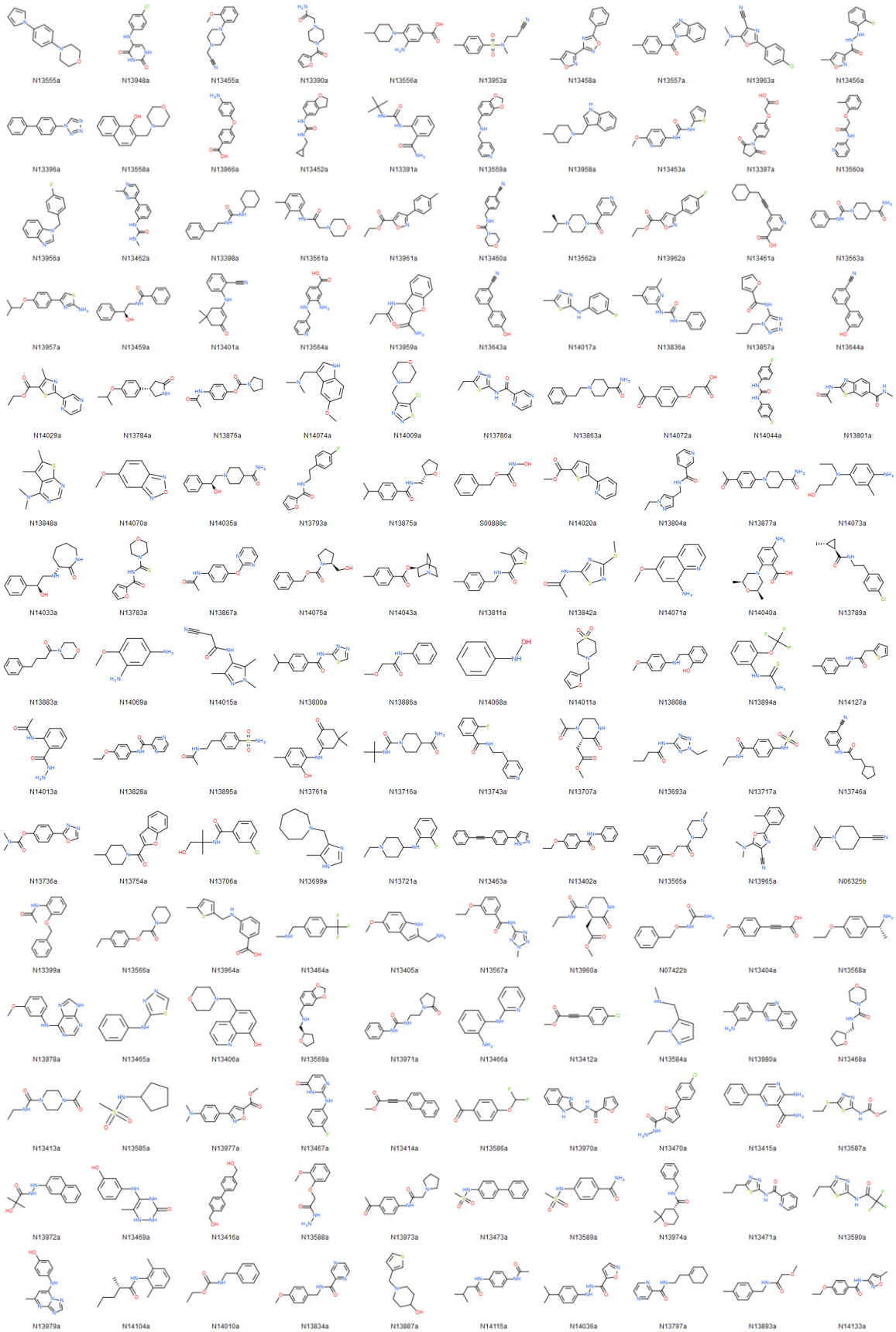
Figure B.9: Subflow used in XPoise to perform a substructure search of a poised fragment.

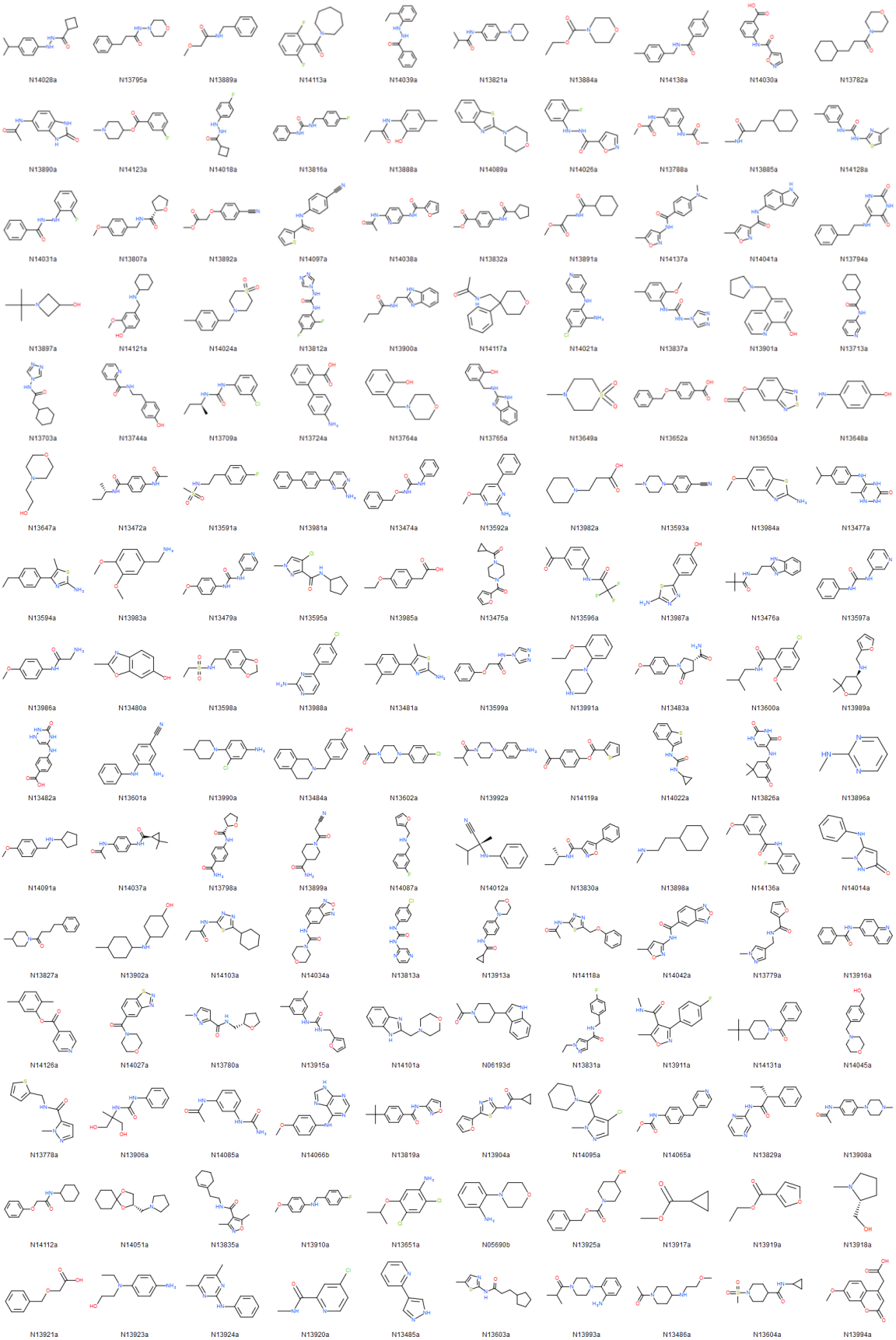
Appendix C: List of DSPL Fragments as 2D Structures

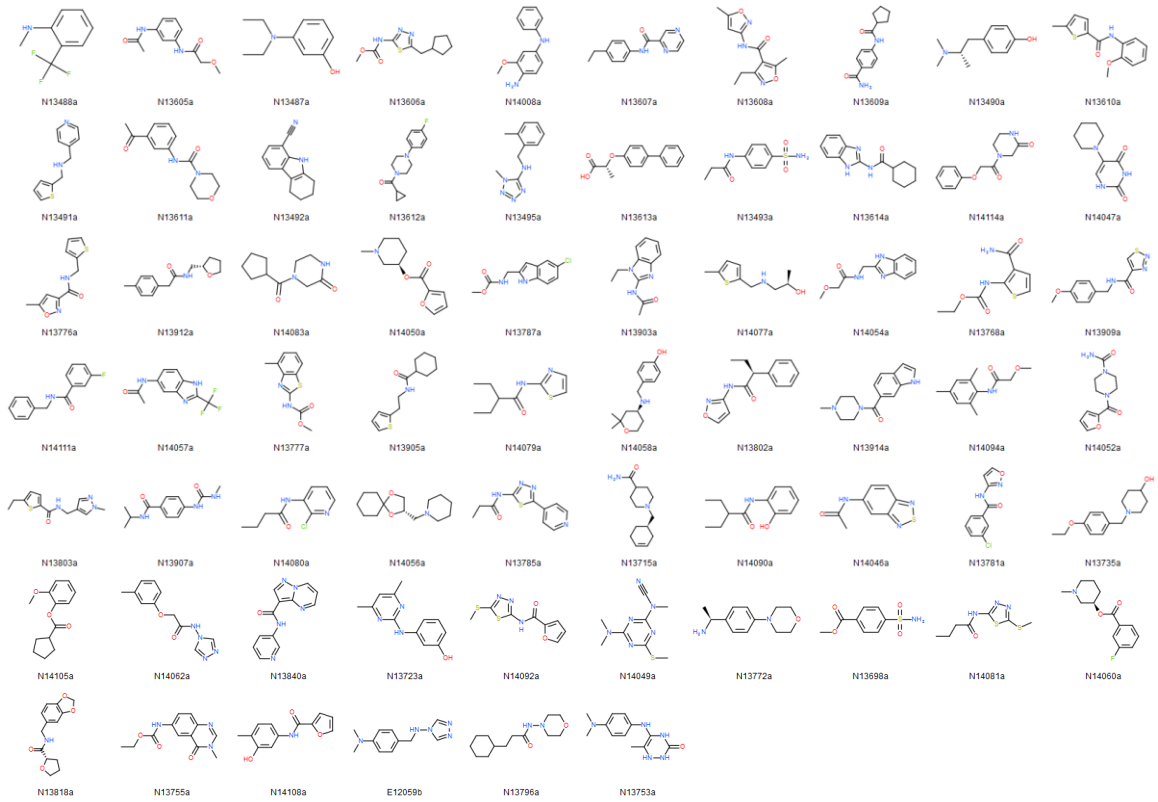



N13654a
N13666a
N14082a
N13999a
N13653a
N13677a
N14130a
N14003a
N08599b
N13683a
N14116a
N14006a
N13860a
N13669a
N14076a
N14002a
N13859a
N13681a
N14098a
N14004a
N13864a
N13676a
N14125a
N14004a
N13879a
N13668a
N14122a
N14007a
N13868a
N13664a
N14088a
S00892b
N13865a
N13673a
N14099a
N13010b
N13845a
N13689a
N14132a
N13998a
N13854a
N13686a
N14093a
E12736c
N13874a
N13733a
N13758a
N13696a
N13702a
N13757a
N13741a
N13749a
N13697a
N13708a
N13752a
N13740a
N13729a
N13429a
N13362a
N13529a
N13581a
N13430a
N13365a
N13530a
N13582a
N13431a
N13363a
N13531a
N06209c
N13366a
N13532a
N13926a
N13433a
N13367a
N13533a
N13927a
N13432a
N13368a
N13534a
N13930a
N13436a
N13369a
N13535a
N13928a
N13437a
N13370a
N13536a
N13931a
N13435a
N08623b
N13537a
N13932a
N13434a
N13372a
N13538a
N13934a
N13442a
N13371a
N13539a
N13939a
N13440a
N13373a
N13540a
N13936a
N13690a
N14135a
N13825a
N13852a
N13680a
N14106a
N13774a
N13873a
N13662a
N14134a
N13824a
N13849a
N13667a
N14109a
N13792a
N13870a
N13688a
N14142a
N13766a
N13861a
N13670a
N14129a
N13817a
N13878a
N13672a
N08253b
N13838a
N13862a
N13691a
N14144a
N13823a
N13847a
N13685a
N14143a
N13771a
N13844a


N13684a
N14145a
N13810a
N13880a
N13692a
N13976a
N13809a
N13881a
N13671a
N13944a
N13767a
N13853a
N13728a
N13704a
N13695a
N13730a
N13718a
N13759a
N13742a
N13712a
N13756a
N13734a
N13720a
N13739a
N13438a
N13374a
N13541a
N13933a
N13439a
N06430b
N13542a
N13938a
N13441a
N13375a
N13543a
N13937a
N13444a
N13376a
N13544a
N13935a
N13443a
N13377a
N13545a
N13946a
N13448a
N13546a
N13940a
N13449a
N13407a
N13547a
N13942a
N13446a
N13408a
N13548a
N13941a
N13450a
N13409a
N13549a
N13949a
N13447a
N13410a
N13550a
N13950a
N13445a
N13411a
N13551a
N13952a
N13457a
N13552a
N13954a
N13682a
N13947a
N13841a
N13866a
N13679a
N13968a
N13815a
N13846a
N13674a
N13967a
N13769a
N13882a
N13665a
N13974a
N13839a
N13872a
N13678a
N13929a
N13814a
N13858a
N13687a
N13969a
N13770a
N13843a
N13663a
N13945a
N13799a
N13871a
N13645a
N13943a
N13833a
N13851a
N13646a
N14032a
N13791a
N13869a
N13642a
N14025a
N13820a
N13856a
N13641a
N14023a
N13790a
N13850a
N13640a
N14019a
N13806a
N13855a
N13694a
N13731a
N13701a
N13750a
N13700a
N13714a
N13737a
N13745a
N13748a
N13725a
N13726a
N13711a
N13451a
N13393a
N13553a
N13955a
N13454a
N13392a
N13554a
N13951a
N06327c
N13395a


N13555a
N13948a
N13455a
N13390a
N13556a
N13953a
N13458a
N13557a
N13963a
N13456a
N13396a
N13558a
N13966a
N13452a
N13391a
N13559a
N13958a
N13453a
N13397a
N13560a
N13956a
N13462a
N13398a
N13561a
N13961a
N13460a
N13562a
N13962a
N13461a
N13563a
N13957a
N13459a
N13401a
N13564a
N13959a
N13643a
N14017a
N13836a
N13857a
N13644a
N14029a
N13784a
N13876a
N14074a
N14009a
N13786a
N13863a
N14072a
N14044a
N13801a
N13848a
N14070a
N14035a
N13793a
N13875a
S00888c
N14020a
N13804a
N13877a
N14073a
N14033a
N13783a
N13867a
N14075a
N14043a
N13811a
N13842a
N14071a
N14040a
N13789a
N13883a
N14069a
N14015a
N13800a
N13886a
N14068a
N14011a
N13808a
N13894a
N14127a
N14013a
N13828a
N13895a
N13761a
N13716a
N13743a
N13707a
N13693a
N13717a
N13746a
N13736a
N13754a
N13706a
N13699a
N13721a
N13463a
N13402a
N13565a
N13965a
N06325b
N13399a
N13566a
N13964a
N13464a
N13405a
N13567a
N13960a
N07422b
N13404a
N13568a
N13978a
N13465a
N13406a
N13569a
N13971a
N13466a
N13412a
N13584a
N13980a
N13468a
N13413a
N13585a
N13977a
N13467a
N13414a
N13586a
N13970a
N13470a
N13415a
N13587a
N13972a
N13469a
N13416a
N13588a
N13973a
N13473a
N13589a
N13974a
N13471a
N13590a
N13979a
N14104a
N14010a
N13834a
N13887a
N14115a
N14036a
N13797a
N13893a
N14133a


N14028a
N13795a
N13889a
N14113a
N14039a
N13821a
N13884a
N14138a
N14030a
N13782a
N13890a
N14123a
N14018a
N13816a
N13888a
N14089a
N14026a
N13788a
N13885a
N14128a
N14031a
N13807a
N13892a
N14097a
N14038a
N13832a
N13891a
N14137a
N14041a
N13794a
N13897a
N14121a
N14024a
N13812a
N13900a
N14117a
N14021a
N13837a
N13901a
N13713a
N13703a
N13744a
N13709a
N13724a
N13764a
N13765a
N13649a
N13652a
N13650a
N13648a
N13647a
N13472a
N13591a
N13981a
N13474a
N13592a
N13982a
N13593a
N13984a
N13477a
N13594a
N13983a
N13479a
N13595a
N13985a
N13475a
N13596a
N13987a
N13476a
N13597a
N13986a
N13480a
N13598a
N13988a
N13481a
N13599a
N13991a
N13483a
N13600a
N13989a
N13482a
N13601a
N13990a
N13484a
N13602a
N13992a
N14119a
N14022a
N13826a
N13896a
N14091a
N14037a
N13798a
N13899a
N14087a
N14012a
N13830a
N13898a
N14136a
N14014a
N13827a
N13902a
N14103a
N14034a
N13813a
N13913a
N14118a
N14042a
N13779a
N13916a
N14126a
N14027a
N13780a
N13915a
N14101a
N06193d
N13831a
N13911a
N14131a
N14045a
N13778a
N13906a
N14085a
N14066b
N13819a
N13904a
N14095a
N14065a
N13829a
N13908a
N14112a
N14051a
N13835a
N13910a
N13651a
N05690b
N13925a
N13917a
N13919a
N13918a
N13921a
N13923a
N13924a
N13920a
N13485a
N13603a
N13993a
N13486a
N13604a
N13994a


N13488a
N13605a
N13487a
N13606a
N14008a
N13607a
N13608a
N13609a
N13490a
N13610a
N13491a
N13611a
N13492a
N13612a
N13495a
N13613a
N13493a
N13614a
N14114a
N14047a
N13776a
N13912a
N14083a
N14050a
N13787a
N13903a
N14077a
N14054a
N13768a
N13909a
N14111a
N14057a
N13777a
N13905a
N14079a
N14058a
N13802a
N13914a
N14094a
N14052a
N13803a
N13907a
N14080a
N14056a
N13785a
N13715a
N14090a
N14046a
N13781a
N13735a
N14105a
N14062a
N13840a
N13723a
N14092a
N14049a
N13772a
N13698a
N14081a
N14060a
N13818a
N13755a
N14108a
E12059b
N13796a
N13753a

Appendix D: Poised Reaction SMARTS used in Chapter 3

Reaction (Rxn ID)	Disconnection Reaction SMARTS
Amides (1)	[#7:1][C;x0:2]=[O:3]>>[#7:1].Cl[C:2]=[O:3]
S_NAr (2)	[c:2][N:1][#6:3]>>[#6:3]-[#7:1].[c:2]Br
Urea (3)	[#7:1][C;x0]([#7:2])=O>>[#7:1].[#7:2]
Suzuki Coupling (4)	[#6;a:1]-[#6;a:2]>>[#6;a:2]Br.[#6;a:1]-[#5](-[#8])-[#8]
Sonogashira (5)	[#6;a:1][C:2]#[C:3]>>[#6;a:1]Br.[C:2]#[C:3]
Sulfonamide (6)	[#7:1][S:2](=[O:3])=[O:4]>>[#7:1].Cl[S:2](=[O:3])=[O:4]
Reductive Amination (7)[1]	[CH1:2]-[#7R0:1]>>[#7:1].[#6:2]=O [CH2R0:2]-[#7:1]>>[#7:1].[#6:2]=O [CH1R0:2]-[#7:1]>>[#7:1].[#6:2]=O [CH2:2]-[#7R0:1]>>[#7:1].[#6:2]=O [CH3:2]-[#7R0:1]>>[#7:1].[#6:2]=O then remove [C:3]-[#7:1]-[C,S:2]=[O,S:4]>>[#7:1]-[C,S:2]=[O,S:4].[C:3]=O
N-Alkylation (8)[1]	[CH0:2]-[#7R0:1]>>[#7:1].[#6:2]Br [CH1:2]-[#7R0:1]>>[#7:1].[#6:2]Br [CH2R0:2]-[#7:1]>>[#7:1].[#6:2]Br [CH1R0:2]-[#7:1]>>[#7:1].[#6:2]Br [CH2:2]-[#7R0:1]>>[#7:1].[#6:2]Br [CH3:2]-[#7R0:1]>>[#7:1].[#6:2]Br [CH0R0:2]-[#7:1]>>[#7:1].[#6:2]Br then remove [C:3]-[#7:1]-[C,S:2]=[O,S:4]>>[#7:1]-[C,S:2]=[O,S:4].[C:3]Br then remove [C,S:1](=[O,S:2])-[N:3]>>[C,S:1](=[O,S:2])-Br.[N:3]
Ether Coupling (9)[1]	[CH0:2]-[#8R0:1]>>[#8:1].[#6:2]Br [CH1:2]-[#8R0:1]>>[#8:1].[#6:2]Br [CH2R0:2]-[#8:1]>>[#8:1].[#6:2]Br [CH1R0:2]-[#8:1]>>[#8:1].[#6:2]Br [CH2:2]-[#8R0:1]>>[#8:1].[#6:2]Br [CH3:2]-[#8R0:1]>>[#8:1].[#6:2]Br [CH0R0:2]-[#8:1]>>[#8:1].[#6:2]Br [n:3][c:2]-[#8:1]>>[#8:1].[n:3][#6:2]Br then remove [C:3]-[#8:1]-[C,S:2]=[O,S:4]>>[#8:1]-[C,S:2]=[O,S:4].[C:3]Br then remove [C,S:1](=[O,S:2])-[O:3]>>[C,S:1](=[O,S:2])-Br.[O:3]
Ester Coupling (10)	[#8:1][C;x0:2]=[O:3]>>[#8:1].Cl[C:2]=[O:3]
Benzimidazole (11)	[#7:1]1[#6:9][#7:2][#6:4]2[#6:3]1[#6:8][#6:7][#6:6][#6:5]2>>[#7:1][c:3]1[c:8][c:7][c:6][c:5][c:4]1[#7:2].Cl[#6:9]=O
Triazole (12)	[#6:6][n:1]1[c:3][c:4][n:2][n:5]1>>[#6:6][#7:1][#7:5]=[#7:2].[C:3]#[C:4]
Benzoxazole (13)	[#7:1]1[#6:9][#8:2][#6:4]2[#6:3]1[#6:8][#6:7][#6:6][#6:5]2>>[#7:1][c:3]1[c:8][c:7][c:6][c:5][c:4]1[#8:2].Cl[#6:9]=O
Indole (14)	[c:10]1[c:9][nH:1][c:3]2[c:8]1[c:7][c:6][c:5][c:4]2>>N[N:1][c:3]1[c:8]([2H])[c:7][c:6][c:5][c:4]1.[C:10][C:9]=O
Oxadiazole (15)	[#6:6][c:4]1[n:5][o:3][c:1][n:2]1>>[O:2]-[C:1]=[O:3].[#6:6][C:4]#[N:5]

[1]Multiple SMARTS used to avoid ring opening disconnections and ensure two snthons are generated

Table B.1: SMARTS strings used to perform disconnection reactions on poised fragments.

Reaction (Rxn ID)	Varied synthon	Reaction Smarts
Amides (1A)	Acids	[Cl,Br,I][C:2]=[O:3].[#7:1]>>[#7:1][C:2]=[O:3] [OH1][C:2]=[O:3].[#7:1]>>[#7:1][C:2]=[O:3]
Amides (1B)[1]	Amines	[NH2:1].Cl[C:2]=[O:3]>>[#7:1][C:2]=[O:3] [NH1:1].Cl[C:2]=[O:3]>>[#7:1][C:2]=[O:3]
S_NAr (2A)[2]	Amines	[#7H2:1].[c:2]Br>>[c:2][PH1:1] [NH1:1].[c:2]Br>>[c:2][PH0:1] then [#15:1]>>[#7:1]
S_NAr (2B)	Aromatic Halides	[c:2]-[Cl,Br,I].[#7:1]>>[c:2]-[#7:1]
Urea (3A)[1]	Amines	[#7H2:1].[#7:2]>>[#7H1:1]C([#7:2])=O [NH1:1].[#7:2]>>[#7H0:1]C([#7:2])=O
Urea (3B)[1]	Isocyanates	[#7:2]=C=O.[#7H2:1]>>[#7:1]-[#6](-[#7:2])=O [#7:2]=C=O.[#7;AH1:1]>>[#7:1]-[#6](-[#7:2])=O
Suzuki Coupling (4A)	Aromatic Halides	[#6;a:2]-[#35,#53].[#6;a:1]-[#5](-[#8])-[#8]>>[#6;a:1]-[#6;a:2]
Suzuki Coupling (4B)	Boronic Acids	[#6;a:1]-[#5](-[#8])-[#8].[#6;a:2]-[#35,#53]>>[#6;a:1]-[#6;a:2]
Sonogashira (5A)	Alkynes	[C:1]#[C:2].[#6;a:3]-[#35,#53]>>[#6;a:3][C:1]#[C:2]
Sonogashira (5B)	Aromatic Halides	[#6;a:1]-[#35,#53].[C:2]#[C:3]>>[#6;a:1][C:2]#[C:3]
Sulfonamide (6A)	Sulfonyl Chlorides	[#17,#35,#53][S:4](=[O:3])=[O:2].[#7:1]>>[#7:1][S:4](=[O:3])=[O:2]
Sulfonamide (6B)[1]	Amines	[#7H2:1].Cl[S:2](=[O:3])=[O:4]>>[#7:1][S:2](=[O:3])=[O:4] [#7H1:1].Cl[S:2](=[O:3])=[O:4]>>[#7:1][S:2](=[O:3])=[O:4]
Reductive Amination (7A)[2,3]	Ketones/ Aldehydes	[#6H1:2]=O.[#7:1]>>[#6:2]-[P:1] [#6:3]-[#6:2](-[#6:4])=O.[#7:1]>>[#6:3]-[#6:2](-[#6:4])-[P:1] then [P:1]>>[N:1]
Reductive Amination (7B)[1,2,3]	Amines	[#7;AH2:1].[#6:4]-[#6:2](-[#6:3])=O>>[#6:4]-[P:2](-[#6:3])-[#7:1] [#7AH1:1].[#6:2]-[#6:3](-[#6:4])=O>>[#6:2]-[P:3](-[#6:4])-[#7:1] [#7AH2:1].[#6:4]-[#6H1:2]=O>>[#6:4]-[P:2]-[#7:1] [#7AH1:1].[#6:2]-[#6H1:3]=O>>[#6:2]-[P:3]-[#7:1] then [P:1]>>[C:1]
N-Alkylation (8A)[1,2]	Amines	[#7H2:1].[#6;A:2]Br>>[#7:1]-[#15:2] [#7H1:1].[#6;A:2]Br>>[#7:1]-[#15:2] then [P:1]>>[C:1]
N-Alkylation (8B)[1,2]	Halides	[C:2][Cl,Br,I].[#7H2:1]>>[#7:1]-[#15:2] [C:2][Cl,Br,I].[#7H1:1]>>[#7:1]-[#15:2] then [P:1]>>[C:1]
Ether Coupling (9A)[4]	Alcohols	[#8H1:1].[#6;A:2]Br>>[#6:2]-[#8:1] [OH1:1].[n:3][#6;a:2]Br>>[O:1]-[#6:2][n:3]
Ether Coupling (9B)[4]	Halides	[#6;A:2][#17,#35,#53].[#8H1:1]>>[#6:2]-[#8:1] [n:3][c:2][#17,#35,#53].[#8H1:1]>>[n:3][c:2][#8:1]
Ester Coupling (10A)	Acids	[Cl,Br,I][C:2]=[O:3].[#8:1]>>[#8:1][C:2]=[O:3] [OH1][C:2]=[O:3].[#8:1]>>[#8:1][C:2]=[O:3]
Ester Coupling (10B)	Alcohols	[OH:1].Cl[C:2]=[O:3]>>[#8:1][C:2]=[O:3]
Benzimidazole (11A)	Acids	Cl[#6:9]=O.[#7:1][c:3]1[c:8][c:7][c:6][c:5][c:4]1[#7:2]>>[nH:1]1[c:9][n:2][c:4]2[c:5][c:6][c:7][c:8][c:3]12 [OH][#6:9]=O.[#7:1][c:3]1[c:8][c:7][c:6][c:5][c:4]1[#7:2]>>[nH:1]1[c:9][n:2][c:4]2[c:5][c:6][c:7][c:8][c:3]12
Benzimidazole (11B)	Diamines	[N:1][c:3]1[c:8][c:7][c:6][c:5][c:4]1[N:2].Cl[#6:9]=O>>[nH:1]1[c:9][n:2][c:4]2[c:5][c:6][c:7][c:8][c:3]12 [#7:1][c:3]1[c:8][c:7][c:6][c:5][c:4]1[#7:2].Cl[#6:9]=O>>[nH:1]1[c:9][n:2][c:4]2[c:5][c:6][c:7][c:8][c:3]12
Triazole (12A)	Alkynes	[C:3]#[C:4][#6:6].[#7:1][#7:5]=[#7:2]>>[#7:1]1[#6H:3]=[#6:4]([#6:6])[#7:2]=[#7:5]1
Triazole (12B)[5]	Azides	[#7:1]=[#7+:2]=[#7-:3]>>[#7:1][#7:2]=[#7:3] then [#7:1][#7:5]=[#7:2].[C:3]#[C:4][#6:6]>>[#7:1]1[#6H:3]=[#6:4]([#6:6])[#7:2]=[#7:5]1
Benzoxazole (13A)	Acids	Cl[#6:9]=O.[#7:1][c:3]1[c:8][c:7][c:6][c:5][c:4]1[#8:2]>>[n:1]1[c:9][o:2][c:4]2[c:5][c:6][c:7][c:8][c:3]12 [OH][#6:9]=O.[#7:1][c:3]1[c:8][c:7][c:6][c:5][c:4]1[#8:2]>>[n:1]1[c:9][o:2][c:4]2[c:5][c:6][c:7][c:8][c:3]12
Benzoxazole (13B)	Aminophenols	[OH:1][c:3]1[c:8][c:7][c:6][c:5][c:4]1[NH2:2].Cl[#6:9]=O>>[o:1]1[c:9][n:2][c:4]2[c:5][c:6][c:7][c:8][c:3]12

Reaction (Rxn ID)	Varied synthon	Reaction Smarts
Indole (14A)[6,7]	Ketones/ Aldehydes	[C:10][C:9]=O.[#7][#7:1][c:3]1[c:8][c:7][c:6][c:5][c:4]1([2H])>>[c:10]1[c:9][nH:1][c:3]2[c:8][c:7][c:6][c:5][c:4]12 N[N:1][c:3]1[c:8][c:7][c:6][c:5][c:4]1>>[N:1][c:3]1[c:8][c:7][c:6][c:5][c:4]1 then [c:10][C:9](=O)[CH2:2][#17,#35,#53].[NH2:1][c:3]1[c:8][c:7][c:6][c:5][c:4]1([2H])>>[c:10][c:9]1[c:2][c:4]2[c:5][c:6][c:7][c:8][c:3]2[nH:1]1
Indole (14B)[6,7]	Hydrazines/ Anilines	[#7][#7:1][c:3]1[c:8][c:7][c:6][c:5][c:4]1.[#6:11][C:10][C:9]=O>>[#6:11][c:10]1[c:9][nH:1][c:3]2[c:8][c:7][c:6][c:5][c:4]12 [c:1][C:2](=[O:3])[CH3:4]>>[c:1][C:2](=[O:3])[C:4]Br then [NH2:1][c:3]1[c:8][c:7][c:6][c:5][c:4]1.[C:9](=O)[C:10][Br]>>[c:9]1[c:10][c:4]2[c:5][c:6][c:7][c:8][c:3]2[nH:1]1
Oxadiazole (15A)	Nitriles	[#6:1][C:3]#[N:2].[OH:5]-[#6:4]=O>>[#6:1][c:3]1[n:2][o:5][c:4]n1
Oxadiazole (15B)	Carboxylic Acids	[OH:5]-[#6:4]=O.[#6:1][C:3]#[N:2]>>[#6:1][c:3]1[n:2][o:5][c:4]n1

[1]Primary and secondary amines reacted separately. [2]Phosphorus used as dumi-atom. [3]Ketones and aldehydes reacted separately. [4]S_N2 and S_NAr encoded separately. [5]Azide converted to [R]NN=N before reaction. [6]Two reactions used for indole synthesis. [7]Deuterium marker used at 2-position to remove symmetry of starting materials.

Table B.2: SMARTS strings used to perform reaction between poised synthons.

Reaction (Rxn ID)	Varied synthon	Starting Material SMARTS to keep	Starting Material SMARTS remove[1]	Product SMARTS remove
Amides (1A)	Acids	[Cl,Br,I]C=O [H]OC=O	[#7H2] [NH1] (keep [#7H1][C,S]=O) C(=O)NC(=O)	
Amides (1B)	Amines	N[H]	Ncn NN	O=CN[C,S]=[O,S] *[n+]*
S_NAr (2A)	Amines	N[H]	Ncn NN	P[C,S]=[O,S] *[n+]*
S_NAr (2B)	Aromatic Halides	c[Cl,Br,I]	[#7H2] [NH1] (keep [#7H1][C,S]=O)	
Urea (3A)	Amines	N[H]	Ncn NN	O=CN[C,S]=[O,S] *[n+]*
Urea (3B)	Isocyanates	N=C=O	[#7H2] [NH1] (keep [#7H1][C,S]=O)	NC(=O)N[C,S](=[O,S]) *[n+]*
Suzuki Coupling (4A)	Aromatic Halides	c[Cl,Br,I]		
Suzuki Coupling (4B)	Boronic Acids	B		
Sonogashira (5A)	Alkynes	C#[CH]		
Sonogashira (5B)	Aromatic Halides	c[Br,I]		
Sulfonamide (6A)	Sulfonyl Chlorides	[Cl,Br,I]S(=O)=O	[#7H2] [NH1] (keep [#7H1][C,S]=O)	
Sulfonamide (6B)	Amines	N[H]	Ncn NN	O=SN[C,S]=[O,S] [N+]
Reductive Amination (7A)	Ketones/ Aldehydes	[CH1]=O (remove [#6]C([#6])=O) [#6]C([#6])=O	[#7H2] [NH1] (keep [#7H1][C,S]=O) NN	P[C,S]=[O,S] N[CH2]N
Reductive Amination (7B)	Amines	N[H]	[CH1]=O [#6]C([#6])=O NN Ncn	PN[C,S]=[O,S] *[N+]*
N-Alkylation (8A)	Amines	N[H]	cO[H] NN Ncn	PN[C,S]=[O,S] *[N+]*

Reaction (Rxn ID)	Varied synthon	Starting Material SMARTS to keep	Starting Material SMARTS remove[1]	Product SMARTS remove
N-Alkylation (8B)	Halides	C[Cl,Br,I]	[#7H2] [NH1] (keep [#7H1][C,S]=O) C=C[Cl,Br,I] [O,S]=[C,S][Cl,Br,I] NN	PN[C,S]=[O,S] *[N+]*
Ether Coupling (9A)	Alcohols	O[H]	ncO[H] cN[H] NN	O[N,S] *[N+]*
Ether Coupling (9B)	Halides	[#6][Cl,Br,I]	O[H] [#6]C(=O)O [#7H2] [NH1] (keep [#7H1][C,S]=O) C=C[Cl,Br,I] [O,S]=[C,S][Cl,Br,I] NN	O[N,S] *[N+]*
Ester Coupling (10A)	Acids	[Cl,Br,I]C=O [H]OC=O	[#7H2] [NH1] (keep [#7H1][C,S]=O) c[OH] [C;X4][OH] C(=O)NC(=O) NN	
Ester Coupling (10B)	Alcohols	[C;X4][OH] c[OH]	[#7H2] [NH1] (keep [#7H1][C,S]=O) Ncn NN	O=CN[C,S]=[O,S] *[n+]*
Benzimidazole (11A)	Acids	[Cl,Br,I]C=O [H]OC=O	[#7H2] [NH1] (keep [#7H1][C,S]=O) NN	
Benzimidazole (11B)	Diamines	N[H] then [NH2]cc[NH] or [NH2]cc[NH2]	NN	
Triazole (12A)	Alkynes	C#[CH]	C#N NN	
Triazole (12B)	Azides	N=[N+]=[N-]	C#C C#N NN	
Benzoxazole (13A)	Acids	[Cl,Br,I]C=O [H]OC=O	[#7H2] [NH1] (keep [#7H1][C,S]=O) NN	
Benzoxazole (13B)	Aminophenols	N[H] then [NH2]cc[OH]	NN	
Indole (14A)	Ketones/ Aldehydes	O=CC	[#7H2] [NH1] (keep [#7H1][C,S]=O) NN	Keep [#6]c1c([#6])[nH]c2ccccc12 Keep [#6]c1[cH][nH]c2ccccc12 Keep [cH]1c([#6])[nH]c2ccccc12 Keep [cH]1[cH][nH]c2ccccc12 [nH]1[c;R1][c;R1]c2ccccc12
Indole (14B)	Hydrazines/ Anilines	c1ccccc1[NH2] (remove NN) c1ccccc1[NH][NH2]		c[NH2]
Oxadiazole (15A)	Nitriles	C#N	NN C=O C[NH2]	c1noc([O,#7,S,Cl,Br,I,F])n1
Oxadiazole (15B)	Carboxylic Acids	[#6]C(=O)[#6] [H]OC=O	[#7H2] [NH1] (keep [#7H1][C,S]=O) C(=O)NC(=O) NN	c1noc([O,#7,S,Cl,Br,I,F])n1

[1]REOS filter applied to starting materials. When a reactive handle was required for the reaction the REOS filter was reconfigured to keep the functional group or the REOS filter was applied to the reaction products.

Table B.3: SMARTS strings used to build synthon libraries and to filter unwanted products.

Appendix E: Experimental Procedures

List of Abbreviations

Ac – Acetyl

Aq. – Aqueous

Bis-Tris – 2,2-*bis*(Hydroxymethyl)-2,2′,2″-nitrilotriethanol

Boc – *tert*-Butoxycarbonyl

Br. – Broad

BSA – Bovine serum albumin

cDNA – Complementary DNA

CHAPS – 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate

DMF – Dimethylformamide

DMSO – Dimethyl sulphoxide

ELSD – Evaporative light scattering detector

Eq – Equivalents

Et – Ethyl

FTIR – Fourier transform infra-red

h - Hours

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

HRMS – High resolution mass spectrometry

LRMS – Low resolution mass spectrometry

M – Molar

Me – Methyl

MeCN – Acetonitrile

mins – Minutes

Mp – Melting point

PEG – Polyethylene glycol

Pr – Propyl

rpm – Revolutions per minute

rt – Room temperature

TCEP – *tris*(2-Carboxyethyl)phosphine

THF – Tetrahydrofuran

TIC – Total ion chromatogram

TLC – Thin layer chromatography

TOF – Time of flight

UV – Ultra violet

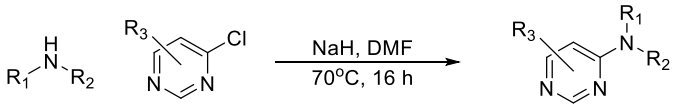
General Chemistry Experimental

All reactions involving moisture sensitive reagents were carried out under a nitrogen atmosphere using standard vacuum line techniques. Water was deionized by an Elga DV 25 system. All other solvents and reagents were used as supplied (analytical or HPLC grade) without prior purification. Organic layers were dried over Na_2SO_4 or $MgSO_4$. Parallel work-ups were carried out using a Radleys stacker and Isolute phase separation cartridges. Thin layer chromatography was performed on aluminium plates coated with 60 F_{254} silica gel. Plates were visualised using UV light (254 nm) or 1% aq. $KMnO_4$. Flash column chromatography was performed on a Biotage Isolera One flash column chromatography platform unless stated. Purification by HPLC was performed using a Waters SFO with 515 HPLC pump and Waters Binary Gradient 2545 device (5%-95% solvent A (18% water, 80% acetonitrile, 2% 0.5 M ammonium acetate pH 6.0) in solvent B (93% water, 5% acetonitrile, 2% ammonium acetate pH 6.0)). Product was detected using a SQ Detector 2 and collected using a Waters Sample Manager 2767. Melting points were recorded on a Stuart SMP40 apparatus or a Barnstead Electrothermal 9100 machine and are uncorrected. IR spectra were recorded on a Nicolet iS5 with an iD7 ATR module, neat. Selected characteristic peaks are reported in cm^{-1}. NMR spectra were recorded on a Bruker Avance spectrometer or Varian spectrometer in the deuterated solvent stated. The field was locked by external referencing to the relevant deuteron resonance. Chemical shifts (δ) are reported in ppm and coupling constants (*J*) in Hz. Spectra were recorded at room temperature unless otherwise stated. Coupling constants (*J*) are quoted in Hz and are recorded to the nearest 0.1 Hz. When peak multiplicities are reported, the following abbreviations are used: s = singlet, d = doublet, t = triplet, m = multiplet, br. = broadened, dd = doublet of doublets, dt = doublet of triplets. *m/z* values are reported in Daltons. Compound purity was assessed by LCMS using either UV absorbance at 254 nm (Waters UV/Visible Detector 2489), ELSD signal (Waters ELS Detector 2424) or ESI+ TIC (SQ Detector 2). The detection method

was chosen which gave the strongest signal. Analytical LCMS was performed on the following system: Kinetex 5µ EVO C18 100A 100 x 3.0 mm column using a linear gradient of solvent A (93 % H_2O, 5 % acetonitrile, and 2 % of 0.5 M ammonium acetate pH 6.0) and solvent B (18 % H_2O, 80 % acetonitrile, and 2 % of 0.5 M ammonium acetate pH 6.0), eluting at a flow rate of 2 mL/min: 5% B over 0.35 min, 5% B to 95% B over 1 min, 95% B over 0.75 min, 95% to 5% B over 0.1 min and 5% B over 0.8 min. Reverse phase HPLC was performed using preparatory scale LCMS on the same system equipped with a Kinetex 5u EVO C18 100A 150 x 21.2 mm column and a Waters 2767 Sample Manager. A linear gradient of solvent A and solvent B, eluting at a flow rate of 20 mL/min: 10% B over 2.5 min, 10% B to 90% B over 2.5 min, 90% B over 7.0 min, 90% to 10% B over 0.5 min and 10% B over 2.5 min. LRMS were recorded by LCMS on a Waters SQ Detector 2; data acquisition and processing was performed using Waters FractionLynx software. HRMS was performed on an Agilent 6530 Accurate Mass Q-TOF; data acquisition and processing was performed using Agilent MassHunter Workstation software.

General Synthetic Procedures

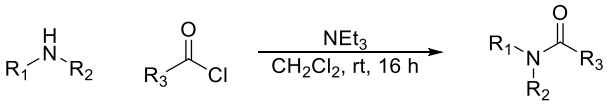
General Procedure 1: *S_NAr between a 4-chloropyrimidine and an amine*



Sodium hydride (60% dispersion in mineral oil, 0.8 mmol, 1.2 eq) was added to a stirred solution of the amine (0.7 mmol, 1.1 eq) in DMF (1 mL). The reaction mixture was stirred for 30 mins at room temperature. The 4-chloropyrimidine (0.7 mmol, 1.0 eq) was added dropwise to the reaction mixture and stirred overnight at 70 °C. The solvent was removed *in vacuo* and the crude product dissolved in CH_2Cl_2 (5 mL) and diluted with aq. NaOH (10%, 5 mL). The organic layer was

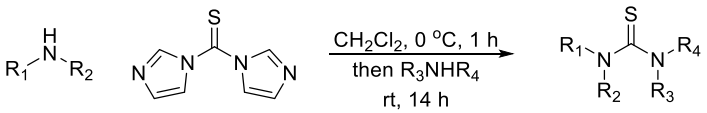
dried and purified by flash column chromatography (10%-80% EtOAc in cyclohexane). Solvent was removed *in vacuo* to afford the desired product.

General Procedure 2: ***Amide coupling between an acyl chloride and an amine***



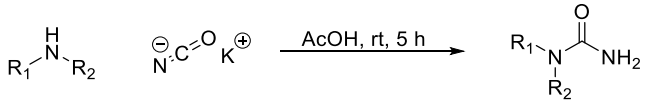
The amine (0.5 mmol, 1.0 eq) and triethylamine (224 µL, 1.6 mmol, 3 eq) were dissolved in CH_2Cl_2 (2 mL). The acyl chloride (0.6 mmol, 1.05 eq) was added and the resulting reaction mixture stirred overnight at room temperature. The crude mixture was diluted with aq. NaOH (10%, 5 mL). The organic layer was separated, washed with water (5 mL) and dried. The solvent was removed *in vacuo* to afford the desired product.

General Procedure 3: ***Thiourea synthesis using 1,1'-Thiocarbonyldiimidazole (Kang, Wang et al. 2009)***



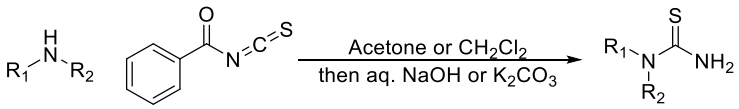
The first amine (0.6 mmol, 1.0 eq) was added to a stirred solution of 1,1'-thiocarbonyldiimidazole (100 mg, 0.6 mmol, 1.0 eq) in CH_2Cl_2 (2 mL). The reaction mixture was stirred for 1-2 h at room temperature. The second amine (0.6 mmol, 1.0 eq) was added to the reaction mixture and stirred overnight at room temperature. The crude product was washed with aq. HCl (1M, 5 mL) and water (5 mL). The organic layer was dried and purified by flash column chromatography (12%-100% EtOAc in Cyclohexane). Solvent was removed *in vacuo* to afford the desired product.

General Procedure 4: ***Urea synthesis using potassium cyanate***



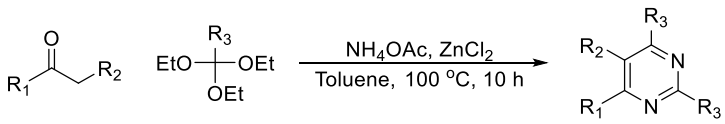
The amine (0.7 mmol, 1.0 eq) was added to a stirred solution of potassium cyanate (60 mg, 0.7 mmol, 1.0 eq) in acetic acid (1 mL, 1.3 mmol, 1.7 eq). The reaction mixture was stirred for 5 h at room temperature. The reaction mixture was filtered and the filtrate purified by reverse phase HPLC. Solvent was removed *in vacuo* to afford the desired product.

General Procedure 5: ***Thiourea synthesis using benzoyl isothiocyanate***



The amine (0.71 mmol, 1.0 eq) was slowly added to a stirred solution of benzoyl isothiocyanate (100µL, 0.74 mmol, 1.05 eq) in acetone (1 mL) or CH_2Cl_2 (1 mL). The reaction mixture was monitored by LCMS and the solvent removed *in vacuo* when all of the amine had been consumed. The resulting solid was diluted in aq. NaOH (10%, 2 mL) or aq. K_2CO_3 (2 mL) and stirred until the carbamothioylbenzamide intermediate had been converted to the terminal thiourea. The pH of the solution was lowered to ~3.5 using aq. HCl, filtered and the filtrand purified by reverse phase HPLC. Solvent was removed *in vacuo* to afford the desired product.

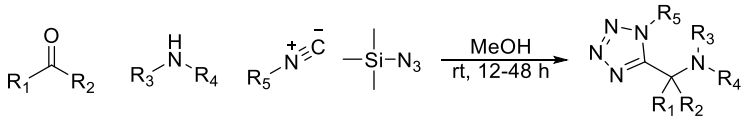
General Procedure 6: ***Pyrimidine synthesis (Sasada, Kobayashi et al. 2009)***



The ketone (1.0 eq) and ammonium acetate (2.0 eq) was added to a stirred solution of zinc chloride (0.2 eq) and triethylorthoformate (12.0 eq) in toluene (2 mL). The reaction mixture was heated to 100 °C for 10 h before being cooled and quenched with aq. $NaHCO_3$ (5 mL). The product

was extracted using CH_2Cl_2 (2 x 5 mL) and purified by flash column chromatography. Pure fractions were collected and the solvent removed *in vacuo* to afford the desired product.

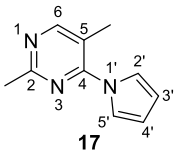
General Procedure 7: ***Ugi 4-CR tetrazole synthesis (Wang, Fang et al. 2013)***



The ketone (or aldehyde) (1.0 eq) was added to a stirred solution of the amine (1.0 eq) in MeOH (2 mL). The reaction was stirred at room temperature for 30-90 min before the isonitrile (1.2 eq) and trimethylsilyl azide (1.5 eq) were added. The reaction was stirred at room temperature for 12-48 h and monitored by LCMS and TLC. The reaction was quenched using aqueous HCl (1.0 M, 5 mL) and extracted using CH_2Cl_2. The resulting mixture was purified by flash column chromatography. Pure fractions were collected and the solvent removed *in vacuo* to afford the desired product.

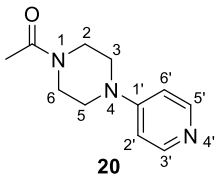
Chemistry Experimental

2,5-Dimethyl-4-(1*H*-pyrrol-1-yl)pyrimidine **17**



Following *General Procedure 1*, pyrrole (43 µL) was treated with sodium hydride (27 mg) then 4-chloro-2,5-dimethylpyrimidine (80 mg) to afford *pyrimidine* **17** as a colourless oil (45 mg, 46%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.40 (s, 1H, C(6)*H*), 7.31 - 7.37 (m, 2H, C(2')*H* & C(5')*H*), 6.25 - 6.34 (m, 2H, C(3')*H* & C(4')*H*), 2.62 (s, 3H, C(2)CH_3), 2.38 (s, 3H, C(5)CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 166.2 (*C*(2)), 161.0 (*C*(6)), 156.2 (*C*(4)), 120.5 (*C*(2') & *C*(5')), 116.3 (*C*(5)), 111.2 (*C*(3') & *C*(4')), 25.5 (C(2)*C*H$_3$), 16.8 (C(5)*C*H$_3$); *m/z* (ESI$^+$) 174 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{10}H_{11}N_3^+$, ([M+H]$^+$) requires 174.1026, found 174.1020.

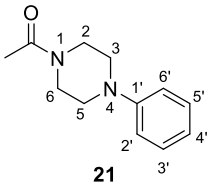
1-(4-(Pyridin-4-yl)piperazin-1-yl)ethan-1-one **20**



20

Following *General Procedure 2*, acetyl chloride (50 µL, 0.70 mmol, 1.05 eq) was treated with 1-(pyridin-4-yl)piperazine (109 µL, 0.67 mmol, 1.0 eq) and triethylamine (187 µL, 2.0 mmol, 2.0 eq) to afford an off-white powder. A portion of the crude product (25 mg) was purified by reverse phase HPLC to afford *amide* **20** as a white powder (11 mg, 8%). Mp 133-134.5 °C; ν_{max} (FTIR) cm^{-1} 2841 (CH), 1617 (C=O), 1592, 1239, 986, 805 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.26 - 8.36 (m, 2H, C(3')*H* & C(5')*H*), 6.67 - 6.75 (m, 2H, C(2')*H* & C(6')*H*), 3.80 (t, *J* = 5.4 Hz, 2H, C(6)H_2), 3.67 (t, *J* = 5.3 Hz, 2H, C(2)H_2), 3.49 (t, *J* = 5.3 Hz, 2H, C(3)H_2), 3.44 (t, *J* = 5.4 Hz, 2H, C(5)H_2), 2.17 (s, 3H, CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 169.3 (*C*=O), 155.2 (*C*(1')), 147.8 (*C*(3') & *C*(5')), 108.2 (*C*(2') & *C*(6')), 45.7 (*C*(3)), 45.6 (*C*(5)), 45.3 (*C*(2)), 40.6 (*C*(6)), 21.3 (*C*H_3); *m/z* (ESI^+) 206 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_{11}H_{16}N_3O^+$, ($[M+H]^+$) requires 206.1288, found 206.1299.

1-(4-Phenylpiperazin-1-yl)ethan-1-one **21**

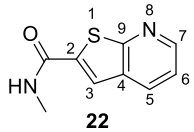


21

Following *General Procedure 2*, acetyl chloride (50 µL, 0.70 mmol, 1.05 eq) was treated with 1-phenylpiperazine (102 µL, 0.67 mmol, 1.0 eq) and triethylamine (187 µL, 2.0 mmol, 2.0 eq) to afford a yellow oil. A portion of the crude product (25 mg) was purified by reverse phase HPLC to afford *amide* **21** as an off-white powder (18 mg, 13%). Mp 85-86.5 °C; ν_{max} (FTIR) cm^{-1} 2817 (CH), 1623 (C=O), 1598, 1427, 1227, 768 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.29 - 7.36 (m, 2H, C(3')*H* & C(5')*H*), 6.91 - 7.06 (m, 3H, C(2')*H*, C(4')*H* & C(6')*H*), 3.78 - 3.87 (m, 2H, C(2)H_2), 3.62 - 3.71 (m, 2H, C(6)H_2), 3.14 - 3.27 (m, 4H, C(3)H_2 & C(5)H_2), 2.17 (s, 3H, CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 169.0 (*C*=O), 150.6 (*C*(1')), 129.3 (*C*(3') & *C*(5')), 121.0 (*C*(4')), 116.9 (*C*(2') & *C*(6')), 50.0 (*C*(5)), 49.7 (*C*(3)), 46.1 (*C*(6)), 41.2

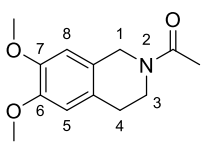
(*C*(2)), 21.4 (*C*H$_3$); *m/z* (ESI$^+$) 205 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{12}H_{17}N_2O^+$, ([M+H]$^+$) requires 205.1335, found 205.1333.

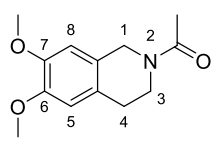
N-Methylthieno[2,3-b]pyridine-2-carboxamide **22**


22

HATU (282 mg, 0.74 mmol, 1.0 eq) was added to a solution of thieno[2,3-*b*]pyridine-2-carboxylic acid (133 mg, 0.74 mmol, 1.0 eq) and *N*-methylmorpholine (163 µL, 1.48 mmol, 2.0 eq) in MeCN (2 mL). The reaction was stirred at room temperature for 2 h before the addition of methanamine hydrochloride (50 mg, 0.74 mmol, 1.0 eq). The reaction was stirred for a further 14 h. Solvent was removed *in vacuo* and the crude solid redissolved in CH_2Cl_2 (5 mL). The resulting solution was washed with aq. NaOH (10%, 10 mL) and purified by flash column chromatography (0%-10% MeOH in CH_2Cl_2). Solvent was removed *in vacuo* to afford *amide* **22** as white crystals (61 mg, 43%). Mp 183-184.5 °C; ν_{max} (FTIR) cm^{-1} 3321 (CONH), 1620 (C=O), 1545 (O=CN), 1308, 790 (aromatic CH), 732; ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.82 (br. d, *J* = 4.20 Hz, 1H, N*H*), 8.65 (dd, *J* = 1.7, 4.6 Hz, 1H, C(7)*H*), 8.37 (dd, *J* = 1.6, 8.2 Hz, 1H, C(5)*H*), 8.01 (s, 1H, C(3)*H*), 7.49 (dd, *J* = 4.5, 8.1 Hz, 1H, C(6)*H*), 2.83 (d, *J* = 4.7 Hz, 3H, C*H*$_3$); ^{13}C NMR (101 MHz, DMSO-d_6) δ 162.0 (*C*=O), 161.6 (*C*(9)), 148.9 (*C*(7)), 140.0 (*C*(2)), 133.7 (*C*(5)), 133.2 (*C*(4)), 123.0 (*C*(3)), 121.0 (*C*(6)), 26.7 (*C*H$_3$); *m/z* (ESI$^+$) 193 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_9H_9N_2OS^+$, ([M+H]$^+$) requires 193.0430, found 193.0430.

1-(6,7-Dimethoxy-3,4-dihydroisoquinolin-2(1H)-yl)ethan-1-one **23**

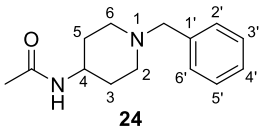

23a


23b

Following *General Procedure 2*, acetyl chloride (50 µL, 0.70 mmol, 1.05 eq) was treated with 6,7-dimethoxy-1,2,3,4-tetrahydroisoquinolin-2-ium chloride (154 mg, 0.67 mmol, 1.0 eq) and triethylamine (187 µL, 2.0 mmol, 2.0 eq) to afford an orange powder. A portion of the crude product (25 mg) was purified by reverse phase HPLC to afford *amide* **23** as an off-white amorphous solid (10 mg, 6%). Rotamers

23a and **23b** were found to be present in a 4:3 ratio by NMR. ν_{max} (FTIR) cm^{-1} 2935 (CH), 1628 (C=O), 1514, 1254 (C-O), 1221 (C-O); **23a** ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.64 (s, 2H, C(5)*H* & C(8)*H*), 4.68 (s, 2H, C(1)H_2), 3.86 - 3.89 (m, 6H, C(6)OCH_3 & C(7)OCH_3), 3.68 (t, *J* = 5.8 Hz, 2H, C(3)H_3), 2.85 (t, *J* = 5.8 Hz, 2H, C(4)H_3), 2.19 (s, 3H, C(=O)CH_3); **23a** ^{13}C NMR (101 MHz, $CDCl_3$) δ 169.5 (*C*=O), 147.9 (*C*(6 or 7)), 147.7 (*C*(6 or 7)), 125.7 (C(4)*C*C(5)), 125.4 (C(1)*C*C(8)), 111.2 (*C*(5)), 109.4 (*C*(8)), 56.0 (C(6)O*C*H_3 & C(7)O*C*H_3), 44.1 (*C*(3)), 43.8 (*C*(1)), 29.0 (*C*(4)), 21.6 (C(=O)*C*H_3); **23b** ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.66 (s, 1H, C(5)*H*), 6.61 (s, 1H, C(8)*H*), 4.57 (s, 2H, C(1)H_2), 3.86 - 3.89 (m, 6H, C(6)OCH_3 & C(7)OCH_3), 3.83 (t, *J* = 5.9 Hz, 2H, C(3)H_3), 2.79 (t, *J* = 5.9 Hz, 2H, C(4)H_3), 2.20 (s, 3H, C(=O)CH_3); **23b** ^{13}C NMR (101 MHz, $CDCl_3$) δ 169.5 (*C*=O), 148.0 (*C*(6 or 7)), 147.7 (*C*(6 or 7)), 127.0 (C(4)*C*C(5)), 124.2 (C(1)*C*C(8)), 111.6 (*C*(5)), 108.9 (*C*(8)), 56.0 (C(6)O*C*H_3 & C(7)O*C*H_3), 47.8 (*C*(1)), 39.5 (*C*(3)), 28.1 (*C*(4)), 21.9 (C(=O)*C*H_3); *m/z* (ESI$^+$) 236 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{13}H_{18}NO_3^+$, ([M+H]$^+$) requires 236.1281, found 236.1276.

N-(1-Benzylpiperidin-4-yl)acetamide **24**

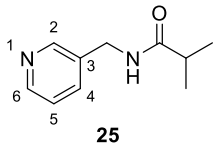


24

Following *General Procedure 2*, acetyl chloride (50 µL, 0.70 mmol, 1.05 eq) was treated with 1-benzylpiperidin-4-amine (137 µL, 0.67 mmol, 1.0 eq) and triethylamine (187 µL, 2.0 mmol, 2.0 eq) to afford a yellow powder. A portion of the crude product (25 mg) was purified by reverse phase HPLC to afford *amide* **24** as white crystals (14 mg, 9%). Mp 138.5-140 °C; ν_{max} (FTIR) cm^{-1} 3261 (CONH), 2938 (CH), 1632 (C=O), 1555 (O=CN), 1364, 738 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.29 - 7.38 (m, 5H, C(2'-6')*H*), 5.55 (br. d, *J* = 7.2 Hz, 1H, N*H*), 3.78 - 3.91 (m, 1H, C(4)*H*), 3.63 (s, 2H, CH_2C(1')), 2.96 (d, *J* = 12.1 Hz, 2H, C(2)H_{EQ} & C(6)H_{EQ}), 2.24 (dt, *J* = 2.3, 12.0 Hz, 2H, C(2)H_{AX} & C(6)H_{AX}), 1.98 (s, 3H, CH_3), 1.90 - 1.97 (m, 2H, C(3)H_{EQ} & C(5)H_{EQ}), 1.61 (dq, *J* = 3.8, 11.8 Hz, 2H, C(3)H_{AX} & C(5)H_{AX}); ^{13}C NMR (101 MHz, $CDCl_3$) δ 169.5 (*C*=O), 136.2 (*C*(1')), 129.6 (*C*(2') & *C*(6')), 128.4 (*C*(3') & *C*(5')), 127.7 (*C*(4')), 62.5 (*C*H_2C(1')),

52.0 (*C*(2) & *C*(6)), 46.1 (*C*(4)), 31.5 (*C*(3) & *C*(5)), 23.5 (*C*H_3); *m/z* (ESI$^+$) 233 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{12}H_{21}N_2O^+$, ([M+H]$^+$) requires 233.1648, found 233.1644.

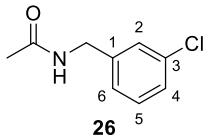
N-(Pyridin-3-ylmethyl)isobutyramide **25**



25

Following *General Procedure 2*, isobutyryl chloride (108 µL, 1.03 mmol, 1.05 eq) was treated with pyridin-3-ylmethanamine (100 µL, 0.98 mmol, 1.0 eq) and triethylamine (274 µL, 2.0 mmol, 2.0 eq) to afford a yellow oil. A portion of the crude product (25 mg) was purified by reverse phase HPLC to afford *amide* **25** as a colourless oil (14 mg, 8%). ν_{max} (FTIR) cm^{-1} 3279 (CONH), 2968 (CH$_3$), 1648 (C=O), 1541 (O=CN), 710 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.48 - 8.54 (m, 2H, C(2)*H* & C(6)*H*), 7.63 (td, *J* = 1.7, 7.8 Hz, 1H, C(4)*H*), 7.27 (dd, *J* = 4.3, 7.8 Hz, 1H, C(5)*H*), 6.11 (br. s, 1H, N*H*), 4.46 (d, *J* = 6.0 Hz, 2H, CH_2), 2.42 (spt, *J* = 6.9 Hz, 1H, C*H*), 1.19 (d, *J* = 7.0 Hz, 6H, C(CH_3)$_2$); ^{13}C NMR (101 MHz, $CDCl_3$) δ 177.1 (*C*=O), 149.0 (*C*(2)), 148.7 (*C*(6)), 135.6 (*C*(4)), 134.4 (*C*(3)), 123.6 (*C*(5)), 40.8 (*C*H$_2$), 35.6 (*C*H), 19.6 (C(*C*H$_3$)$_2$); *m/z* (ESI$^+$) 179 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{10}H_{14}N_2O^+$, ([M+H]$^+$) requires 179.1179, found 179.1179.

N-(3-Chlorobenzyl)acetamide **26**

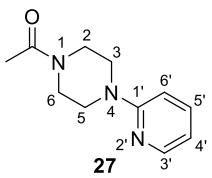


26

Following *General Procedure 2*, acetyl chloride (50 µL, 0.70 mmol, 1.05 eq) was treated with 3-chlorobenzylamine (82 µL, 0.67 mmol, 1.0 eq) and triethylamine (187 µL, 2.0 mmol, 2.0 eq) to afford a yellow oil. A portion of the crude product (25 mg) was purified by reverse phase HPLC to afford *amide* **26** as a white powder (10 mg, 8%). Mp 42.5-44.5 °C; ν_{max} (FTIR) cm^{-1} 3273 (CONH), 1637 (C=O), 1545 (O=CN), 1428, 1372, 781 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.24 - 7.30 (m, 3H, C(2)*H*, C(4)*H* & C(5)*H*), 7.14 - 7.21 (m, 1H, C(6)*H*), 5.95 (br. s, 1H, N*H*), 4.43 (d, *J* = 5.8 Hz, 2H, CH_2), 2.06 (s, 3H, CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 170.1 (*C*=O), 140.3 (C(1)), 134.5 (C(3)), 130.0 (C(5)), 127.8 (C(2 or 4)), 127.7 (C(2 or 4)),

125.9 (C(6)), 43.1 (*C*H_2), 23.2 (*C*H_3); *m/z* (ESI$^+$) 184 ([M(^{35}Cl)+H]$^+$, 100%), 186 ([M(^{37}Cl)+H]$^+$, 45%); HRMS (ESI$^+$) $C_9H_{11}{}^{35}ClNO^+$, ([M+H]$^+$) requires 184.0524, found 184.0525.

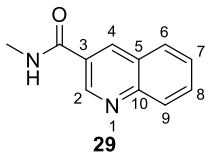
1-(4-(Pyridin-2-yl)piperazin-1-yl)ethan-1-one **27**



27

Following *General Procedure 2*, acetyl chloride (50 µL, 0.70 mmol, 1.05 eq) was treated with 1-(pyridin-2-yl)piperazine (102 µL, 0.67 mmol, 1.0 eq) and triethylamine (187 µL, 2.0 mmol, 2.0 eq) to afford a brown powder. A portion of the crude product (50 mg) was purified by reverse phase HPLC to afford *amide* **27** as a white powder (27 mg, 20%). Mp 88.5-90 °C; ν_{max} (FTIR) cm^{-1} 2849 (CH), 1642 (C=O), 1589, 1239, 977, 778 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.22 (ddd, *J* = 0.8, 1.9, 4.9 Hz, 1H, C(3')*H*), 7.54 (ddd, *J* = 2.0, 7.0, 8.6 Hz, 1H, C(5')*H*), 6.64 - 6.73 (m, 2H, C(4')*H* & C(6')*H*), 3.73 - 3.80 (m, 2H, C(2)H_2), 3.63 - 3.69 (m, 2H, C(5)H_2), 3.58 - 3.63 (m, 2H, C(6)H_2), 3.47 - 3.55 (m, 2H, C(3)H_2), 2.17 (s, 3H, CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 169.2 (*C*=O), 158.9 (*C*(1')), 147.8 (*C*(3')), 137.8 (*C*(5')), 113.9 (*C*(4')), 107.3 (*C*(6')), 46.0 (*C*(6)), 45.3 (*C*(3)), 45.1 (*C*(5)), 41.0 (*C*(2)), 21.4 (*C*H_3); *m/z* (ESI$^+$) 206 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{11}H_{16}N_3O^+$, ([M+H]$^+$) requires 206.1288, found 206.1293.

N-Methylquinoline-3-carboxamide **29**

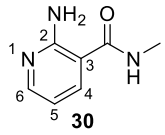


29

HATU (282 mg, 0.74 mmol, 1.0 eq) was added to a solution of quinoline-3-carboxylic acid (128 mg, 0.74 mmol, 1.0 eq) and *N*-methylmorpholine (163 µL, 1.48 mmol, 2.0 eq) in MeCN (2 mL). The reaction was stirred at room temperature for 2 h before the addition of methanamine hydrochloride (50 mg, 0.74 mmol, 1.0 eq). The reaction was stirred for a further 14 h. Solvent was removed *in vacuo* and the crude solid redissolved in CH_2Cl_2 (5 mL). The resulting solution was washed with aq. NaOH (10%, 10 mL) and purified by flash column chromatography (0%-10% MeOH in CH_2Cl_2). Solvent was removed *in vacuo* to afford *amide* **29** as a pink powder (101 mg, 73%). Mp 120-121.5 °C; ν_{max} (FTIR) cm^{-1} 3263 (CONH), 1619 (C=O), 1544

(O=CN), 1307, 787 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.27 (d, *J* = 2.3 Hz, 1H, C(2)*H*), 8.61 (d, *J* = 2.1 Hz, 1H, C(4)*H*), 8.13 (d, *J* = 8.6 Hz, 1H, C(9)*H*), 7.87 (d, *J* = 8.2 Hz, 1H, C(6)*H*), 7.80 (ddd, *J* = 1.5, 7.0, 8.4 Hz, 1H, C(8)*H*), 7.60 (ddd, *J* = 1.0, 7.0, 8.4 Hz, 1H, C(7)*H*), 6.75 (br. s, 1H, N*H*), 3.10 (d, *J* = 4.9 Hz, 3H, CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 166.4 (*C*=O), 149.2 (*C*(10)), 148.1 (*C*(2)), 135.6 (*C*(4)), 131.2 (*C*(8)), 129.3 (*C*(9)), 128.7 (*C*(6)), 127.5 (*C*(7)), 127.2 (*C*(3)), 126.9 (*C*(5)), 27.0 (*C*H_3); *m/z* (ESI$^+$) 187 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{11}H_{11}N_2O^+$, ([M+H]$^+$) requires 187.0866, found 187.0867.

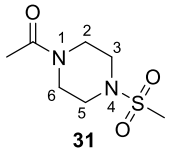
2-amino-*N*-methylnicotinamide **30**



30

HATU (282 mg, 0.74 mmol, 1.0 eq) was added to a solution of 2-aminonicotinic acid (102 mg, 0.74 mmol, 1.0 eq) and *N*-methylmorpholine (163 µL, 1.48 mmol, 2.0 eq) in MeCN (2 mL). The reaction was stirred at room temperature for 2 h before the addition of methanamine hydrochloride (50 mg, 0.74 mmol, 1.0 eq). The reaction was stirred for a further 14 h. Solvent was removed *in vacuo* and the crude solid redissolved in CH_2Cl_2 (5 mL). The resulting solution was washed with aq. NaOH (10%, 10 mL) and purified by flash column chromatography (0%-10% MeOH in CH_2Cl_2). Solvent was removed *in vacuo* to afford *amide* **30** as white crystals (39 mg, 35%). Mp 138.5-140 °C; ν_{max} (FTIR) cm^{-1} 3331 (NH_2), 3132 (CONH), 1627 (C=O), 1532 (O=CN), 1256, 751 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.14 (dd, *J* = 1.7, 4.9 Hz, 1H, C(6)*H*), 7.63 (dd, *J* = 1.7, 7.7 Hz, 1H, C(4)*H*), 6.60 (dd, *J* = 4.9, 7.7 Hz, 1H, C(5)*H*), 6.43 (br. s, 2H, NH_2), 6.25 (br. s, 1H, N*H*), 2.99 (d, *J* = 4.8 Hz, 3H, CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 168.3 (*C*=O), 158.4 (*C*(2)), 151.1 (*C*(6)), 135.6 (*C*(4)), 112.3 (*C*(5)), 110.9 (*C*(3)), 26.7 (*C*H_3); *m/z* (ESI$^+$) 152 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_7H_{10}N_3O^+$, ([M+H]$^+$) requires 152.0818, found 152.0824.

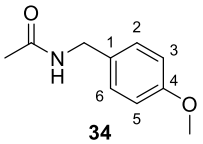
1-(4-(Methylsulfonyl)piperazin-1-yl)ethan-1-one **31**



31

Following *General Procedure 2*, acetyl chloride (50 µL, 0.70 mmol, 1.05 eq) was treated with 1-(methylsulfonyl)piperazine (110 mg, 0.67 mmol, 1.0 eq) and triethylamine (187 µL, 2.0 mmol, 2.0 eq), and the resulting mixture purified by flash column chromatography (0%-10% MeOH in CH_2Cl_2). Solvent was removed *in vacuo* to afford *amide* **31** as a white powder (48 mg, 35%). Mp 167.5-168 °C; ν_{max} (FTIR) cm^{-1} 2917 (CH), 1625 (C=O), 1430 (C=O), 1318 (SO_2N), 1248, 1140 (SO_2N), 973; ^1H NMR (400 MHz, $CDCl_3$) δ 3.71 - 3.81 (m, 2H, C(2)H_2), 3.55 - 3.65 (m, 2H, C(6)H_2), 3.24 (dd, *J* = 4.5, 9.1 Hz, 4H, C(3)H_2 & C(5)H_2), 2.81 (s, 3H, SO_2CH_3), 2.14 (s, 3H, C(=O)CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 169.0 (*C*=O), 45.9 (*C*(6)), 45.6 (*C*(3) & *C*(5)), 41.0 (*C*(2)), 34.8 (SO_2CH_3), 21.3 (C(=O)CH_3); *m/z* (ESI^+) 207 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_7H_{15}N_2O_3S^+$, ($[M+H]^+$) requires 207.0798, found 207.0795.

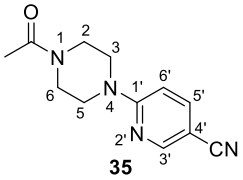
N-(4-Methoxybenzyl)acetamide **34**



34

Following *General Procedure 2*, acetyl chloride (50 µL, 0.70 mmol, 1.05 eq) was treated with 4-methoxybenzylamine (88 µL, 0.67 mmol, 1.0 eq) and triethylamine (187 µL, 2.0 mmol, 2.0 eq) to afford a yellow powder. A portion of the crude product (25 mg) was purified by reverse phase HPLC to afford *amide* **34** as a white powder (16 mg, 13%). Mp 96.5-98 °C; ν_{max} (FTIR) cm^{-1} 3278 (CONH), 1628 (C=O), 1551 (O=CN), 1509, 1229 (C-O), 1029 (CO), 755 (aromatic CH); ^1H NMR (400 MHz, $CDCl_3$) δ 7.18 - 7.25 (m, 2H, C(2)*H* & C(6)*H*), 6.85 - 6.91 (m, 2H, C(3)*H* & C(5)*H*), 5.81 (br. s, 1H, N*H*), 4.38 (d, *J* = 5.4 Hz, 2H, CH_2), 3.82 (s, 3H, OCH_3), 2.03 (s, 3H, C(=O)CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 170.0 (*C*=O), 159.1 (*C*(4)), 130.2 (*C*(1)), 129.3 (*C*(2) & *C*(6)), 114.1 (*C*(3) & *C*(5)), 55.3 (OCH_3), 43.3 (CH_2), 23.2 (C(=O)CH_3); *m/z* (ESI^+) 180 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_{10}H_{13}NNaO_2^+$, ($[M+Na]^+$) requires 202.0838, found 202.0835.

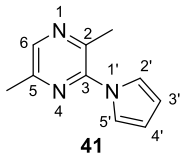
6-(4-Acetylpiperazin-1-yl)nicotinonitrile **35**



35

Following *General Procedure 2*, acetyl chloride (50 µL, 0.70 mmol, 1.05 eq) was treated with 6-(piperazin-1-yl)nicotinonitrile (126 mg, 0.67 mmol, 1.0 eq) and triethylamine (187 µL, 2.0 mmol, 2.0 eq), and the resulting mixture purified by flash column chromatography (0%-10% MeOH in CH_2Cl_2). Solvent was removed *in vacuo* to afford *amide* **35** as white crystals (126 mg, 82%). Mp 144.5-146.5 °C; ν_{max} (FTIR) cm^{-1} 2856 (CH), 2209 (C≡N), 1624 (C=O), 1599 (C=O), 1427, 998; ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.44 (dd, *J* = 0.6, 2.3 Hz, 1H, C(3')*H*), 7.68 (dd, *J* = 2.4, 9.0 Hz, 1H, C(5')*H*), 6.64 (dd, *J* = 0.6, 9.0 Hz, 1H, C(6')*H*), 3.80 - 3.86 (m, 2H, C(5)H_2), 3.74 - 3.79 (m, 2H, C(2)H_2), 3.64 - 3.69 (m, 2H, C(3)H_2), 3.58 - 3.64 (m, 2H, C(6)H_2), 2.17 (s, 3H, CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 169.3 (*C*=O), 158.9 (*C*(1')), 152.4 (*C*(3')), 140.2 (*C*(5')), 118.2 (*C*≡N), 106.0 (*C*(6')), 97.3 (*C*(4')), 45.6 (*C*(6)), 44.5 (*C*(3)), 44.1 (*C*(5)), 40.7 (*C*(2)), 21.4 (*C*H_3); *m/z* (ESI^+) 231 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_{12}H_{15}N_4O^+$, ($[M+H]^+$) requires 231.1240, found 231.1237.

2,5-Dimethyl-3-(1*H*-pyrrol-1-yl)pyrazine **41**

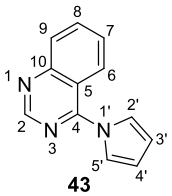


41

A flask was charged with pyrrole (29 µl, 0.41 mmol, 1.0 eq), caesium carbonate (162 mg, 0.50 mmol, 1.2 eq) and 3-chloro-2,5-dimethylpyrazine (0.1 ml, 0.82 mmol, 2.0 eq) in DMF (1 ml). The reaction mixture was stirred for 20 h at 100 °C for 20 h and then cooled to room temperature. The crude mixture was diluted with EtOAc, washed with water and brine, and purified by flash column chromatography (7%-60% EtOAc in cyclohexane). The solvent was removed *in vacuo* to afford *pyrazine* **41** as a colourless oil (20 mg, 28%). ν_{max} (FTIR) cm^{-1} 2923 (CH_3), 1719 1378, 1030, 731 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.32 (s, 1H, C(6)*H*), 7.21 (t, *J* = 2.2 Hz, 2H, C(2')*H* & C(5')*H*), 6.39 (t, *J* = 2.2 Hz, 2H, C(3')*H* & C(4')*H*), 2.67 (s, 3H, C(2)CH_3), 2.57 (s, 3H, C(5)CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 150.1 (*C*(5)), 146.6 (*C*(3)), 143.1 (*C*(2)), 141.1

(*C*(6)), 120.8 (*C*(2') & *C*(5')), 110.5 (*C*(3') & *C*(4')), 21.6 (C(2)*C*H$_3$), 20.8 (C(5)*C*H$_3$); *m/z* (ESI$^+$) 174 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{10}H_{11}N_3^+$, ([M+H]$^+$) requires 174.1026, found 174.1019.

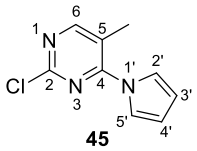
4-(1*H*-Pyrrol-1-yl)quinazoline **43**



43

Following *General Procedure 1*, pyrrole (50 µL, 1.0 eq) was treated with sodium hydride (32 mg, 1.1 eq) then 4-chloroquinazoline (178 mg, 1.5 eq) to afford *benzopyrimidine* **43** as a brown powder (52 mg, 37%). Mp 191-192 °C; ν_{max} (FTIR) cm^{-1} 3132, 1697, 1465, 910, 762 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.17 (s, 1H, C(2)*H*), 8.38 (d, *J* = 8.4 Hz, 1H, C(6)*H*), 8.19 (d, *J* = 8.4 Hz, 1H, C(9)*H*), 7.99 (ddd, *J* = 1.4, 7.0, 8.4 Hz, 1H, C(8)*H*), 7.73 (ddd, *J* = 1.1, 7.2, 8.4 Hz, 1H, C(7)*H*), 7.51 - 7.57 (m, 2H, C(2')*H* & C(5')*H*), 6.51 - 6.56 (m, 2H, C(3')*H* & C(4')*H*); ^{13}C NMR (101 MHz, $CDCl_3$) δ 157.9 (*C*(4)), 153.1 (*C*(2)), 150.8 (*C*(10)), 134.9 (*C*(8)), 128.8 (*C*(7)), 127.5 (*C*(9)), 125.6 (*C*(6)), 122.2 (*C*(2') & *C*(5')), 117.2 (*C*(5)), 113.3 (*C*(3') & *C*(4')); *m/z* (ESI$^+$) 196 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{12}H_9N_3^+$, ([M+H]$^+$) requires 196.0869, found 196.0860.

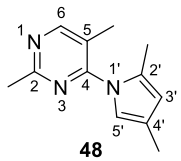
2-Chloro-5-methyl-4-(1*H*-pyrrol-1-yl)pyrimidine **45**



45

Following *General Procedure 1*, pyrrole (50 µL) was treated with sodium hydride (31 mg) then 2,4-dichloro-5-methylpyrimidine (107 mg) to afford *pyrimidine* **45** as a colourless oil (55 mg, 43%). ν_{max} (FTIR) cm^{-1} 1551, 1383, 1155, 729 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.45 (d, *J* = 0.6 Hz, 1H, C(6)*H),* 7.44 - 7.51 (m, 2H, C(2')*H* & C(5')*H*), 6.38 - 6.46 (m, 2H, C(3')*H* & C(4')*H*), 2.53 (s, 3H, C*H*$_3$); ^{13}C NMR (101 MHz, $CDCl_3$) δ 162.9 (*C*(6)), 158.6 (*C*(2)), 157.7 (*C*(4)), 120.7 (*C*(2') & *C*(5')), 117.8 (*C*(5)), 112.3 (*C*(3') & *C*(4')), 17.0 (*C*H$_3$); *m/z* (ESI$^+$) 194 ([M(^{35}Cl)+H]$^+$, 100%), 196 ([M(^{37}Cl)+H]$^+$, 90%); HRMS (ESI$^+$) $C_9H_8{}^{35}ClN_3^+$, ([M+H]$^+$) requires 194.0480, found 194.0478.

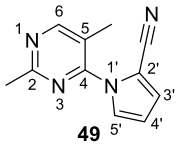
4-(2,4-Dimethyl-1*H*-pyrrol-1-yl)-2,5-dimethylpyrimidine **48**



48

Following *General Procedure 1*, 2,4-dimethyl-1*H*-pyrrole (64 µL) was treated with sodium hydride (27 mg) then 4-chloro-2,5-dimethylpyrimidine (80 mg) to afford *pyrimidine* **48** as a red oil (10 mg, 9%). ν_{max} (FTIR) cm^{-1} 2925 (CH_3), 1554, 1430, 734 (aromatic CH); ^1H NMR (400 MHz, $CDCl_3$) δ 8.57 (s, 1H, C(6)*H*), 6.52 (s, 1H, C(5')*H*), 5.93 (s, 1H, C(3')*H*), 2.71 (s, 3H, C(2)CH_3), 2.26 (s, 3H, C(5)CH_3), 2.22 (s, 3H, C(2')CH_3), 2.10 (s, 3H, C(4')CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 166.2 (*C*(2)), 160.3 (*C*(6)), 158.2 (*C*(4)), 129.7 (*C*(2')), 121.5 (*C*(5)), 120.1 (*C*(4')), 117.7 (*C*(5')), 111.6 (*C*(3')), 25.4 (C(2)*C*H_3), 14.9 (C(5)*C*H_3), 12.9 (C(2')*C*H_3), 11.8 (C(4')*C*H_3); *m/z* (ESI^+) 202 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_{12}H_{15}N_3^+$, ($[M+H]^+$) requires 202.1339, found 202.1329.

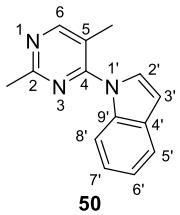
1-(2,5-Dimethylpyrimidin-4-yl)-1*H*-pyrrole-2-carbonitrile **49**



49

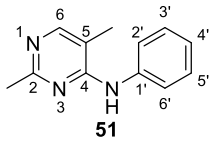
Following *General Procedure 1*, 1-pyrrole-2-carbonitrile (53 µL) was treated with sodium hydride (27 mg) then 4-chloro-2,5-dimethylpyrimidine (80 mg) to afford *pyrimidine* **49** as a white powder (7 mg, 6.3%). Mp 109.5-111 °C; ν_{max} (FTIR) cm^{-1} 2216 (C≡N), 1546, 1447, 742 (aromatic CH); ^1H NMR (400 MHz, $CDCl_3$) δ 8.69 (s, 1H, C(6)*H*), 7.19 (dd, *J* = 1.6, 2.9 Hz, 1H, C(3' or 5')*H*), 7.07 (dd, *J* = 1.6, 3.8 Hz, 1H, C(3' or 5')*H*), 6.44 (dd, *J* = 3.0, 3.8 Hz, 1H, C(4')*H*), 2.78 (s, 3H, C(2)CH_3), 2.36 (s, 3H, C(5)CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 167.0 (*C*(2)), 161.5 (*C*(6)), 155.2 (*C*(4)), 126.2 (*C*(3' or 5')), 123.0 (*C*(3' or 5')), 120.7 (*C*(5)), 113.2 (*C*≡N), 111.4 (*C*(4')), 104.6 (*C*(2')), 25.2 (C(2)*C*H_3), 14.8 (C(5)*C*H_3); *m/z* (ESI^+) 199 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_{11}H_{10}N_4^+$, ($[M+H]^+$) requires 199.0978, found 199.0970.

1-(2,5-Dimethylpyrimidin-4-yl)-1*H*-indole **50**



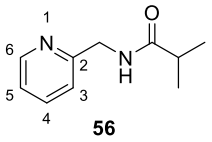
Following *General Procedure 1*, indole (72 mg) was treated with sodium hydride (27 mg) then 4-chloro-2,5-dimethylpyrimidine (80 mg) to afford *pyrimidine* **50** as a pale yellow oil (54 mg, 43%). ν_{max} (FTIR) cm^{-1} 1541, 1429, 1187, 740 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.64 (s, 1H, C(6)*H*), 7.65 - 7.73 (m, 2H, C(5')*H* & C(8')*H*), 7.44 (d, *J* = 3.4 Hz, 1H, C(2')*H*), 7.21 - 7.34 (m, 2H, C(6')*H* & C(7')*H*), 6.76 (d, *J* = 3.4 Hz, 1H, C(3')*H*), 2.78 (s, 3H, C(2)CH_3), 2.34 (s, 3H, C(5)CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 166.4 (*C*(2)), 161.0 (*C*(6)), 157.3 (*C*(4)), 135.6 (*C*(4')), 129.5 (*C*(9')), 126.6 (*C*(2')), 123.2 (*C*(6')), 121.6 (*C*(7')), 121.1 (*C*(8')), 119.8 (*C*(5)), 112.7 (*C*(5')), 105.6 (*C*(3')), 25.5 (C(2)*C*H_3), 15.7 (C(5)*C*H_3); *m/z* (ESI^+) 224 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_{14}H_{13}N_3^+$, ($[M+H]^+$) requires 224.1182, found 224.1170.

2,5-Dimethyl-*N*-phenylpyrimidin-4-amine **51**



Following *General Procedure 1*, aniline (51 µL) was treated with sodium hydride (25 mg) then 4-chloro-2,5-dimethylpyrimidine (80 mg) to afford *pyrimidine* **51** as a yellow oil (21 mg, 18%). ν_{max} (FTIR) cm^{-1} 3278, 1562, 1438, 1014, 749 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.11 (s, 1H, C(6)*H*), 7.65 - 7.73 (m, 2H, C(2')*H* & C(6')*H*), 7.34 - 7.41 (m, 2H, C(3')*H* & C(5')*H*), 7.09 - 7.15 (m, 1H, C(4')*H*), 6.41 (br. s, 1H, N*H*), 2.59 (s, 3H, C(2)CH_3), 2.18 (s, 3H, C(5)CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 165.5 (*C*(2)), 158.7 (*C*(4)), 154.6 (*C*(6)), 138.9 (*C*(1')), 128.9 (*C*(3') & *C*(5')), 123.5 (*C*(4')), 120.6 (*C*(2') & *C*(6')), 111.0 (*C*(5)), 25.5 (C(2)*C*H_3), 13.6 (C(5)*C*H_3); *m/z* (ESI^+) 200 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_{12}H_{13}N_3^+$, ($[M+H]^+$) requires 200.1182, found 200.1173.

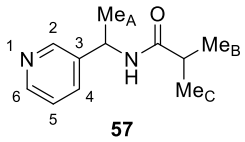
N-(Pyridin-2-ylmethyl)isobutyramide **56**



56

Following *General Procedure 2*, isobutyryl chloride (100 µL, 0.95 mmol, 1.05 eq) was treated with pyridin-2-ylmethanamine (94 µL, 0.91 mmol, 1.0 eq) and triethylamine (253 µL, 2.0 mmol, 1.8 eq) to afford an orange oil. A portion of the crude product (25 mg) was purified by reverse phase HPLC to afford *amide* **56** as an orange oil (17 mg, 10%). ν_{max} (FTIR) cm^{-1} 3292 (CONH), 2967 (CH_3), 1648 (C=O), 1535 (O=CN), 750 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.46 (d, *J* = 4.9 Hz, 1H, C(6)*H*), 7.59 (dt, *J* = 1.8, 7.7 Hz, 1H, C(4)*H*), 7.19 (d, *J* = 7.9 Hz, 1H, C(3)*H*), 7.13 (dd, *J* = 5.0, 7.7 Hz, 1H, C(5)*H*), 6.73 (br. s, 1H, N*H*), 4.48 (d, *J* = 4.9 Hz, 2H, CH_2), 2.41 (spt, *J* = 6.8 Hz, 1H, C*H*), 1.13 (d, *J* = 6.9 Hz, 6H, C(CH_3)$_2$); ^{13}C NMR (101 MHz, $CDCl_3$) δ 177.0 (*C*=O), 156.6 (*C*(2)), 149.0 (*C*(6)), 136.8 (*C*(4)), 122.3 (*C*(5)), 122.1 (*C*(3)), 44.3 (*C*H_2), 35.6 (*C*H), 19.6 (C(*C*H_3)$_2$); *m/z* (ESI$^+$) 179 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{10}H_{14}N_2O^+$, ([M+H]$^+$) requires 179.1179, found 179.1178.

N-(1-(Pyridin-3-yl)ethyl)isobutyramide **57**

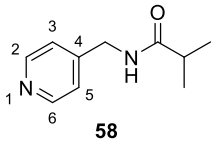


57

Following *General Procedure 2*, isobutyryl chloride (100 µL, 0.95 mmol, 1.05 eq) was treated with 1-(pyridin-3-yl)ethan-1-amine (111 mg, 0.91 mmol, 1.0 eq) and triethylamine (253 µL, 2.0 mmol, 1.8 eq) to afford an off-white powder. A portion of the crude product (25 mg) was purified by reverse phase HPLC to afford *amide* **57** as a white powder (13 mg, 8%). Mp 25.5-27 °C; ν_{max} (FTIR) cm^{-1} 3247 (CONH), 2967 (CH_3), 1639 (C=O), 1545 (O=CN), 710 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.59 (d, *J* = 2.3 Hz, 1H, C(2)*H*), 8.51 (dd, *J* = 1.6, 4.8 Hz, 1H, C(6)*H*), 7.63 (td, *J* = 1.7, 7.9 Hz, 1H, C(4)*H*), 7.27 (ddd, *J* = 0.7, 4.8, 7.9 Hz, 1H, C(5)*H*), 5.88 (br. d, *J* = 5.2 Hz, 1H, N*H*), 5.16 (quin, *J* = 7.2 Hz, 1H, C*H*Me$_A$), 2.38 (spt, *J* = 6.9 Hz, 1H, C(=O)C*H*), 1.52 (d, *J* = 7.0 Hz, 3H, CH_3(Me$_A$)), 1.17 (dd, *J* = 6.9, 9.8 Hz, 6H, C(CH_3)$_2$ (Me$_B$ & Me$_C$)); ^{13}C NMR (101 MHz, $CDCl_3$) δ 176.2 (*C*=O), 148.6 (*C*(6)), 147.8 (*C*(2)), 138.9 (*C*(3)), 133.9 (*C*(4)), 123.5

(*C*(5)), 46.5 (N*C*H), 35.5 (C(=O)*C*H), 21.6 (*C*H$_3$(Me$_A$)), 19.6 (*C*H$_3$(Me$_B$ or Me$_C$)), 19.5 (*C*H$_3$(Me$_B$ or Me$_C$)); *m/z* (ESI$^+$) 193 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{11}H_{16}N_2O^+$, ([M+H]$^+$) requires 193.1335, found 193.1335.

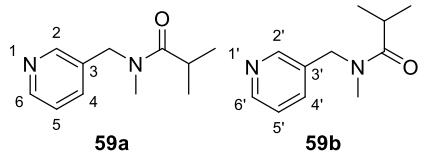
N-(Pyridin-4-ylmethyl)isobutyramide **58**



58

Following *General Procedure 2*, isobutyryl chloride (100 µL, 0.95 mmol, 1.05 eq) was treated with pyridin-4-ylmethanamine (92 µL, 0.91 mmol, 1.0 eq) and triethylamine (253 µL, 2.0 mmol, 1.8 eq) to afford a yellow gum. A portion of the crude product (25 mg) was purified by reverse phase HPLC to afford *amide* **58** as a colourless gum (10 mg, 6%). ν_{max} (FTIR) cm^{-1} 2926 (CH$_3$), 1631 (C=O), 1553 (O=CN), 1428; ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.52 - 8.57 (m, 2H, C(2)*H* & C(6)*H*), 7.18 (d, *J* = 6.0 Hz, 2H, C(3)*H* & C(5)*H*), 6.06 (br. s, 1H, N*H*), 4.47 (d, *J* = 6.1 Hz, 2H, C*H*$_2$), 2.46 (spt, *J* = 6.9 Hz, 1H, C*H*), 1.23 (d, *J* = 6.9 Hz, 6H, C(C*H*$_3$)$_2$); ^{13}C NMR (101 MHz, CDCl$_3$) δ 177.1 (*C*=O), 150.0 (*C*(2) & *C*(6)), 147.7 (*C*(4)), 122.2 (*C*(3) & *C*(5)), 42.3 (*C*H$_2$), 35.6 (*C*H), 19.6 (C(*C*H$_3$)$_2$); *m/z* (ESI$^+$) 179 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{10}H_{14}N_2O^+$, ([M+H]$^+$) requires 179.1179, found 179.1177.

N-Methyl-*N*-(pyridin-3-ylmethyl)isobutyramide **59**

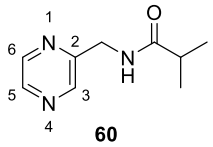


59a **59b**

Following *General Procedure 2*, isobutyryl chloride (100 µL, 0.95 mmol, 1.05 eq) was treated with *N*-methyl-1-(pyridin-3-yl)methanamine (111 µL, 0.91 mmol, 1.0 eq) and triethylamine (253 µL, 2.0 mmol, 1.8 eq) to afford a yellow oil. A portion of the crude product (25 mg) was purified by reverse phase HPLC to afford *amide* **59** as a yellow oil (12 mg, 7%). Rotamers **59a** and **59b** were found to be present in a 3:1 ratio by NMR. ν_{max} (FTIR) cm^{-1} 3278 (CONH), 2968 (CH$_3$), 1649 (C=O), 1541 (O=CN), 1416, 1242; **59a** ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.55 (dd, *J* = 1.0, 4.5 Hz, 1H, C(6)*H*), 8.48 - 8.53 (m, 1H, C(2)*H*), 7.62 (d, *J* = 7.8 Hz, 1H, C(4)*H*), 7.29 (dd, *J* = 4.5, 7.8 Hz, 1H, C(5)*H*), 4.62 (s, 2H, C*H*$_2$), 3.02 (s, 3H,

NC*H*$_3$), 2.88 (spt, *J* = 6.7 Hz, 1H, C*H*), 1.19 (d, *J* = 6.7 Hz, 6H, C(C*H*$_3$)$_2$); **59a** ^{13}C NMR (101 MHz, $CDCl_3$) δ 177.4 (*C*=O), 149.1 (*C*(2)), 148.7 (*C*(6)), 135.9 (*C*(4)), 133.4 (*C*(3)), 123.8 (*C*(5)), 48.6 (*C*H$_2$), 34.8 (NCH$_3$), 30.4 (*C*H), 19.2 (C(*C*H$_3$)$_2$); **59b** ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.59 (d, *J* = 4.5 Hz, 1H, C(6')*H*), 8.48 - 8.53 (m, 1H, C(2')*H*), 7.51 (d, *J* = 7.8 Hz, 1H, C(4')*H*), 7.34 (dd, *J* = 4.5, 7.8 Hz, 1H, C(5')*H*), 4.62 (s, 2H, C*H*$_2$), 2.97 (s, 3H, NC*H*$_3$), 2.88 (spt, *J* = 6.7 Hz, 1H, C*H*), 1.19 (d, *J* = 6.7 Hz, 6H, C(C*H*$_3$)$_2$); **59b** ^{13}C NMR (101 MHz, $CDCl_3$) δ 177.4 (*C*=O), 148.7 (*C*(6')), 148.2 (*C*(2')), 134.0 (*C*(4')), 133.4 (*C*(3')), 123.8 (*C*(5')), 50.8 (*C*H$_2$), 34.8 (NCH$_3$), 30.4 (*C*H), 19.8 (C(*C*H$_3$)$_2$); *m/z* (ESI$^+$) 193 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{11}H_{16}N_2O^+$, ([M+H]$^+$) requires 193.1335, found 193.1334.

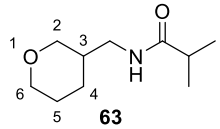
N-(Pyrazin-2-ylmethyl)isobutyramide **60**



60

Following *General Procedure 2*, isobutyryl chloride (100 µL, 0.95 mmol, 1.05 eq) was treated with pyrazin-2-ylmethanamine (99 mg, 0.91 mmol, 1.0 eq) and triethylamine (253 µL, 2.0 mmol, 1.8 eq) to afford a red gum. A portion of the crude product (25 mg) was purified by reverse phase HPLC to afford *amide* **60** as a yellow oil (13 mg, 8%). ν_{max} (FTIR) cm^{-1} 3288 (CONH), 2968 (CH$_3$), 1648 (C=O), 1529 (O=CN), 1017; ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.61 - 8.64 (m, 1H, C(3)*H*), 8.51 - 8.54 (m, 2H, C(5)*H* & C(6)*H*), 6.55 (br. s, 1H, N*H*), 4.64 (d, *J* = 5.3 Hz, 2H, C*H*$_2$), 2.50 (spt, *J* = 6.9 Hz, 1H, C*H*), 1.23 (d, *J* = 6.9 Hz, 6H, C(C*H*$_3$)$_2$); ^{13}C NMR (101 MHz, $CDCl_3$) δ 177.2 (*C*=O), 152.6 (*C*(2)), 144.1 (*C*(3)), 143.7 (*C*(6)), 143.5 (*C*(5)), 42.2 (*C*H$_2$), 35.5 (*C*H), 19.6 (C(*C*H$_3$)$_2$); *m/z* (ESI$^+$) 180 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_9H_{13}N_3O^+$, ([M+H]$^+$) requires 180.1131, found 180.1133.

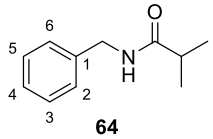
N-((Tetrahydro-2*H*-pyran-3-yl)methyl)isobutyramide **63**



63

Following *General Procedure 2*, isobutyryl chloride (100 µL, 0.95 mmol, 1.05 eq) was treated with (tetrahydro-2H-pyran-3-yl)methanamine (105 mg, 0.91 mmol, 1.0 eq) and triethylamine (253 µL, 2.0 mmol, 1.8 eq) to afford a white powder. A portion of the crude product (25 mg) was purified by reverse phase HPLC to afford *amide* **63** as a white powder (13 mg, 8%). Mp 55.5-57 °C; ν_{max} (FTIR) cm^{-1} 3285 (CONH), 2944 (CH_3), 1644 (C=O), 1551 (O=CN), 1095; ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.53 (br. s, 1H, N*H*), 3.81 - 3.89 (m, 2H, C(2)H_{EQ} & C(6)H_{EQ}), 3.43 (ddd, *J* = 3.6, 9.7, 11.3 Hz, 1H, C(6)H_{AX}), 3.13 - 3.25 (m, 3H, C(2)H_{AX} & NCH_2), 2.36 (spt, *J* = 6.9 Hz, 1H, C(=O)C*H*), 1.76 - 1.88 (m, 2H, C(3)H_{AX} & C(4)H_{AX}), 1.54 - 1.70 (m, 2H, C(5)H_2), 1.22 - 1.34 (m, 1H, C(4)H_{EQ}), 1.17 (d, *J* = 7.0 Hz, 6H, C(CH_3)$_2$); ^{13}C NMR (101 MHz, $CDCl_3$) δ 177.1 (*C*=O), 71.2 (*C*(2)), 68.4 (*C*(6)), 41.4 (NCH_2), 36.5 (*C*(3)), 35.7 (C(=O)C*H*), 27.4 (*C*(4)), 25.1 (*C*(5)), 19.7 (C(CH_3)$_2$); *m/z* (ESI^+) 186 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_{10}H_{19}NO_2^+$, ($[M+H]^+$) requires 186.1489, found 186.1485.

N-Benzylisobutyramide **64**

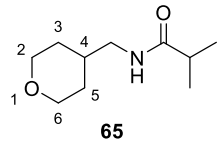


64

Following *General Procedure 2*, isobutyryl chloride (100 µL, 0.95 mmol, 1.05 eq) was treated with benzylamine (99 µL, 0.91 mmol, 1.0 eq) and triethylamine (253 µL, 2.0 mmol, 1.8 eq) to afford a white powder. A portion of the crude product (50 mg) was purified by reverse phase HPLC to afford *amide* **64** as a white powder (28 mg, 17%). Mp 90.5-91.5 °C; ν_{max} (FTIR) cm^{-1} 3283 (CONH), 2971 (CH_3), 1638 (C=O), 1539 (O=CN), 692 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.33 - 7.39 (m, 2H, C(3)*H* & C(5)*H*), 7.26 - 7.32 (m, 3H, C(2)*H*, C(4)*H* & C(6)*H*), 5.78 (br. s, 1H, N*H*), 4.46 (d, *J* = 5.8 Hz, 2H, CH_2), 2.41 (spt, *J* = 6.9 Hz, 1H, C*H*), 1.21 (d, *J* = 6.9 Hz, 6H, C(CH_3)$_2$); ^{13}C NMR (101 MHz, $CDCl_3$) δ 176.8 (*C*=O), 138.5, (*C*(1)), 128.7

(*C*(3) & *C*(5)), 127.8 (*C*(2) & *C*(6)), 127.5 (*C*(4)), 43.5 (*C*H_2), 35.7 (*C*H), 19.6 (C(*C*H_3)$_2$); *m/z* (ESI$^+$) 178 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{11}H_{15}NO^+$, ([M+H]$^+$) requires 178.1226, found 178.1219.

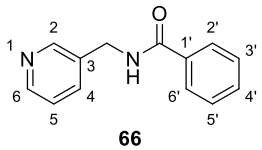
N-((Tetrahydro-2*H*-pyran-4-yl)methyl)isobutyramide **65**



65

Following *General Procedure 2*, isobutyryl chloride (100 µL, 0.95 mmol, 1.05 eq) was treated with (tetrahydro-2H-pyran-4-yl)methanamine (105 mg, 0.91 mmol, 1.0 eq) and triethylamine (253 µL, 2.0 mmol, 1.8 eq) to afford an off-white powder. A portion of the crude product (50 mg) was purified by reverse phase HPLC to afford *amide* **65** as an off-white powder (41 mg, 24%). Mp 62.5-64 °C; ν_{max} (FTIR) cm^{-1} 3288 (CONH), 2918 (CH_2), 1643 (C=O), 1557 (O=CN), 1093, 710; ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.51 (br. s, 1H, N*H*), 3.90 (ddd, *J* = 1.0, 4.0, 11.0 Hz, 2H, C(2)H_{EQ} & C(6)H_{EQ}), 3.30 (dt, *J* = 2.0, 11.8 Hz, 2H, C(2)H_{AX} & C(6)H_{AX}), 3.09 (t, *J* = 6.5 Hz, 2H, NCH_2), 2.28 (spt, *J* = 6.9 Hz, 1H, C(=O)C*H*), 1.62 - 1.76 (m, 1H, C(4)H_{AX}), 1.52 (qd, *J* = 1.8, 12.5 Hz, 2H, C(3)H_{EQ} & C(5)H_{EQ}), 1.24 (dq, *J* = 5.0, 12.5 Hz, 2H, C(3)H_{AX} & C(5)H_{AX}), 1.09 (d, *J* = 7.0 Hz, 6H, C(CH_3)$_2$); ^{13}C NMR (101 MHz, $CDCl_3$) δ 177.1 (*C*=O), 67.6 (*C*(2) & *C*(6)), 45.0 (NCH_2), 35.8 (C(=O)*C*H), 35.3 (*C*(4)), 30.6 (*C*(3) & *C*(5)), 19.7 (C(*C*H_3)$_2$); *m/z* (ESI$^+$) 186 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{10}H_{19}NO_2^+$, ([M+H]$^+$) requires 186.1489, found 186.1484.

N-(Pyridin-3-ylmethyl)benzamide **66**

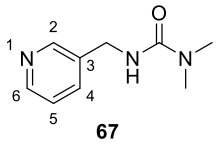


66

Following *General Procedure 2*, benzoyl chloride (120 µL, 1.03 mmol, 1.05 eq) was treated with pyridin-3-ylmethanamine (100 µL, 0.98 mmol, 1.0 eq) and triethylamine (274 µL, 2.0 mmol, 2.0 eq) to afford a red oil. A portion of the crude product (25 mg) was purified by reverse phase HPLC to afford *amide* **66** as a colourless oil (14 mg, 7%). ν_{max} (FTIR) cm^{-1} 3281 (CONH), 1637 (C=O), 1530 (O=CN), 1292, 692 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.60 (s, 1H, C(2)*H*), 8.55 (d, *J* = 4.2 Hz, 1H, C(6)*H*), 7.78 - 7.85 (m, 2H, C(2')*H* & C(6')*H*), 7.73 (d, *J* = 7.8 Hz, 1H, C(4)*H*), 7.50 - 7.56 (m, 1H, C(4')*H*), 7.41 - 7.48 (m, 2H, C(3')*H* & C(5')*H*), 7.29

(dd, *J* = 4.0, 8.2 Hz, 1H, C(5)*H*), 6.73 (br. s, 1H, N*H*), 4.68 (d, *J* = 5.9 Hz, 2H, CH_2); ^{13}C NMR (101 MHz, $CDCl_3$) δ 167.6 (*C*=O), 149.2 (*C*(2)), 149.0 (*C*(6)), 135.8 (*C*(4)), 134.0 (*C*(3) & *C*(1')), 131.8 (*C*(4')), 128.7 (*C*(3') & *C*(5')), 127.0 (*C*(2') & *C*(6')), 123.7 (*C*(5)), 41.5 (*C*H_2); *m/z* (ESI$^+$) 213 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{13}H_{12}N_2O^+$, ([M+H]$^+$) requires 213.1022, found 213.1021.

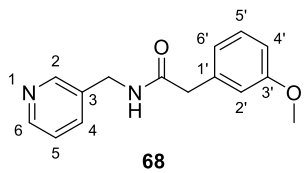
1,1-Dimethyl-3-(pyridin-3-ylmethyl)urea **67**



67

Following *General Procedure 2*, dimethylcarbamic chloride (95 µL, 1.03 mmol, 1.05 eq) was treated with pyridin-3-ylmethanamine (100 µL, 0.98 mmol, 1.0 eq) and triethylamine (274 µL, 2.0 mmol, 2.0 eq) to afford a red oil. A portion of the crude product (25 mg) was purified by reverse phase HPLC to afford *urea* **67** as a yellow powder (16 mg, 9%). Mp 92-93 °C; ν_{max} (FTIR) cm^{-1} 3374 (CONH), 2938 (CH_3), 1627 (C=O), 1515 (O=CN), 715 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.56 (d, *J* = 1.8 Hz, 1H, C(2)*H*), 8.52 (dd, *J* = 1.4, 4.8 Hz, 1H, C(6)*H*), 7.71 (d, *J* = 7.8 Hz, 1H, C(4)*H*), 7.28 (dd, *J* = 4.8, 7.8 Hz, 1H, C(5)*H*), 4.85 (br. s, 1H, N*H*), 4.45 (d, *J* = 5.9 Hz, 2H, CH_2), 2.95 (s, 6H, N(CH_3)$_2$); ^{13}C NMR (101 MHz, $CDCl_3$) δ 158.2 (*C*=O), 149.0 (*C*(2)), 148.5 (*C*(6)), 135.8 (*C*(4)), 135.4 (*C*(3)), 123.6 (*C*(5)), 42.4 (*C*H_2), 36.3 (N(*C*H_3)$_2$); *m/z* (ESI$^+$) 180 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_9H_{13}N_3O^+$, ([M+H]$^+$) requires 180.1131, found 180.1127.

2-(3-Methoxyphenyl)-*N*-(pyridin-3-ylmethyl)acetamide **68**

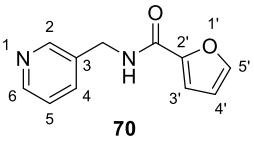


68

Following *General Procedure 2*, 2-(3-methoxyphenyl)acetyl chloride (161 µL, 1.03 mmol, 1.05 eq) was treated with pyridin-3-ylmethanamine (100 µL, 0.98 mmol, 1.0 eq) and triethylamine (274 µL, 2.0 mmol, 2.0 eq) to afford a brown oil. A portion of the crude product (25 mg) was purified by reverse phase HPLC to afford *amide* **68** as a brown oil (10 mg, 4%). ν_{max} (FTIR) cm^{-1} 3270 (CONH), 3053 (CH), 1648 (C=O), 1425 (O=CN), 1255 (C-O), 1030, 710 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.52 (d, *J* = 4.0 Hz, 1H, C(6)*H*), 8.46 (s, 1H, C(2)*H*), 7.56 (d, *J* = 7.7 Hz, 1H, C(4)*H*), 7.29 (t, *J*

= 7.8 Hz, 1H, C(5')*H*), 7.25 (dd, *J* = 4.9, 7.7 Hz, 1H, C(5)*H*), 6.86 (dd, *J* = 2.1, 8.0 Hz, 2H, C(4')*H* & C(6')*H*), 6.78 - 6.83 (m, 1H, C(2')*H*), 5.86 (br. s, 1H, N*H*), 4.44 (d, *J* = 6.1 Hz, 2H, NCH_2), 3.81 (s, 3H, OCH_3), 3.63 (s, 2H, C(=O)CH_2); ^{13}C NMR (101 MHz, $CDCl_3$) δ 171.0 (*C*=O), 160.1 (*C*(3')), 148.9 (*C*(2) & *C*(6)), 136.0 (*C*(1')), 135.3 (*C*(4)), 133.9 (*C*(3)), 130.3 (*C*(5')), 123.6 (*C*(5)), 121.7 (*C*(6')), 115.1 (*C*(2')), 113.0 (*C*(4')), 55.2 (O*C*H_3), 43.8 (C(=O)*C*H_2), 41.1 (N*C*H_2); *m/z* (ESI$^+$) 257 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{15}H_{16}N_2O_2^+$, ([M+H]$^+$) requires 257.1285, found 257.1280.

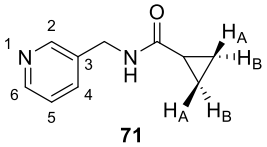
N-(Pyridin-3-ylmethyl)furan-2-carboxamide **70**



70

Following *General Procedure 2*, furan-2-carbonyl chloride (102 µL, 1.03 mmol, 1.05 eq) was treated with pyridin-3-ylmethanamine (100 µL, 0.98 mmol, 1.0 eq) and triethylamine (274 µL, 2.0 mmol, 2.0 eq) to afford a brown solid. A portion of the crude product (25 mg) was purified by reverse phase HPLC to afford *amide* **70** as a colourless oil (13 mg, 7%). ν_{max} (FTIR) cm^{-1} 3273 (CONH), 1641 (C=O), 1523 (O=CN), 1298, 709 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.61 (d, *J* = 2.0 Hz, 1H, C(2)*H*), 8.55 (dd, *J* = 1.5, 4.8 Hz, 1H, C(6)*H*), 7.73 (td, *J* = 1.9, 7.9 Hz, 1H, C(4)*H*), 7.45 (dd, *J* = 0.7, 1.8 Hz, 1H, C(3' or 5')*H*), 7.29 (dd, *J* = 4.8, 7.8 Hz, 1H, C(5)*H*), 7.18 (dd, *J* = 0.7, 3.5 Hz, 1H, C(3' or 5')*H*), 6.83 (br. s, 1H, N*H*), 6.53 (dd, *J* = 1.8, 3.5 Hz, 1H, C(4')*H*), 4.65 (d, *J* = 6.1 Hz, 2H, CH_2); ^{13}C NMR (101 MHz, $CDCl_3$) δ 158.4 (*C*=O), 149.2 (*C*(2)), 149.0 (*C*(6)), 147.6 (*C*(2')), 144.1 (*C*(3' or 5')), 135.8 (*C*(4)), 133.8 (*C*(3)), 123.6 (*C*(5)), 114.8 (*C*(3' or 5')), 112.3 (*C*(4')), 40.6 (*C*H_2); *m/z* (ESI$^+$) 203 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{11}H_{10}N_2O_2^+$, ([M+H]$^+$) requires 203.0815, found 203.0808.

N-(Pyridin-3-ylmethyl)cyclopropanecarboxamide **71**

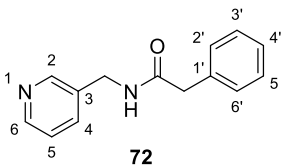


71

Following *General Procedure 2*, cyclopropanecarbonyl chloride (94 µL, 1.03 mmol, 1.05 eq) was treated with pyridin-3-ylmethanamine (100 µL, 0.98 mmol, 1.0 eq) and triethylamine (274 µL, 2.0 mmol, 2.0 eq) to afford a yellow oil. A portion

of the crude product (25 mg) was purified by reverse phase HPLC to afford *amide* **71** as a colourless oil (16 mg, 9%). ν_{max} (FTIR) cm^{-1} 3274 (CONH), 3010 (CH_2), 1644 (C=O), 1542 (O=CN), 710 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.48 - 8.61 (m, 2H, C(2)*H* & C(6)*H*), 7.67 (td, *J* = 1.9, 7.8 Hz, 1H, C(4)*H*), 7.29 (dd, *J* = 4.8, 7.8 Hz, 1H, C(5)*H*), 6.17 (br. s, 1H, N*H*), 4.49 (d, *J* = 6.0 Hz, 2H, NCH_2), 1.40 (tt, *J* = 4.6, 7.9 Hz, 1H, C(=O)C*H*), 0.99 - 1.06 (m, 2H, CH_A), 0.75 - 0.82 (m, 2H, CH_B); ^{13}C NMR (101 MHz, $CDCl_3$) δ 173.7 (*C*=O), 149.1 (*C*(2 or 6)), 148.8 (*C*(2 or 6)), 135.7 (*C*(4)), 134.3 (*C*(3)), 123.6 (*C*(5)), 41.2 (N*C*H_2), 14.7 (C(=O)*C*H), 7.4 (*C*$H_A H_B$); *m/z* (ESI^+) 177 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_{10}H_{12}N_2O^+$, ($[M+H]^+$) requires 177.1022, found 177.1016.

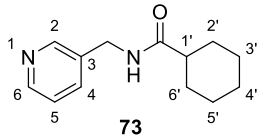
2-Phenyl-*N*-(pyridin-3-ylmethyl)acetamide **72**



72

Following *General Procedure 2*, 2-phenylacetyl chloride (136 µL, 1.03 mmol, 1.05 eq) was treated with pyridin-3-ylmethanamine (100 µL, 0.98 mmol, 1.0 eq) and triethylamine (274 µL, 2.0 mmol, 2.0 eq) to afford a yellow solid. A portion of the crude product (25 mg) was purified by reverse phase HPLC to afford *amide* **72** as a white powder (17 mg, 8%). Mp 100-101.5 °C; ν_{max} (FTIR) cm^{-1} 3277 (CONH), 1642 (C=O), 1549 (O=CN), 1032, 690 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.51 (d, *J* = 4.3 Hz, 1H, C(6)*H*), 8.45 (s, 1H, C(2)*H*), 7.52 - 7.59 (m, 1H, C(4)*H*), 7.21 - 7.41 (m, 6H, C(5)*H*, C(2')*H*, C(3')*H*, C(4')*H*, C(5')*H* & C(6')*H*), 5.84 (br. s, 1H, N*H*), 4.44 (d, *J* = 6.0 Hz, 2H, NCH_2), 3.66 (s, 2H, C(=O)CH_2); ^{13}C NMR (101 MHz, $CDCl_3$) δ 171.1 (*C*=O), 148.9 (*C*(2) & *C*(6)), 135.4 (*C*(4)), 134.5 (*C*(1')), 133.9 (*C*(3)), 129.4 (*C*(2') & *C*(6')), 129.2 (*C*(3') & *C*(5')), 127.5 (*C*(4')), 123.5 (*C*(5)), 43.8 (C(=O)*C*H_2), 41.1 (N*C*H_2); *m/z* (ESI^+) 227 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_{14}H_{14}N_2O^+$, ($[M+H]^+$) requires 227.1179, found 227.1174.

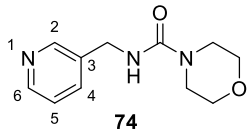
N-(Pyridin-3-ylmethyl)cyclohexanecarboxamide **73**



73

Following *General Procedure 2*, cyclohexanecarbonyl chloride (138 µL, 1.03 mmol, 1.05 eq) was treated with pyridin-3-ylmethanamine (100 µL, 0.98 mmol, 1.0 eq) and triethylamine (274 µL, 2.0 mmol, 2.0 eq) to afford a white powder. A portion of the crude product (25 mg) was purified by reverse phase HPLC to afford *amide* **73** as a white powder (16 mg, 7%). Mp 98-99.5 °C; ν_{max} (FTIR) cm^{-1} 3282 (CONH), 2918 (CH_2), 1639 (C=O), 1545 (O=CN), 711 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.53 (s, 2H, C(2)*H* & C(6)*H*), 7.63 (d, *J* = 7.8 Hz, 1H, C(4)*H*), 7.27 (dd, *J* = 4.7, 8.0 Hz, 1H, C(5)*H*), 5.93 (br. s, 1H, N*H*), 4.47 (d, *J* = 6.0 Hz, 2H, NCH_2), 2.15 (tt, *J* = 3.5, 11.7 Hz, 1H, C(1')*H*), 1.86 - 1.95 (m, 2H, C(2')H_{EQ} & C(6')H_{EQ}), 1.77 - 1.85 (m, 2H, C(3')H_{AX} & C(5')H_{AX}), 1.64 - 1.73 (m, 1H, C(4')H_{EQ}), 1.48 (dq, *J* = 2.9, 12.1 Hz, 2H, C(2')H_{AX} & C(6')H_{AX}), 1.20 - 1.35 (m, 3H, C(3')H_{EQ}, C(4')H_{AX} & C(5')H_{EQ}); ^{13}C NMR (101 MHz, $CDCl_3$) δ 176.1 (*C*=O), 149.1 (*C*(2 or 6)), 148.9 (*C*(2 or 6)), 135.5 (*C*(4)), 134.3 (*C*(3)), 123.6 (*C*(5)), 45.5 (*C*(1')), 40.8 (NCH_2), 29.7 (*C*(2') & *C*(6')), 25.7 (*C*(3') & *C*(5')); *m/z* (ESI$^+$) 219 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{13}H_{18}N_2O^+$, ([M+H]$^+$) requires 219.1492, found 219.1485.

N-(Pyridin-3-ylmethyl)morpholine-4-carboxamide **74**

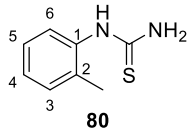


74

Following *General Procedure 2*, morpholine-4-carbonyl chloride (120 µL, 1.03 mmol, 1.05 eq) was treated with pyridin-3-ylmethanamine (100 µL, 0.98 mmol, 1.0 eq) and triethylamine (274 µL, 2.0 mmol, 2.0 eq) to afford a red oil. A portion of the crude product (25 mg) was purified by reverse phase HPLC to afford *urea* **74** as a colourless oil (14 mg, 6%). ν_{max} (FTIR) cm^{-1} 3322 (CONH), 2852 (CH_2), 1624 (C=O), 1529 (O=CN), 1256 (C-O), 1113; ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.56 (d, *J* = 1.8 Hz, 1H, C(2)*H*), 8.53 (dd, *J* = 1.5, 4.8 Hz, 1H, C(6)*H*), 7.70 (td, *J* = 2.0, 7.8 Hz, 1H, C(4)*H*), 7.29 (dd, *J* = 4.5, 7.8 Hz, 1H, C(5)*H*), 4.94 (br. s, 1H, N*H*), 4.47 (d, *J* = 5.6 Hz, 2H, C(3)CH_2), 3.65 - 3.75 (m, 4H, CH_2NCH_2), 3.35 - 3.44 (m, 4H, CH_2OCH_2); ^{13}C NMR

(101 MHz, $CDCl_3$) δ 157.6 (*C*=O), 149.1 (*C*(2)), 148.7 (*C*(6)), 135.8 (*C*(4)), 135.0 (*C*(3)), 123.6 (*C*(5)), 66.4 (CH_2OCH_2), 44.0 (CH_2NCH_2), 42.4 (C(3)CH_2); *m/z* (ESI^+) 222 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_{11}H_{15}N_3O_2^+$, ($[M+H]^+$) requires 222.1237, found 222.1233.

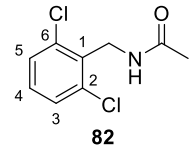
1-(o-tolyl)thiourea **80**



80

Following *General Procedure 3*, 1,1'-thiocarbonyldiimidazole (166 mg, 0.9 mmol, 1.0 eq) was treated with *o*-toluidine (99 µL, 0.9 mmol, 1.0 eq) followed by ammonia methanol solution (133 µL, 7 M, 1.0 eq) to afford *thiourea* **80** as white crystals (102 mg, 66%). Mp 162-163 °C (Lit. (Ubarhande, Thakare *et al.* 2010) 161 °C); ν_{max} (FTIR) cm^{-1} 3422 (NH_2), 3152 (NH), 1615, 1526 (C=S), 1286 (-CSNH-), 759 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.20 (br. s, 1H, N*H*), 7.22 - 7.26 (m, 1H, C(3)*H*), 7.12 - 7.21 (m, 3H, C(4)*H* & C(5)*H* & C(6)*H*), 2.18 (s, 3H, CH_3), (NH_2 not observed); ^{13}C NMR (101 MHz, DMSO-d_6) δ 182.1 (*C*=S), 137.6 (*C*(1) or *C*(2), 134.9 (*C*(1) or *C*(2), 131.0 (*C*(3)), 128.0 (*C*(6)), 127.1 (*C*(4) or *C*(5)), 126.8 (*C*(4) or *C*(5)), 18.1 (CH_3); *m/z* (ESI^+) 167 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_8H_{11}N_2S^+$, ($[M+H]^+$) requires 167.0637, found 167.0633.

N-(2,6-dichlorobenzyl)acetamide **82**

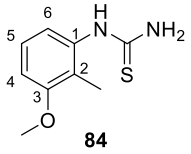


82

Following *General Procedure 2*, acetyl chloride (119 µL, 1.7 mmol, 2.5 eq) was treated with 2,6-dichlorobenzylamine (120 mg, 0.7 mmol, 1.0 eq) and triethylamine (2.9 mL, 20.8 mmol, 30 eq) to afford *amide* **82** as yellow crystals (66 mg, 44%). Mp 182-183 °C; ν_{max} (FTIR) cm^{-1} 3249 (NH), 3077 (CH_3), 1630 (C=O), 1561 (C=O), 1434, 780 (C-Cl or aromatic CH), 766 (C-Cl or aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.02 (br. s, 1H, N*H*), 7.49 (d, *J* = 7.8 Hz, 2H, C(3)*H* & C(5)*H*), 7.32 - 7.40 (m, 1H, C(4)*H*), 4.47 (d, *J* = 4.7 Hz, 2H, CH_2), 1.81 (s, 3H, CH_3); ^{13}C NMR (101 MHz, DMSO-d_6) δ 169.4 (*C*=O), 136.0 (*C*(2) & *C*(6)), 133.9 (*C*(1)), 130.7 (*C*(4)), 129.0 (*C*(3) & *C*(5)), 39.3 (CH_2), 22.7 (CH_3); *m/z* (ESI^+) 218 ($[M(^{35}Cl,^{35}Cl)+H]^+$, 100%), 220

([M(^{35}Cl,^{37}Cl)+H]$^+$, 75%), 222 ([M(^{37}Cl,^{37}Cl)+H]$^+$, 14%); HRMS (ESI$^+$) $C_9H_9{}^{35}Cl_2NONa^+$, ([M+Na]$^+$) requires 239.9953, found 239.9947.

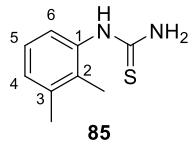
1-(3-Methoxy-2-methylphenyl)thiourea **84**



84

Following *General Procedure 3*, 1,1'-thiocarbonyldiimidazole (100 mg, 0.6 mmol, 1.0 eq) was treated with 2-methyl-3-methoxyaniline (78 mg, 0.6 mmol, 1.0 eq) followed by ammonia methanol solution (81 µL, 7 M, 1.0 eq). The crude reaction mixture was filtered and the filtrate collected to afford *thiourea* **84** as a white powder (79 mg, 71%). Mp 207-209 °C; ν_{max} (FTIR) cm^{-1} 3403 (NH_2), 3139 (NH), 1619, 1461, 1111 (C-O-C), 718 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.26 (br. s, 1H, N*H*), 7.16 (t, *J* = 8.0 Hz, 1H, C(5)*H*), 6.88 (d, *J* = 8.3 Hz, 1H, C(4)*H*), 6.78 (d, *J* = 8.0 Hz, 1H, C(6)*H*), 3.79 (s, 3H, OCH_3), 2.00 (s, 3H, C(2)CH_3), (NH_2 not observed); ^{13}C NMR (101 MHz, DMSO-d_6) δ 181.9 (*C*=S), 158.3 (*C*(3)), 138.1 (*C*(1)), 126.8 (*C*(5)), 123.5 (*C*(2)), 120.2 (*C*(6)), 109.4 (*C*(4)), 56.0 (OCH_3), 11.1 (C(2)CH_3); *m/z* (ESI$^+$) 197 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_9H_{13}N_2OS^+$, ([M+H]$^+$) requires 197.0743, found 197.0742.

1-(2,3-Dimethylphenyl)thiourea **85**

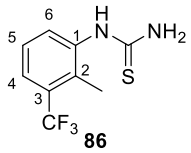


85

Following *General Procedure 3*, 1,1'-thiocarbonyldiimidazole (100 mg, 0.6 mmol, 1.0 eq) was treated with 2,3-dimethylaniline (74 µL, 0.6 mmol, 1.0 eq) followed by ammonia methanol solution (81 µL, 7 M, 1.0 eq) to afford *thiourea* **85** as a white powder (12 mg, 12%). Mp 200-202 °C (Lit. (Vasilev 1966) 172-174 °C); ν_{max} (FTIR) cm^{-1} 3419 (NH_2), 3154 (NH), 1619, 1526 (C=S), 1293 (-CSNH-), 720 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.27 (br. s, 1H, N*H*), 7.06 - 7.10 (m, 2H, C(4)*H* & C(5)*H*), 6.96 - 7.01 (m, 1H, C(6)*H*), 2.25 (s, 3H, C(3)CH_3), 2.07 (s, 3H, C(2)CH_3), (NH_2 not observed); ^{13}C NMR (101 MHz, DMSO-d_6) δ 181.9 (*C*=S), 137.9 (*C*(3)), 137.2 (*C*(1)), 134.0 (*C*(2)), 128.7 (*C*(4)), 126.1 (*C*(5)), 125.8 (*C*(6)), 20.5 (C(3)CH_3), 14.4

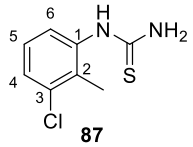
(C(2)*C*H$_3$); *m/z* (ESI$^+$) 181 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_9H_{13}N_2S^+$, ([M+H]$^+$) requires 181.0794, found 181.0790.

1-(2-Methyl-3-(trifluoromethyl)phenyl)thiourea **86**


86

Following *General Procedure 3*, 1,1'-thiocarbonyldiimidazole (100 mg, 0.6 mmol, 1.0 eq) was treated with 2-methyl-3-trifluoromethylaniline (99 mg, 0.6 mmol, 1.0 eq) followed by ammonia methanol solution (81 µL, 7 M, 1.0 eq) to afford *thiourea* **86** as a white powder (44 mg, 33%). Mp 133-134 °C; ν_{max} (FTIR) cm^{-1} 3157 (NH), 1631, 1314, 1122, 1098; ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.36 (br. s, 1H, N*H*), 7.60 (d, *J* = 7.8 Hz, 1H, C(4)*H*), 7.50 (d, *J* = 7.7 Hz, 1H, C(6)*H*), 7.40 (t, *J* = 7.8 Hz, 1H, C(5)*H*), 2.28 (d, *J* (H-F) = 1.2 Hz, 3H, C*H*$_3$), (N*H*$_2$ not observed); ^{13}C NMR (101 MHz, DMSO-d_6) δ 182.7 (*C*=S), 139.7 (*C*(1)), 134.1 (d, *J* (C-F) = 1.5 Hz, *C*(2)), 133.0 (*C*(6)), 128.8 (d, *J* (C-F) = 29.0 Hz, *C*(3)), 127.0 (*C*(5)), 124.9 (d, *J* (C-F) = 273.0 Hz, *C*F$_3$)), 124.5 (d, *J* (C-F) = 5.9 Hz, *C*(4)), 14.1 (d, *J* (C-F) = 2.2 Hz, *C*H$_3$); ^{19}F NMR (376 MHz, DMSO-d_6) δ -59.37 (C*F*$_3$); *m/z* (ESI$^+$) 235 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_9H_{10}F_3N_2S^+$, ([M+H]$^+$) requires 235.0511, found 235.0510.

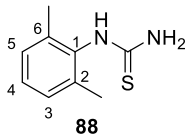
1-(3-Chloro-2-methylphenyl)thiourea **87**


87

Following *General Procedure 3*, 1,1'-thiocarbonyldiimidazole (100 mg, 0.6 mmol, 1.0 eq) was treated with 2-methyl-3-chloroaniline (68 µL, 0.6 mmol, 1.0 eq) followed by ammonia methanol solution (81 µL, 7 M, 1.0 eq) to afford *thiourea* **87** as a white powder (30 mg, 26%). Mp 179-181 °C; ν_{max} (FTIR) cm^{-1} 3431 (NH$_2$), 3147 (NH), 1617, 1523 (C=S), 1294 (-CSNH-), 710 (C-Cl or aromatic CH), 663 (C-Cl or aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.38 (br. s, 1H, N*H*), 7.32 - 7.37 (m, 1H, C(4)*H*), 7.15 - 7.24 (m, 2H, C(5)*H* & C(6)*H*), 2.21 (s, 3H, C*H*$_3$), (N*H*$_2$ not observed); ^{13}C NMR (101 MHz, DMSO-d_6) δ 182.3 (*C*=S), 139.3 (*C*(1)), 134.3 (*C*(3)),

133.7 (*C*(2)), 127.8 (*C*(4)), 127.6 (*C*(5)), 127.5 (*C*(6)), 15.5 (*C*H_3); *m/z* (ESI$^+$) 201 ([M(^{35}Cl)+H]$^+$, 100%), 203 ([M(^{37}Cl)+H]$^+$, 44%); HRMS (ESI$^+$) $C_8H_{10}{}^{35}ClN_2S^+$, ([M+H]$^+$) requires 201.0248, found 201.0243.

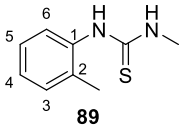
1-(2,6-Dimethylphenyl)thiourea **88**



88

Following *General Procedure 3*, 1,1'-thiocarbonyldiimidazole (100 mg, 0.6 mmol, 1.0 eq) was treated with 2,6-dimethylaniline (70 µL, 0.6 mmol, 1.0 eq) followed by ammonia methanol solution (81 µL, 7 M, 1.0 eq) to afford *thiourea* **88** as a white powder (8 mg, 8%). Mp 174-176 °C (Lit. (Izvekov and Skorikov 1966) 201-202 °C); ν_{max} (FTIR) cm^{-1} 3414 (NH_2), 3143 (NH), 1614, 993, 780 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.16 (br. s, 1H, N*H*), 7.00 - 7.16 (m, 3H, C(3)*H*, C(4)*H* & C(5)*H*), 2.16 (s, 6H, C(2)CH_3 & C(6)CH_3), (NH_2 not observed); ^{13}C NMR (101 MHz, DMSO-d_6) δ 180.8 (*C*=S), 136.6 (*C*(1)), 135.4 (*C*(2) & *C*(6)), 128.7 (*C*(3) & *C*(5)), 127.1 (*C*(4)), 18.1 (*C*H_3); *m/z* (ESI$^+$) 181 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_9H_{13}N_2S^+$, ([M+H]$^+$) requires 181.0794, found 181.0786.

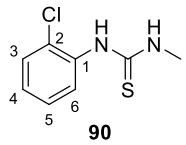
1-Methyl-3-(o-tolyl)thiourea **89**



89

Following *General Procedure 3*, 1,1'-thiocarbonyldiimidazole (100 mg, 0.6 mmol, 1.0 eq) was treated with methanamine hydrochloride (63 mg, 0.9 mmol, 1 eq) and triethylamine (143 µL, 1.0 mmol, 1.1 eq) followed by *o*-toluidine (99 µL, 0.9 mmol, 1.0 eq) to afford *thiourea* **89** as a yellow solid (23 mg, 14%). Mp 161-162 °C (Lit. (Doub, Richardson *et al.* 1958) 163-164 °C); ν_{max} (FTIR) cm^{-1} 3250 (NH), 3150 (NH), 1517 (C=S), 1249 (-CSNH-), 1032, 743 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.42 (br. s, 1H, C(1)N*H*), 7.30 - 7.35 (m, 1H, C(3)*H*), 7.24 - 7.30 (m, 2H, C(4)*H* & C(5)*H*), 7.18 - 7.24 (m, 1H, C(6)*H*), 5.68 (br. s, 1H, N*H* CH_3), 3.13 (d, *J* = 4.4 Hz, 3H, NHCH_3), 2.29 (s, 3H, C(2)CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 182.1 (*C*=S), 135.9 (*C*(2)), 134.3 (*C*(1)), 131.8 (*C*(3)), 128.6 (*C*(4)), 127.6 (*C*(6)), 127.6 (*C*(5)), 32.2 (NH*C*H_3), 17.7 (C(2)*C*H_3); *m/z* (ESI$^+$) 181 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_9H_{13}N_2S^+$, ([M+H]$^+$) requires 181.0794, found 181.0786.

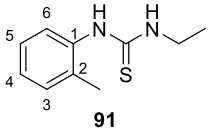
1-(2-Chlorophenyl)-3-methylthiourea **90**



90

Following *General Procedure 3*, 1,1'-thiocarbonyldiimidazole (100 mg, 0.6 mmol, 1.0 eq) was treated with 2-chloroaniline (60 µL, 0.6 mmol, 1.0 eq) and triethylamine (87 µL, 1.0 mmol, 1.1 eq) followed by methanamine hydrochloride (38 mg, 0.6 mmol, 1 eq) to afford *thiourea* **90** as an off-white powder (22 mg, 19%). Mp 158.5-160 °C (Lit. (Yahyazadeh, Aliakbar *et al.* 2010) 164-167 °C); ν_{max} (FTIR) cm^{-1} 3141 (NH), 1514 (-CSNH-), 1474, 1257 (-CSNH-), 1032, 745 (aromatic CH or C-Cl); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.59 (br. s, 1H, C(1)N*H*), 7.52 (dd, *J* = 1.2, 8.0 Hz, 1H, C(3)*H*), 7.47 (br. s, 1H, C(6)*H*), 7.36 (dt, *J* = 1.4, 7.7 Hz, 1H, C(5)*H*), 7.27 (dt, *J* = 1.2, 7.7 Hz, 1H, C(4)*H*), 6.03 (br. s, 1H, N*H*CH_3), 3.18 (d, *J* = 4.7 Hz, 3H, C*H*$_3$); ^{13}C NMR (101 MHz, $CDCl_3$) δ 181.7 (*C*=S), 133.6 (*C*(1)), 130.9 (*C*(3)), 130.1 (*C*(2)), 128.3 (*C*(4)), 128.1 (*C*(5)), 127.0 (*C*(6)), 32.2 (*C*H_3); *m/z* (ESI^+) 201 ([M(^{35}Cl)+H]$^+$, 100%), 203 ([M(^{37}Cl)+H]$^+$, 40%); HRMS (ESI^+) $C_8H_{10}{}^{35}ClN_2S^+$, ([M+H]$^+$) requires 201.0248, found 201.0244.

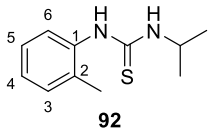
1-Ethyl-3-(o-tolyl)thiourea **91**



91

Following *General Procedure 3*, 1,1'-thiocarbonyldiimidazole (100 mg, 0.6 mmol, 1.0 eq) was treated with *o*-toluidine (60 µL, 0.6 mmol, 1.0 eq) followed by ethanamine tetrahydrofuran solution (284 µL, 2 M, 1.0 eq) to afford *thiourea* **91** as a white powder (82 mg, 74%). Mp 85-87 °C (Lit. (Nair 1965) 83 °C); ν_{max} (FTIR) cm^{-1} 3147 (NH), 1536 (C=S), 1514, 1245 (-CSNH-), 737 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.55 (br. s, 1H, C(1)N*H*), 7.31 - 7.36 (m, 1H, C(3)*H*), 7.25 - 7.30 (m, 2H, C(4)*H* & C(5)*H*), 7.19 - 7.24 (m, 1H, C(6)*H*), 5.62 (br. s., 1H, N*H*CH_2), 3.61 - 3.71 (m, 2H, C*H*$_2$), 2.30 (s, 3H, C(2)C*H*$_3$), 1.17 (t, *J* = 7.2 Hz, 3H, CH_2C*H*$_3$); ^{13}C NMR (101 MHz, $CDCl_3$) δ 180.8 (*C*=S), 135.9 (*C*(2)), 134.3 (*C*(1)), 131.8 (*C*(3)), 128.5 (*C*(4)), 127.6 (*C*(5 or 6)), 127.5 (*C*(5 or 6)), 40.3 (*C*H_2), 17.8 (C(2)*C*H_3), 14.4 (CH_2*C*H_3); *m/z* (ESI^+) 195 ([M+H]$^+$, 100%); HRMS (ESI^+) $C_{10}H_{15}N_2S^+$, ([M+H]$^+$) requires 195.0950, found 195.0944.

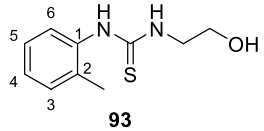
1-Isopropyl-3-(o-tolyl)thiourea **92**



92

Following *General Procedure 3*, 1,1'-thiocarbonyldiimidazole (100 mg, 0.6 mmol, 1.0 eq) was treated with *o*-toluidine (60 µL, 0.6 mmol, 1.0 eq) followed by isopropylamine (49 µL, 0.6 mmol, 1.0 eq) to afford *thiourea* **92** as a white powder (92 mg, 78%). Mp 83-84 °C; ν_{max} (FTIR) cm^{-1} 3234 (NH), 1533 (C=S), 1323, 721 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.42 (br. s, 1H, C(1)N*H*), 7.32 - 7.36 (m, 1H, C(3)*H*), 7.25 - 7.31 (m, 2H, C(4)*H* & C(5)*H*), 7.15 - 7.23 (m, 1H, C(6)*H*), 5.40 (br. d, *J* = 5.1 Hz, 1H, N*H*CH), 4.61 (dd, *J* = 6.6, 14.6 Hz, 1H, C*H*), 2.30 (s, 3H, C(2)CH_3), 1.18 (d, *J* = 6.5 Hz, 6H, CH(CH_3)$_2$); ^{13}C NMR (101 MHz, $CDCl_3$) δ 179.7 (*C*=S), 135.7 (*C*(2)), 134.2 (*C*(1)), 131.8 (*C*(3)), 128.5 (*C*(4)), 127.6 (*C*(5)), 127.4 (*C*(6)), 47.4 (*C*H), 22.4 (CH(*C*H_3)$_2$), 17.7 (C(2)*C*H_3); *m/z* (ESI^+) 209 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_{11}H_{17}N_2S^+$, ($[M+H]^+$) requires 209.1107, found 209.1100.

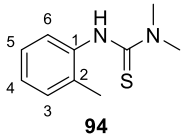
1-(2-Hydroxyethyl)-3-(o-tolyl)thiourea **93**



93

Following *General Procedure 3*, 1,1'-thiocarbonyldiimidazole (166 mg, 0.9 mmol, 1.0 eq) was treated with *o*-toluidine (99 µL, 0.9 mmol, 1.0 eq) followed by 2-aminoethanol (56 µL, 0.9 mmol, 1.0 eq) to afford *thiourea* **93** as white crystals (136 mg, 69%). Mp 1 23-124 °C; ν_{max} (FTIR) cm^{-1} 3376 (OH), 3270 (NH), 3178 (NH), 1541, 1509 (C=S), 1060 (C-O), 740 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.13 (br. s, 1H, C(1)N*H*), 7.39 (br. s, 1H, N*H*CH_2), 7.21 - 7.28 (m, 2H, C(3)*H* & C(6)*H*), 7.09 - 7.21 (m, 2H, C(4)*H* & C(5)*H*), 4.74 (br. s, 1H, O*H*), 3.51 (br. s., 4H, CH_2CH_2OH), 2.17 (s, 3H, CH_3); ^{13}C NMR (101 MHz, DMSO-d_6) δ 181.7 (*C*(8)), 137.6 (*C*(1) or *C*(2)), 134.9 (*C*(1) or *C*(2)), 130.9 (*C*(3)), 128.1 (*C*(6)), 126.8 (*C*(4) or *C*(5)), 126.7 (*C*(4) or *C*(5)), 59.9 (*C*H_2CH_2OH or CH$_2$*C*H_2OH), 47.1 (*C*H_2CH_2OH or CH$_2$*C*H_2OH), 18.1 (*C*H_3); *m/z* (ESI^+) 211 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_{10}H_{15}N_2OS^+$, ($[M+H]^+$) requires 211.0900, found 211.0896.

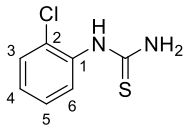
1,1-Dimethyl-3-(o-tolyl)thiourea **94**



94

Following *General Procedure 3*, 1,1'-thiocarbonyldiimidazole (100 mg, 0.6 mmol, 1.0 eq) was treated with *o*-toluidine (60 µL, 0.6 mmol, 1.0 eq) followed by dimethylamine hydrochloride (46 mg, 0.6 mmol, 1.0 eq) and triethylamine (87 µL, 0.6 mmol, 1.1 eq) to afford *thiourea* **94** as a white powder (72 mg, 65%). Mp 137-138 °C (Lit. (Ho, Yang *et al.* 1960) 137-138.5 °C); ν_{max} (FTIR) cm^{-1} 3250 (NH), 1524 (C=S), 1332, 724 (724 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.79 (br. s, 1H, N*H*), 7.18 - 7.23 (m, 1H, C(3)*H*), 7.11 - 7.18 (m, 2H, C(4)*H* & C(5)*H*), 7.02 - 7.07 (m, 1H, C(6)*H*), 3.27 (s, 6H, N(CH_3)$_2$), 2.17 (s, 3H, C(2)CH_3); ^{13}C NMR (101 MHz, DMSO-d_6) δ 181.9 (*C*=S), 140.3 (*C*(1)), 136.2 (*C*(2)), 130.4 (*C*(3)), 129.5 (*C*(6)), 126.7 (*C*(4)), 126.3 (*C*(5)), 41.2 (N(CH_3)$_2$), 18.4 (C(2)CH_3); *m/z* (ESI$^+$) 195 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{10}H_{15}N_2S^+$, ([M+H]$^+$) requires 195.0950, found 195.0944.

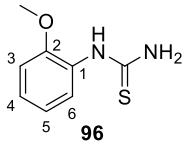
1-(2-Chlorophenyl)thiourea **95**



95

Following *General Procedure 3*, 1,1'-thiocarbonyldiimidazole (100 mg, 0.6 mmol, 1.0 eq) was treated with 2-chloroaniline (60 µL, 0.6 mmol, 1.0 eq) followed by ammonia methanol solution (81 µL, 7 M, 1.0 eq) to afford *thiourea* **95** as a white powder (31 mg, 29%). Mp 114.5-116 °C (Lit. (Hay, Turcotte *et al.* 2010) 141-143 °C); ν_{max} (FTIR) cm^{-1} 3413 (NH_2), 3122 (NH), 1614, 1511 (C=S), 1293 (-CSNH-), 723 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.31 (s, 1H, N*H*), 7.65 (dd, *J* = 1.5, 8.0 Hz, 1H, C(6)*H*), 7.49 (dd, *J* = 1.4, 8.0 Hz, 1H, C(3)*H*), 7.33 (dt, *J* = 1.5, 7.7 Hz, 1H, C(5)*H*), 7.24 (dt, *J* = 1.4, 7.5 Hz, 1H, C(4)*H*), (NH_2 not observed); ^{13}C NMR (101 MHz, DMSO-d_6) δ 182.5 (*C*=S), 136.5 (*C*(1)), 129.9 (*C*(3)), 129.7 (*C*(6)), 129.6 (*C*(2)), 127.7 (*C*(4)), 127.6 (*C*(5)); *m/z* (ESI$^+$) 187 ([M(^{35}Cl)+H]$^+$, 100%), 189 ([M(^{37}Cl)+H]$^+$, 33%); HRMS (ESI$^+$) $C_7H_8{}^{35}ClN_2S^+$, ([M+H]$^+$) requires 187.0091, found 187.0087.

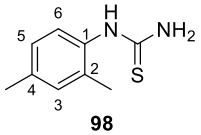
1-(2-Methoxyphenyl)thiourea **96**



96

Following *General Procedure 3*, 1,1′-thiocarbonyldiimidazole (100 mg, 0.6 mmol, 1.0 eq) was treated with 2-methoxyaniline (64 µL, 0.6 mmol, 1.0 eq) followed by ammonia methanol solution (81 µL, 7 M, 1.0 eq) to afford *thiourea* **96** as a white powder (20 mg, 19%). Mp 155.5-157 °C (Lit. (Hay, Turcotte *et al.* 2010) 156 °C); ν_{max} (FTIR) cm^{-1} 3431 (NH_2), 3327, 3159 (NH), 1596, 1526 (C=S), 1253 (-CSNH-), 1011 (COC), 765 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.02 (br. s, 1H, C(1)N*H*), 7.82 (br. d, *J* = 7.5 Hz, 1H, C(6)*H*), 7.10 - 7.17 (m, 1H, C(4)*H*), 7.04 (dd, *J* = 1.2, 8.2 Hz, 1H, C(3)*H*), 6.91 (dt, *J* = 1.3, 7.6 Hz, 1H, C(5)*H*), 3.82 (s, 3H, OCH_3), (NH_2 not observed); ^{13}C NMR (101 MHz, DMSO-d_6) δ 181.8, (*C*=S), 152.2 (*C*(2)), 128.0 (*C*(1)), 126.2 (*C*(4)), 126.0 (*C*(6)), 120.3 (*C*(5)), 111.9 (*C*(3)), 56.0 (O*C*H_3); *m/z* (ESI^+) 183 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_8H_{11}N_2OS^+$, ($[M+H]^+$) requires 183.0587, found 183.0591.

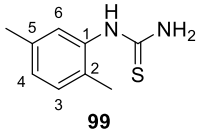
1-(2,4-Dimethylphenyl)thiourea **98**



98

Following *General Procedure 3*, 1,1′-thiocarbonyldiimidazole (100 mg, 0.6 mmol, 1.0 eq) was treated with 2,4-dimethylaniline (70 µL, 0.6 mmol, 1.0 eq) followed by ammonia methanol solution (81 µL, 7 M, 1.0 eq) to afford *thiourea* **98** as a white powder (33 mg, 32%). Mp 185-187 °C (Lit. (Karba 1967) 188-190 °C); ν_{max} (FTIR) cm^{-1} 3421 (NH_2), 3156 (NH), 1614, 1528 (C=S), 1298 (-CSNH-); ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.14 (br. s, 1H, N*H*), 7.02 - 7.08 (m, 2H, C(3 or 5)*H* & C(6)*H*), 6.97 - 7.02 (m, 1H, C(3 or 5)*H*), 2.26 (s, 3H, C(2)CH_3 or C(4)CH_3), 2.14 (s, 3H, C(2)CH_3 or C(4)CH_3), (NH_2 not observed); ^{13}C NMR (101 MHz, DMSO-d_6) δ 182.1 (*C*=S), 136.3 (*C*(2 or 4)), 134.9 (*C*(1)), 134.8 (*C*(2 or 4)), 131.6 (*C*(3 or 5)), 127.9 (*C*(6)), 127.4 (*C*(3 or 5)), 21.0 (C(2)*C*H_3 or C(4)*C*H_3), 18.0 (C(2)*C*H_3 or C(4)*C*H_3); *m/z* (ESI^+) 181 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_9H_{13}N_2S^+$, ($[M+H]^+$) requires 181.0794, found 181.0790.

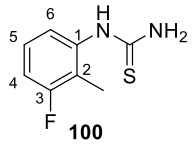
1-(2,5-Dimethylphenyl)thiourea **99**



99

Following *General Procedure 3*, 1,1'-thiocarbonyldiimidazole (100 mg, 0.6 mmol, 1.0 eq) was treated with 2,5-dimethylaniline (71 µL, 0.6 mmol, 1.0 eq) followed by ammonia methanol solution (81 µL, 7 M, 1.0 eq) to afford *thiourea* **99** as a white powder (45 mg, 44%). Mp 143-144 °C (Lit. (Karba 1967) 145 °C); ν_{max} (FTIR) cm^{-1} 3424 (NH_2), 3152 (NH), 1617, 1525 (C=S), 1295 (-CSNH-), 818 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.18 (br. s, 1H, N*H*), 7.12 (d, *J* = 7.5 Hz, 1H, C(3)*H*), 7.00 (s, 2H, C(3)*H* & C(6)*H*), 2.26 (s, 3H, C(5)CH_3), 2.13 (s, 3H, C(2)CH_3), (NH_2 not observed); ^{13}C NMR (101 MHz, DMSO-d_6) δ 181.9 (*C*=S), 137.2 (*C*(1)), 135.9 (*C*(5)), 131.7 (*C*(2)), 130.8 (*C*(3)), 128.4 (*C*(4 or 6)), 127.8 (*C*(4 or 6)), 20.9 (C(5)*C*H_3), 17.6 (C(2)*C*H_3); *m/z* (ESI$^+$) 181 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_9H_{13}N_2S^+$, ([M+H]$^+$) requires 181.0794, found 181.0787.

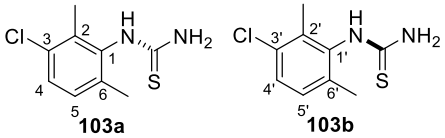
1-(3-Fluoro-2-methylphenyl)thiourea **100**



100

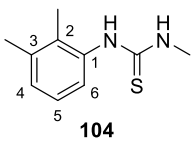
Following *General Procedure 3*, 1,1'-thiocarbonyldiimidazole (100 mg, 0.6 mmol, 1.0 eq) was treated with 2-methyl-3-fluoroaniline (65 µL, 0.6 mmol, 1.0 eq) followed by ammonia methanol solution (81 µL, 7 M, 1.0 eq) to afford *thiourea* **100** as a white powder (47 mg, 45%). Mp 169-170 °C; ν_{max} (FTIR) cm^{-1} 3444 (NH_2), 3128 (NH), 2923, 1619, 1513 (C=S), 1462, 713 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.34 (br. s, 1H, N*H*), 7.18 - 7.25 (m, 1H, C(5)*H*), 7.03 - 7.11 (m, 2H, C(4)*H* & C(6)*H*), 2.09 (d, *J* (H-F) = 2.1 Hz, 3H, CH_3), (NH_2 not observed); ^{13}C NMR (101 MHz, DMSO-d_6) δ 182.3 (*C*=S), 161.4 (d, *J* (C-F) = 255.0 Hz, *C*(3)), 139.4 (d, *J* (C-F) = 6.6 Hz, *C*(1)), 127.3 (d, *J* (C-F) = 10.3 Hz, *C*(5)), 124.1 (d, *J* (C-F) = 2.9 Hz, *C*(6)), 122.5 (d, *J* (C-F) = 17.6 Hz, *C*(2)), 113.6 (d, *J* (C-F) = 22.0 Hz, *C*(4)), 10.2 (d, *J* (C-F) = 4.4 Hz, *C*H_3); ^{19}F NMR (376 MHz, DMSO-d_6) δ -115.49 (C(4)*F*); *m/z* (ESI$^+$) 185 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_8H_{10}FN_2S^+$, ([M+H]$^+$) requires 185.0543, found 185.0539.

1-(3-Chloro-2,6-dimethylphenyl)thiourea **103**



Following *General Procedure 3*, 1,1'-thiocarbonyldiimidazole (100 mg, 0.6 mmol, 1.0 eq) was treated with 3-chloro-2,6-dimethylanilne (88 mg, 0.6 mmol, 1.0 eq) followed by ammonia methanol solution (81 µL, 7 M, 1.0 eq) to afford *thiourea* **103** as a white powder (44 mg, 36%). Rotamers **103a** and **103b** were found to be present in a 3:2 ratio by NMR. Absolute stereochemistry was not determined. Mp 156-157.5 °C; ν_{max} (FTIR) cm^{-1} 3417 (NH_2), 3128 (NH), 1614, 1517 (C=S), 1286 (-CSNH-), 1070, 809 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6, 343 K) δ 9.00 (br. s, 1H, N*H*), 7.27 (d, *J* = 8.2 Hz, 1H, C(4)*H*), 7.11 (d, *J* = 8.2 Hz, 1H, C(5)*H*), 7.05 (br. s, 2H, NH_2), 2.23 (s, 3H, C(2)CH_3), 2.18 (s, 3H, C(6)H_3); **103a** ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.29 (br. s, 1H, N*H*), 7.31 (d, *J* = 8.2 Hz, 1H, C(4)*H*), 7.14 (d, *J* = 8.2 Hz, 1H, C(5)*H*), 2.19 (s, 3H, C(2)CH_3), 2.14 (s, 3H, C(6)CH_3), (NH_2 not observed); **103b** ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.98 (br. s, 1H, N*H*), 7.26 (d, *J* = 8.1 Hz, 1H, C(4')*H*), 7.09 (d, *J* = 8.1 Hz, 1H, C(5')*H*), 2.19 (s, 3H, C(2')CH_3), 2.14 (s, 3H, C(6')CH_3), (NH_2 not observed); **103a** ^{13}C NMR (101 MHz, DMSO-d_6) δ 181.0 (*C*=S), 136.9 (*C*(1)), 135.9 (*C*(2)), 134.8 (*C*(6)), 131.7 (*C*(3)), 129.4 (*C*(5)), 128.5 (*C*(4)), 18.0 (C(6)*C*H$_3$), 15.7 (C(2)*C*H$_3$); **103b** ^{13}C NMR (101 MHz, DMSO-d_6) δ 183.7 (*C*=S), 139.4 (*C*(1')), 136.4 (*C*(2')), 135.2 (*C*(6')), 131.0 (*C*(3')), 128.7 (*C*(5')), 127.6 (*C*(4')), 18.3 (C(6')*C*H$_3$), 15.8 (C(2')*C*H$_3$); *m/z* (ESI$^+$) 215 ([M(^{35}Cl)+H]$^+$, 100%), 217 ([M(^{37}Cl)+H]$^+$, 45%); HRMS (ESI$^+$) $C_9H_{12}{}^{35}ClN_2S^+$, ([M+H]$^+$) requires 215.0404, found 215.0411.

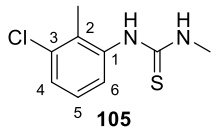
1-(2,3-Dimethylphenyl)-3-methylthiourea **104**



Following *General Procedure 3*, 1,1'-thiocarbonyldiimidazole (100 mg, 0.6 mmol, 1.0 eq) was treated with 2,3-dimethylaniline (74 µL, 0.6 mmol, 1.0 eq) followed by methanamine hydrochloride (38 mg, 0.6 mmol, 1.0 eq) and triethylamine (87 µL, 0.62 mmol, 1.1 eq) to afford *thiourea* **104** as a white powder (37 mg, 33%). Mp 173-174 °C (Lit. (Berger

and Borgaes 1968) 177-178 °C); ν_{max} (FTIR) cm^{-1} 3178 (NH), 1514 (-CSNH-), 1259 (C=S), 1222 (-NC(=S)N-), 752 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.45 (br. s, 1H, C(1)N*H*), 7.06 - 7.14 (m, 2H, C(4)*H* & C(5)*H*), 6.98 (dd, *J* = 2.1, 6.9 Hz, 1H, C(6)*H*), 5.53 (br. s, 1H, N*H*Me), 3.04 (d, *J* = 4.8 Hz, 3H, NHCH_3), 2.25 (s, 3H, C(3)CH_3), 2.11 (s, 3H, C(2)CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 182.1 (*C*=S), 139.3 (*C*(3)), 134.8 (*C*(2)), 134.1 (*C*(1)), 130.2 (*C*(4)), 126.9 (*C*(5)), 125.4 (*C*(6)), 32.1 (NH*C*H_3), 20.4 (C(3)*C*H_3), 14.1 (C(2)*C*H_3); *m/z* (ESI^+) 195 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_{10}H_{15}N_2S^+$, ($[M+H]^+$) requires 195.0950, found 195.0955.

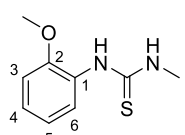
1-(3-Chloro-2-methylphenyl)-3-methylthiourea **105**



105

Following *General Procedure 3*, 1,1'-thiocarbonyldiimidazole (100 mg, 0.6 mmol, 1.0 eq) was treated with 3-chloro-2-methylaniline (68 µL, 0.6 mmol, 1.0 eq) followed by methanamine hydrochloride (38 mg, 0.6 mmol, 1.0 eq) and triethylamine (87 µL, 0.62 mmol, 1.1 eq) to afford *thiourea* **105** as a white powder (53 mg, 44%). Mp 157-158.5 °C; ν_{max} (FTIR) cm^{-1} 3148 (NH), 1498 (-CSNH-), 1256 (C=S), 1221 (-NC(=S)N-), 798 (aromatic CH or C-Cl), 759 (aromatic CH or C-Cl); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.62 (br. s, 1H, C(1)N*H*), 7.42 (d, *J* = 8.0 Hz, 1H, C(6)*H*), 7.23 (t, *J* = 7.9 Hz, 1H, C(5)*H*), 7.17 (d, *J* = 7.6 Hz, 1H, C(4)*H*), 5.62 (br. s, 1H, N*H*Me), 3.14 (d, *J* = 4.8 Hz, 3H, NHCH_3), 2.35 (s, 3H, C(2)CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 182.1 (*C*=S), 136.4 (*C*(1)), 135.6 (*C*(3)), 134.8 (*C*(2)), 129.6 (*C*(6)), 127.8 (*C*(5)), 126.3 (*C*(4)), 32.3 (NH*C*H_3), 15.2 (C(2)*C*H_3); *m/z* (ESI^+) 215 ($[M(^{35}Cl)+H]^+$, 100%), 217 ($[M(^{37}Cl)+H]^+$, 40%); HRMS (ESI^+) $C_9H_{12}{}^{35}ClN_2S^+$, ($[M+H]^+$) requires 215.0404, found 215.0408.

1-(2-Methoxyphenyl)-3-methylthiourea **106**

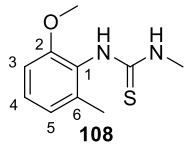


106

Following *General Procedure 3*, 1,1'-thiocarbonyldiimidazole (100 mg, 0.6 mmol, 1.0 eq) was treated with 2-methoxyaniline (64 µL, 0.6 mmol, 1.0 eq) followed by methanamine hydrochloride (38 mg, 0.6 mmol, 1.0 eq) and triethylamine (87 µL,

0.62 mmol, 1.1 eq) to afford *thiourea* **106** as a white powder (52 mg, 47%). Mp 141.5-142.5 °C (Lit. (Yahyazadeh, Aliakbar *et al.* 2010) 137-139 °C); ν_{max} (FTIR) cm^{-1} 3168 (NH), 1499 (-CSNH-), 1245 (C=S), 1025 (COC), 752 (aromatic CH), 759 (aromatic CH or C-Cl); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.55 (br. s, 1H, C(1)N*H*), 7.32 (br. s, 1H, C(6)*H*), 7.26 (dt, *J* = 1.7, 7.9 Hz, 1H, C(4)*H*), 6.96 - 7.04 (m, 2H, C(3)*H* & C(5)*H*), 6.18 (br. s, 1H, N*H*CH_3), 3.87 (s, 3H, OC*H*$_3$), 3.17 (d, *J* = 4.7 Hz, 3H, NHC*H*$_3$); ^{13}C NMR (101 MHz, $CDCl_3$) δ 181.7 (*C*=S), 152.7 (*C*(2)), 127.7 (*C*(4)), 125.3 (*C*(1)), 125.0 (*C*(6)), 121.0 (*C*(5)), 112.1 (*C*(3)), 55.7 (O*C*H_3), 32.2 (NH*C*H_3); *m/z* (ESI^+) 197 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_9H_{13}N_2OS^+$, ($[M+H]^+$) requires 197.0743, found 197.0754.

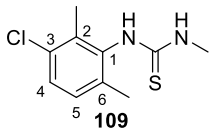
1-(2-Methoxy-6-methylphenyl)-3-methylthiourea **108**



108

Following *General Procedure 3*, 1,1′-thiocarbonyldiimidazole (100 mg, 0.6 mmol, 1.0 eq) was treated with 2-methoxy-6-methylaniline (78 mg, 0.6 mmol, 1.0 eq) followed by methanamine hydrochloride (38 mg, 0.6 mmol, 1.0 eq) and triethylamine (87 µL, 0.62 mmol, 1.1 eq) to afford *thiourea* **108** as an off-white powder (57 mg, 48%). Mp 138.5-139.5 °C; ν_{max} (FTIR) cm^{-1} 3361 (NH), 1473 (-CSNH-), 1280 (C=S), 1242 (-NC(=S)N-), 1082 (COC), 774 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.17 (t, *J* = 8.0 Hz, 1H, C(4)*H*), 7.07 (br. s, 1H, C(1)N*H*), 6.81 (d, *J* = 7.7 Hz, 1H, C(5)*H*), 6.75 (d, *J* = 8.2 Hz, 1H, C(3)*H*), 5.49 (br. s, 1H, N*H*Me), 3.74 (s, 3H, OC*H*$_3$), 3.04 (d, *J* = 4.8 Hz, 3H, NHC*H*$_3$), 2.19 (s, 3H, C(6)C*H*$_3$); ^{13}C NMR (101 MHz, $CDCl_3$) δ 182.4 (*C*=S), 155.5 (*C*(2)), 138.1 (*C*(6)), 129.3 (*C*(4)), 122.9 (*C*(5)), 109.5 (*C*(3)), 55.8 (O*C*H_3), 32.2 (NH*C*H_3), 17.8 (C(6)*C*H_3), (*C*(1) not observed); *m/z* (ESI^+) 211 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_{10}H_{15}N_2OS^+$, ($[M+H]^+$) requires 211.0900, found 211.0901.

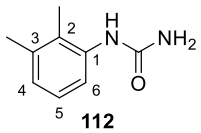
1-(3-Chloro-2,6-dimethylphenyl)-3-methylthiourea **109**



109

Following *General Procedure 3*, 1,1′-thiocarbonyldiimidazole (100 mg, 0.6 mmol, 1.0 eq) was treated with 3-chloro-2,6-dimethylaniline (88 mg, 0.6 mmol, 1.0 eq) followed by methanamine hydrochloride (38 mg, 0.6 mmol, 1.0 eq) and triethylamine (87 µL, 0.62 mmol, 1.1 eq) to afford *thiourea* **109** as a white powder (16 mg, 12%). Mp 162-163 °C; ν_{max} (FTIR) cm^{-1} 3165 (NH), 1507 (-CSNH-), 1257 (C=S), 1057, 824 (aromatic CH), 715 (C-Cl); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.57 (br. s, 1H, C(1)N*H*), 7.34 (d, *J* = 8.3 Hz, 1H, C(4)*H*), 7.12 (d, *J* = 8.3 Hz, 1H, C(5)*H*), 5.36 (br. s, 1H, N*H*Me), 3.12 (d, *J* = 4.7 Hz, 3H, NHCH_3), 2.33 (s, 3H, C(2)CH_3), 2.26 (s, 3H, C(6)CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 181.9 (*C*=S), 136.1 (*C*(2)), 135.8 (*C*(6)), 133.2 (*C*(3)), 129.7 (*C*(4)), 129.3 (*C*(5)), 32.2 (NH*C*H_3), 18.1 (C(6)*C*H_3), 15.5 (C(2)*C*H_3); *m/z* (ESI^+) 229 ([M(^{35}Cl)+H]$^+$, 100%), 231 ([M(^{37}Cl)+H]$^+$, 35%); HRMS (ESI^+) $C_{10}H_{15}{}^{35}ClN_2S^+$, ([M+H]$^+$) requires 229.0561, found 229.0560.

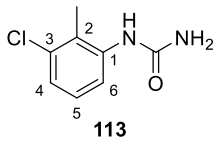
1-(2,3-Dimethylphenyl)urea **112**



112

Following *General Procedure 4*, 2,3-dimethylaniline (96 µL, 0.7 mmol, 1.0 eq) was treated with potassium cyanate and acetic acid to afford *urea* **112** as a pink powder (59 mg, 49%). Mp 194-195.5 °C (Lit. (Friedrich and Pohloudek-Fabini 1967) 219-221 °C); ν_{max} (FTIR) cm^{-1} 3436 (NH_2), 3307 (NH), 1646 (C=O), 1544, 1349, 763 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.70 (s, 1H, N*H*), 7.48 (d, *J* = 8.1 Hz, 1H, C(6)*H*), 6.97 (t, *J* = 7.8 Hz, 1H, C(5)*H*), 6.83 (d, *J* = 7.3 Hz, 1H, C(4)*H*), 5.90 (s, 2H, NH_2), 2.23 (s, 3H, C(3)CH_3), 2.08 (s, 3H, C(2)CH_3); ^{13}C NMR (101 MHz, DMSO-d_6) δ 156.8 (*C*=O), 138.2 (*C*(1)), 136.8 (*C*(3)), 127.5 (*C*(2)), 125.5 (*C*(5)), 124.8 (*C*(4)), 120.8 (*C*(6)), 20.8 (C(3)*C*H_3), 14.0 (C(2)*C*H_3); *m/z* (ESI^+) 165 ([M+H]$^+$, 100%); HRMS (ESI^+) $C_9H_{12}N_2ONa^+$, ([M+Na]$^+$) requires 187.0842, found 187.0847.

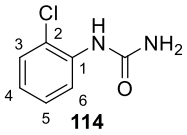
1-(3-Chloro-2-methylphenyl)urea **113**



113

Following *General Procedure 4*, 3-chloro-2-methylaniline (60 mg, 0.7 mmol, 1.0 eq) was treated with potassium cyanate and acetic acid to afford *urea* **113** as a white powder (71 mg, 52%). Mp 269-272 °C; ν_{max} (FTIR) cm^{-1} 3424 (NH_2), 3307 (NH), 1647 (C=O), 1545, 1349, 777 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.90 (br. s, 1H, C(1)N*H*), 7.72 (dd, *J* = 1.6, 7.7 Hz, 1H, C(6)*H*), 7.05 - 7.14 (m, 2H, C(4)*H* & C(5)*H*), 6.08 (br. s, 2H, NH_2), 2.24 (s, 3H, CH_3); ^{13}C NMR (101 MHz, DMSO-d_6) δ 156.5 (*C*=O), 140.2 (*C*(1) or *C*(3)), 133.8 (*C*(1) or *C*(3)), 127.2 (*C*(5)), 126.2 (*C*(2)), 123.4 (*C*(4)), 121.0 (*C*(6)), 15.1 (*C*H_3); *m/z* (ESI$^+$) 185 ([M(^{35}Cl)+H]$^+$, 100%), 187 ([M(^{37}Cl)+H]$^+$, 30%); HRMS (ESI$^+$) $C_8H_{10}{}^{35}ClN_2O^+$, ([M+H]$^+$) requires 185.0476, found 185.0481.

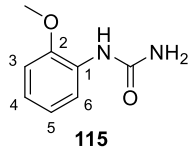
1-(2-Chlorophenyl)urea **114**



114

Following *General Procedure 4*, 2-chloroaniline (78 µL, 0.7 mmol, 1.0 eq) was treated with potassium cyanate and acetic acid to afford *urea* **114** as a white powder (53 mg, 42%). Mp 180-181 °C (Lit. (Gold-Aubert and Toribio 1963) 175 °C); ν_{max} (FTIR) cm^{-1} 3426 (NH_2), 3310 (NH), 1648 (C=O), 1538, 1352, 732 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.14 (dd, *J* = 1.5, 8.4 Hz, 1H, C(6)*H*), 8.03 (br. s, 1H, C(1)N*H*), 7.39 (dd, *J* = 1.5, 8.0 Hz, 1H, C(3)*H*), 7.23 (dt, *J* = 1.5, 7.8 Hz, 1H, C(5)*H*), 6.95 (dt, *J* = 1.5, 7.8 Hz, 1H, C(4)*H*), 6.38 (br. s, 2H, NH_2); ^{13}C NMR (101 MHz, DMSO-d_6) δ 156.0 (*C*=O), 137.2 (*C*(1)), 129.5 (*C*(3)), 127.8 (*C*(5)), 123.0 (*C*(4)), 121.8 (*C*(2)), 121.5 (*C*(6)); *m/z* (ESI$^+$) 171 ([M(^{35}Cl)+H]$^+$, 100%), 173 ([M(^{37}Cl)+H]$^+$, 30%); HRMS (ESI$^+$) $C_7H_7{}^{35}ClN_2ONa^+$, ([M+Na]$^+$) requires 193.0139, found 193.0141.

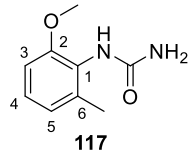
1-(2-Methoxyphenyl)urea **115**



115

Following *General Procedure 4*, 2-methoxyaniline (83 µL, 0.7 mmol, 1.0 eq) was treated with potassium cyanate and acetic acid to afford *urea* **115** as a white powder (56 mg, 46%). Mp 141-143 °C; ν_{max} (FTIR) cm^{-1} 3470 (NH_2), 3326 (NH), 1664 (C=O), 1530, 1457, 1251, 736 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.08 (dd, *J* = 1.9, 7.8 Hz, 1H, C(6)*H*), 7.93 (br. s, 1H, C(1)N*H*), 6.95 (dd, *J* = 1.6, 7.9 Hz, 1H, C(3)*H*), 6.80 - 6.90 (m, 2H, C(4)*H* & C(5)*H*), 6.18 (br. s, 2H, NH_2), 3.83 (s, 3H, CH_3); ^{13}C NMR (101 MHz, DMSO-d_6) δ 156.5 (*C*=O), 147.8 (*C*(2)), 129.9 (*C*(1)), 121.5 (*C*(4)), 120.9 (*C*(5)), 118.6 (*C*(6)), 111.0 (*C*(3)), 56.1 (*C*H_3); *m/z* (ESI^+) 167 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_8H_{10}N_2O_2Na^+$, ($[M+Na]^+$) requires 189.0635, found 189.0632.

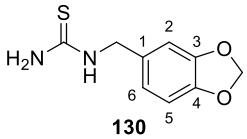
1-(2-Methoxy-6-methylphenyl)urea **117**



117

Following *General Procedure 4*, 2-methoxy-6-methylaniline (60 mg, 0.7 mmol, 1.0 eq) was treated with potassium cyanate and acetic acid to afford *urea* **117** as a white powder (25 mg, 19%). Mp 197.5-199 °C; ν_{max} (FTIR) cm^{-1} 3453 (NH_2), 3262 (NH), 1654 (C=O), 1535, 1079 (COC), 767 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.32 (br. s, 1H, C(1)N*H*), 7.05 (t, *J* = 7.9 Hz, 1H, C(4)*H*), 6.82 (d, *J* = 8.1 Hz, 1H, C(3)*H*), 6.78 (d, *J* = 7.6 Hz, 1H, C(5)*H*), 5.73 (br. s, 2H, NH_2), 3.75 (s, 3H, OCH_3), 2.15 (s, 3H, C(6)CH_3); ^{13}C NMR (101 MHz, DMSO-d_6) δ 157.1 (*C*=O), 154.8 (*C*(2)), 137.0 (*C*(6)), 126.7 (*C*(1)), 126.1 (*C*(4)), 122.5 (*C*(5)), 109.1 (*C*(3)), 55.9 (O*C*H_3), 18.7 (C(6)*C*H_3); *m/z* (ESI^+) 181 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_9H_{12}N_2O_2Na^+$, ($[M+Na]^+$) requires 203.0791, found 203.0796.

1-(Benzo[*d*][1,3]dioxol-5-ylmethyl)thiourea **130**

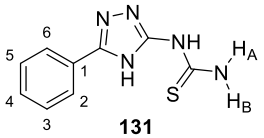


130

Following *General Procedure 5*, benzo[*d*][1,3]dioxol-5-ylmethanamine (88 µL) was treated with benzoyl isothiocyanate in acetone at 65 °C for 3 h,

then with aq. NaOH (5 M) for 2 h at 25 °C to afford *thiourea* **130** as a white powder (6 mg, 3%). Mp 153-153.5 °C; ν_{max} (FTIR) cm^{-1} 3417 (NH_2), 3171, 1624 (NH_2), 1489 (CSNH), 1234 (CSNH), 1033 (C=O), 742 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.80 (br. s, 1H, N*H*), 6.88 (s, 3H, NH_2 & C(2)*H*), 6.84 (d, *J* = 8.0 Hz, 1H, C(5)*H*), 6.77 - 6.81 (m, 1H, C(6)*H*), 5.97 (s, 2H, OCH_2O), 4.49 (br. d, *J* = 4.20 Hz, 1H, C(1)CH_2); ^{13}C NMR (101 MHz, DMSO-d_6) δ 147.6 (*C*(3)), 146.6 (*C*(4)), 133.5 (*C*(1)), 121.1 (*C*(6)), 108.5 (*C*(2) & *C*(5)), 101.3 (OCH_2O), 47.5 (C(1)CH_2), (*C*=S not observed); *m/z* (ESI^+) 211 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_9H_{11}N_2O_2S^+$, ($[M+H]^+$) requires 211.0536, found 211.0537.

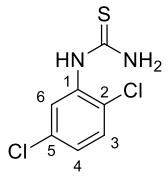
1-(5-Phenyl-4*H*-1,2,4-triazol-3-yl)thiourea **131**



131

Following *General Procedure 5*, 5-phenyl-4*H*-1,2,4-triazol-3-amine (113 mg) was treated with benzoyl isothiocyanate in acetone at 65 °C for 3 h, then with aq. NaOH (5 M) for 2 h at room temperature to afford *thiourea* **131** as a red powder (14 mg, 9.1%). Mp 235-237.5 °C; ν_{max} (FTIR) cm^{-1} 3088, 1644 (NH_2), 1570 (CSNH), 1096, 787 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 14.15 (br. s, 1H, triazole N*H*), 10.87 (br. s, 1H, N*H*), 9.18 (br. s, 1H, NH_A), 8.79 (br. s, 1H, NH_B), 7.87 - 8.10 (m, 2H, C(2)*H* and C(6)*H*), 7.47 - 7.59 (m, 3H, C(3)*H*, C(4)*H* and C(5)*H*); ^{13}C NMR (101 MHz, DMSO-d_6) δ 179.8 (*C*=S), 130.7 (*C*(4)), 129.5 (*C*(3) and *C*(5)), 126.5 (*C*(2) and *C*(6)), (*C*(1) and triazole carbons not observed); *m/z* (ESI^+) 220 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_9H_{10}N_5S^+$, ($[M+H]^+$) requires 220.0651, found 220.0647.

1-(2,5-Dichlorophenyl)thiourea **133**

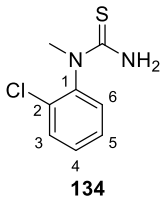


133

Following *General Procedure 5*, 2,5-dichloroaniline (115 mg) was treated with benzoyl isothiocyanate in acetone at 65 °C for 3 h, then with aq. NaOH (5 M) for 2 h at room temperature to afford *thiourea* **133** as an off-white powder (26 mg, 17%). Mp 182-183 °C; ν_{max} (FTIR) cm^{-1} 3162, 1600 (NH_2), 1544 (CSNH), 1398, 1329 (CSNH), 813 (C-Cl); ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.39 (s, 1H, N*H*), 7.89 (d, *J* = 2.6 Hz, 1H, C(6)*H*), 7.53 (d, *J* = 8.5

Hz, 1H, C(3)*H*), 7.29 (dd, *J* = 2.6, 8.7 Hz, 1H, C(4)*H*), (NH_2 not observed); ^{13}C NMR (101 MHz, DMSO-d_6) δ 182.4 (*C*=S), 137.9 (*C*(1)), 131.4 (*C*(5)), 131.1 (*C*(3)), 128.6 (*C*(6)), 127.7 (*C*(2)), 127.1 (*C*(4)); *m/z* (ESI$^+$) 221 ([M(^{35}Cl,^{35}Cl)+H]$^+$, 100%), 223 ([M(^{35}Cl,^{37}Cl)+H]$^+$, 80%), 225 ([M(^{37}Cl,^{37}Cl)+H]$^+$, 10%); HRMS (ESI$^+$) $C_7H_8{}^{35}Cl_2N_2S^+$, ([M+H]$^+$) requires 220.9702, found 220.9702.

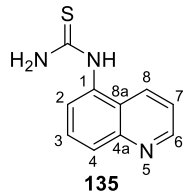
1-(2-Chlorophenyl)-1-methylthiourea **134**



134

Following *General Procedure 5*, 2-chloro-*N*-methylaniline (87 µL) was treated with benzoyl isothiocyanate in acetone at 65 °C for 3 h, then with aq. NaOH (5 M) for 14 h at 90 °C to afford *thiourea* **134** as an amorphous white solid (5 mg, 4%). ν_{max} (FTIR) cm^{-1} 3243, 3137, 1608 (NH_2), 1473 (CSNH), 1359 (CSNH), 1053, 984, 733 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.49 - 7.66 (m, 1H, C(3)*H*), 7.32 - 7.47 (m, 3H, C(4)*H*, C(5)*H* & C(6)*H*), 5.58 (br. s, 2H, NH_2), 3.57 - 3.75 (m, 3H, CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 182.5 (*C*=S), 139.6 (*C*(1)), 132.3 (*C*(2)), 131.4 (*C*(3)), 130.6 (*C*(4)), 128.9 (*C*(5) & *C*(6)), 42.4 (*C*H_3); *m/z* (ESI$^+$) 201 ([M(^{35}Cl)+H]$^+$, 100%), 203 ([M(^{37}Cl)+H]$^+$, 35%); HRMS (ESI$^+$) $C_8H_{10}{}^{35}ClN_2S^+$, ([M+H]$^+$) requires 201.0248, found 201.0246.

1-(Quinolin-5-yl)thiourea **135**

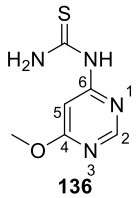


135

Following *General Procedure 5*, quinolin-5-amine (102 mg) was treated with benzoyl isothiocyanate in acetone at 25 °C for 1 h, then with aq. NaOH (10%) for 2 h at 25 °C to afford *thiourea* **135** as a white powder (25 mg, 17%). Mp 219-220.5 °C; ν_{max} (FTIR) cm^{-1} 3059, 1648 (NH_2), 1477 (CSNH), 1254 (CSNH), 1087, 803, 724 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.83 (br. s, 1H, N*H*), 8.93 (dd, *J* = 1.6, 4.2 Hz, 1H, C(6)*H*), 8.23 - 8.30 (m, 1H, C(8)*H*), 7.95 (d, *J* = 8.6 Hz, 1H, C(4)*H*), 7.76 (dd, *J* = 7.5, 8.4 Hz, 1H, C(3)*H*), 7.60 (dd, *J* = 4.2, 8.6 Hz, 1H, C(7)*H*), 7.55 (dd, *J* = 1.0, 7.3 Hz, 1H, C(2)*H*), (NH_2 not observed); ^{13}C NMR (101 MHz, DMSO-d_6) δ 183.1 (*C*=S), 151.0 (*C*(6)), 148.9 (*C*(4a)), 135.3 (*C*(1)), 132.0 (*C*(8)), 129.6 (*C*(3)),

128.2 (*C*(4)), 125.6 (*C*(2)), 125.5 (*C*(8a)), 121.9 (*C*(7)); *m/z* (ESI$^+$) 204 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{10}H_{10}N_3S^+$, ([M+H]$^+$) requires 204.0590, found 204.0589.

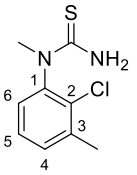
1-(6-Methoxypyrimidin-4-yl)thiourea **136**



136

Following *General Procedure 5*, 6-methoxypyrimidin-4-amine (89 mg) was treated with benzoyl isothiocyanate in CH_2Cl_2 at 45 °C for 14 h, then with aq. K_2CO_3 (3.6 M) for 14 h at 82 °C to afford *thiourea* **136** as a white powder (4 mg, 3.1%). Mp 218-219.5 °C; ν_{max} (FTIR) cm^{-1} 3094, 1575 (CSNH), 1204 (CSNH), 1047, 816; ^{1}H NMR (400 MHz, DMSO-d_6) δ 10.65 (br. s, 1H, C(6)N*H*), 10.25 (br. s, 1H, N(H_A or H_B)), 9.15 (br. s, 1H, N(H_A or H_B)), 8.54 (d, *J* = 0.9 Hz, 1H, C(2)*H*), 6.54 (d, *J* = 0.9 Hz, 1H, C(5)*H*), 3.89 (s, 3H, CH_3); ^{13}C NMR (101 MHz, DMSO-d_6) δ 181.5 (*C*=S), 170.3 (*C*(4)), 160.1 (*C*(6)), 157.4 (*C*(2)), 91.9 (*C*(5)), 54.5 (CH_3); *m/z* (ESI$^+$) 185 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_6H_8N_4OS^+$, ([M+H]$^+$) requires 185.0492, found 185.0485.

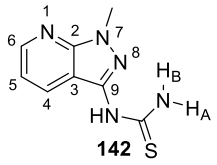
1-(2-Chloro-3-methylphenyl)-1-methylthiourea **139**



139

Following *General Procedure 5*, 3-chloro-*N*,2-dimethylaniline (115 mg) was treated with benzoyl isothiocyanate in CH_2Cl_2 at 45 °C for 14 h, then with aq. K_2CO_3 (3.6 M) for 4 h at 82 °C to afford *thiourea* **139** as an amorphous white solid (5 mg, 4%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.46 (d, *J* = 8.0 Hz, 1H, C(4)*H* or C(6)*H*), 7.28 (t, *J* = 7.8 Hz, 1H, C(5)*H*), 7.17 (d, *J* = 7.7 Hz, 1H, C(4)*H* or C(6)*H*), 5.53 (br. s, 2H, NH_2), 3.62 (s, 3H, NCH_3), 2.31 (s, 3H, C(3)CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 182.3 (*C*=S, by HMBC), 142.4 (*C*(1), by HMBC), 136.6 (*C*(3), by HMBC), 134.4 (*C*(2), by HMBC), 130.2 (*C*(4) or *C*(6)), 128.4 (*C*(5)), 125.4 (*C*(4) or *C*(6)), 42.8 (NCH_3), 14.7 (C(3)CH_3); *m/z* (ESI$^+$) 215 ([M(^{35}Cl)+H]$^+$, 100%), 217 ([M(^{37}Cl)+H]$^+$, 40%); HRMS (ESI$^+$) $C_9H_{12}{}^{35}ClN_2S^+$, ([M+H]$^+$) requires 215.0404, found 215.0407.

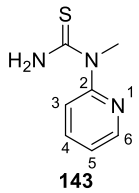
1-(1-Methyl-1*H*-pyrazolo[3,4-*b*]pyridin-3-yl)thiourea **142**



142

Following *General Procedure 5*, 1-methyl-1*H*-pyrazolo[3,4-*b*]pyridin-3-amine (117 mg) was treated with benzoyl isothiocyanate in CH_2Cl_2 at 45 °C for 14 h, then with aq. K_2CO_3 (3.6 M) for 14 h at 82 °C to afford *thiourea* **142** as an off-white powder (6.5 mg, 4.0%). Mp 239.5-240 °C; ν_{max} (FTIR) cm^{-1} 3292, 3145, 3060, 1602 (NH_2), 1550 (CSNH), 1280 (CSNH), 763 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 11.05 (br. s, 1H, C(9)N*H*), 9.13 (br. s, 1H, N(H_A or H_B)), 8.85 (br. s, 1H, N(H_A or H_B)), 8.54 - 8.65 (m, 2H, C(4)*H* and C(6)*H*), 7.20 (dd, *J* = 4.5, 8.1 Hz, 1H, C(5)*H*), 3.97 (s, 3H, CH_3); ^{13}C NMR (101 MHz, DMSO-d_6) δ 179.8 (*C*=S), 150.8 (*C*(6)), 149.7 (*C*(2)), 140.6 (*C*(9)), 131.3 (*C*(4)), 116.6 (*C*(5)), 106.3 (*C*(3)), 33.8 (*C*H_3); *m/z* (ESI^+) 208 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_8H_{10}N_5S^+$, ($[M+H]^+$) requires 208.0651, found 208.0651.

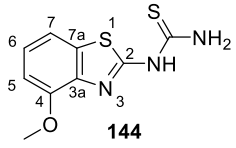
1-Methyl-1-(pyridin-2-yl)thiourea **143**



143

Following *General Procedure 5*, *N*-methylpyridin-2-amine (84 mg) was treated with benzoyl isothiocyanate in CH_2Cl_2 at 45 °C for 4 h, then with aq. K_2CO_3 (3.6 M) for 14 h at 82 °C to afford *thiourea* **143** as an amorphous white solid (5 mg, 5%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.37 (dd, *J* = 2.0, 5.0 Hz, 1H, C(6)*H*), 7.82 (ddd, *J* = 2.0, 7.4, 8.5 Hz, 1H, C(4)*H*), 7.20 (d, *J* = 8.4 Hz, 1H, C(3)*H*), 7.15 (ddd, *J* = 0.7, 5.0, 7.4 Hz, 1H, C(5)*H*), 3.89 (s, 3H, CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 185.4 (*C*=S, by HMBC), 156.5 (*C*(2)), 146.4 (*C*(6)), 139.1 (*C*(4)), 119.8 (*C*(5)), 115.0 (*C*(3)), 41.2 (*C*H_3); *m/z* (ESI^+) 168 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_7H_{10}N_3S^+$, ($[M+H]^+$) requires 168.0590, not found.

1-(4-Methoxybenzo[d]thiazol-2-yl)thiourea **144**

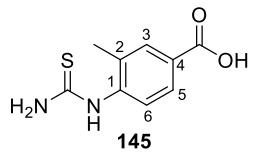


144

Following *General Procedure 5*, 4-methoxybenzo[*d*]thiazol-2-amine (128 mg) was treated with benzoyl isothiocyanate in acetone at 65 °C for 16 h,

then aq. NaOH (10%) for 2 h at 80 °C to afford *thiourea* **144** as a white powder (69 mg, 41%). Mp 222-223 °C; ν_{max} (FTIR) cm^{-1} 3184, 3011, 1562 (NH_2), 1527 (CSNH), 1215 (C-O), 1041, 728 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 11.91 (br. s, 1H, N*H*), 9.07 (br. s, 2H, NH_2), 7.49 (d, *J* = 8.0 Hz, 1H, C(7)*H*), 7.25 (t, *J* = 8.1 Hz, 1H, C(6)*H*), 7.00 (d, *J* = 7.6 Hz, 1H, C(5)*H*), 3.91 (s, 3H, CH_3); ^{13}C NMR (101 MHz, DMSO-d_6) δ 179.9 (*C*(2) or *C*=S), 159.9 (*C*(2) or *C*=S), 151.7 (*C*(4)), 138.9 (*C*(3a)), 132.1 (*C*(7a)), 125.2 (*C*(6)), 113.9 (*C*(7)), 108.3 (*C*(5)), 56.2 (CH_3); *m/z* (ESI^+) 240 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_9H_{10}N_3OS_2^+$, ($[M+H]^+$) requires 240.0260, not found.

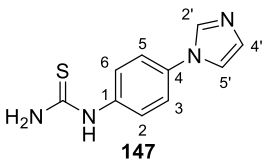
3-Methyl-4-thioureidobenzoic acid **145**



145

Following *General Procedure 5*, 4-amino-3-methylbenzoic acid (107 mg) was treated with benzoyl isothiocyanate in acetone at 25 °C for 1 h, then with aq. NaOH (10%) for 2 h at 25 °C to afford *thiourea* **145** as an amorphous solid (4 mg, 3%). ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.53 (br. s, 1H, N*H*), 7.78 (s, 1H, C(3)*H*), 7.71 (dd, *J* = 1.6, 8.2 Hz, 1H, C(5)*H*), 7.39 (d, *J* = 8.0 Hz, 1H, C(6)*H*), 2.24 (s, 3H, CH_3), (NH_2 not observed); ^{13}C NMR (101 MHz, DMSO-d_6) δ 182.1 (*C*=O or *C*=S), 133.8 (*C*(2)), 131.9 (*C*(3)), 127.7 (*C*(5)), 127.1 (*C*(6)), 18.2 (CH_3), (*C*(1) & *C*(4) not observed); *m/z* (ESI^+) 211 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_9H_{11}N_2O_2S^+$, ($[M+H]^+$) requires 211.0536, found 211.0531.

1-(4-(1*H*-Imidazol-1-yl)phenyl)thiourea **147**

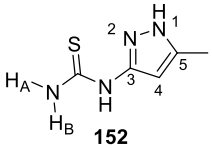


147

Following *General Procedure 5*, 4-(1*H*-imidazol-1-yl)aniline (111 mg) was treated with benzoyl isothiocyanate in CH_2Cl_2 at 45 °C for 14 h, then with aq. K_2CO_3 (3.6 M) for 4 h at 82 °C to afford *thiourea* **147** as a grey powder (5 mg, 3%). Mp 214.5-215 °C; ν_{max} (FTIR) cm^{-1} 3289, 3111, 2994, 1636 (NH_2), 1522 (CSNH), 1244 (CSNH), 1064, 820 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 10.16 (br. s, 1H, N*H*), 8.20 - 8.23 (m, 1H, C(2')*H*), 7.69 - 7.72 (m, 1H, C(4' or 5')*H*), 7.56 - 7.67 (m, 4H, C(2)*H*, C(3)*H*, C(5)*H* and C(6)*H*), 7.09 - 7.12 (m,

1H, C(4' or 5')*H*); ^{13}C NMR (101 MHz, DMSO-d_6) δ 181.8 (*C*=S), 138.7 (*C*(1)), 135.9 (*C*(2')), 133.3 (*C*(4)), 130.2 (*C*(4' or 5')), 124.2 (*C*(2) and *C*(6)), 121.0 (*C*(3) and *C*(5)), 118.5 (*C*(4' or 5')); *m/z* (ESI$^+$) 219 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{10}H_{13}N_4S^+$, ([M+H]$^+$) requires 219.0699, found 219.0703.

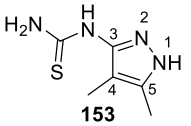
1-(5-Methyl-1H-pyrazol-3-yl)thiourea **152**



152

Following *General Procedure 5*, 3-methyl-1H-pyrazol-5-amine (69 mg) was treated with benzoyl isothiocyanate in acetone at 25 °C for 1 h, then with aq. NaOH (10%) for 2 h at 25 °C to afford *thiourea* **152** as an amorphous white solid (13 mg, 11%). ν_{max} (FTIR) cm^{-1} 3143 (NH_2), 1546 (CSNH), 1243 (CSNH), 998, 776 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 12.16 (s, 1H, N(1)*H*), 10.20 (s, 1H, C(3)N*H*), 9.05 (br. s, 1H, H_A or H_B), 8.38 (br. s, 1H, H_A or H_B), 5.67 (s, 1H, C(4)*H*), 2.18 (s, 3H, CH_3); ^{13}C NMR (101 MHz, DMSO-d_6) δ 178.8 (*C*=S), 150.2 (*C*(3)), 139.5 (*C*(5)), 93.7 (*C*(4)), 10.8 (*C*H$_3$); *m/z* (ESI$^+$) 157 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_5H_9N_4S^+$, ([M+H]$^+$) requires 157.0542, not found.

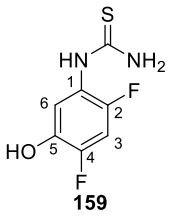
1-(4,5-Dimethyl-1H-pyrazol-3-yl)thiourea **153**



153

Following *General Procedure 5*, 3,4-dimethyl-1*H*-pyrazol-5-amine (74 mg) was treated with benzoyl isothiocyanate in CH_2Cl_2 at 45 °C for 14 h, then with aq. K_2CO_3 (3.6 M) for 4 h at 82 °C to afford *thiourea* **153** as a white powder (12 mg, 10%). Mp 209.5-210.5 °C; ν_{max} (FTIR) cm^{-1} 3266, 3120, 1618 (NH_2), 1542 (CSNH), 1067, 793; ^{1}H NMR (400 MHz, $CDCl_3$) δ 2.24 (s, 3H, C(5)CH_3), 1.93 (s, 3H, C(4)CH_3), (pyrazole N*H*, C(3)N*H* and NH_2); ^{13}C NMR (101 MHz, $CDCl_3$) δ 147.9 (*C*(3)), 137.0 (*C*(4 or 5)), 100.5 (*C*(4 or 5)), 9.6 (*C*(5)*C*H$_3$), 6.4 (*C*(4)*C*H$_3$), (*C*=S not observed); *m/z* (ESI$^+$) 171 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_6H_{11}N_4S^+$, ([M+H]$^+$) requires 171.0699, found 171.0694.

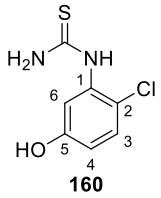
1-(2,4-Difluoro-5-hydroxyphenyl)thiourea **159**



159

Following *General Procedure 5*, 5-amino-2,4-difluorophenol (103 mg) was treated with benzoyl isothiocyanate in CH_2Cl_2 at 45 °C for 14 h, then with aq. K_2CO_3 (3.6 M) for 4 h at 82 °C to afford *thiourea* **159** as a white powder (4 mg, 3%). Mp >250 °C; ν_{max} (FTIR) cm^{-1} 3271, 3172, 1588 (CSNH), 1376, 775 (aromatic CH); ^{1}H NMR (400 MHz, CD_3OD) δ 7.76 (dd, *J* = 6.5, 10.8 Hz, 1H, C(6)*H*), 7.45 (dd, *J* = 6.5, 10.3 Hz, 1H, C(3)*H*), (O*H*, N*H* and NH_2 not observed); ^{13}C NMR (101 MHz, CD_3OD) δ 182.5 (*C*=S), 169.4 (*C*(5)), 155.7 (*C*(2 or 4)), 151.0 (*C*(2 or 4)), 128.1 (*C*(1)), 116.5 (*C*(3)), 113.4 (*C*(6)); ^{19}F NMR (376 MHz, CD_3OD) δ -119.22 (C(2 or 4)*F*), -119.26 (C(2 or 4)*F*); *m/z* (ESI^+) 205 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_7H_7N_2OS^+$, ($[M+H]^+$) requires 205.0061, not found.

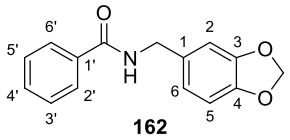
1-(2-Chloro-5-hydroxyphenyl)thiourea **160**



160

Following *General Procedure 5*, 3-amino-4-chlorophenol (102 mg) was treated with benzoyl isothiocyanate in CH_2Cl_2 at 45 °C for 14 h, then with aq. K_2CO_3 (3.6 M) for 4 h at 82 °C to afford *thiourea* **160** as a yellow powder (7 mg, 4.9%). Mp 205-206 °C; ν_{max} (FTIR) cm^{-1} 3154 (OH), 1641 (NH_2), 1598, 1430 (CSNH), 1220 (CSNH), 825 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.74 (br. s, 1H, O*H*), 9.19 (br. s, 1H, N*H*), 7.24 (d, *J* = 8.8 Hz, 1H, C(3)*H*), 7.11 (d, *J* = 2.8 Hz, 1H, C(6)*H*), 6.64 (dd, *J* = 2.9, 8.7 Hz, 1H, C(4)*H*), (NH_2 not observed); ^{13}C NMR (101 MHz, DMSO-d_6) δ 182.1 (*C*=S), 156.7 (*C*(5)), 136.8 (*C*(1)), 130.2 (*C*(3)), 118.8 (*C*(2)), 115.9 (*C*(6)), 114.9 (*C*(4)); *m/z* (ESI^+) 203 ($[M(^{35}Cl)+H]^+$, 100%), 205 ($[M(^{37}Cl)+H]^+$, 40%); HRMS (ESI^+) $C_7H_8{}^{35}ClN_2OS^+$, ($[M+H]^+$) requires 203.0040, found 203.0040.

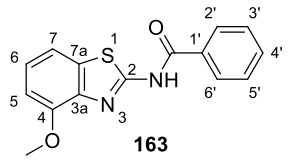
N-(Benzo[d][1,3]dioxol-5-ylmethyl)benzamide **162**



162

Following *General Procedure 5*, 4- benzo[*d*][1,3]dioxol-5-ylmethanamine (121 mg) was treated with benzoyl isothiocyanate in acetone at 25 °C for 1 h, then with aq. NaOH (10%) for 2 h at 25 °C. The reaction byproduct was isolated from the filtrate, purified by reverse phase HPLC and the solvent removed *in vacuo* to afford *amide* **162** as a yellow solid (6 mg, 3%). Mp 91-92.5 °C; ν_{max} (FTIR) cm^{-1} 3303 (CONH), 1629 (C=O), 1542 (O=CN), 1481 (C-O), 1227, 1033; ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.98 (t, *J* = 5.9 Hz, 1H, N*H*), 7.84 - 7.91 (m, 2H, C(2')*H* & C(6')*H*), 7.50 - 7.56 (m, 1H, C(4')*H*), 7.43 - 7.50 (m, 2H, C(3')*H* & C(5')*H*), 6.89 (d, *J* = 1.5 Hz, 1H, C(2)*H*), 6.86 (d, *J* = 7.8 Hz, 1H, C(5)*H*), 6.80 (dd, *J* = 1.5, 7.9 Hz, 1H, C(6)*H*), 5.98 (s, 2H, OCH_2O), 4.39 (d, *J* = 6.0 Hz, 2H, CH_2C(1)); ^{13}C NMR (101 MHz, DMSO-d_6) δ 166.6 (*C*=O), 147.7 (*C*(3)), 146.5 (*C*(4)), 134.8 (*C*(1')), 134.1 (*C*(1)), 131.7 (*C*(4')), 128.8 (*C*(3') & *C*(5')), 127.7 (*C*(2') & *C*(6')), 120.9 (*C*(6)), 108.5 (*C*(5)), 108.4 (*C*(2)), 101.3 (O*C*H_2O), 42.9 (*C*H_2C(1)); *m/z* (ESI$^+$) 256 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{15}H_{14}NO_3^+$, ([M+H]$^+$) requires 256.0968, found 256.0970.

N-(4-Methoxybenzo[d]thiazol-2-yl)benzamide **163**

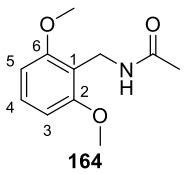


163

Following *General Procedure 5*, 4-methoxybenzo[*d*]thiazol-2-amine (128 mg) was treated with benzoyl isothiocyanate at 65 °C for 16 h followed by aq. NaOH (10%) for 2h at 80 °C. The reaction byproduct was isolated from the filtrate and the solvent removed *in vacuo* to afford *amide* **163** as a yellow powder (55 mg, 27%). Mp 232-233.5 °C; ν_{max} (FTIR) cm^{-1} 3649, 2969 (CH), 1530 (O=CN), 1420, 1249, 704 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.03 - 8.16 (m, 2H, C(2')*H* & C(6')*H*), 7.24 - 7.37 (m, 3H, C(3')*H*, C(4')*H* & C(5')*H*), 7.17 (dd, *J* = 0.7, 7.7 Hz, 1H, C(7)*H*), 6.85 (t, *J* = 7.8 Hz, 1H, C(6)*H*), 6.68 (d, *J* = 7.8 Hz, 1H, C(5)*H*), 3.81 (s, 3H, CH_3), (N*H* not observed); ^{13}C NMR (101 MHz, DMSO-d_6) δ 171.0 (*C*=O), 150.8 (*C*(4)), 141.1 (*C*(3a)), 140.9 (*C*(1')), 134.8 (*C*(7a)), 129.7 (*C*(4')), 128.6 (*C*(2') & *C*(6')), 127.8 (*C*(3') &

C(5')), 120.8 (*C*(6)), 113.8 (*C*(7)), 107.1 (*C*(5)), 56.2 (*C*H_3), (*C*(2) not observed); *m/z* (ESI$^+$) 285 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{15}H_{13}N_2O_2S^+$, ([M+H]$^+$) requires 285.0692, found 285.0699.

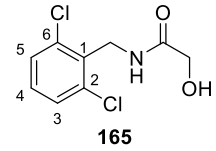
N-(2,6-Dimethoxybenzyl)acetamide **164**



164

Following *General Procedure 2*, acetyl chloride (40 µL, 0.6 mmol, 1.05 eq) was treated with 2,6-dimethoxybenzylamine (90 mg, 0.5 mmol, 1.0 eq) and triethylamine (149 µL, 1.0 mmol, 2.0 eq) to afford *amide* **164** as a yellow solid (93 mg, 79%). Mp 120-122 °C; ν_{max} (FTIR) cm^{-1} 3299 (NH), 1634 (C=O), 1594, 1474, 1114 (C-O-C), 773 (C-Cl or aromatic CH), 720 (C-Cl or aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.15 (t, *J* = 8.4 Hz, 1H, C(4)*H*), 6.49 (d, *J* = 8.4 Hz, 2H, C(3)*H* & C(5)*H*), 5.75 (br. s, 1H, N*H*), 4.46 (d, *J* = 5.3 Hz, 2H, CH_2), 3.77 (s, 6H, C(2)OCH_3 & C(6)OCH_3), 1.87 (s, 3H, C(=O)CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 169.4 (*C*=O), 158.6 (*C*(2) & *C*(6)), 129.0 (*C*(4)), 114.1 (*C*(1)), 103.8 (*C*(3) & *C*(5)), 55.9 (C(2)O*C*H_3 & C(6)O*C*H_3), 32.6 (*C*H_2), 23.5 (*C*H_3); *m/z* (ESI$^+$) 210 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{11}H_{15}NO_3Na^+$, ([M+Na]$^+$) requires 232.0944, found 232.0939.

N-(2,6-Dichlorobenzyl)-2-hydroxyacetamide **165**

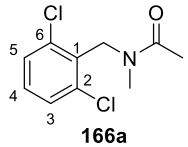


165

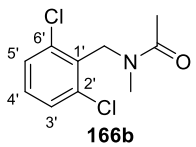
Glycolic acid (43 mg, 0.6 mmol, 1.0 eq) was treated with *N,N*-dimthylpyridin-4-amine (7 mg, 0.06 mmol, 0.1 eq) and N-Ethyl-N'-(3-dimethylaminopropyl)carbodiimide hydrochloride (88 mg, 0.5 mmol, 0.8 eq) and added to a stirred solution of 2,6-dichlorobenzylamine (100 mg, 0.6 mmol, 1 eq) and triethylamine (79 µL, 0.6 mmol, 1.0 eq) in DMF (2 mL). The reaction mixture was stirred overnight at room temperature. The reaction mixture was diluted with EtOAc and washed with water (2 x 20 mL) and brine (20 mL). The crude mixture was purified by flash column chromatography (0-10% MeOH in CH_2Cl_2). Pure fractions were collected and the solvent removed *in vacuo* to afford *amide* **165** as a white powder (20 mg, 15%). Mp 112-113 °C; ν_{max} (FTIR) cm^{-1} 3271 (NH), 1647 (C=O), 1099 (C-O), 785 (C-

Cl or aromatic CH), 764 (C-Cl or aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.33 (d, *J* = 7.8 Hz, 2H, C(3)*H* & C(5)*H*), 7.20 (dd, *J* = 7.4, 8.5 Hz, 1H, C(4)*H*), 6.71 (br. s, 1H, N*H*), 4.81 (d, *J* = 5.6 Hz, 2H, CH_2NH), 4.12 (s, 2H, CH_2OH), 2.76 (br. s., 1H, O*H*); ^{13}C NMR (101 MHz, $CDCl_3$) δ 171.0 (*C*=O), 136.2 (*C*(2) & *C*(6)), 133.1 (*C*(1)), 129.8 (*C*(4)), 128.5 (*C*(3) & *C*(5)), 62.2 (*C*H_2OH), 38.7 (*C*H_2NH); *m/z* (ESI$^+$) 234 ([M(^{35}Cl,^{35}Cl)+H]$^+$, 100%), 236 ([M(^{35}Cl,^{37}Cl)+H]$^+$, 65%); HRMS (ESI$^+$) $C_9H_9{}^{35}Cl_2NO_2Na^+$, ([M+Na]$^+$) requires 255.9903, found 255.9907.

N-(2,6-Dichlorobenzyl)-*N*-methylacetamide **166**



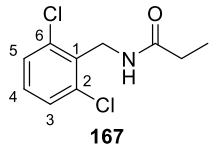
166a



166b

Following *General Procedure 2*, acetyl chloride (40 µL, 0.6 mmol, 1.05 eq) was treated with *N*-methyl-2,6-dichlorobenzylamine hydrochloride (121 mg, 0.5 mmol, 1.0 eq) and triethylamine (224 µL, 1.6 mmol, 3.0 eq) to afford *amide* **166** as a yellow solid (103 mg, 83%). Rotamers **166a** and **166b** were found to be present in a 2:1 ratio by NMR. Mp 71-72 °C; ν_{max} (FTIR) cm^{-1} 1641 (C=O), 1435, 1400, 787 (C-Cl or aromatic CH), 755 (C-Cl or aromatic CH); **166a** ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.24 - 7.32 (m, 2H, C(3)*H* & C(5)*H*), 7.09 - 7.19 (m, 1H, C(4)*H*), 4.90 (s, 2H, CH_2), 2.69 (s, 3H, NCH_3), 2.06 (s, 3H, C(=O)CH_3); **166a** ^{13}C NMR (101 MHz, $CDCl_3$) δ 170.4 (*C*=O), 136.9 (*C*(2) & *C*(6)), 132.2 (*C*(1)), 129.5 (*C*(4)), 128.6 (*C*(3) & *C*(5)), 44.7 (*C*H_2), 33.7 (N*C*H_3), 21.8 (C(=O)*C*H_3); **166b** ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.24 - 7.32 (m, 2H, C(3')*H* & C(5')*H*), 7.09 - 7.19 (m, 1H, C(4')*H*), 4.74 (s, 2H, CH_2), 2.61 (s, 3H, NCH_3), 2.26 (s, 3H, C(=O)CH_3); **166b** ^{13}C NMR (101 MHz, $CDCl_3$) δ 170.4 (*C*=O), 136.7 (*C*(2') & *C*(6')), 131.2 (*C*(1')), 130.0 (*C*(4')), 128.9 (*C*(3') & *C*(5')), 48.7 (*C*H_2), 30.4 (N*C*H_3), 22.0 (C(=O)*C*H_3); *m/z* (ESI$^+$) 232 ([M(^{35}Cl,^{35}Cl)+H]$^+$, 100%), 234 ([M(^{35}Cl,^{37}Cl)+H]$^+$, 75%) ([M+H]$^+$, 100%), 236 ([M(^{37}Cl,^{37}Cl)+H]$^+$, 12%); HRMS (ESI$^+$) $C_{10}H_{11}{}^{35}Cl_2NONa^+$, ([M+Na]$^+$) requires 254.0110, found 254.0106.

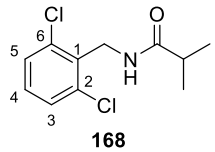
N-(2,6-Dichlorobenzyl)propionamide **167**



167

Following *General Procedure 2*, propionyl chloride (52 µL, 0.6 mmol, 1.05 eq) was treated with 2,6-dichlorobenzylamine (100 mg, 0.6 mmol, 1.0 eq) and triethylamine (158 µL, 1.1 mmol, 2.0 eq) to afford *amide* **167** as a white powder (108 mg, 82%). Mp 148-149 °C; ν_{max} (FTIR) cm^{-1} 3281 (NH), 1635 (C=O), 1537, 1434, 778 (C-Cl or aromatic CH), 765 (C-Cl or aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.26 (d, *J* = 8.0 Hz, 2H, C(3)*H* & C(5)*H*), 7.08 - 7.14 (m, 1H, C(4)*H*), 5.59 (br. s, 1H, N*H*), 4.69 (d, *J* = 5.4 Hz, 2H, C(1)CH_2), 2.15 (q, *J* = 7.6 Hz, 2H, CH_2CH$_3$), 1.09 (t, *J* = 7.6 Hz, 3H, CH$_2$CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 173.2 (*C*=O), 136.2 (*C*(2) & *C*(6)), 133.7 (*C*(1)), 129.6 (*C*(4)), 128.5 (*C*(3) & *C*(5)), 39.2 (C(1)*C*H$_2$), 29.5 (*C*H$_2$CH$_3$), 9.7 (CH$_2$*C*H$_3$); *m/z* (ESI$^+$) 232 ([M(^{35}Cl,^{35}Cl)+H]$^+$, 100%), 234 ([M(^{35}Cl,^{37}Cl)+H]$^+$, 73%), 236 ([M(^{37}Cl,^{37}Cl)+H]$^+$, 11%); HRMS (ESI$^+$) $C_{10}H_{11}{}^{35}Cl_2NONa^+$, ([M+Na]$^+$) requires 254.0110, found 254.0105.

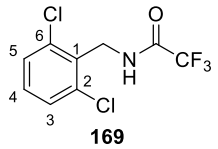
N-(2,6-Dichlorobenzyl)isobutyramide **168**



168

Following *General Procedure 2*, isobutyryl chloride (62 µL, 0.6 mmol, 1.05 eq) was treated with 2,6-dichlorobenzylamine (100 mg, 0.6 mmol, 1.0 eq) and triethylamine (158 µL, 1.1 mmol, 2.0 eq) to afford *amide* **168** as a white powder (114 mg, 82%). Mp 149-150 °C; ν_{max} (FTIR) cm^{-1} 3241 (NH), 2967 (aliphatic CH), 1639 (C=O), 1537, 1435, 779 (C-Cl or aromatic CH), 767 (C-Cl or aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.25 (d, *J* = 7.8 Hz, 2H, C(3)*H* & C(5)*H*), 7.08 - 7.14 (m, 1H, C(4)*H*), 5.61 (br. s, 1H, N*H*), 4.67 (d, *J* = 5.4 Hz, 2H, CH_2), 2.28 (spt, *J* = 6.9 Hz, 1H, C*H*), 1.08 (d, *J* = 6.9 Hz, 6H, CH(CH_3)$_2$); ^{13}C NMR (101 MHz, $CDCl_3$) δ 176.4 (*C*=O), 136.2 (*C*(2) & *C*(6)), 133.7 (*C*(1)), 129.5 (*C*(4)), 128.5 (*C*(3) & *C*(5)), 39.3 (C(1)*C*H$_2$), 35.6 (*C*H), 19.6 (CH(*C*H$_3$)$_2$); *m/z* (ESI$^+$) 246 ([M(^{35}Cl,^{35}Cl)+H]$^+$, 100%), 248 ([M(^{35}Cl,^{37}Cl)+H]$^+$, 74%), 250 ([M(^{37}Cl,^{37}Cl)+H]$^+$, 12%); HRMS (ESI$^+$) $C_{11}H_{13}{}^{35}Cl_2NONa^+$, ([M+Na]$^+$) requires 268.0266, found 268.0263.

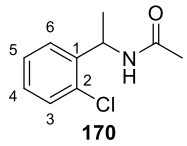
N-(2,6-Dichlorobenzyl)-2,2,2-trifluoroacetamide **169**



169

Ethyl 2,2,2-trifluoroacetate (135 µL, 1.1 mmol, 2.0 eq) and triethylamine (79 µL, 0.6 mmol, 1.0 eq) were added to a stirred solution of 2,6-dichlorobenzylamine (100 mg, 0.6 mmol, 1.0 eq) in MeOH (1 mL). The reaction mixture was stirred overnight at room temperature. The solvent was removed *in vacuo* and the crude solid dissolved in EtOAc. The solution was washed with water (2 x 20 mL) and brine (20 mL). The organic layer was dried and the solvent removed *in vacuo* to afford *amide* **169** as a white powder (103 mg, 67%). Mp 115-117 °C; ν_{max} (FTIR) cm^{-1} 3276 (NH), 1693 (C=O), 1167 (C-F), 787 (aromatic CH or C-Cl); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.33 - 7.41 (m, 2H, C(3)*H* & C(5)*H*), 7.22 - 7.30 (m, 1H, C(4)*H*), 6.50 (br. s, 1H, N*H*), 4.89 (d, *J* = 5.6 Hz, 2H, CH_2); ^{13}C NMR (101 MHz, $CDCl_3$) δ 156.9 (*C*=O), 136.3 (*C*(2) & *C*(6)), 131.4 (*C*(1)), 130.5 (*C*(4)), 128.7 (*C*(3) & *C*(5)), 39.4 (*C*H_2), (*C*F_3 not observed); ^{19}F NMR (376 MHz, $CDCl_3$) δ -75.69 (C*F*$_3$); *m/z* not observed in (ESI^+) or (ESI^-); HRMS (ESI^-) $C_9H_5{}^{35}Cl_2F_3NO^-$, ($[M-H]^-$) requires 269.9706, found 269.9711.

N-(1-(2-Chlorophenyl)ethyl)acetamide **170**

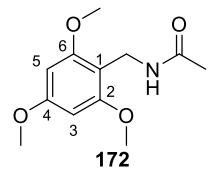


170

Following *General Procedure 2*, acetyl chloride (40 µL, 0.6 mmol, 1.05 eq) was treated with 1-(2-chlorophenyl)ethanamine (83 mg, 0.5 mmol, 1.0 eq) and triethylamine (149 µL, 1.0 mmol, 2.0 eq) to afford *amide* **170** as a white powder (92 mg, 87%). Mp 109-110 °C; ν_{max} (FTIR) cm^{-1} 3269 (NH), 1644 (C=O), 1558, 756 (C-Cl or aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.23 - 7.30 (m, 2H, C(3)*H* & C(6)*H*), 7.10 - 7.19 (m, 2H, C(4)*H* & C(5)*H*), 5.92 (br. s, 1H, N*H*), 5.32 (quin, *J* = 7.3 Hz, 1H, C(1)C*H*), 1.92 (s, 3H, C(=O)CH_3), 1.42 (d, *J* = 7.0 Hz, 3H, CHCH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 169.0 (*C*=O), 140.4 (*C*(1)), 132.9 (*C*(2)), 130.2 (*C*(3)), 128.5 (*C*(4)), 127.3 (*C*(6)), 127.1 (*C*(5)), 47.3 (C(1)*C*H), 23.3 (C(=O)*C*H_3), 20.9 (CH*C*H_3); *m/z* (ESI^+) 198 ($[M(^{35}Cl)+H]^+$,

100%), 200 ([M(^{37}Cl)+H]$^+$, 34%); HRMS (ESI$^+$) $C_{10}H_{12}{}^{35}ClNONa^+$, ([M+Na]$^+$) requires 220.0500, found 220.0492.

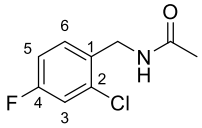
N-(2,4,6-Trimethoxybenzyl)acetamide **172**



172

Following *General Procedure 2*, acetyl chloride (40 µL, 0.6 mmol, 1.05 eq) was treated with 2,4,6-trimethoxybenzylamine hydrochloride (125 mg, 0.5 mmol, 1.0 eq) and triethylamine (224 µL, 1.6 mmol, 3.0 eq) to afford *amide* **172** as a white powder (107 mg, 83%). Mp 153-154 °C; ν_{max} (FTIR) cm^{-1} 3310 (NH), 1643 (C=O), 1598, 1130 (C-O-C); ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.06 (s, 2H, C(3)*H* & C(5)*H*), 5.63 (br. s, 1H, N*H*), 4.37 (d, *J* = 5.0 Hz, 2H, CH_2), 3.75 (s, 6H, C(2)OCH_3 & C(6)OCH_3), 3.74 (s, 3H, C(4)OCH_3), 1.86 (s, 3H, C(=O)CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 169.4 (*C*=O), 160.9 (*C*(4)), 159.3 (*C*(2) & *C*(6)), 106.7 (*C*(1)), 90.6 (*C*(3) & *C*(5)), 55.8 (C(2)O*C*H$_3$ & C(6)O*C*H$_3$), 55.4 (C(4)O*C*H$_3$), 32.4 (*C*H$_2$), 23.5 (C(=O)*C*H$_3$); *m/z* (ESI$^+$) 240 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{12}H_{17}NO_4Na^+$, ([M+Na]$^+$) requires 262.1050, found 262.1049.

N-(2-Chloro-4-fluorobenzyl)acetamide **173**

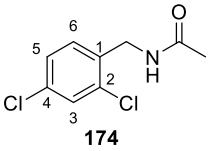


173

Following *General Procedure 2*, acetyl chloride (40 µL, 0.6 mmol, 1.05 eq) was treated 2-chloro-4-fluorobenzylamine (86 mg, 0.5 mmol, 1.0 eq) and triethylamine (149 µL, 1.0 mmol, 2.0 eq) to afford *amide* **173** as a white powder (89 mg, 82%). Mp 80-81 °C; ν_{max} (FTIR) cm^{-1} 3285 (NH), 1639 (C=O), 1549, 1491, 903; ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.31 (dd, *J* = 6.1, 8.6 Hz, 1H, C(6)*H*), 7.05 (dd, *J* = 2.6, 8.4 Hz, 1H, C(3)*H*), 6.88 (dt, *J* = 2.6, 8.3 Hz, 1H, C(5)*H*), 5.88 (br. s, 1H, N*H*), 4.40 (d, *J* = 6.1 Hz, 2H, CH_2), 1.94 (s, 3H, CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 170.0 (*C*=O), 161.9 (d, *J* (C-F) = 249.0 Hz, *C*(4)), 134.2 (d, *J* (C-F) = 10.3 Hz, *C*(2)), 131.8 (d, *J* (C-F) = 3.7 Hz, *C*(1)), 131.5 (d, *J* (C-F) = 8.8 Hz, *C*(6)), 116.9 (d, *J* (C-F) = 25.0 Hz, *C*(3)), 114.3 (d, *J* (C-F) = 22.0 Hz, *C*(5)), 41.0 (*C*H$_2$), 23.2 (*C*H$_3$); ^{19}F NMR (376 MHz, $CDCl_3$) δ -112.46 (C(4)*F*); *m/z* (ESI$^+$)

202 ([M(^{35}Cl)+H]$^+$, 100%), 204 ([M(^{37}Cl)+H]$^+$, 33%); HRMS (ESI$^+$) $C_9H_{10}{}^{35}ClFNO^+$, ([M+H]$^+$) requires 202.0435, not found.

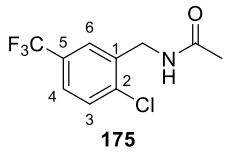
N-(2,4-Dichlorobenzyl)acetamide **174**



174

Following *General Procedure 2*, acetyl chloride (40 µL, 0.6 mmol, 1.05 eq) was treated with 2,4-dichlorobenzylamine (72 µL, 0.5 mmol, 1.0 eq) and triethylamine (149 µL, 1.0 mmol, 2.0 eq) to afford *amide* **174** as a white powder (98 mg, 84%). Mp 88-89 °C (Lit. (Li, Meng *et al.* 2009) 88-90 °C); ν_{max} (FTIR) cm^{-1} 3288 (NH), 1639 (C=O), 1543, 848 (aromatic CH), 708 (C-Cl); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.31 (d, *J* = 2.1 Hz, 1H, C(3)*H*), 7.27 (d, *J* = 8.3 Hz, 1H, C(6)*H*), 7.15 (dd, *J* = 2.1, 8.3 Hz, 1H, C(5)*H*), 5.87 (br. s, 1H, N*H*), 4.40 (d, *J* = 6.1 Hz, 2H, CH_2), 1.95 (s, 3H, CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 169.3 (*C*=O), 133.7 (*C*(1)), 133.5 (*C*(2)), 133.4 (*C*(4)), 130.5 (*C*(6)), 128.6 (*C*(3)), 126.7 (*C*(5)), 40.3 (*C*H$_2$), 22.5 (*C*H$_3$); *m/z* (ESI$^+$) 218 ([M(^{35}Cl,^{35}Cl)+H]$^+$, 100%), 220 ([M(^{35}Cl,^{35}Cl)+H]$^+$, 73%), 222 ([M(^{37}Cl,^{37}Cl)+H]$^+$, 14%); HRMS (ESI$^-$) $C_9H_8{}^{35}Cl_2NO^-$, ([M-H]$^-$) requires 215.9988, found 216.0001.

N-(2-Chloro-5-(trifluoromethyl)benzyl)acetamide **175**

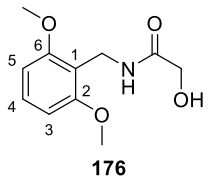


175

Following *General Procedure 2*, acetyl chloride (40 µL, 0.6 mmol, 1.05 eq) was treated with 2-chloro-5-trifluoromethylbenzylamine (75 µL, 0.5 mmol, 1.0 eq) and triethylamine (149 µL, 1.0 mmol, 2.0 eq) to afford *amide* **175** as a yellow solid (98 mg, 65%). Mp 86-87 °C; ν_{max} (FTIR) cm^{-1} 3282 (NH), 1644 (C=O), 1326, 1119, 1080; ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.56 (s, 1H, C(6)*H*), 7.41 - 7.43 (m, *J* = 1.2 Hz, 2H, C(3)*H* & C(4)*H*), 5.96 (br. s, 1H, N*H*), 4.49 (d, *J* = 6.1 Hz, 2H, CH_2), 1.98 (s, 3H, CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 170.1 (*C*=O), 137.4 (d, *J* (C-F) = 22.0 Hz, *C*(5)), 136.8 (*C*(1)), 130.0 (*C*(3)), 129.4 (*C*(2)), 126.7 (d, *J* (C-F) = 3.7 Hz, *C*(6)), 125.7 (d, *J* (C-F) = 3.7 Hz, *C*(4)), 123.6 (d, *J* (C-F) = 272.0 Hz, *C*F$_3$), 41.3 (*C*H$_2$), 23.2 (*C*H$_3$); ^{19}F NMR (376 MHz,

$CDCl_3$) δ -62.53 (C*F*$_3$); *m/z* (ESI$^+$) 252 ([M(^{35}Cl)+H]$^+$, 100%), 254 ([M(^{37}Cl)+H]$^+$, 31%); HRMS (ESI$^-$) $C_{10}H_9{}^{35}ClF_3NO^-$, ([M-H]$^-$) requires 250.0252, found 250.0256.

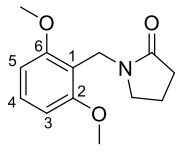
N-(2,6-Dimethoxybenzyl)-2-hydroxyacetamide **176**



176

Compound **271** (150 mg, 0.32 mmol, 1 eq) was treated with tetra-*N*-butylammonium fluoride (1M in THF, 0.49 mL, 0.49 mmol, 1.5 eq) in THF (5 mL). The reaction mixture was stirred for 1 h at room temperature. The solvent was removed *in vacuo* and the crude mixture was purified by flash column chromatography (0-10% MeOH in CH_2Cl_2). Pure fractions were collected and the solvent removed *in vacuo* to afford *amide* **176** as a white powder (55 mg, 75%). Mp 126.5-128 °C; ν_{max} (FTIR) cm^{-1} 3395 (NH), 3277 (OH), 1646 (C=O), 1474 (OH), 1116 (COC); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.24 (t, *J* = 8.4 Hz, 1H, C(4)*H*), 6.73 (br. s, 1H, N*H*), 6.58 (d, *J* = 8.3 Hz, 2H, C(3)*H* & C(5)*H*), 4.61 (d, *J* = 5.6 Hz, 2H, CH_2NH), 4.06 (d, *J* = 5.3 Hz, 2H, CH_2OH), 3.86 (s, 6H, OCH_3), 2.91 (t, *J* = 5.4 Hz, 1H, O*H*); ^{13}C NMR (101 MHz, $CDCl_3$) δ 170.9 (*C*=O), 158.6 (*C*(2) & *C*(6)), 129.2 (*C*(4)), 113.6 (*C*(1)), 103.8 (*C*(3) & *C*(5)), 62.1 (*C*H_2OH), 55.9 (O*C*H_3), 32.0 (*C*H_2NH); *m/z* (ESI$^+$) 226 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{11}H_{15}NO_4Na^+$, ([M+Na]$^+$) requires 248.0893, found 248.0894.

1-(2,6-Dimethoxybenzyl)pyrrolidin-2-one **177**

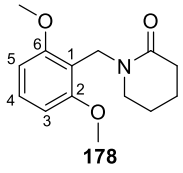


177

2,6-Dimethoxybenzylamine (50 mg, 0.3 mmol, 1.0 eq) was treated with triethylamine (46 µL, 0.33 mmol, 1.1 eq) in MeCN (3 mL) before the addition of ethyl 4-bromobutanoate (43 µL, 0.3 mmol, 1.0 eq) and triethylamine (46 µL, 0.33 mmol, 1.1 eq). The reaction mixture was stirred for 3 days at 70 °C. The solvent was removed *in vacuo.* The crude mixture was dissolved in CH_2Cl_2 and washed with NaOH (10%, 5 mL). The resulting solution was purified by flash column chromatography (0-10% MeOH in CH_2Cl_2). Pure fractions were collected and the solvent removed *in vacuo* to afford *lactam* **177** as a red powder

(4 mg, 6%). Mp 106-107.5 °C; ν_{max} (FTIR) cm^{-1} 1675 (C=O), 1259, 1114 (COC); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.16 (t, *J* = 8.4 Hz, 1H, C(4)*H*), 6.48 (d, *J* = 8.3 Hz, 2H, C(3)*H* & C(5)*H*), 4.49 (s, 2H, C(1)CH_2), 3.74 (s, 6H, OCH_3), 3.05 (t, *J* = 7.0 Hz, 2H, NCH_2CH_2), 2.29 (t, *J* = 8.1 Hz, 2H, C(=O)CH_2), 1.80 (tt, *J* = 7.0, 8.1 Hz, 2H, NCH_2CH_2); ^{13}C NMR (101 MHz, $CDCl_3$) δ 174.3 (*C*=O), 159.2 (*C*(2) & *C*(6)), 129.2 (*C*(4)), 112.0 (*C*(1)), 103.6 (*C*(3) & *C*(5)), 55.8 (O*C*H_3), 46.1 (N*C*H_2CH_2), 34.7 (C(1)*C*H_2), 31.2 (C(=O)*C*H_2), 17.7 (NCH_2*C*H_2); *m/z* (ESI^+) 236 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_{13}H_{17}NO_3Na^+$, ($[M+Na]^+$) requires 258.1101, found 258.1102.

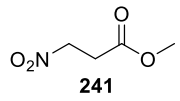
1-(2,6-Dimethoxybenzyl)piperidin-2-one **178**



178

2,6-Dimethoxybenzylamine (100 mg, 0.6 mmol, 1.0 eq) was treated with triethylamine (92 µL, 0.66 mmol, 1.1 eq) in MeCN (12 mL) before the addition of ethyl 5-bromopentanoate (96 µL, 0.6 mmol, 1.0 eq) and triethylamine (92 µL, 0.66 mmol, 1.1 eq). The reaction mixture was stirred for 3 days at 70 °C. 1,8-Diazabicycloundec-7-ene (99 µL, 0.66 mmol, 1.1 eq) was added to the reaction and stirred for a further 3 days. The solvent was removed *in vacuo.* The crude mixture was dissolved in CH_2Cl_2 and washed with NaOH (10%, 5mL). The resulting solution was purified by flash column chromatography (0-10% MeOH in CH_2Cl_2). Pure fractions were collected and the solvent removed *in vacuo* to afford *lactam* **178** as a colourless oil (27 mg, 18%). ν_{max} (FTIR) cm^{-1} 2936, 1594 (C=O), 1257, 1112 (COC); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.24 (t, *J* = 8.3 Hz, 1H, C(4)*H*), 6.56 (d, *J* = 8.3 Hz, 2H, C(3)*H* & C(5)*H*), 4.78 (s, 2H, C(1)CH_2), 3.82 (s, 6H, OCH_3), 3.02 (t, *J* = 5.9 Hz, 2H, NCH_2CH_2), 2.42 (t, *J* = 6.8 Hz, 2H, C(=O)CH_2CH_2), 1.63 - 1.77 (m, 4H, NCH_2CH_2 & C(=O)CH_2CH_2); ^{13}C NMR (101 MHz, $CDCl_3$) δ 169.2 (*C*=O), 159.5 (*C*(2) & *C*(6)), 129.0 (*C*(4)), 112.6 (*C*(1)), 103.6 (*C*(3) & *C*(5)), 55.8 (O*C*H_3), 45.4 (N*C*H_2CH_2), 38.1 (C(1)*C*H_2), 32.6 (C(=O)*C*H_2CH_2), 23.2 (NCH_2*C*H_2), 21.4 (C(=O)CH_2*C*H_2); *m/z* (ESI^+) 250 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_{14}H_{19}NO_3Na^+$, ($[M+Na]^+$) requires 272.1257, found 272.1257.

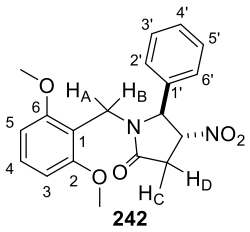
Methyl 3-nitropropanoate **241** (Silva, Costa *et al.* 2001)



241

Sulfuric acid (224 µL, 4.2 mmol, 1.0 eq) was added to a solution of 3-nitropropanoic acid (500 mg, 4.2 mmol, 1.0 eq) in MeOH (10 mL). The reaction mixture was stirred at room temperature for 48 h. Solvent was removed *in vacuo* and the resulting crude mixture diluted with aq. NaOH. The organics were extracted with CH_2Cl_2 to afford *ester* **241** as a pale yellow liquid (416 mg, 74%). ν_{max} (FTIR) cm^{-1} 2958 (C-H), 1733 (C=O), 1552 (C-NO_2), 1371 ($COOCH_3$), 1177; ^1H NMR (400 MHz, $CDCl_3$) δ 4.68 (t, *J* = 6.2 Hz, 2H, O_2NCH_2), 3.78 (s, 3H, CH_3), 3.02 (t, *J* = 6.2 Hz, 2H, $CH_2C(=O)$); ^{13}C NMR (101 MHz, $CDCl_3$) δ 169.8 (*C*=O), 69.7 (O_2NCH_2), 52.4 (CH_3), 30.9 ($CH_2C(=O)$).

1-(2,6-Dimethoxybenzyl)-4-nitro-5-phenylpyrrolidin-2-one **242**

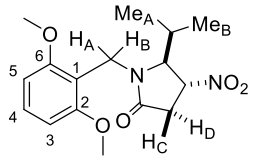


242

2,6-Dimethoxybenzylamine (102 mg, 0.61 mmol, 1.5 eq) and benzaldehyde (62 µL, 0.61 mmol, 1.5 eq) were added to a stirred solution of *ester* **241** (54 mg, 0.41 mmol, 1.0 eq) and benzoic acid (74 mg, 0.61 mmol, 1.5 eq) in toluene (2.5 mL). The reaction mixture was degassed (vacuum) and purged three times with nitrogen. The reaction was stirred at 70 °C for 14 hours. Upon completion the reaction mixture was concentrated under vacuum. The residue was dissolved in CH_2Cl_2 and washed three times with aq. K_2CO_3 (1 M), dried over Na_2SO_4, filtered and the filtrate concentrated *in vacuo*. The crude product was purified by flash column chromatography (15-100% EtOAc in cyclohexane). Solvent was removed *in vacuo* to afford *lactam* **242** as a yellow powder (78 mg, 54%). Mp 88-90 °C; ν_{max} (FTIR) cm^{-1} 2969, 1691 (C=O), 1552 (C-NO_2), 1254 (C-O), 1113, 703 (aromatic CH); ^1H NMR (400 MHz, $CDCl_3$) δ 7.35 - 7.42 (m, 3H, C(3')*H*, C(4')*H* and C(5')*H*), 7.21 (t, *J* = 8.4 Hz, 1H, C(4)*H*), 7.07 - 7.13 (m, 2H, C(2')*H* and C(6')*H*), 6.45 (d, *J* = 8.4 Hz, 2H, C(3)*H* and C(5)*H*), 5.02 (d, *J* = 13.8 Hz, 1H, C(H_A or H_B)), 4.78 (td, *J* = 2.4, 7.8 Hz, 1H, C*H*NO_2), 4.74 (d, *J* = 2.1 Hz, 1H, C*H*C(1')), 4.17 (dd, *J* =

1.0, 13.5 Hz, 1H, C(H_A or H_B)), 3.60 (s, 6H, OCH_3), 3.17 (dd, *J* = 2.8, 18.3 Hz, 1H,C(H_C or H_D)), 3.04 (ddd, *J* = 1.0, 7.9, 18.0 Hz, 1H, C(H_C or H_D)); ^{13}C NMR (101 MHz, $CDCl_3$) δ 169.6 (*C*=O), 159.0 (*C*(2) and *C*(6)), 137.4 (*C*(1')), 129.8 (*C*(4)), 129.1 (*C*(3') and *C*(5')), 128.7 (*C*(4')), 125.9 (*C*(2') and *C*(6')), 109.9 (*C*(1)), 103.2 (*C*(3) and *C*(5)), 85.0 (*C*NO_2), 65.8 (*C*HC(1')), 55.4 (O*C*H_3), 34.5 (C(=O)*C*H_2), 34.1 (C(1)*C*H_2); *m/z* (ESI$^+$) 357 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{19}H_{20}N_2NaO_5^+$, ([M+Na]$^+$) requires 379.1264, found 379.1261.

1-(2,6-Dimethoxybenzyl)-5-isopropyl-4-nitropyrrolidin-2-one **243**

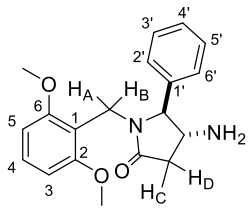


243

2,6-Dimethoxybenzylamine (1.3 g, 7.8 mmol, 1.5 eq) and isobutyraldehyde (710 µL, 7.8 mmol, 1.5 eq) were added to a stirred solution of *ester* **241** (690 mg, 5.2 mmol, 1.0 eq) and benzoic acid (950 mg, 7.8 mmol, 1.5 eq) in toluene (20 mL). The reaction mixture was degassed (vacuum) and purged three times with nitrogen. The reaction was stirred at 70 °C for 14 hours. Upon completion the reaction mixture was concentrated under vacuum. The residue was dissolved in CH_2Cl_2 and washed three times with aq. Na_2CO_3 (1 M), dried over Na_2SO_4, filtered and the filtrate concentrated *in vacuo*. The crude product was purified by flash column chromatography (0-60% EtOAc in cyclohexane). Solvent was removed *in vacuo* to afford *lactam* **243** as an off-white powder (1.05 g, 63%). Mp 140.5-142 °C; ν_{max} (FTIR) cm^{-1} 2963, 1699 (C=O), 1551 (C-NO_2), 1257 (C-O), 1110; ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.25 (t, *J* = 8.4 Hz, 1H, C(4)*H*), 6.55 (d, *J* = 8.3 Hz, 2H, C(3)*H* and C(5)*H*), 5.09 (d, *J* = 14.1 Hz, 1H, C($H_A$or H_B)), 4.75 (td, *J* = 1.7, 8.4 Hz, 1H, C*H*NO_2), 4.19 (dd, *J* = 1.2, 14.1 Hz, 1H, C($H_A$or H_B)), 3.82 (s, 6H, OCH_3), 3.49 (dd, *J* = 1.4, 3.5 Hz, 1H, C*H*^{i}Pr), 3.10 (dd, *J* = 1.8, 18.4 Hz, 1H, CH_D), 2.82 (ddd, *J* = 1.4, 8.5, 18.4 Hz, 1H, CH_C), 2.33 (dspt, *J* = 3.6, 7.0 Hz, 1H, C*H*$(CH_3)_2$), 0.99 (d, *J* = 7.0 Hz, 3H, Me$_A$), 0.81 (d, *J* = 6.85 Hz, 3H, Me$_B$); ^{13}C NMR (101 MHz, $CDCl_3$) δ 169.3 (*C*=O), 159.1 (*C*(2) and *C*(6)), 129.7 (*C*(4)), 110.0 (*C*(1)), 103.5 (*C*(3) and *C*(5)), 78.4 (*C*NO_2), 67.3 (*C*^{i}Pr), 55.7 (O*C*H_3), 35.9 (*C*H_2C(=O)),

33.5 (C(1)*C*H$_2$), 28.1 (*C*(CH$_3$)$_2$), 18.2 (Me$_A$), 15.0 (Me$_B$); *m/z* (ESI$^+$) 323 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{16}H_{22}N_2NaO_5^+$, ([M+Na]$^+$) requires 345.1414, found 345.1421.

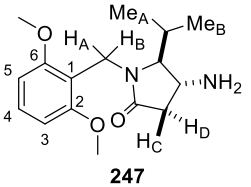
4-Amino-1-(2,6-dimethoxybenzyl)-5-phenylpyrrolidin-2-one **246**



246

Raney nickel solution (~20 mg, excess) was added to a solution of *lactam* **242** (60 mg, 0.17 mmol, 1.0 eq) in MeOH (5 ml). The mixture was flushed with H_2 and stirred under H_2 at room temperature for 26 h. The black mixture was filtered through celite, and the celite was washed with additional MeOH (100 mL). The pale yellow solution was concentrated *in vacuo* to afford *amine* **246** as a yellow gum (52 mg, 95%). ν_{max} (FTIR) cm^{-1} 2935, 1671 (C=O), 1593 (NH$_2$), 1475, 1112 (C-O), 700 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.20 - 7.32 (m, 3H, C(3')*H*, C(4')*H* and C(5')*H*), 7.16 (t, *J* = 8.3 Hz, 1H, C(4)*H*), 6.90 - 7.02 (m, 2H, C(2')*H* and C(6')*H*), 6.50 (d, *J* = 8.4 Hz, 2H, C(3)*H* and C(5)*H*), 4.71 (d, *J* = 13.6 Hz, 1H, C(*H*$_A$ or *H*$_B$)), 3.95 (d, *J* = 13.5 Hz, 1H, C(*H*$_A$ or *H*$_B$)), 3.77 (d, *J* = 3.6 Hz, 1H, C(1')C*H*), 3.51 (s, 6H, OC*H*$_3$), 3.06 (quin, *J* = 3.8 Hz, 1H, C*H*NH$_2$), 2.58 (dd, *J* = 6.8, 16.5 Hz, 1H, C(*H*$_C$ or *H*$_D$)), 1.97 (dd, *J* = 4.8, 16.5 Hz, 1H, C(*H*$_C$ or *H*$_D$)), (N*H*$_2$ not observed); ^{13}C NMR (101 MHz, DMSO-d_6) δ 172.9 (*C*=O), 159.0 (*C*(2) and *C*(6)), 140.8 (*C*(1')), 129.6 (*C*(4)), 128.6 (*C*(3') and *C*(5')), 127.5 (*C*(4')), 126.5 (*C*(2') and *C*(6')), 111.2 (*C*(1)), 104.0 (*C*(3) and *C*(5)), 70.9 (C(1')*C*H), 55.9 (O*C*H$_3$), 54.9 (*C*NH$_2$), 39.7 (*C*H$_C$H$_D$), 33.5 (*C*H$_A$H$_B$); *m/z* (ESI$^+$) 327 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{19}H_{23}N_2O_3^+$, ([M+H]$^+$) requires 327.1703, found 327.1699.

4-Amino-1-(2,6-dimethoxybenzyl)-5-isopropylpyrrolidin-2-one **247**

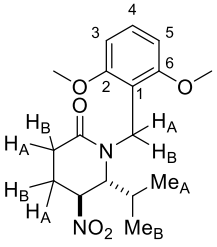


247

Raney nickel solution (~20 mg, excess) was added to a solution of *lactam* **243** (1.0 g, 3.1 mmol, 1.0 eq) in MeOH (50 ml). The mixture was flushed with H_2 and stirred under H_2 at room temperature for 26 h. The black mixture was filtered through celite, and the celite was washed with additional MeOH (100 mL).

The pale yellow solution was concentrated *in vacuo* to afford *amine* **247** as a yellow gum (quantitative), which was used for without further purification. A portion of the product (50 mg) was purified by reverse phase HPLC to afford *amine* **247** as a colourless oil. ν_{max} (FTIR) cm^{-1} 2959, 1664 (C=O), 1594 (NH_2), 1475, 1253 (C-O), 1112; ^1H NMR (400 MHz, $CDCl_3$) δ 7.24 (t, J = 8.4 Hz, 1H, C(4)*H*), 6.57 (d, J = 8.3 Hz, 2H, C(3)*H* and C(5)*H*), 5.08 (d, J = 13.9 Hz, 1H, C(H_A or H_B)), 4.20 (dd, J = 1.3, 14.0 Hz, 1H, C(H_A or H_B)), 3.84 (s, 6H, OCH_3), 3.25 (d, J = 7.2 Hz, 1H, C*H*NH_2), 2.76 - 2.81 (m, 1H, C*H*^{i}Pr), 2.74 (ddd, J = 1.3, 7.2, 17.4 Hz, 1H, CH_C), 2.10 - 2.18 (m, 1H, C*H*$(CH_3)_2$), 2.09 (dd, J = 1.41, 17.42 Hz, 1H, CH_D), 0.93 (d, J = 6.97 Hz, 3H, Me_A), 0.78 (d, J = 6.97 Hz, 3H, Me_B); ^{13}C NMR (101 MHz, $CDCl_3$) δ 172.4 (*C*=O), 159.2 (*C*(2) and *C*(6)), 129.3 (*C*(4)), 111.8 (*C*(1)), 103.7 (*C*(3) and *C*(5)), 72.1 (*C*H^{i}Pr), 55.8 (O*C*H_3), 45.2 (*C*HNH_2), 42.3 (*C*H_CH_D), 33.4 (C(1)*C*H_2), 28.2 (*C*HMe_2), 18.7 (Me_A), 15.8 (Me_B); *m/z* (ESI^+) 293 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_{16}H_{25}N_2O_3^+$, ($[M+H]^+$) requires 293.1860, found 293.1854.

1-(2,6-Dimethoxybenzyl)-6-isopropyl-5-nitropiperidin-2-one **249**

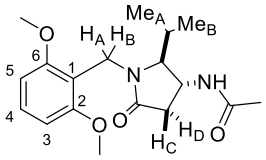


249

2,6-Dimethoxybenzylamine (84 mg, 0.50 mmol, 1.0 eq) and isobutyraldehyde (46 µL, 0.50 mmol, 1.0 eq) were added to a stirred solution of methyl 4-nitrobutanoate (64 µl, 0.50 mmol, 1.0 eq) and benzoic acid (92 mg, 0.75 mmol, 1.5 eq) in toluene (3.0 mL). The reaction mixture was degassed (vacuum) and purged three times with nitrogen. The reaction was stirred at 70 °C for 48 hours. Upon completion the reaction mixture was concentrated under vacuum. The residue was dissolved in CH_2Cl_2 and washed three times with aq. Na_2CO_3 (1 M), dried over Na_2SO_4, filtered and the filtrate concentrated *in vacuo*. The crude product was purified by flash column chromatography (10-80% EtOAc in cyclohexane). Solvent was removed *in vacuo* to afford *lactam* **249** as an amorphous yellow solid (28 mg, 17%). ν_{max} (FTIR) cm^{-1} 2968, 1648 (C=O), 1594, 1550 (C-

NO_2), 1260 (C-O), 1111, 780 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.20 (t, *J* = 8.4 Hz, 1H, C(4)*H*), 6.48 (d, *J* = 8.4 Hz, 2H, C(3)*H* and C(5)*H*), 5.42 (d, *J* = 13.8 Hz, 1H, C(1)CH_A), 4.46 - 4.56 (m, 1H, C*H*NO_2), 4.04 (d, *J* = 13.9 Hz, 1H, C(1)CH_A), 3.72 (s, 6H, OCH_3), 3.58 (dd, *J* = 2.6, 7.7 Hz, 1H, C*H*^{i}Pr), 2.59 - 2.70 (m, 1H, C$H_A$$H_B$C(=O)), 2.29 - 2.41 (m, 2H, CH$_A$$H_B$C(=O) and C$H_A$$H_BCHNO_2$), 2.08 - 2.19 (m, 1H, CH$_A$$H_BCHNO_2$), 1.98 - 2.06 (m, 1H, C*H*Me_2), 0.99 (d, *J* = 6.9 Hz, 3H, Me_A), 0.90 (d, *J* = 6.7 Hz, 3H, Me_B); ^{13}C NMR (101 MHz, $CDCl_3$) δ 169.1 (*C*=O), 159.4 (*C*(2) and *C*(6)), 129.8 (*C*(4)), 110.7 (*C*(1)), 103.4 (*C*(3) and *C*(5)), 80.8 (*C*HNO_2), 61.7 (*C*H^{i}Pr), 55.5 (OCH_3), 37.7 (C(1)CH_2), 31.9 (*C*HMe_2), 27.4 (*C*H_2C(=O)), 21.8 (*C*H_2CHNO_2), 19.1 (Me_A), 18.7 (Me_B); *m/z* (ESI$^+$) 337 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{17}H_{25}N_2O_5^+$, ([M+H]$^+$) requires 337.1758, found 337.1749.

N-(1-(2,6-Dimethoxybenzyl)-2-isopropyl-5-oxopyrrolidin-3-yl)acetamide **250**

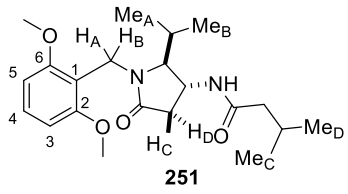


250

Following *General Procedure 2*, acetyl chloride (13 µL, 0.18 mmol, 1.05 eq) was treated with *amine* **247** (50 mg, 0.17 mmol, 1.0 eq) and triethylamine (72 µL, 0.51 mmol, 3.0 eq) and the resulting mixture purified by flash column chromatography (0-8% MeOH in CH_2Cl_2) to afford *amide* **250** as off-white crystals (36 mg, 63%). Mp 166-167 °C; ν_{max} (FTIR) cm^{-1} 3286 (CONH), 2953, 1654 (lactam C=O), 1260 (C-O), 1114, 732 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.28 (t, *J* = 8.4 Hz, 1H, C(4)*H*), 6.61 (d, *J* = 8.4 Hz, 2H, C(3)*H* and C(5)*H*), 5.62 (br. d, *J* = 7.8 Hz, 1H, N*H*), 5.07 (d, *J* = 14.1 Hz, 1H, C(H_A or H_B)), 4.29 (t, *J* = 7.8 Hz, 1H, C*H*NH), 4.24 (dd, *J* = 1.1, 14.0 Hz, 1H, C(H_A or H_B)), 3.86 (s, 6H, OCH_3), 2.84 (d, *J* = 4.4 Hz, 1H, C*H*^{i}Pr), 2.80 (ddd, *J* = 1.2, 7.4, 17.8 Hz, 1H, CH_C), 2.14 (dd, *J* = 1.0, 17.6 Hz, 1H, CH_D), 2.03 - 2.11 (m, 1H, C*H*Me_2), 1.85 (s, 3H, C(=O)CH_3), 0.97 (d, *J* = 7.0 Hz, 3H, Me_A), 0.85 (d, *J* = 7.0 Hz, 3H, Me_B); ^{13}C NMR (101 MHz, $CDCl_3$) δ 172.1 (lactam *C*=O), 168.8 (NH*C*(=O)CH_3), 159.1 (*C*(2) and *C*(6)), 129.5 (*C*(4)), 112.0 (*C*(1)), 104.2 (*C*(3) and *C*(5)), 69.8 (*C*^{i}Pr), 56.0 (O*C*H_3), 44.5 (*C*HNH), 39.2 (*C*$H_C$$H_D$), 34.0 (*C*$H_A$$H_B$), 28.9 (*C*H$Me_2$), 23.3 (C(=O)*C*$H_3$), 18.8 ($Me_A$), 16.5 ($Me_B$); *m/z*

(ESI$^+$) 335 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{18}H_{26}N_2NaO_4^+$, ([M+Na]$^+$) requires 357.1782, found 357.1785.

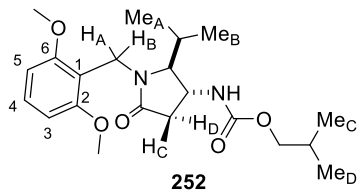
N-(1-(2,6-Dimethoxybenzyl)-2-isopropyl-5-oxopyrrolidin-3-yl)-3-methylbutanamide **251**



251

Following *General Procedure 2*, 3-methylbutanoyl chloride (22 µL, 0.18 mmol, 1.05 eq) was treated with *amine* **247** (50 mg, 0.17 mmol, 1.0 eq) and triethylamine (72 µL, 0.51 mmol, 3.0 eq) and the resulting mixture purified by flash column chromatography (0-10% MeOH in CH_2Cl_2) to afford *amide* **251** as an off-white powder (52 mg, 81%). Mp 160.5-162 °C; ν_{max} (FTIR) cm^{-1} 3294 (CONH), 2959, 1655 (lactam C=O), 1258 (C-O), 1114, 779 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.18 (t, *J* = 8.3 Hz, 1H, C(4)*H*), 6.50 (d, *J* = 8.3 Hz, 2H, C(3)*H* and C(5)*H*), 5.49 (br. d, *J* = 7.80 Hz, 1H, N*H*), 4.99 (d, *J* = 14.1 Hz, 1H, C(H_A or H_B)), 4.22 (t, *J* = 7.8 Hz, 1H, C*H*NH), 4.12 (dd, *J* = 1.0, 14.0 Hz, 1H, C(H_A or H_B)), 3.76 (s, 6H, OCH_3), 2.76 (d, *J* = 4.6 Hz, 1H, C*H*^{i}Pr), 2.72 (ddd, *J* = 1.0, 7.6, 17.6 Hz, 1H, CH_C), 1.97 - 2.08 (m, 2H, CH_D and C*H*Me$_A$Me$_B$), 1.83 - 1.96 (m, 1H, C*H*Me$_C$Me$_D$), 1.75 - 1.81 (m, 2H, CH_2^{i}Pr), 0.89 (d, *J* = 7.1 Hz, 3H, Me$_A$), 0.72 - 0.81 (m, 9H, Me$_B$, Me$_C$ and Me$_D$); 13C NMR (101 MHz, $CDCl_3$) δ 172.0 (lactam *C*=O), 171.3 (NH*C*(=O)CH$_2$), 159.1 (*C*(2) and *C*(6)), 129.5 (*C*(4)), 111.9 (*C*(1)), 104.1 (*C*(3) and *C*(5)), 70.0 (*C*H^{i}Pr), 56.0 (O*C*H$_3$), 46.0 (*C*H$_2$^{i}Pr), 44.1 (*C*HNH), 39.2 (*C*H$_C$H$_D$), 33.7 (*C*H$_A$H$_B$), 28.7 (*C*Me$_A$Me$_B$), 26.1 (*C*Me$_C$Me$_D$), 22.4 (Me$_C$ and Me$_D$), 18.8 (Me$_A$), 16.3 (Me$_B$); *m/z* (ESI$^+$) 377 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{21}H_{33}N_2O_4^+$, ([M+H]$^+$) requires 377.2435, found 377.2419.

Isobutyl (1-(2,6-dimethoxybenzyl)-2-isopropyl-5-oxopyrrolidin-3-yl)carbamate **252**

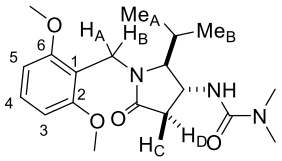


252

Following *General Procedure 2*, isobutyl carbonochloridate (23 µL, 0.18 mmol, 1.05 eq) was treated with *amine* **247** (50 mg, 0.17 mmol, 1.0 eq) and triethylamine (72 µL, 0.51 mmol, 3.0 eq) and the resulting mixture purified by flash column chromatography (0-10% MeOH in CH_2Cl_2) to afford

amide **252** as an orange gum (33 mg, 49%). ν_{max} (FTIR) cm^{-1} 3299 (CONH), 2960, 1663 (lactam C=O), 1247 (C-O), 1113, 777 (aromatic CH); ^1H NMR (400 MHz, $CDCl_3$) d 7.26 (t, *J* = 8.4 Hz, 1H, C(4)*H*), 6.59 (d, *J* = 8.4 Hz, 2H, C(3)*H* and C(5)*H*), 5.06 (d, *J* = 14.1 Hz, 1H, C(H_A or H_B)), 4.78 (br. d, *J* = 7.4 Hz, 1H, N*H*), 4.24 (d, *J* = 14.2 Hz, 1H, C(H_A or H_B)), 4.03 (t, *J* = 7.4 Hz, 1H, C*H*NH), 3.85 (s, 6H, OCH_3), 3.72 - 3.84 (m, 2H, $OCH_2{}^iPr$), 2.85 (d, *J* = 4.5 Hz, 1H, C*H*^{i}Pr), 2.80 (dd, *J* = 7.4, 17.8 Hz, 1H, CH_C), 2.20 (d, *J* = 17.6 Hz, 1H, CH_D), 2.00 - 2.13 (m, 1H, C*H*Me_AMe_B), 1.80 - 1.94 (m, 1H, C*H*Me_CMe_D), 0.96 (d, *J* = 6.9 Hz, 3H, Me_A), 0.90 (d, *J* = 6.7 Hz, 6H, Me_C and Me_D), 0.85 (d, *J* = 6.97 Hz, 3H, Me_B); ^{13}C NMR (101 MHz, $CDCl_3$) δ 171.9 (lactam *C*=O), 159.1 (*C*(2) and *C*(6)), 129.4 (*C*(4)), 111.5 (*C*(1)), 103.9 (*C*(3) and *C*(5)), 71.0 ($OCH_2{}^iPr$), 69.8 (*C*H^{i}Pr), 55.8 (OCH_3), 46.2 (*C*HNH), 39.5 (CH_CH_D), 34.1 (CH_AH_B), 28.9 (CMe_AMe_B), 28.0 (CMe_CMe_D), 19.0 (Me_C and Me_D), 18.8 (Me_A), 16.6 (Me_B), (*N.B.* Carbamate *C* not observed); *m/z* (ESI^+) 393 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_{21}H_{33}N_2O_5^+$, ($[M+H]^+$) requires 393.2384, found 393.2366.

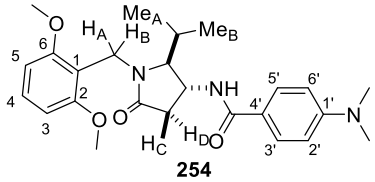
3-(1-(2,6-Dimethoxybenzyl)-2-isopropyl-5-oxopyrrolidin-3-yl)-1,1-dimethylurea **253**



Following *General Procedure 2*, dimethylcarbamic chloride (17 µL, 0.18 mmol, 1.05 eq) was treated with *amine* **247** (50 mg, 0.17 mmol, 1.0 eq) and triethylamine (72 µL, 0.51 mmol, 3.0 eq) and the resulting mixture purified by flash column chromatography (0-10% MeOH in CH_2Cl_2) to afford *amide* **253** as a white powder (34 mg, 55%). Mp 197-199 °C; ν_{max} (FTIR) cm^{-1} 3305 (CONH), 2936, 1680 (urea C=O), 1645 (lactam C=O), 1244 (C-O), 1115, 784 (aromatic CH); ^1H NMR (400 MHz, $CDCl_3$) δ 7.25 (t, *J* = 8.4 Hz, 1H, C(4)*H*), 6.58 (d, *J* = 8.3 Hz, 2H, C(3)*H* and C(5)*H*), 5.08 (d, *J* = 14.0 Hz, 1H, C(H_A or H_B)), 4.43 (br. d, *J* = 8.0 Hz, 1H, N*H*), 4.17 - 4.25 (m, 2H, C(H_A or H_B) and C*H*NH), 3.84 (s, 6H, OCH_3), 2.87 (d, *J* = 4.2 Hz, 1H, C*H*^{i}Pr), 2.79 - 2.85 (m, 1H, CH_C), 2.78 (s, 6H, $N(CH_3)_2$), 2.16 (d, *J* = 17.7 Hz, 1H, CH_D), 2.05 - 2.12 (m, 1H, C*H*Me_2), 0.99 (d, *J* = 7.1 Hz, 3H, Me_A), 0.86 (d, *J* = 6.9 Hz, 3H, Me_B); ^{13}C NMR (101 MHz,

$CDCl_3$) δ 172.4 (lactam *C*=O), 159.2 (*C*(2) and *C*(6)), 157.2 (N*C*(=O)N), 129.4 (*C*(4)), 112.5 (*C*(1)), 104.4 (*C*(2) and *C*(6)), 70.3 (*C*H^{i}Pr), 56.2 (OCH_3), 45.8 (*C*HNH), 40.2 (*C*$H_C$$H_D$), 35.9 (N(*C*$H_3$)$_2$), 33.7 (*C*$H_A$$H_B$), 28.9 (C*H*$Me_2$), 18.9 ($Me_A$), 16.5 ($Me_B$); *m/z* (ESI$^+$) 364 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{19}H_{30}N_3O_4^+$, ([M+H]$^+$) requires 364.2231, found 364.2216.

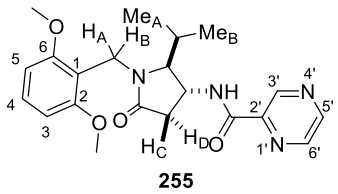
N-(1-(2,6-Dimethoxybenzyl)-2-isopropyl-5-oxopyrrolidin-3-yl)-4-(dimethylamino)benzamide **254**



254

Following *General Procedure 2*, 4-(dimethylamino)benzoyl chloride (33 mg, 0.18 mmol, 1.05 eq) was treated with *amine* **247** (50 mg, 0.17 mmol, 1.0 eq) and triethylamine (72 µL, 0.51 mmol, 3.0 eq) and the resulting mixture purified by flash column chromatography (0-10% MeOH in CH_2Cl_2). Fractions containing the product were further purified using a Biotage SNAP KP-NH column (100% MeOH) and the solvent removed *in vacuo* to afford *amide* **254** as a white powder (59 mg, 78%). Mp 216.5-218.5 °C; ν_{max} (FTIR) cm^{-1} 3348 (CONH), 2932, 1661 (lactam C=O), 1605 (amide C=O), 1508, 1108, 767 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.50 (d, *J* = 9.0 Hz, 2H, C(3')*H* and C(5')*H*), 7.25 (t, *J* = 8.4 Hz, 1H, C(4)*H*), 6.64 (d, *J* = 9.0 Hz, 2H, C(2')*H* and C(6')*H*), 6.58 (d, *J* = 8.3 Hz, 2H, C(3)*H* and C(5)*H*), 6.03 (br. d, *J* = 8.0 Hz, 1H, N*H*), 5.10 (d, *J* = 14.0 Hz, 1H, C(H_A or H_B)), 4.50 (t, *J* = 7.6 Hz, 1H, C*H*NH), 4.27 (dd, *J* = 1.0, 14.1 Hz, 1H, C(H_A or H_B)), 3.83 (s, 6H, OCH_3), 3.05 (s, 6H, N(CH_3)$_2$), 2.97 (d, *J* = 4.3 Hz, 1H, C*H*^{i}Pr), 2.89 (ddd, *J* = 1.0, 7.4, 17.6 Hz, 1H, CH_C), 2.27 (d, *J* = 17.6 Hz, 1H, CH_D), 2.06 - 2.17 (m, 1H, C*H*Me_2), 1.02 (d, *J* = 7.0 Hz, 3H, Me_A), 0.91 (d, *J* = 6.9 Hz, 3H, Me_B); ^{13}C NMR (101 MHz, $CDCl_3$) δ 172.3 (lactam *C*=O), 166.4 (C(4')*C*=O), 159.2 (*C*(2) and *C*(6)), 152.5 (*C*(1')), 129.5 (*C*(4)), 128.2 (*C*(3') and *C*(5')), 120.9 (*C*(4')), 112.2 (*C*(1)), 110.9 (*C*(2') and *C*(6')), 104.3 (*C*(3) and *C*(5)), 69.9 (*C*H^{i}Pr), 56.2 (OCH_3), 44.8 (*C*HNH), 40.1 (N(*C*H_3)$_2$), 39.6 (*C*$H_C$$H_D$), 34.0 (*C*$H_A$$H_B$), 29.1 (*C*H$Me_2$), 18.9 ($Me_A$), 16.6 ($Me_B$); *m/z* (ESI$^+$) 440 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{25}H_{34}N_3O_4^+$, ([M+H]$^+$) requires 440.2544, found 440.2535.

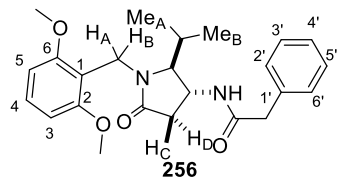
N-(1-(2,6-Dimethoxybenzyl)-2-isopropyl-5-oxopyrrolidin-3-yl)pyrazine-2-carboxamide **255**



255

Following *General Procedure 2*, pyrazine-2-carbonyl chloride (26 mg, 0.18 mmol, 1.05 eq) was treated with *amine* **247** (50 mg, 0.17 mmol, 1.0 eq) and triethylamine (72 µL, 0.51 mmol, 3.0 eq) and the resulting mixture purified by flash column chromatography (0-10% MeOH in CH_2Cl_2). Fractions containing the product were washed with saturated aq. $NaHCO_3$ and water, and the organic layer concentrated *in vacuo* to afford *amide* **255** as an orange gum (42 mg, 62%). ν_{max} (FTIR) cm^{-1} 2959, 1667 (lactam C=O), 1257, 1111, 728 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.27 (d, *J* = 1.4 Hz, 1H, C(3')*H*), 8.68 (d, *J* = 2.5 Hz, 1H, C(6')*H*), 8.42 (dd, *J* = 1.5, 2.4 Hz, 1H, C(5')*H*), 7.75 (br. d, *J* = 8.4 Hz, 1H, N*H*), 7.11 (t, *J* = 8.4 Hz, 1H, C(4)*H*), 6.45 (d, *J* = 8.4 Hz, 2H, C(3)*H* and C(5)*H*), 5.00 (d, *J* = 14.1 Hz, 1H, C(H_A or H_B)), 4.44 (t, *J* = 8.1 Hz, 1H, C*H*NH), 4.20 (dd, *J* = 1.0, 14.1 Hz, 1H, C(H_A or H_B)), 3.78 (s, 6H, OCH_3), 2.91 (d, *J* = 3.6 Hz, 1H, C*H*^{i}Pr), 2.84 (ddd, *J* = 1.0, 7.4, 17.8 Hz, 1H, CH_C), 2.22 (dd, *J* = 1.0, 17.8 Hz, 1H, CH_D), 1.98 - 2.11 (m, 1H, C*H*Me_2), 0.91 (d, *J* = 7.0 Hz, 3H, Me_A), 0.83 (d, *J* = 6.9 Hz, 3H, Me_B); ^{13}C NMR (101 MHz, $CDCl_3$) δ 171.8 (lactam *C*=O), 161.8 (C(2')*C*=O), 159.1 (*C*(2) and *C*(6)), 147.4 (*C*(6')), 144.5 (*C*(3')), 144.0 (*C*(2')), 142.3 (*C*(5')), 129.3 (*C*(4)), 111.4 (*C*(1)), 103.7 (*C*(3) and *C*(5)), 69.6 (*C*H^{i}Pr), 55.9 (O*C*H_3), 44.1 (*C*HNH), 39.3 (*C*H_CH_D), 34.1 (*C*H_AH_B), 28.9 (*C*HMe_2), 18.8 (Me_A), 16.3 (Me_B); *m/z* (ESI$^+$) 399 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{21}H_{27}N_4O_4^+$, ([M+H]$^+$) requires 399.2027, found 399.2015.

N-(1-(2,6-Dimethoxybenzyl)-2-isopropyl-5-oxopyrrolidin-3-yl)-2-phenylacetamide **256**

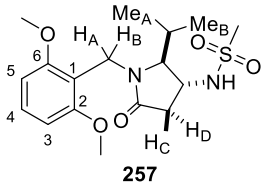


256

Following *General Procedure 2*, 2-phenylacetyl chloride (24 µL, 0.18 mmol, 1.05 eq) was treated with *amine* **247** (50 mg, 0.17 mmol, 1.0 eq) and triethylamine (72 µL, 0.51 mmol, 3.0 eq) and the resulting mixture purified by flash column chromatography (0-10% MeOH in CH_2Cl_2). Fractions

containing the product were further purified using a Biotage SNAP KP-NH column (100% MeOH) and the solvent removed *in vacuo* to afford *amide* **256** as a yellow gum (38 mg, 54%). ν_{max} (FTIR) cm^{-1} 3281 (CONH), 2958, 1649 (lactam C=O), 1246, 1112, 785 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.13 (br. d, *J* = 6.8 Hz, 1H, N*H*), 7.16 - 7.24 (m, 3H, C(4)*H*, C(3')*H* and C(5')*H*), 7.11 - 7.16 (m, 1H, C(4')*H*), 7.04 - 7.08 (m, 2H, C(2')*H* and C(6')*H*), 6.59 (d, *J* = 8.4 Hz, 2H, C(3)*H* and C(5)*H*), 4.72 (d, *J* = 13.7 Hz, 1H, C(H_A or H_B)), 3.87 - 4.02 (m, 2H, C(H_A or H_B) and C*H*NH), 3.67 (s, 6H, OCH_3), 3.24 (d, *J* = 1.8 Hz, 2H, C(1')CH_2), 2.90 (t, *J* = 3.0 Hz, 1H, C*H*^{i}Pr), 2.50 (dd, *J* = 8.2, 17.6 Hz, 1H, CH_C), 1.87 - 2.02 (m, 2H, CH_D and C*H*Me_2), 0.72 (d, *J* = 7.1 Hz, 3H, Me_A), 0.61 (d, *J* = 7.0 Hz, 3H, Me_B); ^{13}C NMR (101 MHz, DMSO-d_6) δ 171.5 (lactam *C*=O), 169.8 (NH*C*=O), 159.2 (*C*(2) and *C*(6)), 136.7 (*C*(1')), 129.8 (*C*(4)), 129.3 (*C*(2') and *C*(6')), 128.6 (*C*(3') and *C*(5')), 126.7 (*C*(4')), 111.6 (*C*(1)), 104.4 (*C*(3) and *C*(5)), 68.6 (*C*H^{i}Pr), 56.2 (O*C*H_3), 43.0 (*C*HNH), 42.5 (C(1')*C*H_2), 38.1 (*C*$H_C$$H_D$), 33.2 (*C*$H_A$$H_B$), 27.8 (*C*H$Me_2$), 18.7 ($Me_A$), 15.4 ($Me_B$); *m/z* (ESI$^+$) 411 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{24}H_{31}N_2O_4^+$, ([M+H]$^+$) requires 411.2278, found 411.2263.

N-(1-(2,6-Dimethoxybenzyl)-2-isopropyl-5-oxopyrrolidin-3-yl)methanesulfonamide **257**



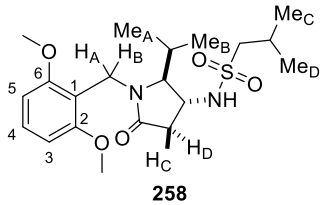
257

Following *General Procedure 2*, methanesulfonyl chloride (14 µL, 0.18 mmol, 1.05 eq) was treated with *amine* **247** (50 mg, 0.17 mmol, 1.0 eq) and triethylamine (72 µL, 0.51 mmol, 3.0 eq). The crude product was washed with saturated aq. NH_4Cl and the resulting mixture purified by flash column chromatography (0-10% MeOH in CH_2Cl_2) to afford *amide* **257** as an orange powder (37 mg, 58%). Mp 184-185 °C; ν_{max} (FTIR) cm^{-1} 3108 (SO_2NH), 2944, 1656 (lactam C=O), 1324 (SO_2N), 1113, 783 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.29 (t, *J* = 8.40 Hz, 1H, C(4)*H*), 6.63 (d, *J* = 8.4 Hz, 2H, C(3)*H* and C(5)*H*), 5.09 (d, *J* = 14.1 Hz, 1H, C(H_A or H_B)), 4.61 (br. d, *J* = 9.0 Hz, 1H, N*H*), 4.25 (dd, *J* = 1.0, 14.0 Hz, 1H, C(H_A or H_B)), 3.87 (s, 6H, OCH_3), 3.83 (d, *J* = 8.1 Hz, 1H, C*H*NH), 2.96 (d, *J* = 5.0

Hz, 1H, C*H*^{i}Pr), 2.85 (ddd, *J* = 1.0, 7.0, 17.6 Hz, 1H, CH_C), 2.76 (s, 3H, SO_2CH_3), 2.31 (d, *J* = 17.6 Hz, 1H, CH_D), 1.97 - 2.13 (m, 1H, C*H*Me_2), 1.01 (d, *J* = 7.0 Hz, 3H, Me_A), 0.87 (d, *J* = 6.9 Hz, 3H, Me_B); ^{13}C NMR (101 MHz, $CDCl_3$) δ 171.3 (*C*=O), 159.0 (*C*(2) and *C*(6)), 129.7 (*C*(4)), 111.6 (*C*(1)), 104.3 (*C*(3) and *C*(5)), 70.0 (*C*H^{i}Pr), 56.1 (O*C*H_3), 49.4 (*C*HNH), 41.9 (SO_2CH_3), 39.8 (CH_CH_D), 34.3 (CH_AH_B), 29.2 (*C*HMe_2), 18.9 (Me_A), 17.0 (Me_B); *m/z* (ESI$^+$) 371 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{17}H_{27}N_2O_5S^+$, ([M+H]$^+$) requires 371.1635, found 371.1623.

N-(1-(2,6-Dimethoxybenzyl)-2-isopropyl-5-oxopyrrolidin-3-yl)-2-methylpropane-1-sulfonamide

258

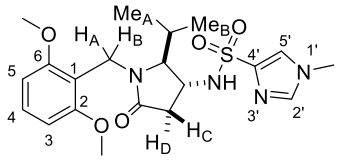


258

Following *General Procedure 2*, 2-methylpropane-1-sulfonyl chloride (23 µL, 0.18 mmol, 1.05 eq) was treated with *amine* **247** (50 mg, 0.17 mmol, 1.0 eq) and triethylamine (72 µL, 0.51 mmol, 3.0 eq). The crude product was washed with saturated aq. NH_4Cl and the resulting mixture purified by flash column chromatography (0-10% MeOH in CH_2Cl_2) to afford *amide* **258** as an orange powder (48 mg, 68%). Mp 152.5-153.5 °C; ν_{max} (FTIR) cm^{-1} 2956, 1658 (lactam C=O), 1319 (SO_2N), 1110, 775 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.29 (t, *J* = 8.4 Hz, 1H, C(4)*H*), 6.62 (d, *J* = 8.4 Hz, 2H, C(3)*H* and *C*(5)*H*), 5.09 (d, *J* = 14.1 Hz, 1H, C(H_A or H_B)), 4.51 (d, *J* = 9.3 Hz, 1H, N*H*), 4.24 (dd, *J* = 1.1, 14.1 Hz, 1H, C(H_A or H_B)), 3.86 (s, 6H, OCH_3), 3.78 - 3.84 (m, 1H, C*H*NH), 2.96 (d, *J* = 5.0 Hz, 1H, C*H*^{i}Pr), 2.83 (ddd, *J* = 1.2, 7.1, 17.6 Hz, 1H, CH_C), 2.72 (dd, *J* = 6.6, 7.7 Hz, 2H, SO_2CH_2), 2.31 (d, *J* = 17.6 Hz, 1H, CH_D), 1.97 - 2.14 (m, 2H, C*H*Me_AMe_B and C*H*Me_CMe_D), 1.05 (d, *J* = 6.7 Hz, 3H, Me_C), 1.01 (d, *J* = 6.7 Hz, 2H, Me_D), 1.00 (d, *J* = 7.0 Hz, 1H, Me_A), 0.86 (d, *J* = 7.0 Hz, 3H, Me_B); ^{13}C NMR (101 MHz, $CDCl_3$) δ 171.3 (*C*=O), 159.0 (*C*(2) and *C*(6)), 129.7 (*C*(4)), 111.5 (*C*(1)), 104.3 (*C*(3) and *C*(5)), 70.4 (*C*H^{i}Pr), 62.0 (SO_2CH_2), 56.0 (OCH_3), 49.2 (*C*HNH), 40.1 (CH_CH_D), 34.3 (CH_AH_B), 29.3 (*C*HMe_AMe_B), 24.9 (*C*HMe_CMe_D), 22.6 (Me_C or Me_D), 22.5 (Me_C or Me_D), 19.0 (Me_A), 17.0 (Me_B);

m/z (ESI$^+$) 413 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{20}H_{33}N_2O_5S^+$, ([M+H]$^+$) requires 413.2105, found 413.2088.

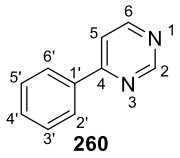
N-(1-(2,6-Dimethoxybenzyl)-2-isopropyl-5-oxopyrrolidin-3-yl)-1-methyl-1*H*-imidazole-4-sulfonamide **259**



259

Following *General Procedure 2*, 1-methyl-1*H*-imidazole-4-sulfonyl chloride (32 mg, 0.18 mmol, 1.05 eq) was treated with *amine* **247** (50 mg, 0.17 mmol, 1.0 eq) and triethylamine (72 µL, 0.51 mmol, 3.0 eq). The crude product was washed with saturated aq. NH_4Cl and the resulting mixture purified by flash column chromatography (0-10% MeOH in CH_2Cl_2) to afford *amide* **259** as a white powder (12 mg, 16%). Mp 224-226 °C; ν_{max} (FTIR) cm^{-1} 3123 (SO_2NH), 2962, 1677 (lactam C=O), 1324 (SO_2N), 1110, 739 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.32 (d, *J* = 1.3 Hz, 1H, C(5')*H*), 7.31 (d, *J* = 1.4 Hz, 1H, C(2')*H*), 7.18 (t, *J* = 8.4 Hz, 1H, C(4)*H*), 6.50 (d, *J* = 8.4 Hz, 2H, C(3)*H* and C(5)*H*), 5.05 (d, *J* = 8.9 Hz, 1H, N*H*), 4.94 (d, *J* = 14.2 Hz, 1H, C(H_A or H_B)), 4.09 (dd, *J* = 1.0, 14.2 Hz, 1H, C(H_A or H_B)), 3.77 (s, 6H, OCH_3), 3.66 - 3.72 (m, 1H, C*H*NH), 3.65 (s, 3H, N(1')CH_3), 2.96 (dd, *J* = 1.4, 4.3 Hz, 1H, C*H*^{i}Pr), 2.47 (ddd, *J* = 1.1, 7.5, 17.8 Hz, 1H, CH_C), 2.12 (dd, *J* = 1.7, 17.7 Hz, 1H, CH_D), 1.92 - 2.03 (m, 1H, C*H*Me$_2$), 0.78 (d, *J* = 7.0 Hz, 3H, Me$_A$), 0.67 (d, *J* = 6.9 Hz, 3H, Me$_B$); ^{13}C NMR (101 MHz, $CDCl_3$) δ 171.6 (*C*=O), 159.0 (*C*(2) and *C*(6)), 140.8 (*C*(4')), 139.1 (*C*(5')), 129.4 (*C*(4)), 123.6 (*C*(2')), 111.3 (*C*(1)), 104.0 (*C*(3) and *C*(5)), 69.5 (*C*H^{i}Pr), 56.0 (O*C*H$_3$), 48.5 (*C*HNH), 39.2 (*C*H$_C$H$_D$), 34.0 (*C*H$_A$H$_B$ and N(1')*C*H$_3$), 28.7 (*C*HMe$_2$), 18.4 (Me$_A$), 16.2 (Me$_B$); *m/z* (ESI$^+$) 437 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{20}H_{29}N_4O_5S^+$, ([M+H]$^+$) requires 437.1853, found 437.1847.

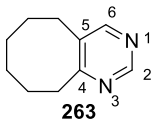
4-Phenylpyrimidine **260**



260

Following *General Procedure 6*, acetophenone (49 µL, 0.4 mmol, 1.0 eq) was treated with ammonium acetate (64 mg, 0.8 mmol, 2.0 eq), triethylorthoformate (208 µL, 1.2 mmol, 3 eq) and zinc chloride (6 mg, 0.04 mmol, 0.1 eq). The crude product was purified by flash column chromatography (0%-30% EtOAc in cyclohexane) to afford *pyrimidine* **260** (23 mg, 36%) as an off-white powder. Mp 67-68 °C (Lit. (Sasada *et al.* 2009) 57.8-58.5 °C); ν_{max} (FTIR) cm^{-1} 1567, 1541, 1390, 741 (aromatic CH), 687 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.29 (d, *J* = 1.1 Hz, 1H, C(2)*H*), 8.78 (d, *J* = 5.4 Hz, 1H, C(5)*H*), 8.05 - 8.16 (m, 2H, C(2')*H* & C(6')*H*), 7.75 (dd, *J* = 1.3, 5.4 Hz, 1H, C(6)*H*), 7.48 - 7.57 (m, 3H, C(3')*H*, C(4')*H* & C(5')*H*); ^{13}C NMR (101 MHz, $CDCl_3$) δ 164.3 (*C*(4)), 158.7 (*C*(2)), 157.0 (*C*(5)), 136.3 (*C*(1')), 131.3 (*C*(4')), 129.1 (*C*(3') & *C*(5')), 127.2 (*C*(2') & *C*(6')), 117.1 (*C*(6)); *m/z* (ESI^+) 157 ($[M+H]^+$, 100%); HRMS (ESI^+) $C_{10}H_9N_2^+$, ($[M+H]^+$) requires 157.0760, found 157.0754.

5,6,7,8,9,10-Hexahydrocycloocta[*d*]pyrimidine **263**

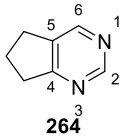


263

Following *General Procedure 6*, cyclooctanone (50 mg, 0.4 mmol, 1.0 eq) was treated with ammonium acetate (61 mg, 0.8 mmol, 2.0 eq), triethylorthoformate (780 µL, 7.1 mmol, 12 eq) and zinc chloride (11 mg, 0.08 mmol, 0.2 eq). The crude product was purified by flash column chromatography (50%-100% EtOAc in *n*-hexane) to afford *pyrimidine* **263** (8 mg, 12%) as a yellow oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.98 (s, 1H, C(2)*H*), 8.40 (s, 1H, C(6)*H*), 2.88 - 2.96 (m, 2H, C(4)CH_2), 2.70 - 2.81 (m, 2H, C(5)CH_2), 1.79 - 1.87 (m, 2H, C(4)CH_2CH_2), 1.68 - 1.77 (m, 2H, C(5)CH_2CH_2), 1.32 - 1.48 (m, 4H, C(4)$(CH_2)_2CH_2CH_2(CH_2)_2$C(5)); ^{13}C NMR (101 MHz, $CDCl_3$) δ 169.4 (*C*(4)), 156.9 (*C*(2)), 156.2 (*C*(6)), 134.0 (*C*(5)), 33.9 (C(4)CH_2), 31.8 (C(5)CH_2CH_2), 29.9 (C(4)CH_2CH_2), 28.9 (C(5)CH_2), 25.7 (C(5)$(CH_2)_2CH_2$), 25.6 (C(4)$(CH_2)_2CH_2$); *m/z* (ESI^+) 181 ($[M+H]^+$,

100%); *m/z* (ESI$^+$) 163 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{10}H_{15}N_2^+$, ([M+H]$^+$) requires 163.1230, not found.

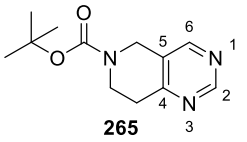
6,7-Dihydro-5*H*-cyclopenta[*d*]pyrimidine **264**



264

Following *General Procedure 6*, cyclopentanone (53 µL, 0.6 mmol, 1.0 eq) was treated with ammonium acetate (92 mg, 1.2 mmol, 2.0 eq), triethylorthoformate (791 µL, 4.8 mmol, 12 eq) and zinc chloride (16 mg, 0.1 mmol, 0.2 eq). The crude product was purified by flash column chromatography (0%-100% EtOAc in cyclohexane) to afford *pyrimidine* **264** (4 mg, 5%) as a yellow oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.97 (s, 1H, C(2)*H*), 8.53 (s, 1H, C(6)*H*), 2.90 - 3.10 (m, 4H, C(4)CH_2 a C(5)CH_2), 2.16 (quin, *J* = 7.7 Hz, 2H, $CH_2CH_2CH_2$); ^{13}C NMR (101 MHz, $CDCl_3$) δ 174.8 (*C*(4)), 156.9 (*C*(2 or 6)), 152.0 (*C*(2 or 6)), 135.0 (*C*(5)), 34.1 (C(4)*C*H_2 or C(5)*C*H_2), 28.4 (C(4)*C*H_2 or C(5)*C*H_2), 22.5 ($CH_2CH_2CH_2$); *m/z* (ESI$^+$) 121 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_7H_9N_2^+$, ([M+H]$^+$) requires 121.0760, not found.

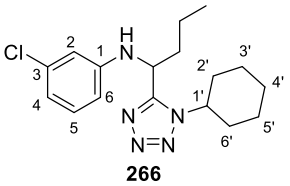
tert-Butyl 7,8-dihydropyrido[4,3-*d*]pyrimidine-6(5*H*)-carboxylate **265**



265

Following *General Procedure 6*, 1-boc-4-piperidone (50 mg, 0.25 mmol, 1.0 eq) was treated with ammonium acetate (39 mg, 0.5 mmol, 2.0 eq), triethylorthoformate (501 µL, 3.0 mmol, 12 eq) and zinc chloride (7 mg, 0.05 mmol, 0.2 eq). The crude product was purified by flash column chromatography (40%-100% EtOAc in *n*-hexane) to afford *pyrimidine* **265** (22 mg, 38%) as a yellow oil. ν_{max} (FTIR) cm^{-1} 2980, 1693 (C=O), 1399 (CMe_3), 1231, 1152; ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.01 (s, 1H, C(2)*H*), 8.48 (s, 1H, C(6)*H*), 4.61 (s, 2H, C(5)CH_2N), 3.77 (t, *J* = 6.0 Hz, 2H, NC$H_2$$CH_2$), 2.94 - 3.01 (m, 2H, C(4)C$H_2$$CH_2$), 1.50 (s, 9H, C(C$H_3$)$_3$); ^{13}C NMR (101 MHz, $CDCl_3$) δ 163.4 (*C*(4)), 156.8 (*C*(2)), 154.5 (*C*(6)), 127.3 (*C*(5)), 80.7 (*C*(CH_3)$_3$), 42.7 (C(5)*C*H_2N), 40.6 (N*C*$H_2$$CH_2$), 31.6 (C(4)*C*$H_2$$CH_2$), 28.4 (C(*C*$H_3$)$_3$), (*C*=O not observed); *m/z* (ESI$^+$) 236 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{12}H_{18}N_3O_2^+$, ([M+H]$^+$) requires 236.1394, found 236.1391.

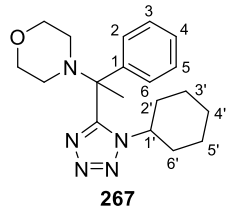
3-Chloro-*N*-(1-(1-cyclohexyl-1*H*-tetrazol-5-yl)butyl)aniline **266**



266

Following *General Procedure 7*, butyrylaldehyde (71 µL, 0.8 mmol, 1.0 eq) was treated with 3-chloroaniline (83 µL, 0.8 mmol, 1.0 eq), cyclohexyl isonitrile (117 µL, 0.9 mmol, 1.2 eq) and trimethylsilyl azide (156 µL, 1.2 mmol, 1.5 eq). The crude product was purified by flash column chromatography (0%-50% EtOAc in cyclohexane) to afford *tetrazole* **266** (112 mg, 43%) as a white powder. Mp 107-109 °C (Lit. (Brigas 2004) 134-135 °C); ν_{max} (FTIR) cm^{-1} 3342 (NH), 2971, 2939, 1597, 1484, 764 (aromatic CH), 684 (C-Cl); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.06 (t, *J* = 8.1 Hz, 1H, C(5)*H*), 6.73 (ddd, *J* = 0.7, 1.8, 8.0 Hz, 1H, C(4)*H*), 6.61 (t, *J* = 2.1 Hz, 1H, C(2)*H*), 6.44 - 6.52 (m, 1H, C(6)*H*), 4.79 (q, *J* = 7.3 Hz, 1H, NHC*H*), 4.33 (tt, *J* = 3.9, 11.6 Hz, 1H, C(1')*H*), 4.22 (d, *J* = 8.1 Hz, 1H, N*H*), 1.69 - 2.09 (m, 9H, mixture of aliphatic CH_2), 1.24 - 1.52 (m, 5H, mixture of aliphatic CH_2), 0.96 (t, *J* = 7.3 Hz, 3H, CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 154.9 (*C*CH), 147.3 (*C*(1)), 135.2 (*C*(3)), 130.6 (*C*(5)), 119.2 (*C*(4)), 113.4(*C*(2)), 112.2 (*C*(6)), 58.4 (*C*(1')), 49.1 (NH*C*H), 36.9 (NHCH*C*H_2), 33.3 (*C*(2' or 6'), 33.2 (*C*(2' or 6')), 25.4 (*C*(3') & *C*(5')), 24.8 (*C*(4')), 19.1 (*C*H_2CH$_3$), 13.7 (CH$_2$*C*H$_3$); *m/z* (ESI^+) 334 ([M(^{35}Cl)+H]$^+$, 100%), 336 ([M^{37}Cl+H]$^+$, 37%); HRMS (ESI^+) $C_{17}H_{25}{}^{35}ClN_5{}^+$, ([M+H]$^+$) requires 334.1793, found 334.1791.

4-(1-(1-Cyclohexyl-1*H*-tetrazol-5-yl)-1-phenylethyl)morpholine **267**

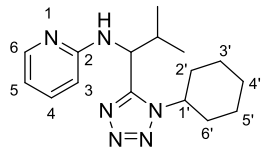


267

Following *General Procedure 7*, acetophenone (97 µL, 0.8 mmol, 1.0 eq) was treated with morpholine (73 µL, 0.8 mmol, 1.0 eq), cyclohexyl isonitrile (124 µL, 1.0 mmol, 1.2 eq) and trimethylsilyl azide (166 µL, 1.2 mmol, 1.5 eq). The reaction mixture was filtered and the filtrate washed with water to afford *tetrazole* **267** 86 mg, 30%) as a white powder. Mp 208-210 °C (Lit. (Ugi 1961) 213-215 °C); ν_{max} (FTIR) cm^{-1} 2927, 2854, 1116, 742 (aromatic CH), 707 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.42 - 7.48 (m, 2H, C(2)*H*

& C(6)*H*), 7.30 - 7.37 (m, 2H, C(3)*H* & C(5)*H*), 7.21 - 7.26 (m, 1H, C(4)*H*), 4.79 - 4.92 (m, 1H, C(1')*H*), 3.83 - 3.91 (m, 2H, O(CH_{AQ})$_2$), 3.74 - 3.83 (m, 2H, O(CH_{EQ})$_2$), 2.79 - 2.89 (m, 2H, N(CH_{AQ})$_2$), 2.24 (d, *J* = 10.0 Hz, 2H, N(CH_{EQ})$_2$), 2.00 - 2.12 (m, 4H, CH_3 & C(2')H_{AX}), 1.91 - 2.00 (m, 1H, C(3')H_{AX}), 1.76 (d, *J* = 12.5 Hz, 1H, C(2')H_{EQ}), 1.68 (d, *J* = 12.7 Hz, 1H, C(4')H_{EQ}), 1.57 - 1.64 (m, 1H, C(5')H_{EQ}), 1.29 - 1.44 (m, 2H, C(3')H_{EQ} & C(6')H_{AX}), 1.14 - 1.27 (m, 1H, C(4')H_{AX}), 1.00 - 1.13 (m, 1H, C(5')H_{AX}), 0.38 (d, *J* = 12.5 Hz, 1H, C(6')H_{EQ}); ^{13}C NMR (101 MHz, $CDCl_3$) δ 156.3 (*C*CCH*3*), 140.8 (*C*(1)), 128.7 (*C*(3) & *C*(5)), 127.7 (*C*(4)), 127.0 (*C*(2) & *C*(6)), 67.4 (O(*C*H_2)$_2$), 63.7 (*C*CH_3), 58.2 (*C*(1')), 48.5 (N(CH_2)$_2$), 32.6 (*C*(2' or 6')), 32.5 (*C*(2' or 6')), 25.9 (*C*(3' or 5')), 25.3 (*C*(3' or 5')), 24.8 (*C*(3')), 18.1 (*C*H_3); *m/z* (ESI$^+$) 342 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{19}H_{27}N_5ONa^+$, ([M+Na]$^+$) requires 364.2108, found 364.3109.

N-(1-(1-Cyclohexyl-1*H*-tetrazol-5-yl)-2-methylpropyl)pyridin-2-amine **268**

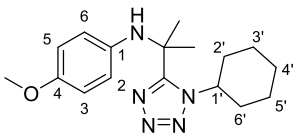


268

Following *General Procedure 7*, isobutyraldehyde (97 µL, 1.1mmol, 1.0 eq) was treated with pyridin-2-amine (100 mg, 1.1 mmol, 1.0 eq), cyclohexyl isonitrile (159 µL, 1.3 mmol, 1.2 eq) and trimethylsilyl azide (212 µL, 1.6 mmol, 1.5 eq). The crude product was purified by flash column chromatography (0%-50% EtOAc in cyclohexane) to afford *tetrazole* **268** (25 mg, 8%) as a white powder. Mp 165-167 °C (Lit. (Ugi 1962) 166-168 °C); ν_{max} (FTIR) cm^{-1} 3257 (NH), 2934, 1606, 1482, 771, 736; ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.99 (dd, *J* = 1.2, 4.9 Hz, 1H, C(6)*H*), 7.34 (ddd, *J* = 1.8, 7.0, 8.5 Hz, 1H, C(4)*H*), 6.51 - 6.59 (m, 1H, C(5)*H*), 6.42 (d, *J* = 8.3 Hz, 1H, C(3)*H*), 5.09 - 5.31 (m, 2H, N*H* & NC*H*), 4.68 - 4.80 (m, 1H, C(1')*H*), 2.42 (quind, *J* = 6.7, 8.8 Hz, 1H, C*H*(CH_3)$_2$), 1.88 - 2.11 (m, 6H, C(2')H_2, C(3')H_{AX}, C(5')H_{AX} & C(6')H_2), 1.69 - 1.82 (m, 1H, C(4')H_{AX}), 1.24 - 1.55 (m, 3H, C(3')H_{EQ}, C(4')H_{EQ} & C(5')H_{EQ}), 1.16 (d, *J* = 6.6 Hz, 3H, C(CH_3)), 0.88 (d, *J* = 6.6 Hz, 3H, C(CH_3)); ^{13}C NMR (101 MHz, $CDCl_3$) δ 157.2 (*C*(2)), 155.9 (NHCH*C*), 147.2 (*C*(6)), 137.3 (*C*(4)), 113.9 (*C*(5)), 109.4 (*C*(3)), 57.7 (*C*(1')), 50.5 (NH*C*H), 33.3 (*C*(2'

or 6')), 32.9 (*C*(2' or 6')), 32.7 (*C*$(CH_3)_2$), 25.4 (*C*(3' or 5')), 25.4 (*C*(3' or 5')), 24.9 (*C*(4')), 19.7 (C(*C*H_3)), 19.5 (C(*C*H_3)); *m/z* (ESI$^+$) 301 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{16}H_{24}N_6Na^+$, ([M+Na]$^+$) requires 323.1955, found 323.1958.

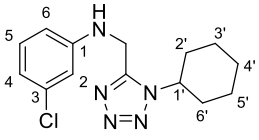
N-(2-(1-Cyclohexyl-1*H*-tetrazol-5-yl)propan-2-yl)-4-methoxyaniline **269**



269

Following *General Procedure 7*, acetone (60 µL, 0.8 mmol, 1.0 eq) was treated with 4-methoxyaniline (100 mg, 0.8 mmol, 1.0 eq), cyclohexyl isonitrile (121 µL, 1.0 mmol, 1.2 eq) and trimethylsilyl azide (162 µL, 1.2 mmol, 1.5 eq). The reaction mixture was filtered and the filtrate washed with water to afford *tetrazole* **269** (187 mg, 73%) as an off-white powder. Mp 132-133 °C (Lit. (Ugi 1961) 130-132 °C); ν_{max} (FTIR) cm^{-1} 3390 (NH), 2937, 1517, 1507, 1247, 815 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.61 - 6.66 (m, 2H, C(3)*H* & C(5)*H*), 6.17 - 6.23 (m, 2H, C(2)*H* & C(6)*H*), 5.06 (s, 1H, C(1')*H*), 3.69 (s, 3H, OCH_3), 3.48 - 3.62 (m, 1H, N*H*), 1.88 - 2.02 (m, 2H, C(2')H_{AX} & C(6')H_{AX}), 1.80 - 1.88 (m, 2H, C(3')H_{AX} & C(5')H_{AX}), 1.77 (s, 6H, C(CH_3)$_2$), 1.66 - 1.75 (m, 3H, C(2')H_{EQ}, C(4')H_{AX} & C(6')H_{EQ}), 1.27 (br. s, 3H, C(3')H_{EQ}, C(4')H_{EQ} & C(5')H_{EQ}); ^{13}C NMR (101 MHz, $CDCl_3$) δ 157.8 (*C*C(CH$_3$)$_2$), 153.3 (*C*(4)), 138.4 (*C*(1)), 116.9 (*C*(2) & *C*(6)), 114.7 (*C*(3) & *C*(5)), 59.1 (*C*(1')), 55.5 (O*C*H$_3$), 52.4 (*C*(CH$_3$)$_2$), 33.3 (*C*(2') & *C*(6')), 28.6 (C(*C*H$_3$)$_2$), 25.6 (*C*(3') & *C*(5')), 24.9 (*C*(4')); *m/z* (ESI$^+$) 316 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{17}H_{25}N_5ONa^+$, ([M+Na]$^+$) requires 338.1951, found 338.1951.

3-Chloro-*N*-((1-cyclohexyl-1*H*-tetrazol-5-yl)methyl)aniline **270**

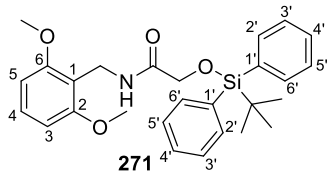


270

Following *General Procedure 7*, paraformaldehyde (24 mg, 0.8 mmol, 1.0 eq) was treated with 3-chloroaniline (83 µL, 0.8 mmol, 1.0 eq), cyclohexyl isonitrile (117 µL, 0.9 mmol, 1.2 eq) and trimethylsilyl azide (156 µL, 1.2 mmol, 1.5 eq). The reaction mixture was filtered and the filtrate washed with water to afford *tetrazole* **270** (150 mg, 66%) as a white powder. Mp 148-149 °C; ν_{max} (FTIR) cm^{-1} 3378 (NH), 2938,

1598, 762 (aromatic CH), 681 (C-Cl); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.13 (t, *J* = 8.0 Hz, 1H, C(5)*H*), 6.75 - 6.83 (m, 1H, C(4 or 6)*H*), 6.71 (t, *J* = 2.1 Hz, 1H, C(2)*H*), 6.59 - 6.65 (m, 1H, C(4 or 6)*H*), 4.53 - 4.61 (m, 2H, NHCH_2), 4.50 (d, *J* = 5.4 Hz, 1H, N*H*), 4.23 - 4.36 (m, 1H, C(1')*H*), 1.91 - 2.09 (m, 6H, C(2')H_2, C(3')H_{AX}, C(5')H_{AX} & C(6')H_2), 1.72 - 1.81 (m, 1H, C(4')H_{AX}), 1.25 - 1.48 (m, 3H, C(3')H_{EQ}, C(4')H_{EQ} & C(5')H_{EQ}); ^{13}C NMR (101 MHz, $CDCl_3$) δ 151.4 (*C*CH_2NH), 147.8 (*C*(1)), 135.3 (*C*(3)), 130.5 (*C*(5)), 119.3 (*C*(4 or 6)), 113.2 (*C*(2)), 111.9 (*C*(4 or 6)), 58.4 (*C*(1')), 38.0 (NH*C*H_2), 32.9 (*C*(2') & *C*(6')), 25.3 (*C*(3') & *C*(5')), 24.8 (*C*(4')); *m/z* (ESI$^+$) 292 ([M(^{35}Cl)+H]$^+$, 100%), 294 ([M(^{37}Cl)+H]$^+$, 37%); HRMS (ESI$^+$) $C_{24}H_{18}{}^{35}ClN_5Na^+$, ([M+Na]$^+$) requires 314.1143, found 314.1144.

2-((*tert*-Butyldiphenylsilyl)oxy)-*N*-(2,6-dimethoxybenzyl)acetamide **271**

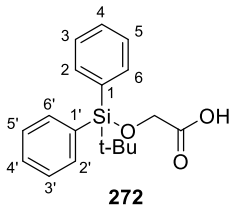


271

1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU) (242 mg, 0.64 mmol, 1.0 eq) was added to a stirred solution of compound **272** (200 mg, 0.64 mmol. 1.0 eq) and *N*-methylmorpholine (140 µL, 1.28 mmol, 2.0 eq) in acetonitrile (5 mL). The reaction was stirred at room temperature for 3 h before the addition of 2,6-dimethoxybenzylamine (106 mg, 0.64 mmol, 1.0 eq). The reaction was stirred for a further 1 h. Solvent was removed *in vacuo* and the crude solid redissolved in CH_2Cl_2 (5 mL). The resulting solution was washed with aq. NaOH (10%, 10 mL) and purified by flash column chromatography (7%-60% EtOAc in cyclohexane). Solvent was removed *in vacuo* to afford *amide* **271** (171 mg, 58%) as white crystals. Mp 128.5-130 °C; ν_{max} (FTIR) cm^{-1} 3427 (NH), 2854, 1676 (C=O), 1471, 1113 (COC), 774 (aromatic CH), 704 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.57 - 7.63 (m, 4H, C(2')*H* & C(6')*H*), 7.53 (br. t, *J* = 5.9 Hz, 1H, N*H*), 7.41 - 7.47 (m, 2H, C(4')*H*), 7.32 - 7.38 (m, 4H, C(3')*H* & C(5')*H*), 7.27 (t, *J* = 8.0 Hz, 1H, C(4)*H*), 6.60 (d, *J* = 8.4 Hz, 2H, C(3)*H* & C(5)*H*), 4.68 (d, *J* = 5.9 Hz, 2H, CH_2NH), 4.12 (s, 2H, C(=O)CH_2), 3.82 (s, 6H, OCH_3), 1.08 (s, 9H, C(CH_3)$_3$); ^{13}C NMR (101 MHz, $CDCl_3$)

δ 169.9 (*C*=O), 158.6 (*C*(2) & *C*(6)), 135.3(*C*(2') & *C*(6')), 132.3 (*C*(1')), 130.0 (*C*(4')), 128.9 (*C*(4)), 127.9 (*C*(3') & *C*(5')), 114.2 (*C*(1)), 103.8 (*C*(3) & *C*(5)), 64.0 (C(=O)*C*H_2), 55.9 (O*C*H_3), 31.7 (*C*H_2NH), 26.7 (C(*C*H_3)$_3$), 19.1 (*C*(CH$_3$)$_3$); *m/z* (ESI$^+$) 464 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{27}H_{34}NO_4Si^+$, ([M+H]$^+$) requires 464.2252, found 464.2250.

2-((*tert*-Butyldiphenylsilyl)oxy)acetic acid **272**



272

tert-Butylchlorodiphenylsilane (513 µL, 2.0 mmol, 1.5 eq) was added to a stirred solution of glycolic acid (100 mg, 1.3 mmol, 1.0 eq) and imidazole (179 mg, 2.6 mmol, 2.0 eq) in DMF (1.5 mL). The reaction was stirred at room temperature for 2 h before quenching with water (5 mL). The reaction was washed with aq. HCl (0.5 M, 5 mL) and extracted using EtOAc (10 mL). The organic layer was dried and purified by flash column chromatography (7%-60% EtOAc in cyclohexane). Solvent was removed *in vacuo* to afford *acid* **272** (244 mg, 59%) as a colourless oil. ν_{max} (FTIR) cm^{-1} 3426, 2856 (OH), 1676 (C=O), 1426, 1112, 703 (aromatic CH); ^{1}H NMR (400 MHz, DMSO-d_6) δ 12.56 (br. s, 1H, COO*H*), 7.63 - 7.68 (m, 4H, C(2)*H*, C(6)*H*, C(2')*H* & C(6')*H*), 7.41 - 7.51 (m, 6H, C(3)*H*, C(4)*H*, C(5)*H*, C(3')*H*, C(4')*H* & C(5')*H*), 4.19 (s, 2H, CH_2), 1.02 (s, 9H, C(CH_3)$_3$); ^{13}C NMR (101 MHz, DMSO-d_6) δ 172.3 (*C*=O), 135.5 (*C*(2), *C*(6), *C*(2') and *C*(6')), 133.1 (*C*(1) and *C*(1')), 130.4 (*C*(4) and *C*(4')), 128.4 (*C*(3), *C*(5), *C*(3') and *C*(5')), 62.0 (*C*H_2), 27.0 (C(*C*H_3)$_3$), 19.3 (*C*(CH$_3$)$_3$); *m/z* (ESI$^+$) 315 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{18}H_{22}O_3SiNa^+$, ([M+Na]$^+$) requires 337.1230, found 337.1231.

Protein Expression and Purification

cDNA encoding BRD1, PHIP(2) (Filippakopoulos, Picaud *et al.* 2012) and JMJD2D (Bavetsias, Lanigan *et al.* 2016) were cloned, expressed and purified as previously described.

AlphaScreen Assay

PHIP(2) and BRD1 assays were performed as described previously (Philpott, Yang *et al.* 2011) with minor modifications from the manufacturer's protocol (PerkinElmer, USA). All reagents were diluted in 25 mM HEPES, 100 mM NaCl, 0.1 % BSA, pH 7.4 supplemented with 0.05 % CHAPS and allowed to equilibrate to room temperature prior to addition to plates. An 11-point 1:2 serial dilution of the ligands was prepared on low-volume 384-well plates (ProxiPlateTM-384 Plus, PerkinElmer, USA), using LabCyte Echo liquid handler. Plates were filled with 12 µL/well with a mix of the assay buffer, biotinylated peptide [H-YSGRGKacGGKacGLGKacGGAKacRHRK(Biotin)-OH] and His-tagged protein to achieve final assay concentrations of 50 nM. Plates were sealed and incubated for a further 30 minutes, before the addition of 8 µl of the mixture of streptavidin-coated donor beads (12.5 µg/ml) and nickel chelate acceptor beads (12.5 µg/ml) under low light conditions. Plates were foil-sealed to protect from light, incubated at room temperature for 60 minutes and read on a PHERAstar FS plate reader (BMG Labtech, Germany) using an AlphaScreen 680 excitation/570 emission filter set. IC_{50} values were calculated in Prism 5 (GraphPad Software, USA) after normalization against corresponding DMSO controls and are given as the final concentration of compound in the 20 µl reaction volume.

JMJD2D enzyme activity was measured using an AlphaScreen assay that monitored the demethylation of a biotinylated trimethylated H3K9 peptide using a H3K9 dimethyl specific antibody and appropriate donor and acceptors beads from PerkinElmer (Kawamura, Tumber *et al.* 2010). Compounds were dispensed using an ECHO 550 acoustic dispenser (Labcyte Inc., Sunnyvale, CA, USA) directly into 384-well Proxiplates (no. 6008289, PerkinElmer, Waltham, MA, USA) to give 1% (v/v) DMSO final assay concentration. The enzyme reaction was performed in assay buffer which consisted of HEPES (50 mM, pH 7.5), BSA (0.1% w/v), sodium ascorbate (100

μM), 2OG (10 μM), ammonium iron(II) sulfate hexahydrate (1 μM), and Tween 20 (0.01% v/v). Compounds were pre-incubated for 15 min with 5 μL of 4 nM JMJD2D enzyme prepared in-house before the addition of 5 μL of 60 nM peptide (ARTKQTARK(Me_3)STGGKAPRKQLA-GGK- biotin, Genecust). The plate was sealed and centrifuged at 1000 rpm for 1 min before being left for 5 min at room temperature. The reaction was stopped with the addition of 5 μL of 30 mM EDTA before the addition of 5 μL of AlphaScreen detection reagents (no. 6760606M, PerkinElmer) and antibody. The detection reagents containing protein A acceptor beads, streptavidin donor beads (20 μg/mL final concentration), and an antidimethyl H3K9 antibody (no. Ab1220, 50 ng/mL final concentration) were preincubated for 2 h in detection buffer containing HEPES (50 mM, pH 7.5), BSA (0.1% w/v), and Tween 20 (0.01% v/v) prior to addition to the plate. Plates were incubated at room temperature for 1 h in the dark before being read on a PHERAstar FS plate reader (BMG Labtech, Germany) using an AlphaScreen 680 excitation/570 emission filter set. IC_{50} values were calculated in Prism 6 after normalisation against corresponding DMSO controls.

Crystallisation

Rod-shaped BRD1 crystals were obtained by mixing 75nl of 10mg/ml protein in 0.01M HEPES pH7.5, 500 mM NaCl, 5 % Glycerol with 75nl of reservoir solution containing 0.1M bis-tris pH 7.0 and 33% PEG3350. Crystals appeared overnight from sitting drop plates at 20 °C. BRD1 crystallized in space group P212121 with typical unit cell dimensions of a = 54 Å, b = 56 Å, c = 101 Å, corresponding to two BRD1 molecules in the asymmetric unit.

PHIPA(2) Brd was crystallized by mixing 100nl of 13mg/ml protein in 20mM HEPES pH7.5, 500mM NaCl, 5% Glycerol with 100nl of reservoir solution containing 0.1M HEPES pH 7.5, 0.15M Magnesium Chloride, 34% PEG3350. Crystals appeared overnight from sitting drop plates at 4°C.

PHIP(2) Brd crystallized in space group P21212 with typical unit cell dimensions of a=60Å, b=92Å, c=24Å, corresponding to one PHIP(2) Brd molecule in the asymmetric unit.

Prism-shaped JMJD2D crystals were obtained by mixing 135 nl of 19 mg/ml protein in 0.02 M HEPES pH7.5, 0.5 M NaCl, 5% glycerol, 0.5 µM TCEP with 65nl of reservoir solution containing 0.1 M HEPES pH 7.0, 0.15 M ammonium sulfate and 25% PEG3350. 2-OG was added when desired to give a final concentration between 1-3 µM. Crystals appeared overnight from sitting drop plates at 20°C. JMJD2D crystallized in space group P 41 21 2 with typical unit cell dimensions of a = 72 Å, b = 72 Å, c = 151 Å, corresponding to a single JMJD2D molecule in the asymmetric unit.

SETDB1 (construct is APC043:C09, His-tag is uncut, MW 26293.2, A280 (1 mg/ml) = 2.201) was thawed on ice from -80C to 4 °C. The protein was already at the correct concentration for crystallisation, approximately 5 mg/ml. Trypsin had already been added to the protein sample at 1:1000 ratio. The mother liquor is Red Wings condition H04 (25% PEG3350, 0.2M $LiSO_4$, 0.1 M Bis-Tris, pH 6.5), supplied in same shipment, keep at 4 °C. Set up 1ul drops mixing the protein with mother liquor 1:1 using the Mosquito. For the Swissci 3 drop plates, use a reservoir volume of 45ul. After sealing the plate, store at ~18 °C for optimal crystal growth. This protocol only yielded ~3 crystals across all plates set up. These 3 crystals were harvested from their drops and added to 1ml of mother liquor. This seed mix was then vigourously vortexed for 5 mins. Working quickly, the plate seals were removed and 100nl of seed mix was added to each drop using the Mosquito. This resulted in ~1/3 of the drops growing crystals within the next 24-48 hours.

Compound Soaking

All compounds were dissolved in ethylene glycol at a nominal concentration of 400mM. It should be noted that not all compounds could be dissolved at such high concentrations, but in this case

we assumed that the solution was saturated. 600nl of each compound was mixed with 600nl of reservoir solution and the mixture was added to the crystals by using a Mosquito® crystallization robot (TTP Labtech). The plates were resealed and incubated for at least 12 hours at 4°C before the crystals were mounted in nylon loops and immediately flash frozen in liquid nitrogen.

Data Collection and Structure Solution

All datasets were collected on beamline I04-1 at the Diamond Light Source. Data were integrated and scaled with XIA2 (Winter 2010) which is part of the Diamond Light Source autoprocessing pipeline. Initial electron density maps were calculated with DIMPLE (Winn, Ballard *et al.* 2011) and inspected with COOT (Emsley, Lohkamp *et al.* 2010). Once a hit was identified, further rounds of refinement with REFMAC (Murshudov, Skubak *et al.* 2011) and manual rebuilding with COOT were carried out. ACEDRG (Winn, Ballard *et al.* 2011) was used to generate compound coordinates and restraint files. The quality of the final models was validated with MOLPROBITY (Chen, Arendall *et al.* 2010).

Appendix F: Microbial Biotransformation Methods and Materials

Screening

Compound was fed at 100 mg/L to a panel of selected bacterial species (Table D.1). Strains were initially recovered from cryo-storage directly into shake-flasks (50 ml medium) and fungal strains were resuscitated via an agar culture step before inoculation into flasks of appropriate growth medium. Substrate was added to strains after a specific pre-growth period to a final substrate concentration of 100 mg/L derived from a DMSO stock at 25 mg/mL and fermentations were sampled at three time-points: 24, 48 and 120 h. Samples were extracted with an equal volume of acetonitrile, vortex-mixed, centrifuged to clarify and analysed by LC-DAD-MS using a screening gradient. Samples were assessed using LCMS ESI+ chromatograms.

Micro-organism Description	177 ([M+H]+ 236m/z)
Bact-Sp1	none/low
Bact-Sp2	none/low
Bact-Sp41	2 products: 252 m/z both
Bact-Sp43	none/low
Bact-Sp45	none/low
Bact-Sp48	none/low
Bact-Sp49	none/low
Bact-Sp50	none/low
Bact-Sp52	none/low
Bact-Sp58	none/low
Bact-Sp59	2 products: 252 m/z both
Bact-Sp60	none/low
Bact-Sp61D	none/low
Bact-Sp61D4	none/low
Bact-Sp62	none/low
Bact-Sp131	none/low
Bact-Sp145	none/low
Bact-Sp146	none/low
Bact-Sp147	none/low
Bact-Sp159	none/low
Bact-Sp160	none/low
Bact-Sp162	none/low
Bact-Sp179	none/low

Table D.1: Summary of bacterial screening results.

Production and Purification

Confirmation of production

Sp59 was recovered from cryovial stocks, and both inoculated into 2 Erlenmeyer flasks containing M3G medium each. After 2 days growth at 27°C cultures of each species were used to inoculate 6 wells in a 24 well MicroBioreactor each containing 2.5 ml of M3G medium. The blocks were fermented at 27°C and 300 rpm. The first pair of wells was dosed at 24 h with 100 mg/L of compound **177**. The second pair of wells was dosed at 48 h and the third pair dosed at 72 h. All cultures were sampled and analysed at various time-points during the fermentation as described in the results.

Scaled-up fermentation

After the target metabolite production was re-confirmed in Sp59, the reaction was scaled up in multiple MicroBioreactors to 480mL total volume. Sp59 was recovered from cryovial stocks and inoculated into Erlenmeyer flasks containing M3G medium each and grown for 2 days at 27°C. Cultures were inoculated into twelve 250 ml Erlenmeyer flasks, each containing 50 ml of M3G medium and fermented at 27°C, 200 rpm for 48 h before being transferred to 8 MicroBioreactors (2.5 ml/well) and dosed with compound **177** at a final concentration of ca. 100 mg/L. Representative MicroBioreactor wells were time-coursed after dosing to assess the production of the target metabolites. Based upon time course analysis, cultures were harvested after 48 h before further processing.

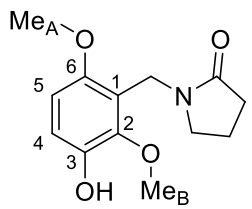
Product purification

The MicroBioreactor reactions were stopped by addition of an equal volume of MeCN, combined and centrifuged with the biomass extracted with a further 2x 400 ml MeCN. The biomass MeCN

extract was combined with the supernatant and the organic solvent removed under vacuum. The residual aqueous was back extracted to EtOAc (3 x 500 ml) with the EtOAc removed under vacuum. The EtOAc extract was fractionated by PHPLC over a Waters Symmetry Shield RP8 19 x 100 mm column with a RP8 19x10 mm guard column, eluting with a flow rate of 17 ml/min and a linear gradient starting from 90/5/5% (H_2O/MeCN/200 mM aq. NH_4 formate), held for 1 mins, then increased to 15/80/5% over 10 mins (t = 11 mins), further increased to 0/95/5% over 1 mins (12 mins total), followed by wash and re-equilibration steps. The targets eluted between 6.5 and 7.5 mins. The target fraction was then purified further over a Waters Atlantis ODS T3 5µm OBD 10x250mm column, eluting with a flow rate of 5ml/min and starting at 67/28/5% (H_2O/MeCN/2% aq. formic acid) then increasing to 66/29/5% over the next 11 mins (t=11 mins), increasing again to 0/95/5% over 1 mins (12 mins total), followed by wash and re-equilibration steps. Target 1 eluted between 6.5 and 7.5 mins (Batch ID: SB44/066-2, 5.5 mg, compound **234**) and target 2 eluted between 7.5 and 8.5 mins (Batch ID: SB44/066-04, 5.2 mg, compound **235**), both of which were >95% by ELSD analysis.

Compound Characterisation

1-(3-Hydroxy-2,6-dimethoxybenzyl)pyrrolidin-2-one **234**

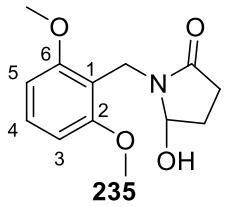


234

Mp 105-107 °C; ν_{max} (FTIR) cm^{-1} 2930, 1636 (C=O), 1263 (C-O), 1100, 802 (aromatic CH); ^1H NMR (400 MHz, $CDCl_3$) δ 6.91 (d, *J* = 8.9 Hz, 1H, C(4)*H*), 6.60 (d, *J* = 8.8 Hz, 1H, C(5)*H*), 5.54 (br. s, 1H, O*H*), 4.60 (s, 2H, C(1)CH_2), 3.83 (s, 3H, Me_A or Me_B), 3.79 (s, 3H, Me_A or Me_B), 3.19 (t, *J* = 7.1 Hz, 2H, NC$H_2$$CH_2$), 2.40 (t, *J* = 8.1 Hz, 2H, C(=O)CH_2), 1.92 (quin, *J* = 7.6 Hz, 2H, NCH_2CH_2); ^{13}C NMR (101 MHz, $CDCl_3$) δ 174.6 (*C*=O), 152.4 (*C*(2) or *C*(6)), 147.0 (*C*(2) or *C*(6)), 143.2 (*C*(3)), 117.7 (*C*(1)), 115.0 (*C*(4)), 107.0 (*C*(5)), 61.7 (Me_A or Me_B), 56.1 (Me_A or Me_B), 46.3 (N*C*$H_2$$CH_2$), 35.5 (C(1)*C*$H_2$),

31.1 (C(=O)*C*H$_2$), 17.7 (NCH$_2$*C*H$_2$); *m/z* (ESI$^+$) 252 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{13}H_{18}NO_4^+$, ([M+H]$^+$) requires 252.1230, found 252.1218.

1-(2,6-Dimethoxybenzyl)-5-hydroxypyrrolidin-2-one **235**



235

(Absolute stereochemistry not determined) Mp 110-112 °C; ν_{max} (FTIR) cm^{-1} 3355 (OH), 2937, 1662 (C=O), 1251 (C-O), 1108, 783 (aromatic CH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.30 (t, *J* = 8.5 Hz, 1H, C(4)*H*), 6.63 (d, *J* = 8.4 Hz, 2H, C(3)*H* and C(5)*H*), 5.11 (d, *J* = 13.7 Hz, 1H, C(1)C*H*$_A$), 4.91 (d, *J* = 6.6 Hz, 1H, C(OH)*H*), 4.14 (d, *J* = 13.6 Hz, 1H, C(1)C*H*$_B$), 3.89 (s, 6H, OC*H*$_3$), 2.56 - 2.68 (m, 1H, C(=O)C*H*$_A$), 2.30 - 2.40 (m, 1H, C(=O)C*H*$_B$), 2.05 - 2.18 (m, 1H, C(=O)CH$_2$C*H*$_A$), 1.82 - 1.91 (m, 1H, C(=O)CH$_2$C*H*$_B$); ^{13}C NMR (101 MHz, $CDCl_3$) δ 174.0 (*C*=O), 159.0 (*C*(2) and *C*(6)), 129.7 (*C*(4)), 111.6 (*C*(1)), 104.2 (*C*(3) and *C*(5)), 81.9 (*C*OH), 56.2 (O*C*H$_3$), 31.9 (C(1)*C*H$_2$), 29.5 (C(=O)*C*H$_2$), 26.0 (C(OH)*C*H$_2$); *m/z* (ESI$^+$) 252 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{13}H_{17}NNaO_4^+$, ([M+Na]$^+$) requires 274.1050, found 274.1044.

Technical Details

Fermentation

Medium: M3G; Strain: Sp59 is a proprietary Actinomycete strain (Streptomyces sp.).

Chromatography details

System 1: Pumps and autosampler: Waters Acquity H-Class UPLC; Detection: Waters Acquity PDA; Detection (MS): Waters Acquity QDA. System 2: Pumps and autosampler: Waters Alliance 2796-HT; Detection: Waters 2996 photodiodearray UV detector; Detection (MS): Waters Acquity SQD; Detection (ELSD): PolymerLabs ELS ICE2100. *Analysis Gradient A:* column: Waters BEH shield RP18 1.7 µm 2.1 mm x 50 mm length; solvents: A: H_2O + 0.1% formic acid, B: Acetonitrile + 0.1% formic

acid; gradient (A%/B%): t = 0mins: 98/2 to 50/50 over 5 mins, then to 2/98 over 0.5 mins (t = 5.5 mins), held for 1 mins (t = 6.5 mins), back to 98/2 over 0.5 mins (t = 7mins), followed by re-equilibration for a further 1 mins (t = 8mins) all at 0.5 ml/min. *Analysis Gradient B:* column: Waters SymmetryShield RP8 3.5 µm 4.6 mm i.d. 7.5 cm length; solvents: A: H_2O, B: Acetonitrile, C: 200 mM aq. NH_4 formate, 2% formic acid; gradient (A%/B%/C%): t = 0mins: 85/10/5 to 45/50/5 over 10 mins (t = 10mins), then increased to 0/95/5 over 1 mins (t = 11mins), returned to 85/10/5 over 0.5 mins (t = 11.5mins), followed by re-equilibration for a further 3.5 mins (t = 15 mins) all at 1.0 ml/min.