Design, Synthesis and Evaluation of Novel Small Molecule Inhibitors of the Histone Methyltransferase DOT1L and Ubiquitination Facilitator Keap1

Sophie S. Spurr

A thesis submitted for the award of doctor of philosophy

School of Pharmacy, University College London

2017

Declaration

I, Sophie S. Spurr confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Sophie S. Spurr

Abstract

This thesis details the design, synthesis and evaluation of novel small molecule inhibitors of the histone methyltransferase DOT1L and the ubiquitination facilitator Keap1. The thesis is in two parts as outlined below.

Part 1: The first part of this thesis details efforts towards identification of novel small molecule inhibitors of DOT1L, a histone methyltransferase which has been implicated in the development and proliferation of mixed lineage leukaemias (MLL). This work aims to optimise the drug-like properties of published DOT1L inhibitors while retaining potency, through further exploration of the nucleobase template. Structure-activity relationships (SARs) identified polar substituents and small heterocycles as favourable replacements for the halogen in 5-ITC, a small molecule inhibitor of DOT1L. Alternative nucleobase templates also demonstrated comparable DOT1L inhibition. To demonstrate proof of concept, a polar nitrile substituent was translated into the inhibitor Br-SAH as a direct replacement of the bromide. Activity was retained and a crystal structure obtained which demonstrated the nitrile occupied the same hydrophobic pocket. This work also demonstrated the use of a nitrile as a non-traditional replacement for heavy halogen atoms.

Part 2: The second part of this thesis details identification of novel inhibitors of the Keap1-Nrf2 protein-protein interactions (PPI) using an approach based on kinetic target-guided synthesis (kTGS). Keap1 is a dimeric cytoplasmic protein that mediates the ubiquitination of Nrf2, a transcription factor that acts as a regulator of cellular antioxidant responses. Disruption of the PPI between Keap1 and Nrf2 has been shown to have a therapeutic benefit in diseases associated with oxidative stress and inflammation as well as providing a potential route to chemopreventative agents for cancer. Biased kTGS was applied as proof of concept. A biased ligand was designed and screened against a focused library of azides. The 1,3-dipolar cycloaddition products formed in the presence of the Keap1 Kelch domain were evaluated and validated through chemical synthesis and screening. A novel triazole structure was identified with improved activity over the initial biased fragment thus demonstrating kTGS as a valid approach for identifying novel inhibitors of the Keap1-Nrf2 PPI interaction.

Impact statement

This thesis contributes to the academic area of research of drug discovery, particularly in relation to the development of novel small molecule inhibitors of the histone methyltransferase DOT1L and the ubiquitination facilitator Keap1. The findings from this body of work could ultimately be beneficial to public health.

Inhibition of DOT1L has been shown to have a therapeutic benefit in treatment of paediatric patients harbouring MLL-rearranged leukaemia. The aim of this research was to discover new inhibitors of DOT1L activity with improved pharmacokinetic properties. A strategy was adopted to identify novel functional groups which could exploit the hydrophobic cavity adjacent to the adenosine nucleobase. This work led to the identification of a nanomolar DOT1L inhibitor, CN-SAH, with an improved drug-like template. This substitution can be translated to other published inhibitor scaffolds, with the effect of improving the pharmacokinetic properties. Disruption of the PPI between Keap1 and Nrf2 has been shown to have a therapeutic benefit in diseases associated with oxidative stress and inflammation as well as providing a potential route to chemopreventative agents for cancer. The aim of this research was to discover novel small molecule inhibitors of the Keap1-Nrf2 PPI using a fragment-based drug discovery approach, kTGS. This work led to the identification of a novel triazole structure demonstrating kTGS as a valid approach for identifying novel inhibitors of the Keap1-Nrf2 PPI interaction.

The work undertaken provides a contribution to knowledge through publication in *Bioorganic Medicinal Chemistry Letters* (impact factor 2.973, 2014). It is anticipated the results gained toward inhibition of Keap1 will provide future publications.

Acknowledgments

First, I am very grateful to Professor Paul Fish for the opportunity to undertake this PhD. I would like to thank all my supervisors, Professor Paul Fish, Dr Geoff Wells and Dr Jess Healy for their fantastic support and guidance. I would like to acknowledge the encouragement and camaraderie I received from everyone – past and present – in the Fish, Wells and Healy groups. I would particularly like to thank Dr Carrie-Anne Molyneaux, Dr Niall Igoe and Dr Elliott Bayle for their support.

I am also thankful to my collaborators at the Structural Genomics Consortium, particularly Dr Matthieu Schapira for his helpful advice. I am also grateful to my funding body, the UCL Drug Discovery PhD studentship programme.

I would like to thank Dr Jamie Baker for his support at the start of my PhD during the 6-month project rotation. I would like to mention Dr Rosemary Huckvale, Dr Samantha Gibson and Dr Rachel Lanigan-Mills for their 'liquid encouragement' (although I think my liver has suffered from our prosecco days!).

Last but not least, I would like to thank my family and my friends, particularly Tom Gockelen-Kozlowski, Richard Oliver, Tom Skelton and Neil Thompson, for their unending support through the tears, merriment, presentation practise, pub trips and endless viewings of Murder She Wrote episodes. Finally, I give a particular mention to my partner, Dr Andrew Wright for being by my side.

Abbreviations

AAC	azide alkyne cycloaddition
AChE	acetylcholinesterase
AcOH	acetic acid
AD	Alzheimer's disease
ALL	acute lymphoblastic leukaemia
AML	acute myeloid leukaemia
ARE	antioxidant response element
BCL-2	B-cell lymphoma 2
BcL-X_L	B-cell lymphoma-extra large
BSA	Bovine serum albumin
BTB	broad complex, tram-track and bric-a-brac
bZIP	basic leucine zipper
CAAC	Cu catalysed AAC
CARMI	coactivator associated arginine methyltransferase 1
cLogP	calculated LogP
CNC	cap'n'collar
COPD	chronic obstructive pulmonary disease
CTR	C-terminal region
DBCL	diffuse large B cell lymphomas
DBPS	Dulbecco's phosphate buffer solution
DCC	dynamic combinatorial chemistry
DCL	dynamic combinatorial library
DOT1L	disruptor of telomeric silencing 1-like
DSF	differential scanning fluorimetry

E. coli	*Escherichia coli*
EMT	epithelial-mesenchymal transition
EtOAc	ethyl acetate
EtOH	ethanol
EZH1	enhancer of zeste homolog 1
FBDD	fragment-based drug discovery
FDA	food and drug administration
FITC	fluorescein isothiocyanate
FP	fluorescence polarisation
GCL	glutamate-cysteine ligase
GSH	glutathione
GST	glutathione *S*-transferase
GWAS	genome-wide analysis
Halo	halogen
H-bonds	hydrogen bonds
HMDS	hexamethyldisilazane
HMDT	histone demethyltransferase
HMPA	hexamethylphosphoramide
HMT	histone methyltransferase
HO-1	heme-oxygenase-1
HTS	high throughput screen
IBD	inflammatory bowel disease
IC_{50}	half maximal inhibitory concentration
in-situ	in situation
IPTG	isopropyl β-D-1-thiogalactopyranoside
IVR	intervening region

Keap1	Kelch-like ECH-associated protein 1
kTGS	kinetic target-guided synthesis
LB	Luria-Bertani
LE	ligand efficiency
LipE	ligand lipophilic efficiency
LogP	log of lipophilicity
*m*CPBA	*meta*-chloroperbenzoic acid
MOA	mechanism of action
MLL	mixed lineage leukaemia
MLPCN	molecular libraries probe production centres network
mol%	mole percent
Mtase	methyltransferase
NAD(P)H	nicotinamide amine dinucleotide phosphate
Neh	Nrf2-ECH homology
NMSF	National Mass Spectrometry Facility
NQO1	NAD(P)H quinone oxidoreductase 1
Nrf2	nuclear factor erythroid 2-related factor 2
NTR	N-terminal region
OA	osteoarthritis
PA	product amplification
PAGE	polyacrylamide gel electrophoresis
PD	Parkinson's disease
PDB	protein databank
PEG	polyethylene glycol
PKMT	protein lysine methyltransferase
PMT	protein methyltransferase

PPI	protein-protein interaction
PRMT	protein arginine methyltransferase
PSA	polar surface area
PTM	post-translational modification
qPCR	quantitative polymerase chain reaction
RAAC	Ru catalysed AAC
RNS	reactive nitrogen species
ROS	reactive oxygen species
SAH	*S*-adenosylhomocysteine
SAM	*S*-adenosylmethionine
SBDD	structure-based drug design
SDS	sodium dodecyl sulfate
SET	su(var)3–9, enhancer-of-zeste, trithorax
SGC	Structural Genomics Consortium
SIM	single ion monitoring
SPR	surface plasmon resonance
TAT	transactivating transcriptional activator
tBHQ	tert-Butylhydroquinone
TBTA	2,2,2-trichloroacetimidate
TF	triazole formation
TGS	target-guided synthesis
T_m	protein melt point
Tris	tris(hydroxymethyl)aminomethane
Ub	ubquitination

Contents

List of tables

List of figures

List of schemes

PART 1

Design, Synthesis and Evaluation of Novel Small Molecule Inhibitors of the Histone Methyltransferase DOT1L

CHAPTER 1
Introduction

1.1 Genetics

The world of genetics has been puzzling scientists for hundreds of years. Innovations from many areas of research have contributed to the development of our modern understanding of the mechanisms and machinery involved in this field.

Genes are composed of units or segments of deoxyribonucleic acid (DNA). This codes for the form and function of an organism and carries genetic information from generation to generation. It comprises of a double-stranded helix, consisting of a sugar-phosphate backbone and pairs of nucleobases which are denoted by a single letter code: A (Adenine), G (Guanine), C (Cytosine) and T (Thymine). Variation is introduced through unique ordering of the four nucleobases. DNA is stored in the nucleus of every cell. Linear DNA is several orders of magnitude larger than the width of a cell nucleus. To be contained within the walls of the cell nucleus, DNA is compressed and packaged into structures called chromosomes (Figure 1.1).

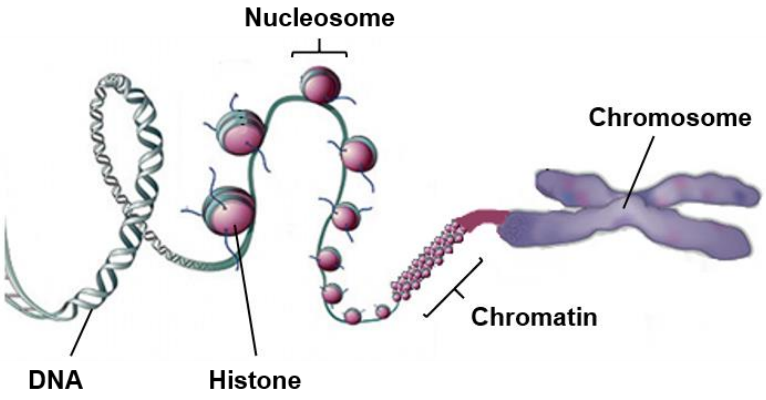


Figure 1.1 Packaging of DNA into the cell nucleus using histone to form chromosomes. Reproduced with amendments from Baylin *et al*.[1]

Every human cell contains 46 chromosomes, 23 inherited from each parent. To form chromosomes, 146 base pairs of DNA are wrapped around a core protein complex consisting of histone proteins (H2A/B, H3, H4), resembling thread around a spool, to form a repeating unit called a nucleosome (Figure 1.2). Each nucleosome consists of a core (H3/H4)2 tetramer and two H2A/H2B dimers which form an octomeric complex, histone H1 acts as a linker between units.[2] These repeating nucleosome units couple to form higher order structures called chromatin which are further compacted into

chromosomes. The genetic information, or more specifically the genes carried within this structure, are transcribed into proteins.

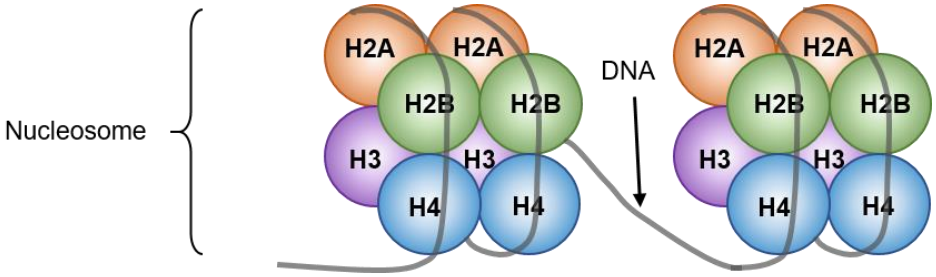


Figure 1.2 Schematic representation of nucleosome. DNA (black line) is wrapped around the core histone proteins (H2A, H2B, H3 and H4) which form an octomeric complex. Reproduced with amendments from Starkman *et al*.[3]

There are many genetic factors which can impact upon gene expression, including but not limited to molecular machinery such as transcription factors or hormones. Factors other than those originating from the genetic code can also alter gene expression. Environmental stimuli, both external and internal, can have a significant impact on gene expression. This connection was demonstrated in the 1940s when Conrad Waddington investigated the effect of heat shock on late stage *Drosophila* larvae. His findings showed remarkable changes to their initial development and subsequent genetic growth which in some cases resurfaced in later generations. Waddington coined the term 'Epigenetics' to describe his observations.[4] At this point in time the molecular nature of genes and genetics was not fully understood. Although his definition did not include any molecular insight, it did nonetheless provide a conceptual model on how genes might interact with external surroundings to produce a phenotype. A more detailed understanding of the mechanisms involved in epigenetics has since been established and is discussed in the following section.

1.2 Epigenetics

'Epigenetics' literally translates as 'above genetics', meaning the underlying genetic sequence, i.e. the ordering of the nucleoside units remains the same but the sequence is *interpreted* differently. A broad definition which encompasses this is changes to gene expression caused by mechanisms other than changes to the underlying genetic code. More specifically epigenetics is "the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states".[5]

This concept can be illustrated by a grammatical analogy, shown below, in which words represent the nucleobase units of DNA and a comma represents the epigenetic mark*:

'Let's eat Grandma'

'Let's eat, Grandma'.

Thus, the comma has changed the meaning of the sentence from suggesting we should eat Grandma, to inviting Grandma to eat with us. In other words, the order of the words in the sentence, i.e. the sequence of DNA, has not been altered.

As not all mechanisms defined as epigenetic are known to be passed on to the next generation, there is still debate over whether heritability should be included. Environmental stimuli which impact on gene expression are evident in everyday life and throughout history. For example, the Dutch famine which occurred over a 6-month period in the winter between 1944-1945 represents a defined population who all suffered from a period of malnutrition at the same time. A knowledge of their medical history allowed analysis of the long-term effects of the famine on this population. Findings suggested that depending on the stage of pregnancy during the famine, the resulting child would suffer from higher or lower than normal obesity rates. Furthermore the subsequent offspring of this generation, in some cases, also presented with the same effects.[6]

A striking visual example of epigenetics in action is provided by monozygotic (identical) twins.[7] Identical twins develop from one zygote that has split to form two embryos, meaning they have the same genetic code, hence why they are identical. However, they age differently. This is due to the impact of environmental factors — smoking, exposure to UV sunlight and radiation, pollution, infections, alcohol, diet and exercise — causing the genetic code (the DNA) to be *interpreted* differently in each twin. Looking at this on a molecular level, young twins are epigenetically indistinguishable whereas older twins show significant differences in the amount and distribution of epigenetic markers* which results in altered gene expression. This observation is called 'epigenetic drift'.[8]

* Explained more fully in 1.2.2 Epigenetic post-translational modifications

1.2.1 Epigenetics and potential in therapeutics

Epigenetic regulation is required for normal cellular development and function.[9] However, studies have shown it can also be responsible for the initiation and propagation of disease states across many major disease areas including cancer, neurological disorders, anti-infectives, metabolic disorders, immune disorders, pain and inflammation. These studies resulted in a growing interest in development of small molecule and peptide inhibitors to evaluate the biological functions and disease associations of the epigenetic mechanisms and thus as potential novel therapeutic targets. This relatively new field has progressed rapidly, providing promise of potential therapeutic benefit across many disease states.[10]

There are two main mechanisms by which epigenetic factors regulate gene expression: DNA methylation (specifically Cytosine methylation) and post-translational modifications (PTMs) usually on the *N*-termini of histone tails.‡ Combinations of these modifications or 'marks' form the 'epigenetic code'.[11] The code creates layers of regulatory control through single modifications and a considerable degree of 'cross-talk' which ensures correct gene function.[12, 13] Modifications, which are located on the same histone molecule, are referred to as *'cis*' cross-talk. Whereas '*trans*' cross-talk occurs if the modifications are on different histones or nucleosomes, or due to DNA methylation.[14] This thesis focuses on the inhibition and disease associations of epigenetic post-translational modifications (PTMs).

1.2.2 Epigenetic post-translational modifications

PTMs are dynamic and reversible. The modifications or marks include acetylation, methylation, phosphorylation and ubiquitination. The marks are mediated by specific families of enzymes termed 'readers', 'writers' and 'erasers' which facilitate the addition, interpretation and removal of marks onto histone tails respectively (Figure 1.3). These enzymes regulate and direct the complex expression of the genome either indirectly through the recruitment of position-specific binding modules, or directly by affecting the packaging of chromatin.[15] Chromatin packaging impacts upon the transcription of genes as it controls the accessibility of the DNA. Highly compact chromatin is called heterochromatin and is associated with diminished transcription.

‡ Eukaryotic histones have two domains, a central globular histone fold domain and less structured *N*-terminal tail regions. In addition, histones H2A and H2B contain unstructured *C*-terminal tails. The globular histone fold domains are the major structural components of the nucleosome core. The tail regions, in contrast, protrude out of the nucleosome.

Conversely, euchromatin is a more loosely packed structure and is associated with the most active portion of the genome (Figure 1.4).

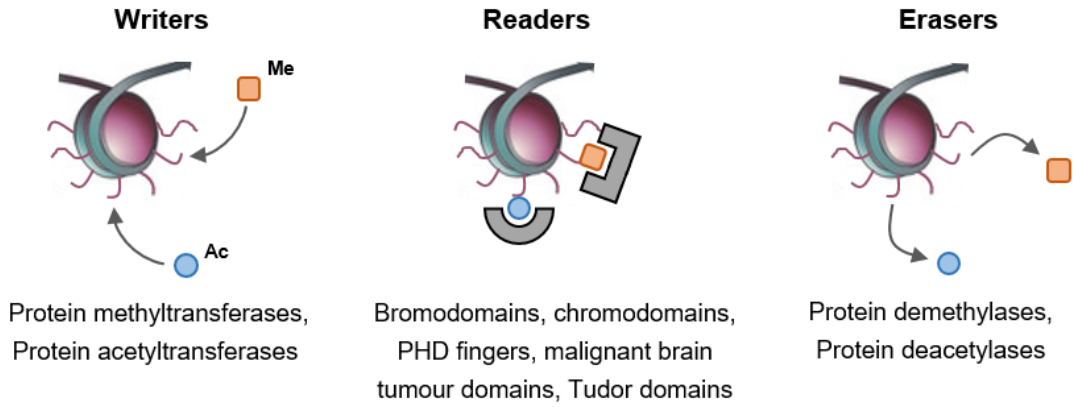


Figure 1.3 Classification of enzyme families involved in epigenetic mechanisms termed 'writers', 'readers' and 'erasers'. The 'marks' are represented by a blue circle and orange rectangle. The nucleosome is represented by a purple sphere with DNA represented by green cord. The 'reader' family are presented by a grey semi-circle and grey rectangle and the 'writer' and 'eraser' families are represented by arrows. Examples of enzyme families from each class have been provided. Reproduced with amendments from Arrowsmith *et al.*[10]

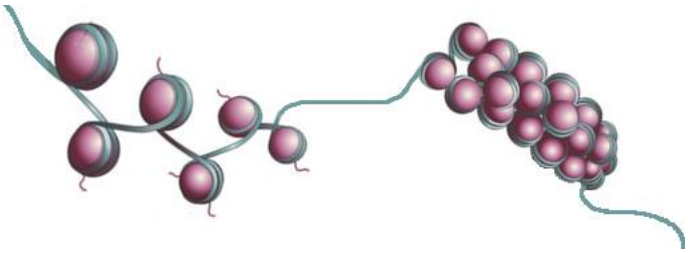


Figure 1.4 Graphical representation of (left) loose chromatin, 'euchromatin' and (right) compact chromatin 'heterochromatin'. Reproduced with amendments from Baylin *et al*.[1]

The writer family of protein methyltransferases (PMTs) constitute one of the largest classes of enzymes involved in post-translational epigenetic modification of histones and will be the focus of this thesis.

1.2.2.1 Protein methyltransferases

PMTs are involved in the methylation of mainly H3 and H4 histone proteins and are also known to modify many non-histone substrates. The PMT family share a catalytic mechanism in which a methyl group is transferred from the universal methylation co-factor *S*-adenosylmethionine (SAM) to the side chain of lysine and arginine residues producing the methylated substrate and by-product *S*-adenosylhomocysteine (SAH)

(Figure 1.5).[16] The substrate and co-factor binding sites are connected by the 'lysine channel', a narrow hydrophobic channel through which methylation occurs. The catalytic process can occur multiple times on one substrate in order to achieve higher methylation states.

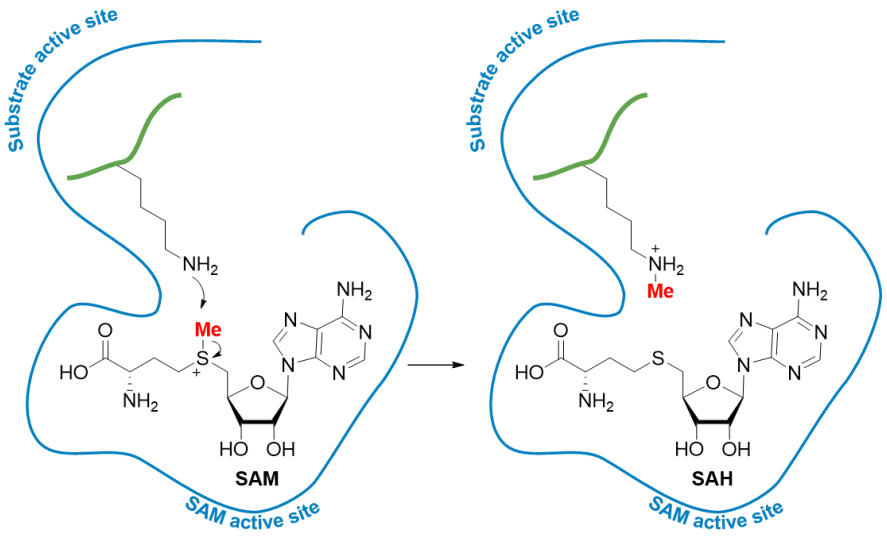


Figure 1.5 PMTs (blue line) catalyse the methylation of lysine and arginine substrates via an S_N2 methyl transfer from the co-factor *S*-adenosylmethionine (SAM) producing the methylated substrate and the by-product *S*-adenosylhomocysteine (SAH).

There are 64 known PMTs, which can be subdivided into two families, protein lysine methyltransferases (PKMTs) and protein arginine methyltransferases (PRMTs).

PKMTs catalyse the mono-, di- and trimethylation of specific lysine residues on histone tails and non-histone residues.[17] Almost all PKMTs contain a SET (**S**u(var)3–9, **E**nhancer-of-zeste, **T**rithorax) domain apart from one enzyme, DOT1L. The PKMTs are categorised by their structure into two classifications, SET-domain containing PKMTs and non-SET domain containing PKMTs. As DOT1L is the only PKMT without a SET domain, it is the sole member of the non-SET domain containing PKMT. The conserved SET domain is approximately 130 amino acids in length and was originally identified in *Drosophila* genes. SET domain containing PKMTs are sub-categorised according to sequence into five classifications: SUV, SET1, SET2, EZ and RIZ. An alternative classification system has also been suggested, which divides the PKMT family into eight classes, KMT1-8, based on function and reaction type rather than structure.[18] Both classifications are employed, however the former will be utilised for the purposes of this discussion.

PRMTs can mono- or dimethylate the guanidine group on the side chain of arginine residues. Dimethylation leads to two different methylation patterns: symmetrical or

unsymmetrical methylation.[19] PRMTs are subdivided into three classes depending on the nature of their associated methylation function: type I, type II and type III (Figure 1.6). All classes of PRMTs catalyse monomethylation of arginine residues. However, both type I and type II PRMTs also catalyse asymmetrical dimethylation and symmetrical dimethylation respectively. PRMT7 is the only known member of the type III PRMTs. It is known to only catalyse monomethylation of arginine residues. All PRMTs contain a conserved core region of 130 amino acids and typically have additions to the *N*-terminal.[20]

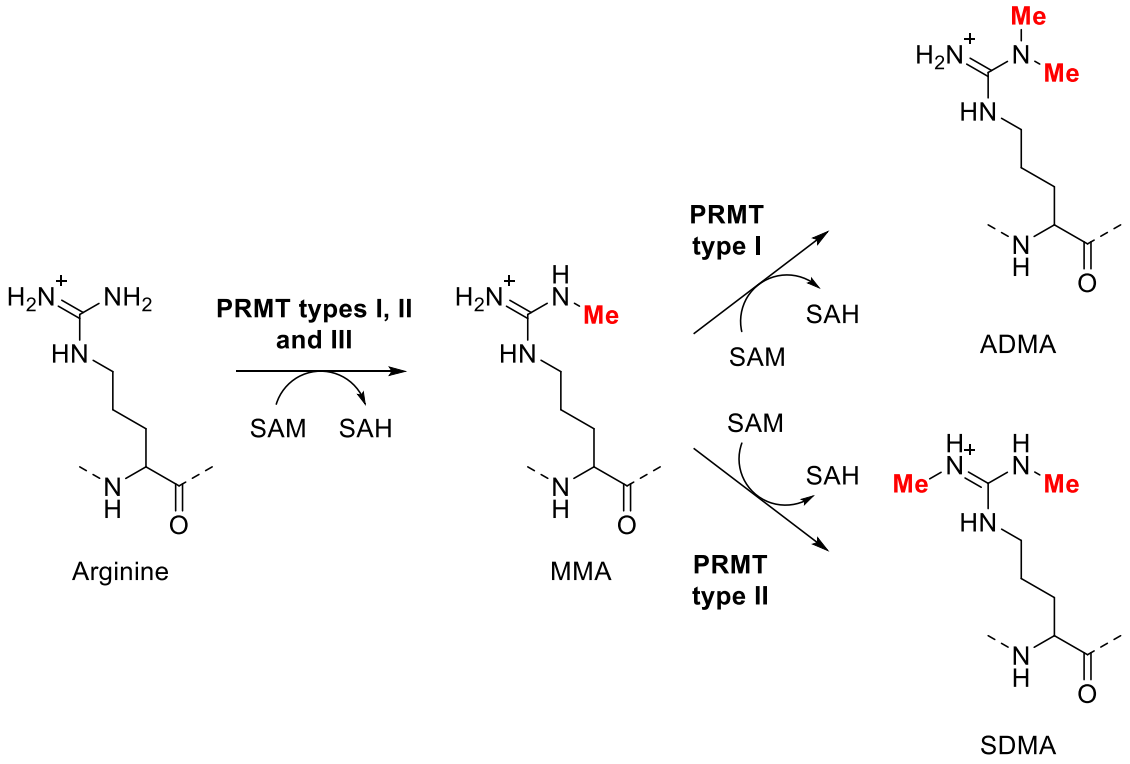


Figure 1.6 Mono, asymmetric and symmetric methylation of the arginine side chains by PRMTs.

It was assumed for many years that histone methylation was irreversible as it was observed to be a stable mark propagated through multiple cell divisions. However, it has been shown that histone methylation, like other PTMs, is an actively regulated and reversible process. Each lysine or arginine methylation mark creates a specific signal which is recognised by reader proteins (for example Tudor domains or PHD domains) and removed by histone demethylases (for example Jumonji domain containing). Histone methylation has been shown to either repress or activate gene expression depending on the location and residue which is marked. For example, arginine methylation on H3 and H4 leads to transcriptional activation whereas lysine methylation on H3 and H4 can lead to both transcriptional activation and repression. Furthermore, different levels of histone methylation have been shown to have varying effects; mono-

methylation of H3K9 (H3K9Me) is an activating mark whereas di- and trimethylation of H3K9 (H3K9Me2, H3K9Me3) are both linked to gene repression.[21, 22] This observation could be explained by the fact histone methylation affects the bulkiness and hydrophobicity but does not alter the charge of the modified residues. It is therefore less likely to direct the nucleosomal interactions required for chromatin folding but affects the recognition of the methylated substrate by the reader enzymes through disruption of the protein-protein interactions (PPI).

Although maintenance of methylation marks by PMTs is required for normal cellular development, aberrant activity associated with these mechanisms can lead to diseases including cancer. As such, molecular targeting of specific histone methyltransferases has emerged as an attractive novel therapeutic avenue.

1.2.2.2 Inhibition of histone methyltransferases

Molecular targeting of specific histone methyltransferases has emerged as a fast-paced area of research over the past decade. Since the first PRMT and PKMT inhibitors were identified in 2004 and 2005 respectively, multiple PMT inhibitors have been disclosed with high potency and selectivity (Figure 1.7).[23] The reported inhibitors act through five mechanisms of action: (i) substrate-competitive inhibitors (ii) SAM-competitive inhibitors (iii) allosteric inhibitors (iv) inhibition via covalent modification and (v) indirect inhibition. Due to the expanse of this field, a synopsis of key developments is given below.

Two sites readily targeted for small molecule inhibition in PMTs comprise the substrate site and cofactor binding site. Molecular targeting of PMTs at these sites can be a challenge due to the high sequence homology, and the distinct and conserved patterns of H-bonding interactions between an enzyme and co-factor observed in PRMTs and PKMTs. Despite these challenges, many selective and potent substrate-competitive and SAM-competitive inhibitors have been identified and co-crystallised with their targets.

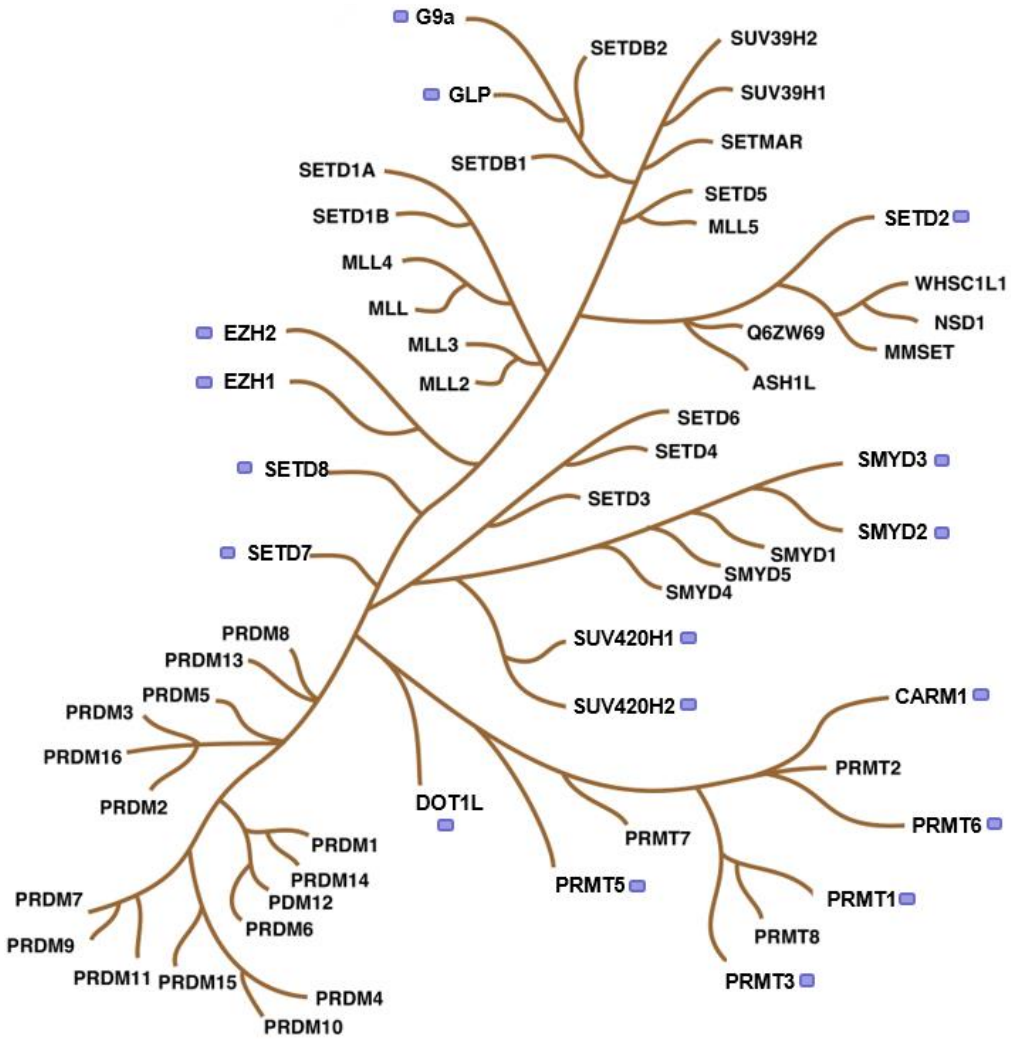


Figure 1.7 Phylogenetic tree of protein methyltransferases. Proteins are grouped based on similarities in their amino acid sequences. It is often more difficult to obtain selectivity for proteins on the same branch as they are structurally similar. Blue squares indicate where an inhibitor of the target protein has been identified. Reproduced with amendments from Kaniskan *et al*.[23]

SAM-competitive inhibitors have made the most progress towards therapeutic use in humans with inhibitors of two targets, EZH2 and DOT1L, in clinical trials. EZH2 catalyses the trimethylation of H3K27. Over-expression and abnormal or hypermethylation of H3K27, driven by point mutations occurring at Tyr641 and Ala677, have been implicated in multiple cancers including: breast, prostrate, lymphoma, myeloma and leukaemia.[24] Towards inhibition of EZH2, three selective and potent SAM-competitive inhibitors have progressed to the clinic, GSK126 (**1**), EPZ-6438 (**2**) and CPI-1205 (**3**) (Figure 1.8). All three scaffolds have high potency (low nanomolar) and selectivity over both the closely related PKMT EZH1 (>26-fold) and other methyltransferases or related enzymes, as well as receptors and ion channels. GSK126 (**1**) entered phase I clinical trials in 2014 for patients with relapsed diffuse large B cell lymphomas (DBCL), transformed follicular lymphoma and other Non-

Hodgkin's lymphomas, solid tumours, and multiple myeloma. GSK126 (**1**) is not orally bioavailable and was administered via intravenous infusion.[23] However, the study was recently terminated due to 'insufficient evidence of clinical activity' at maximal dose [NCT02082977].[25] Conversely, both EPZ-6438 (**2**) and CPI-1205 (**3**) are orally bioavailable. EPZ-6438 (**2**), developed by Epizyme, is currently in a two-part phase I/II clinical study for patients with DBCL and follicular lymphoma and is also in additional clinical studies for different diseases [NCT01897571].[26] The first part of the study is completed and EPZ-6438 shows a favourable safety and tolerability profile.[27] CPI-1205 (**3**) was disclosed by Bradley *et al*. and is the third EZH2 inhibitor to reach human clinical trials.[28] It is currently in phase I clinical trials for patients with B-cell lymphoma [NCT02395601].[29]

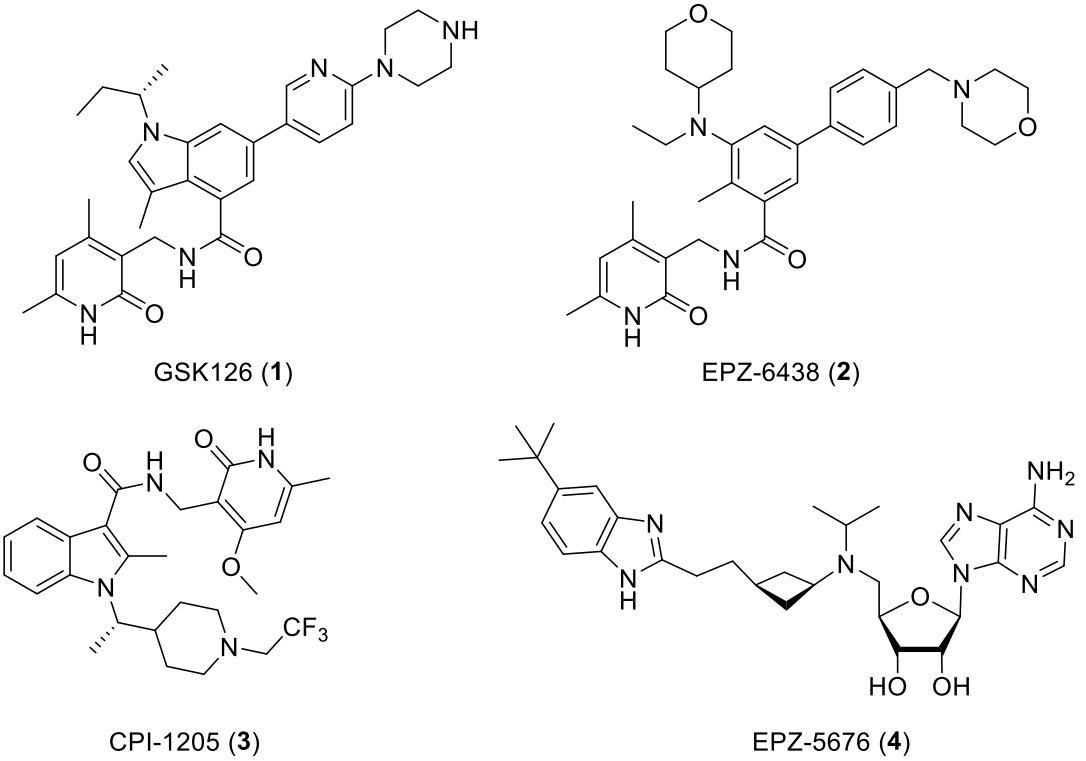


Figure 1.8 Structures of EZH2 **1**–**3** and DOT1L **4** inhibitors.

DOT1L is a non-SET domain containing PKMT which methylates H3K79. It has been implicated in the proliferation of MLL mediated leukaemias amongst other diseases. Epizyme developed the potent and selective SAM-competitive DOT1L inhibitor EPZ-5676 (**4**) (Figure 1.8). It was the first PMT to reach clinical trials and has recently completed phase I clinical trials for patients with MLL-rearranged (MLL-r) leukaemia [NCT02141828].[30] It is not orally bioavailable and is administered intravenously.

Substrate-competitive inhibitors have been extensively developed towards both PRMTs (including PRMT1, PRMT5, CARMI, PRMT6) and PKMTs (including G9a/GLP, SETD7, SETD8, SMYD2 and SMYD3).[23] Recently, a PRMT5 substrate-competitive inhibitor, GSK3326595 **5**, was reported to have entered clinical trials [NCT02783300][31] for patients with solid tumours and Non-Hodgkin's Lymphoma (Figure 1.9).[19] As previously noted, selective inhibition of PMTs can be difficult to achieve due to high sequence homology. This is particularly true for PKMTs G9a and GLP which share 80% sequence homology. Despite this similarity, G9a and GLP possess distinct physiological and pathophysiological functions.[32] Although multiple potent and selective G9a/GLP inhibitors have been published, including UNC0642 (**6**) (GLP/G9a IC_{50} = <2 nM, >2,000-fold selective over other PMTs) which is suitable for *in vivo* studies, they all have comparable inhibitory activity for both G9a and GLP (Figure 1.9). This issue was recently addressed with the disclosure of a potent GLP selective inhibitor, MS012 (**7**) (IC_{50} = 7 nM) (Figure 1.9). Despite possessing the quinazoline scaffold like its predecessors and binding in virtually identical binding orientations, it is 140-fold selective for GLP over G9a (Figure 1.9). Additional structural studies are currently being conducted to establish the structural basis for the observed selectivity.[33]

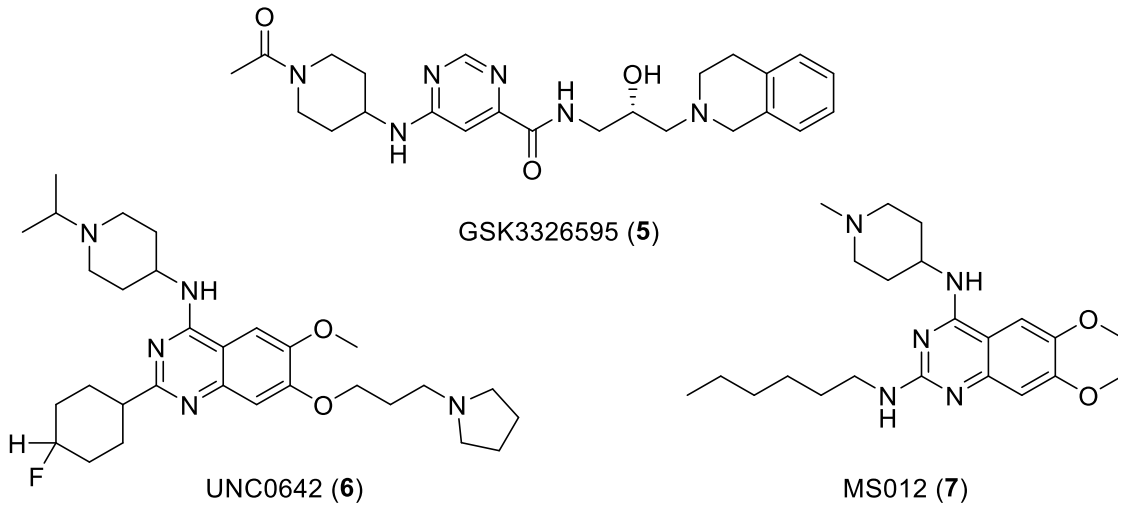
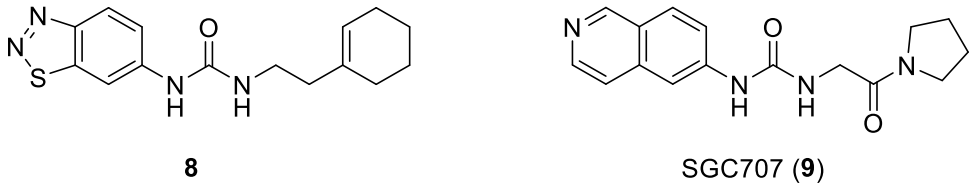


Figure 1.9 Structures of substrate competitive inhibitors for PRMT5 **5**, GLP/G9a **6** and selective GLP inhibitor **7**.

More recently three alternative mechanisms of action to attain inhibitory activity against PMTs have been disclosed via (i) allosteric inhibition, (ii) covalent modification and (iii) disruption of the PPI. Siarheyeva *et al*. reported the first allosteric inhibitor of PMTs, **8** which exhibited micromolar affinity for PRMT3 and was identified from a screen of 16,000 compounds (Figure 1.10).[34] From this micromolar inhibitor, Kaniskan *et al*. developed SGC707 (**9**) (IC_{50} = 31 nM) using SAR (Figure 1.10).[35] The allosteric

inhibition mechanism was confirmed through mechanism of action (MOA) studies where SGC707 (**9**) displayed a non-competitive inhibition pattern towards both the substrate and co-factor binding sites. A co-crystal structure of the PRMT3-SGC707 (**9**) complex (PDB code: 4RYL) demonstrated the inhibitor binds to an allosteric site in the protein which disrupts the catalytically active state. SGC707 (**9**) has been shown to be cell active and also applicable in *in vivo* mouse studies making it a useful biochemical tool for probing PRMT3 disease associations and biological functions.

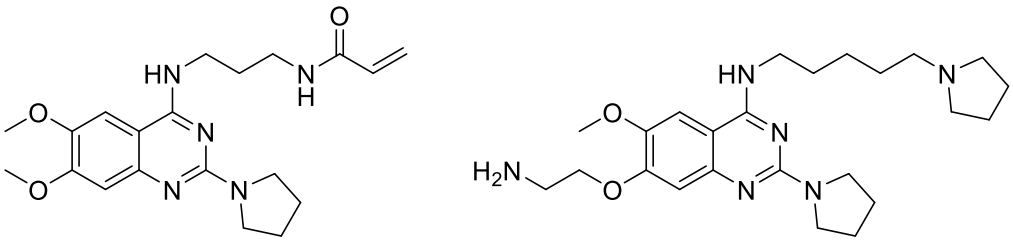


8 SGC707 (**9**)

Figure 1.10 Structures of allosteric PRMT3 inhibitors **8** and **9**.

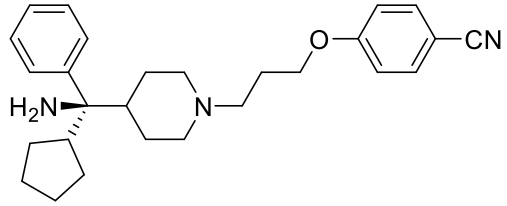
Butler *et al*. identified a covalent inhibitor of the PKMT SETD8, MS453 (**10**) (IC_{50} 0.8 µM) (Figure 1.11).[36] The covalent inhibitor was designed from a published substrate-competitive analogue MS2177 (**11**) (IC_{50} = 1.9 µM) (Figure 1.11). The co-crystal structure of MS2177 (**11**) in complex with SETD8 (PDB code: 5T5G) revealed a cysteine residue (C311) in close proximity to the substrate binding pocket. To exploit residue C311 for covalent modification, an electrophilic acrylamide moiety was installed resulting in the covalent inhibitor MS453 (**10**). Although potent, this inhibitor has poor cell permeability and so is not suitable for cellular studies. Nevertheless, it has shown that inhibition of PKMT by covalent modification of an active cysteine residue is possible and paves the way for future covalent PKMT inhibitors.

Finally, the PKMT MLL has been targeted for inhibition indirectly using PPI inhibitors. The interaction between MLL and menin, a highly specific fusion partner for MLL, has been shown to be critical for acute leukaemias bearing chromosomal rearrangements, thus disruption of this PPI interaction is an attractive therapeutic strategy. He *et al.* disclosed MIV-6r (**12**) as a PPI inhibitor targeting the menin-MLL PPI interaction (IC_{50} = 56 nM) (Figure 1.11).[37] MIV-6*R* was developed from a micromolar HTS screening hit of ~288,000 compounds. Through anti-proliferative studies, using MV4-11 and MOLM-13 cell lines for AML, they showed MIV-6*R* (**12**) effectively arrests cell growth (GI_{50} = 1.1 µM). Furthermore, through co-crystal structures they demonstrated MIV-6*R* (**12**) mimics key binding interactions of MLL with menin (PDB code: 4G08) thus demonstrating its MOA.



MS453 (**10**)

MS2177 (**11**)



MIV-6*R* (**12**)

Figure 1.11 Structures of the covalent SETD8 inhibitor **10**, substrate-competitive SETD8 inhibitor **11** and PPI inhibitor for MLL **12**.

In summary, nearly a third of known PMTs can be potently and selectively inhibited via a variety of MOA including substrate-competitive, SAM-competitive, allosteric, covalent inhibition and PPI inhibition. Although this is significant progress, there are still many PMT targets which lack selective inhibitors or for which inhibitors have not yet been identified. Furthermore, although numerous inhibitors have reached clinical trials many lack good physicochemical properties and are not orally bioavailable, for example EZH2 inhibitor GSK126 (**1**) and DOT1L inhibitor EPZ-5676 (**4**), leaving room for further optimisation. This thesis will focus on the design and synthesis of selective and potent inhibitors of the PKMT DOT1L with improved physicochemical properties.

CHAPTER 2
DOT1L

2.1 Function and role

Disruptor **O**f **T**elomeric silencing 1-**L**ike (DOT1L) is a PKMT and as previously indicated, does not contain the conserved SET domain required for catalytic activity present in all other known PKMTs. Instead, DOT1L adopts a folding topology which is also observed in the PRMT family and DNA methyltransferases (DNMTs). DOT1L is responsible for mono-, di- and trimethylation of histone 3 lysine 79 (H3K79).[38] Knockout studies with mice suggested DOT1L was the sole methyltransferase (MTase) acting on H3K79. However, it was recently suggested the multiple myeloma SET domain (MMSET) isoform, interleukin-5 response element II binding protein (RE-ITBP), also methylates this site.[39] H3K79 is located on the nucleosomal surface as opposed to the *N*-terminal histone tails where epigenetic modifications usually occur.[40] The location of H3K79, coupled with the observation methylation only takes place in *in vitro* studies if the full nucleosome complex is present, suggests DOT1L recognises other features of the nucleosome and thus *trans*-histone cross talk plays a role in regulating DOT1L activity. Indeed, it has been shown ubiquitination of H2B at K120 (H2BK120-Ub) is required for stimulation of human DOT1L (*h*DOT1L) activity coordinating the active site of DOT1L toward the location of H3K79.[41] There is evidence to suggest H3K79 methylation is subject to dynamic regulation which indicates H3K79 methylation is reversible. However, to date, there are no known H3K79 demethyltransferases (HDMTs).[40]

Genome-wide analysis (GWAS) of the mammalian genome has shown H3K79 methylation is broadly associated with actively transcribed genes.[21] Studies in *Drosophila* have demonstrated DOT1L plays an important role in early embryonic development and has further roles in cardiac and kidney function.[42] H3K79 methylation has also been associated with transcriptional regulation, DNA repair, cell cycle regulation and hematopoiesis.[43, 44]

2.2 Topography of the DOT1L catalytic site

Full length human DOT1L (*h*DOT1L) consists of 1,537 amino acids, of which residues 1-416 comprise the catalytic domain. A co-crystal structure of SAM in complex with the

catalytic domain of *h*DOT1L (PDB code: 1NW3), revealed the *h*DOT1L catalytic domain to be an elongated structure consisting of an *N*-terminal domain (residues 1-123) and *C*-terminal domain (residues 141-332) connected via an L-EF loop (residues 122–140).[45] The L-EF loop, or activation loop, encloses the SAM-binding pocket and comprises one of the five segments of the protein which form the SAM binding site (D1, D2, I, II and I') (Figure 2.1).

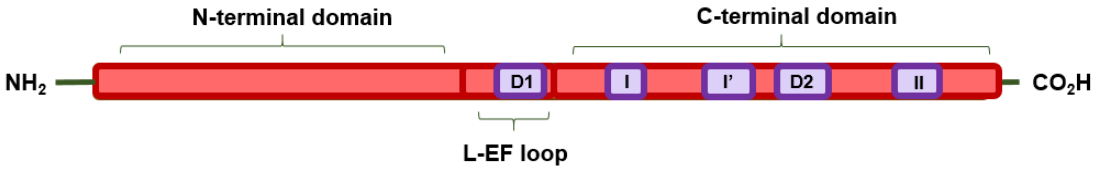


Figure 2.1 Domain structure of DOT1L. The five SAM binding sites, D1, D2, I, II and I' are depicted by purple boxes.

The adenine moiety of SAM sits in a hydrophobic portion of the binding site at the surface of the protein, while the methionine moiety sits deeper in the binding site in a narrower, negatively charged pocket. DOT1L-SAM binding is governed by hydrophobic or van der Waals interactions and specific H-bond interactions. The H-bonding network between DOT1L and SAM closely reflects the network observed for the PRMTs, with the exception of the loss of the carboxylate H-bond on the methionine moiety (Figure 2.2).[46] Set-domain containing PKMT form two H-bonds with the 6-amino adenosine moiety as opposed to one H-bond observed between SAM and DOT1L.

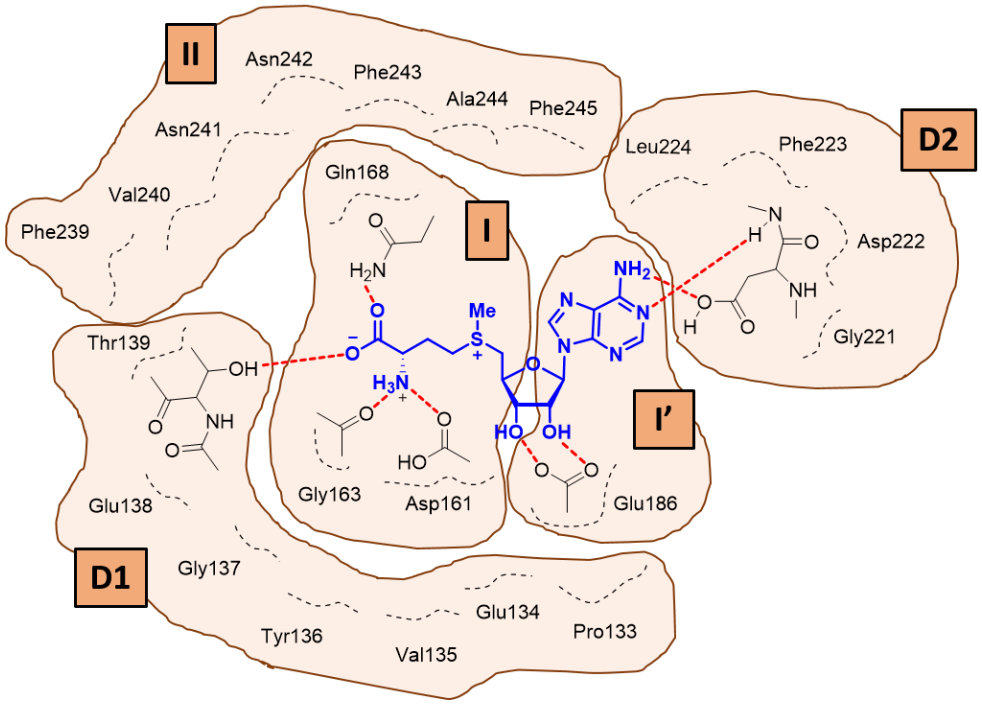


Figure 2.2 Schematic diagram showing SAM-DOT1L interactions in the DOT1L catalytic site. The five segments of DOT1L are labelled and shaded in orange. Most amino acids are involved in van der Waals interactions. Specific H-bond interactions represented by red dotted lines. Reproduced from Min *et al*.[45]

The substrate-binding site, formed from the substrate binding loop (residues 301-311), is connected to the SAM binding site pocket via a 4-Å transfer channel through which methylation occurs. Adjacent to the cofactor-binding site is a small hydrophobic cavity or pocket specific to DOT1L, composed of residues F245, V249, L224 and P133.

2.3 Disease association

DOT1L has been associated with the proliferation of various disease states including several cancer types. DOT1L cooperates with the c-Myc-p300 complex which epigenetically activates epithelial-mesenchymal transition (EMT) regulators in breast cancer progression.[47] DOT1L was also found to play an important role in the canonical Wnt signalling cascade forming a molecular complex involving TCF transcription factors. Wnt signalling is required for joint development and homeostasis, increased activation in progressive osteoarthritis (OA) is detrimental to the joint.[48] More extensive therapeutic investigation has focused on the association of DOT1L with the proliferation of mixed lineage leukaemia (MLL) mediated leukaemias - acute myeloid (AML), acute lymphoblastic (ALL) and mixed lineage leukaemia (MLL); all of which are aggressive blood cancers which have a low patient survival rate. The MLL protein is a methyltransferase that is involved in the positive regulation of Hox gene expression and methylation of histone H3 lysine residue 4 (H3K4). The function of the MLL protein can be disrupted due to inter-chromosomal translocation (11q23), resulting in fusion of the MLL *N*-terminus to the *C*-terminus of one of more than 60 partner proteins and the loss of the H3K4 methyl transferase domain. Many of the MLL fusion partners interact with and recruit DOT1L, which is subsequently directed to maintain leukaemogenesis through aberrant H3K79 methylation.[49] The current treatment for patients with leukaemia harbouring MLL translocations has a highly unfavourable prognosis. The therapeutic significance of DOT1L in established MLL-r leukaemia has been validated by genetic and chemical approaches.

2.4 Inhibition of DOT1L[§]

DOT1L is an attractive drug target: (i) it is the only known H3K79 PKMT, (ii) it has a unique non-SET catalytic domain, and (iii) it has a known role in promoting leukaemogenesis. These factors present a strong opportunity to deliver a novel therapeutic agent. At present, all published efforts at small molecule inhibition of

[§] Since completion of this study further progress towards DOT1L inhibition has been made, for a full account please refer to 4.7 Epilogue.

DOT1L are structural mimetics of the cofactor SAM, although it has been indicated that the substrate binding site is theoretically druggable.[50] A synopsis of published SAM-competitive inhibitors is given below.

Epizyme was the first to report a selective and potent DOT1L inhibitor, EPZ-4777 (**13**) (IC_{50} = 0.4 nM) (Figure 2.3).[51] EPZ4-777 (**13**) was also the first reported SAM-competitive inhibitor of a PMT with *in vivo* efficacy. The selectivity of EPZ-4777 (**13**) was attributed to the large conformational change induced by the urea side chain as observed in the co-crystal structure of the DOT1L:EPZ-4777 (**13**) complex (PDB code: 4ER3). Although potent and selective, EPZ-4777 (**13**) required subcutaneously implanted osmotic mini-pumps to achieve *in vivo* inhibition due to poor pharmacokinetic properties.[52] An analogue with improved pharmacokinetics, EPZ-5676 (**4**), was identified through structure-guided drug discovery.[53] EPZ-5676 (**4**) was reported to be highly potent (K_i = 0.08 nM) and selective (>37,000-fold) over a panel of PMTs. EPZ-5676 (**4**) (also known as Pinometostat) is currently undergoing phase I clinical trials and is the first PMT to do so [NCT02141828].[53] Although EPZ-5676 (**4**) exhibits improved pharmacokinetics relative to EPZ-4777 (**13**), non-clinical *in vivo* studies indicate EPZ-5676 (**4**) still has moderate to high clearance and low oral bioavailability, most likely due to undesirable pharmacokinetic properties in the form of high metabolic instability and lipophilicity.

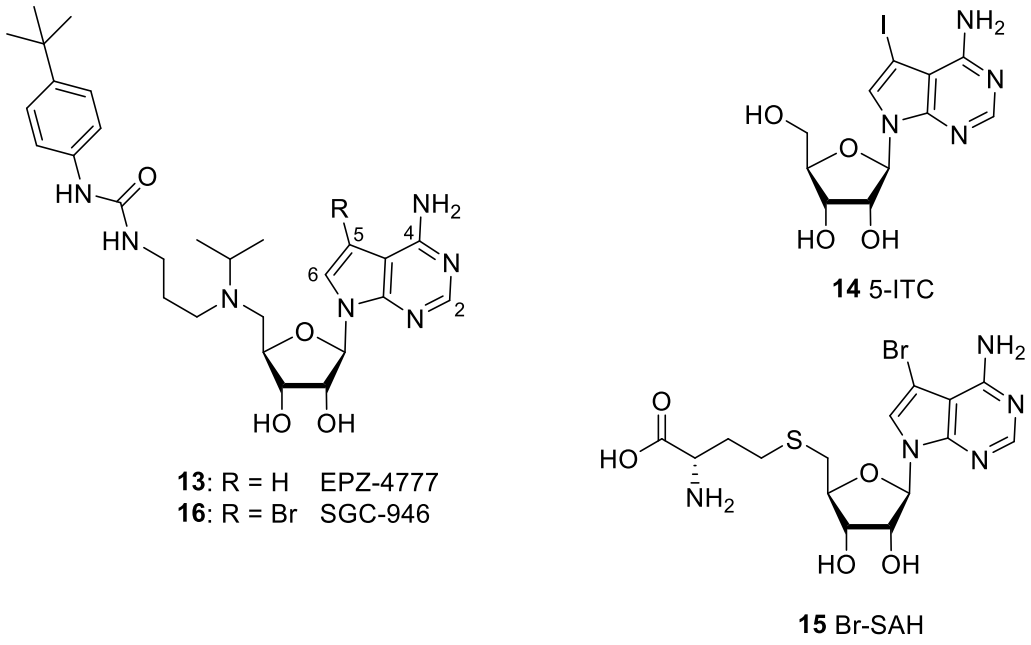


Figure 2.3 Structures of published DOT1L inhibitors **13**-**16**.

Indeed, metabolite data from EPZ-5676 (**4**) shows the appearance of the adenine nucleobase by LRMS.[54] EPZ-5676 (**4**) is administered in the clinic by continuous intravenous infusion due to low oral bioavailability.

It has been shown that occupation of the DOT1L hydrophobic cleft, which is adjacent to the nucleobase core of the SAM binding site, can significantly increase potency as well as selectivity of the inhibitor. From a screen of a library of 3,120 kinase inhibitors, the Structural Genomics Consortium (SGC) identified a small molecule inhibitor 5-iodotubercidin (**14**, 5-ITC) which exhibited a surprising[**] micromolar activity against DOT1L (IC_{50} = 18 μM) (Figure 2.3).[55] Co-crystal structures of 5-ITC (**14**) and DOT1L indicated that the iodo substituent occupies the hydrophobic cavity (PDB code: 3UWP). This lipophilic group could disrupt the critical H-bonding required for effective binding in other PKMTs thus providing a rationale for the observed selectivity. Using this key observation, the SGC subsequently designed Br-SAH (**15**), a potent (IC_{50} = 77 nM) and selective (30-fold over a panel of PMTs) biochemical tool (Figure 2.3).[36] The 5-Br substituent was shown to occupy the hydrophobic cleft in a co-crystal structure demonstrating proof of concept (PDB code: 3SX0). From these observations, the 5-Br substituent was translated into a cell penetrant analogue SGC-946 (**16**) (IC_{50} = 0.3 nM), an SGC-Epizyme hybrid of EPZ-4777 (**13**) with a 5-H to 5-Br substitution (Figure 2.3).[56] SGC-946 (**16**) displayed improved *in vitro* potencies compared to EPZ-4777 (**13**), it was almost 10-fold more potent than EPZ-4777 (**13**) at reducing H3K79 methylation levels in MCF10A cells (IC_{50} = 8.8 nM *c.f.* 84 nM respectively).

Two groups independently identified compound **17** (Figure 2.4); an analogue of the nucleoside inhibitor **13** designed by Epizyme with a 5-CH to 5-N substitution. Compound **17** was found to be both potent (K_i = 0.46 - 1.99 nM) and selective towards inhibition of DOT1L over a small panel of PMTs (>100 μM for CARM1, PRMT1 and SUV39H1).[56, 57] Yu *et al*. argued the structural modification away from the adenosine core in EPZ-4777 (**13**) did not serve as an advantage as it complicated the synthetic effort and impaired solubility.[56] Yao *et al*. also exploited the hydrophobic cavity, showing *N*-substitution of the 6-NH_2 (adenosine numbering system) on SAH could also be tolerated by methyl (**18**), allyl (**19**) and benzyl (**20**) moieties (K_i = 0.76, 12 and 22 nM respectively) (Figure 2.4).[57] The same group also published compound **21**, a potent (IC_{50} = 38 nM) and selective (>29-fold over CARM1, PRMT1 and G9a) DOT1L inhibitor with a covalent MOA. It was hypothesised the compound would undergo intramolecular cyclisation to form a reactive aziridinium intermediate, which may then covalently bind

[**] Comparable ligand efficiency (LE): 5-ITC (**14**) = 0.35 vs EPZ-4777 (**13**) = 0.34

to the -NH_2 group of H3K79. The activity of this DOT1L inhibitor in cell-based assays was not reported.[58]

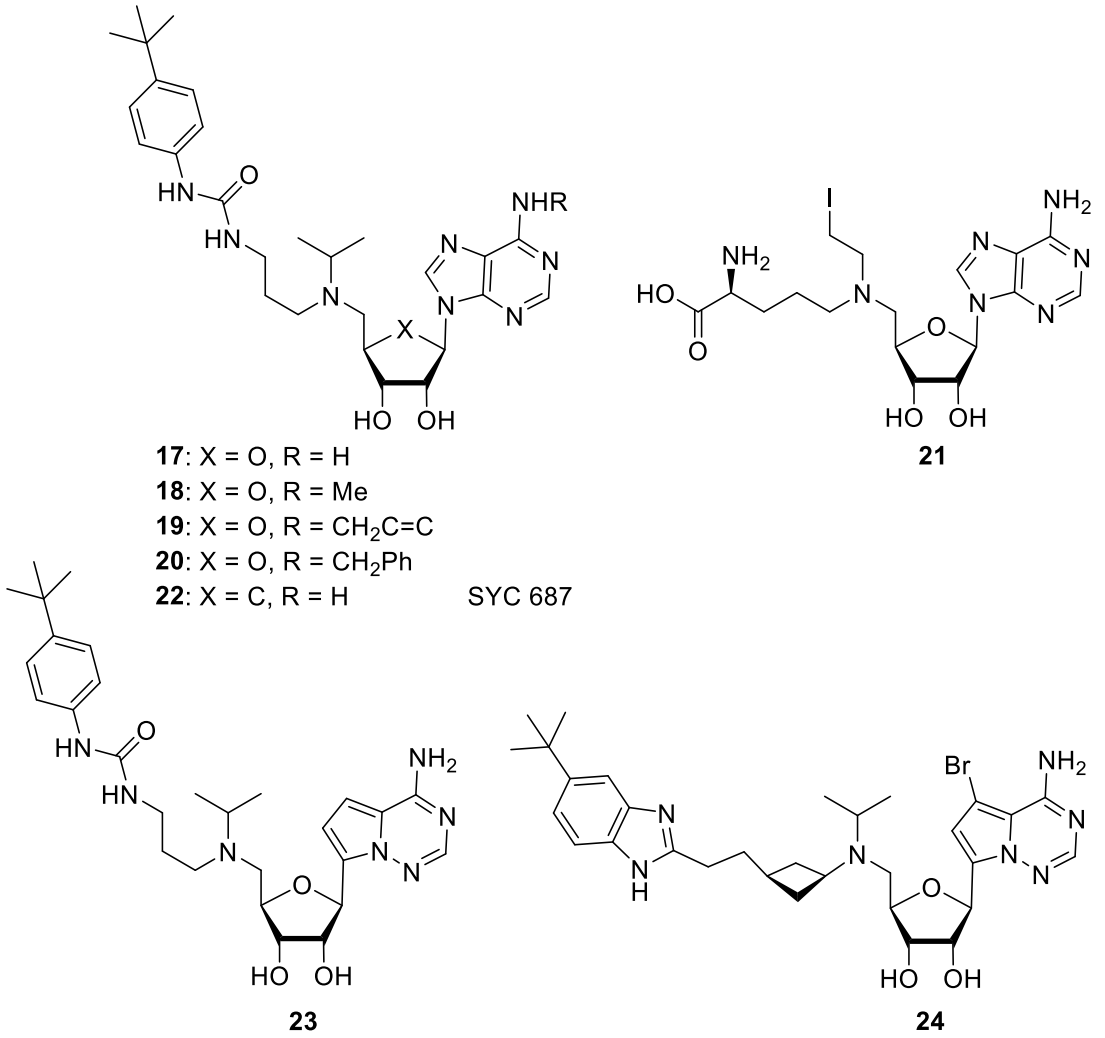


Figure 2.4 Structures of published DOT1L inhibitors **17**-**24**.

The west side or 'tail' of these inhibitors no longer resembles the methionine amino acid structure found in the natural co-factor SAM. The nucleoside core on the other hand still bears great structural similarity to the metabolically labile SAM structure. It is present in many naturally occurring structures and has been associated with a broad range of biological activities, receiving attention as potential antiviral and anticancer agents.[59, 60] Hence, this structure is recognised by many metabolic enzymes and so is metabolically labile, as demonstrated by the metabolite data for EPZ-5676 (**4**). Two structures have been published which move away from this nucleoside core in an attempt to find templates with more favourable pharmacokinetic properties. First, Deng *et al*. reported a carbocyclic compound SYC687 (**22**) which features an O-deletion in the ribofuranose ring.[61] It retains potency (K_i = 1.1 nM) and selectivity (45-fold over

PRMT1, CARM1 and SUV39H1) for DOT1L and more importantly has high metabolic stability in human plasma and liver microsomes. This demonstrates the beneficial pharmacokinetic effects gained on changing the metabolically labile ribofuranose ring.

Second, Moon *et al*. demonstrated that potency can be retained while altering the adenosine nucleobase through movement of the nitrogen atoms generating **23** and **24** (IC_{50} = 0.4 nM and 3.4 nM respectively) (Figure 2.4).[62] These compounds were also shown to be selective versus a panel of PKMTs and PRMTs (IC_{50} = >10 μM). The activity of these DOT1L inhibitors in cell-based assays was not reported.

Bradner *et al.* noted current biochemical assays to assess DOT1L activity are not easily accessible.[63] For example, current assays often require radioligands and highly purified histone proteins as substrates. Thus, they developed two new biological assays to aid identification of novel DOT1L inhibitors: an AlphaScreen nanomaterial proximity assay and a fluorescent polarisation assay (FP). These assays were developed using two tagged DOT1L ligands through *N*-substitution at 6-NH_2 on the adenine base, polyethylene glycol (PEG) linked biotin **25** and a thiourea coupled fluorescein (FITC) **26** (Figure 2.5). They also noted a possible hurdle in identifying inhibitors via cell-based assays is perhaps, in part, due to the relationship between ubiquitination of H2B and methylation of H3K79 catalysed by DOT1L.

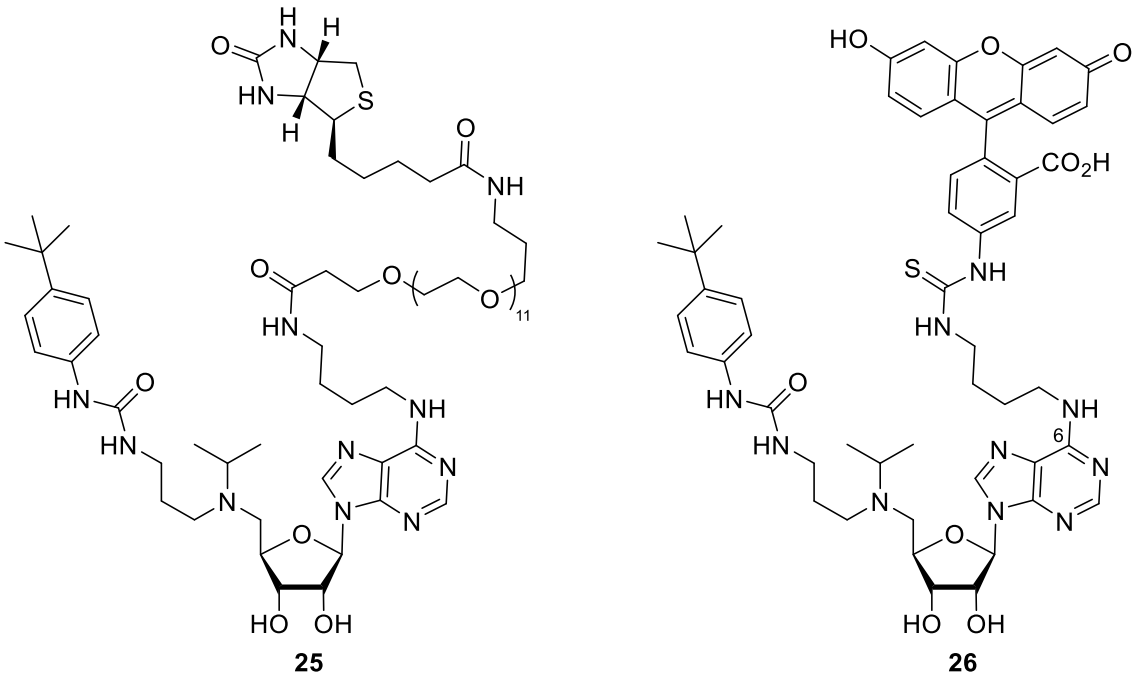


Figure 2.5 Structures of *N*-substituted DOT1L ligands, PEG linked biotin **25** and a thiourea linked fluorescein **26**.

In summary, several potent and selective SAM-competitive DOT1L inhibitors have been reported which use the adenosine core as a template. Also, a milestone has been reached toward inhibition of PMTs with the first-in-class DOT1L inhibitor EPZ-5676 (**4**), which is in clinical trials. Despite this progress, many of these inhibitors lack metabolic stability due to resemblance to the metabolically labile adenosine core. Attempts to move away from this core have shown promise, although there is room for further optimisation of this series of inhibitors.

2.5 Aim

Current published DOT1L inhibitors are potent and selective. However they lack good pharmacokinetic properties. The aim of this thesis was to improve the physicochemical properties of these published DOT1L inhibitors, while maintaining potency and selectivity. It was hypothesised this aim could be achieved by moving away from the metabolically labile adenosine core and reducing the lipophilicity thus moving to a more 'drug-like' template. It was envisaged that this could be accomplished either through single point transformations to the nucleoside core or through exploitation of the DOT1L hydrophobic pocket. It was hypothesised that through incorporation of these transformations, a novel therapeutic inhibitor with *in vivo* efficacy and drug-like pharmacokinetics could be delivered. This method involved three stages:

(i) structure-activity guided design – achieved using a suitable model template to investigate substitutions into the hydrophobic cavity and single point transformations on the nucleoside core

(ii) translate suitable substitutions into a biochemical tool to deliver proof of concept

(iii) incorporate the substitutions into the published DOT1L inhibitors to improve the drug-like properties of this template and thus deliver a novel DOT1L inhibitor suitable for *in vivo* use by oral delivery.

CHAPTER 3
Structure-activity guided design

3.1 Investigations into potential substitution sites

Through combined analysis of structural literature reviews of the DOT1L active site[50] and published crystal structures of DOT1L inhibitors from the protein data bank (PDB), key interactions important to binding were identified. From this analysis, heteroatoms that might be redundant or less important to binding were noted. A summary of the key interactions and potential sites for alteration are depicted in Figure 3.1.

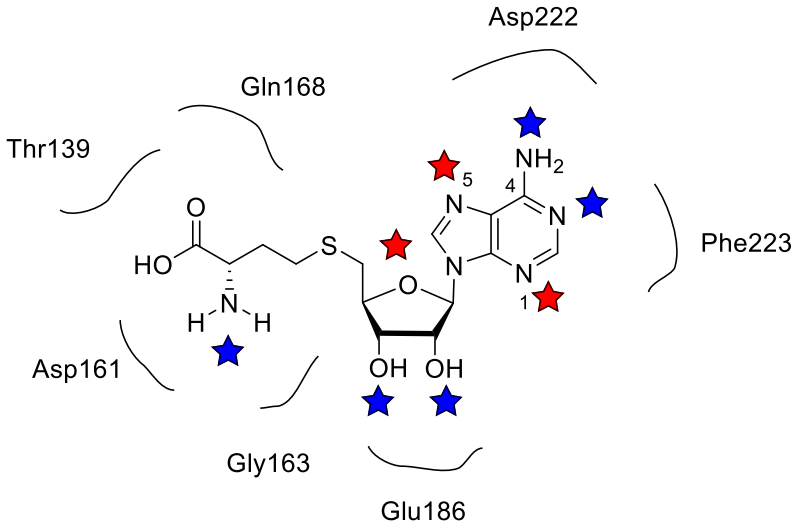


Figure 3.1 Summary of conserved hydrogen bonds in DOT1L of published small molecule inhibitors. Blue star = conserved. Three letter codes denote the side chain amino acid H-bond interactions. Red star = key areas which could be exploited to improve pharmacokinetics.

Blue stars highlight conserved or important H-bonding interactions from DOT1L structures. Three areas, highlighted by red stars, were identified on the core nucleoside structure which could be exploited to improve the pharmacokinetics of published DOT1L inhibitors: (i) deletion of N-5 allows the attractive option of substitution at this site allowing for further exploration of the hydrophobic cavity, (ii) N-1 deletion and (iii) O-deletion in the ribofuranose ring could both move the inhibitor template away from the metabolically labile nucleoside core towards a more drug-like template.

A small library of nucleoside target compounds was selected to explore the effect of these potential substitution sites on inhibition of DOT1L (Figure 3.2). Compounds were either purchased from commercial sources (**14** and **27**-**34**) or prepared via short

synthetic sequences (**35**-**42**). Investigation of a carbocycle template (O-deletion in the ribofuranose) was also sought as it was hypothesised it would confer metabolic stability through removal of the labile glycosidic linkage. However, it was not possible to obtain a commercially available or easily synthetically tractable carbocycle structure to explore this ribofuranose substitution.

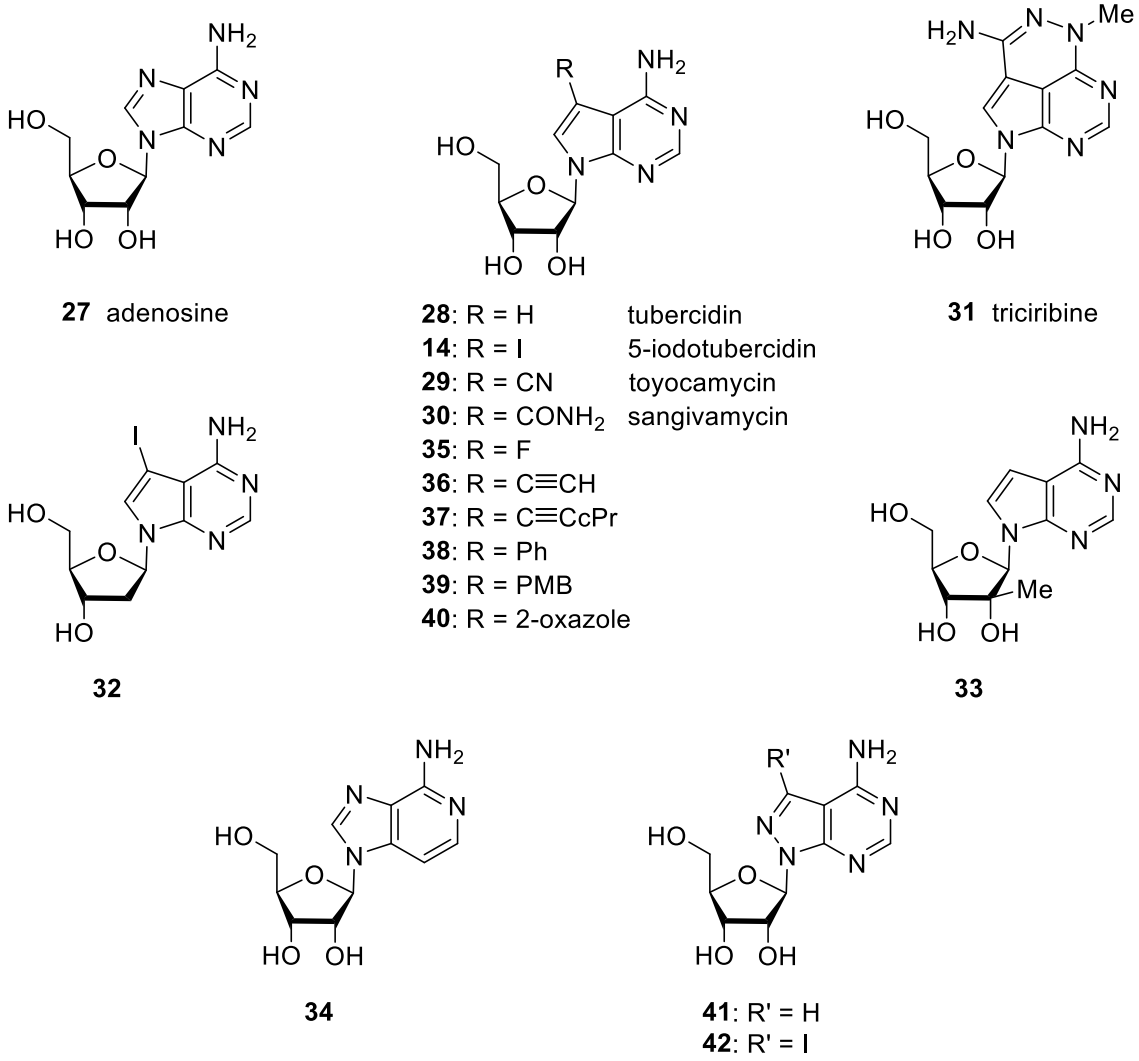
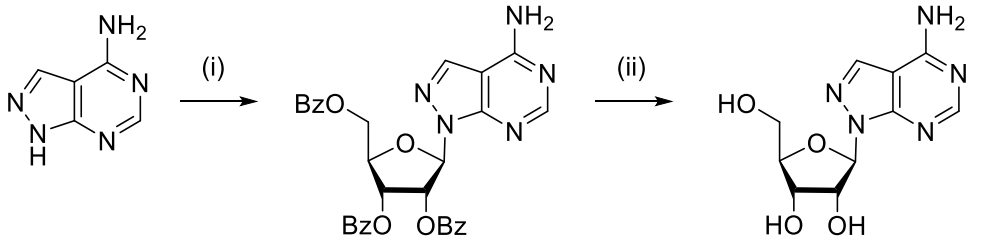


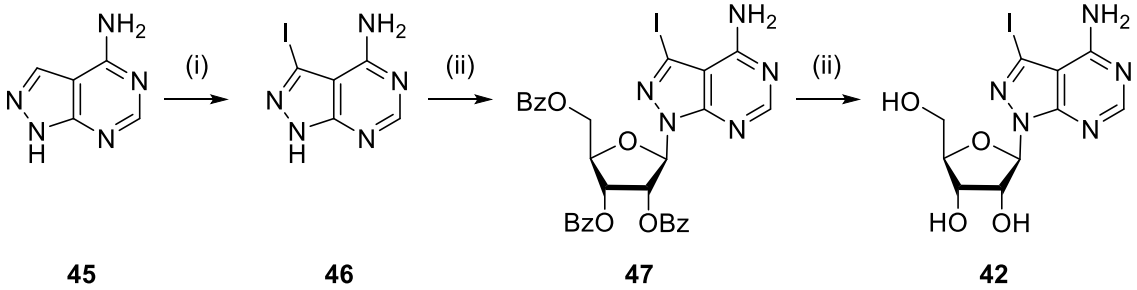
Figure 3.2 Structures of commercially available natural products and synthetic compounds. Compounds **14** and **27**-**34** were purchased from commerical vendors. Compounds **35**-**40** were synthesised by Dr Elliott Bayle. Screening was carried out by members of the molecular biophysics team at the SGC (Toronto).

3.1.1 Synthesis of target compounds

The pyrazolopyrimidine compound **41** and its 5-iodo analogue **42** were prepared following via short synthetic routes as outlined in the literature.[64] They were both isolated in moderate yields (Scheme 3.1 and Scheme 3.2). The remaining synthetically prepared compounds, **35**-**40** were synthesised by Dr Elliott Bayle, a summary of the synthetic procedures and yields are outlined in Appendix 1.



Scheme 3.1 Synthetic route to pyrazolo analogue **41**. *Reagents and conditions*: (i) 1-*O*-Acetyl-2,3,5-tri-*O*-benzoyl-β-D-ribofuranose, anhydrous $MeNO_2$, reflux, 30 min, *then* $BF_3.OEt_2$, reflux, 3 hrs (42%); (ii) NaOMe, MeOH, rt, 1 hr *then* 3 hrs reflux (26%).



Scheme 3.2 Synthetic route to pyrazolo analogue **42**. *Reagents and conditions*: (i) NIS, DMF, 80 °C, overnight (78%); (ii) 1-*O*-Acetyl-2,3,5-tri-*O*-benzoyl-β-D-ribofuranose, anhydrous $MeNO_2$, rt, 15 min, *then* $BF_3.OEt_2$, reflux, 3 hrs (44%); (iii) NaOMe, MeOH, rt, 1 hr *then* 3 hrs reflux (54%).

3.1.2 Evaluation of target compounds

DOT1L inhibition data for the target compounds was obtained by the SGC (Toronto) and were assessed using assay conditions as reported by Yu *et al*. with minor modifications.[55] The results are summarised in Table 3.1 and SAR are evaluated below.

(i) Exploration of the DOT1L hydrophobic cavity (C-5)

The SARs were initially focused on exploring substitution at the 5-position. Exploitation of this cavity has been exemplified through the nucleoside fragment 5-ITC (**14**) (published IC_{50} = 18.2 μM) and DOT1L inhibitor Br-SAH (**15**) (IC_{50} = 77 nM).[13] In this study, beneficial occupation of the hydrophobic pocket located at the 5-position on the nucleobase was exemplified through the loss of activity observed via comparison of 5-

ITC (**14**, 5-I) (IC_{50} = 7 µM) versus two natural product analogues adenosine (**27**) (26% @ 200 µM) and tubercidin (**28**, 5-H) (41% @ 200 µM).

Table 3.1 DOT1L inhibitory activity of target compounds

#	%inhibition @ 200 or 250 µM	IC_{50} (µM)
14		7 ± 2
27	26% ± 4% @ 200 µM	
28	41% ± 4% @ 200 µM	
29		44 ± 10
30		26 ± 7
31	<20% @ 200 µM	
32	<20% @ 200 µM	
33	<20% @ 200 µM	
34	<20% @ 200 µM	
35	52% ± 5% @ 200 µM	
36		41 ± 12
37	<20% @ 200 µM	
38	44% ± 5% @ 200 µM	
39	42% ± 1% @ 250 µM	
40		43 ± 8
41	<20% @ 200 µM	
42		22 ± 2

Inhibition data was obtained by SGC (Toronto). Compounds were evaluated by 11-point dose-response curves with test concentrations from 0.2–200 µM, except for compound **39** 0.5–500 µM. Assay conditions were used as reported by Yu *et al*[55] with minor modifications: DOT1L (0.2 nM), [^{3}H]-SAM (0.7 µM), chicken nucleosome (0.1 µM), buffer: 20mM Tris pH 8, 5mM DTT, 0.01% Triton X-100. Experiments were performed in triplicate with data presented as the mean ±SD.

Intriguingly, two natural products toyocamycin (**29**) and sangivamycin (**30**)**,** which both bear a polar substituent at the 5-position were found to have promising activity (IC_{50} = 44 µM and IC_{50} = 26 µM respectively). The polar cyclic substituent oxazole **40**, a structurally rigid surrogate for the 5-carboxamide substituent of sangivamycin (**30**)**,** was found to retain potency (IC_{50} = 43 µM). Furthermore the 5-ethyne analogue **36** showed promising activity (IC_{50} = 41 µM) and was equipotent to toyocamycin (**29**). These results suggest the hydrophobic cavity can tolerate both lipophilic and polar substituents with retention in potency with respect to 5-ITC (**14**).

Extension of the alkyne resulted in the inactive cyclopentaneacetylene **37** (<20% @ activity of the 200 µM). The bulky and rigid natural product triciribine (**31**) was also found to be inactive (<20% @ 200 M). However other larger substituents, such as a phenyl **38** and *p*-methoxybenzyl substitution **39** (42–44% @ 200 µM) were found to be equipotent to tubercidin (**28**). This indicated the pocket is not sufficiently flexible and smaller individual substituents are required at C-4 and C-5.

(ii) Exploration of analogous nitrogenous bases

Two scaffolds were evaluated as potential DOT1L inhibitors which exhibit an analogous nitrogenous base. First, the importance of N-1 in the adenine base was observed through the loss of activity of compound (**34**) (<20% inhibition @ 200 µM) versus adenosine (**27**) (26% @ 200 µM). Second, an analogous nitrogenous base, a pyrazolo[2,3-*d*]pyrimidine scaffold which explores substitution of the 6-CH for 6-N, was found to exhibit comparable potency to the pyrrolo[2,3-*d*]pyrimidine scaffold (**14** vs **42**). In fact, compound **42**, was found to be one of the most potent inhibitors from this screen.

(iii) Exploration of the ribose scaffold

Limited SAR of the ribofuranse scaffold indicated the 2'-OH is important for binding as shown through comparison of **32** (<20% @ 200 µM) versus 5-ITC **14** (IC_{50} = 7 µM). A commercially available analogue with a 3'-OH deletion was not available. Compound **33**, which was selected to investigate potential areas of substitution around the ribofuranose ring was found to be inactive (<20% @ 200 µM).

In summary, it was found through deletion of 2'-OH **32** and N-1 **34** that both groups significantly contribute to binding and so are not viable options for deletion; the binding contribution from 3'-OH remains untested. An analogous nucleobase bearing an additional nitrogen, pyrazolo[2,3-*d*]pyrimidine scaffold **42**, was found to be equipotent to the pyrrolo[2,3-*d]*pyrimidine scaffold **14**, producing one of the most active compounds from this screen. Evaluation of substitutions on the nucleobase at the 5-position into the hydrophobic cavity found the pocket is inflexible in both width and length indicating a small substituent is required. Of the groups screened, three were found to exhibit promising activities: toyocamycin (**29**), sangivamycin (**30**) and 5-ethylene **36**. These substitutions demonstrate the hydrophobic cavity is capable of accommodating both lipophilic and polar groups.

The replacement of a lipophilic iodide for polar substituents in the form of a nitrile or primary carboxamide was unexpected as they are not traditional bioisosteric transformations. This is explored further in the next section.

3.2 Non-traditional isosteres

A bioisosteric transformation allows alteration of the chemical properties of a compound (polarity, acidity etc.) through alternative structural motifs while retaining broadly similar biological properties.[65] As previously noted, the replacement of a lipophilic iodide for polar substituents in the form of a nitrile or primary carboxamide was unexpected as they are not traditional bioisosteric transformations. Although aromatic chloro-nitrile substitutions have been shown to consistently improve the medicinal chemical properties of drug candidates (reviewed by Jones *et al*.)[66], there are fewer reports for other halo-nitrile transformations, for example, Iodo-nitrile, bromo-nitrile or fluoro-nitrile. Examples of the non-traditional iodo-nitrile substitution is limited to lead optimisation of imidazole[67] and pyrimidone[68] scaffolds for treating heart failure. Bromo-nitrile substitutions have also been cited in an investigation into cardiac activity, as part of a series of analogues of bipyridines.[69] This transformation has also been cited in identification of a non-steroidal inhibitor for treating oestrogen-dependent disease.[70] Fluoro-nitrile substitutions have been cited as part of an investigation into the identification of azabicyclic inhibitors for senile dementia.[71] Aromatic halo-carboxamide bioisosteric substitutions have not been explored.

Bioisosteric substitutions are used to manipulate the drug-like properties of hit or lead compounds. The substitution of a lipophilic halogen for a more polar nitrile is favourable as it reduces the overall lipophilicity of the compound. Lipophilicity is a key physicochemical property that plays an important role in drug discovery and design as it impacts upon ADMET (Absorption, Distribution, Metabolism, Excretion and Toxicity) properties and the overall suitability of drug candidates.[72] For example, highly lipophilic compounds tend to be associated with promiscuous binding (which can lead to off-target effects) and metabolic instability (which can result in high clearance due to metabolism by Cyp450 enzymes). Potency is a significant factor when evaluating compounds in a drug discovery programme. To normalise the observed potency with changes in lipophilicity, Leeson *et al*.[73] developed the ligand lipophilic efficiency (LipE) parameter, which directly evaluates potency (IC_{50}) and lipophilicity (P);

$$LipE = pIC_{50} - LogP$$

Thus, for a 1 nM inhibitor with a LogP of 3, the LipE is 6, while for a 10 nM inhibitor with a LogP of 3 the LipE is 5. The optimum range for LipE is considered to be between 5 and 7. LipE is particularly useful when analysing the effect on potency or lipophilicity for single point transformations. For example, in this study the IC_{50} remains equipotent when the aromatic iodide in 5-ITC (**14**) is replaced with a nitrile to form toyocamycin (**29**) or a carboxamide to form sangivamycin (**30**), using the substituents Rekker values (fragmental lipophilic contribution) the change in lipophilicity and thus ΔLipE can be evaluated.[74] For example, substituting an iodo (+1.12) for a nitrile (-0.57) or primary carboxamide group (-1.49) delivers a decrease in lipophilicity and thus an increase in LipE by 1.6 units and 2.6 units respectively.

These substitutions (iodo-nitrile or iodo-carboxamide) could be applied to the published inhibitors to move them towards a more drug-like template. As discussed, current published inhibitors are metabolically labile. This is due to their structural similarities to the adenosine core (which results in metabolism by plasma enzymes) and is also due to high lipophilicity (which is associated with high clearance due to metabolism by Cyp450 enzymes). These substitutions would not only produce a more drug-like template through reduction in lipophilicity but could also yield a more metabolically stable template as they also alter the adenosine core, the latter potentially disrupting recognition by Cyp450 enzymes. Importantly, the substitutions also retain affinity to DOT1L.

To further assess the potential impact of these substitutions on the structure of the published inhibitors, an electrostatic potential map was generated using Avogadro.[75] The electrostatic maps allow evaluation of the differences in polarisation on the aromatic core between an aromatic iodide versus an aromatic nitrile or an aromatic carboxamide (Figure 3.3). The electrostatic maps indicate the main difference in electronics is at the site of the substituent. This infers neither the iodo-nitrile or iodo-carboxamide substitutions should have an impact nor negative effect on any existing binding interactions which occur between residues on the pyrrolo[2,3-*d*]pyrimidine ring and the DOT1L hydrophobic pocket. This observation is in-line with analysis carried out by Jones *et al*.[66] when analysing an aromatic chloride to nitrile transformation, thus extending the potential use of nitriles as a bioisostere for multiple halogens, i.e. a halo-bioisostere.

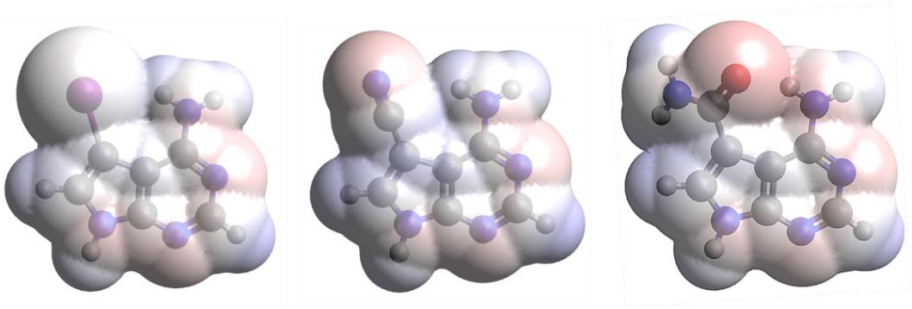


Figure 3.3 Electrostatic potential map of (left) 5-I pyrrolo[2,3-*d*]pyrimidine (**14**) versus (middle) 5-CN pyrrolo[2,3-*d*]pyrimidine (**29**) and (right) 5-$CONH_2$ pyrrolo[2,3-*d*]pyrimidine (**30**). The site of the substituent is the main difference in electronics. Blue = highest, Red = lowest electrostatic potential energy. Images generated using Avogadro.[75]

Furthermore, due to the hydrophobic effect, halogens prefer more buried pockets than nitriles.[66] As the hydrophobic cavity in DOT1L is near the surface, it is likely to be a more favourable environment for a nitrile substituent.

3.3 Summary and conclusions

In conclusion, it has been shown through the literature that occupation of the DOT1L hydrophobic cavity provides both selectivity and improved potency for DOT1L over PKMTs through disruption of conserved H-bonds.[55] The SARs and subsequent analysis of hits (specifically in relation to lipophilicity) undertaken in this study, identified two polar 5-substitutions into the DOT1L hydrophobic cavity (nitrile **29** and carboxamide **30**) as promising hits to be taken forward. Both substitutions have equipotent affinity for DOT1L in comparison to 5-ITC (**14**) and, importantly, could also introduce potential improvement to the pharmacokinetic properties of the drug candidates through reduction in lipophilicity and formation of a more metabolically stable template. Thus, improvement to the pharmacokinetic properties of published inhibitors could be gained through a single point transformation into the hydrophobic cavity.

Furthermore, the results from this study indicate the iodo-nitrile (**14** vs **29**) or iodo-carboxamide (**14** vs **30**) substitutions could be valuable non-traditional or bioisosteric transformations. Bioisosteric substitutions are an effective tool to guide drug discovery. The results from this study could extend the application of bioisosteric substitutions in drug discovery.

To establish proof of concept, the iodo-nitrile substitution was selected to initially translate into a biochemical tool. The nanomolar DOT1L inhibitor Br-SAH (**15**) published by the SGC was identified as the template from which to translate the nitrile

moiety (Figure 3.4).[55] The synthesis and evaluation of the resulting analogue, CN-SAH (**48**) is discussed in the following chapter.[76]

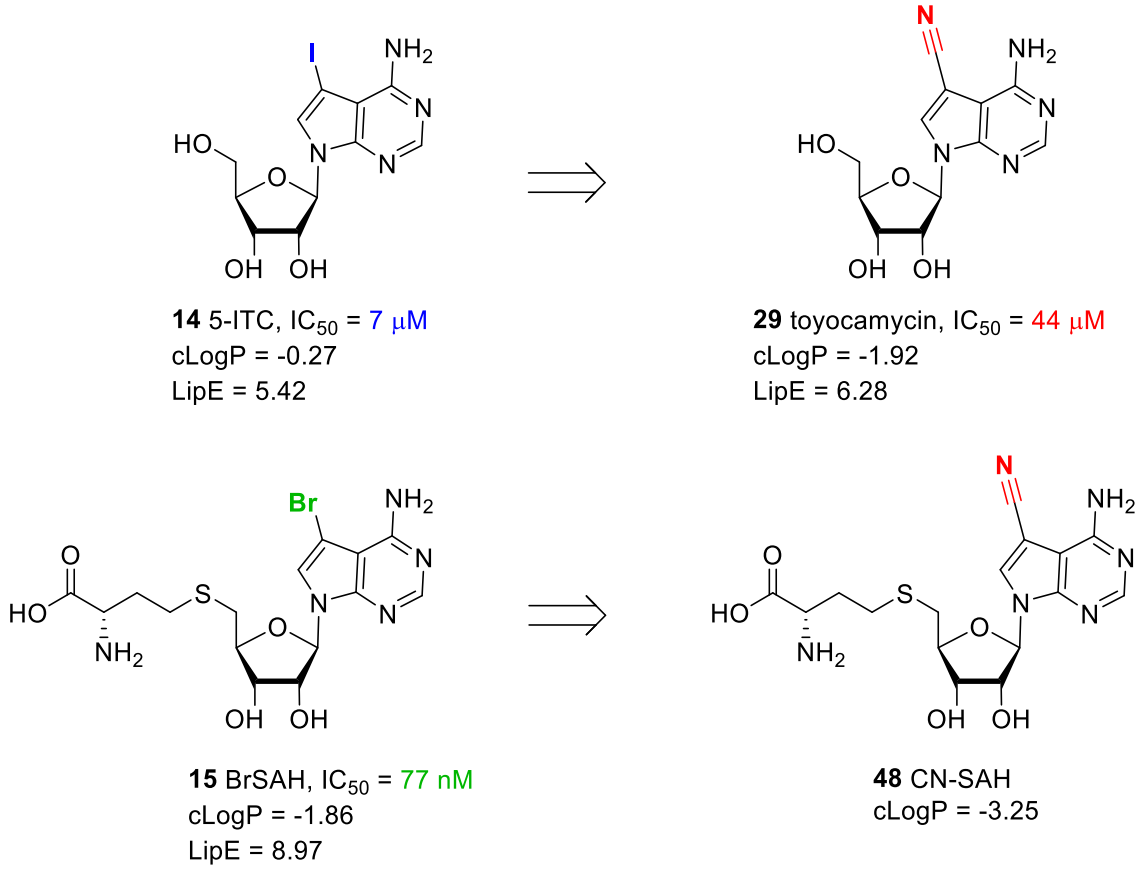
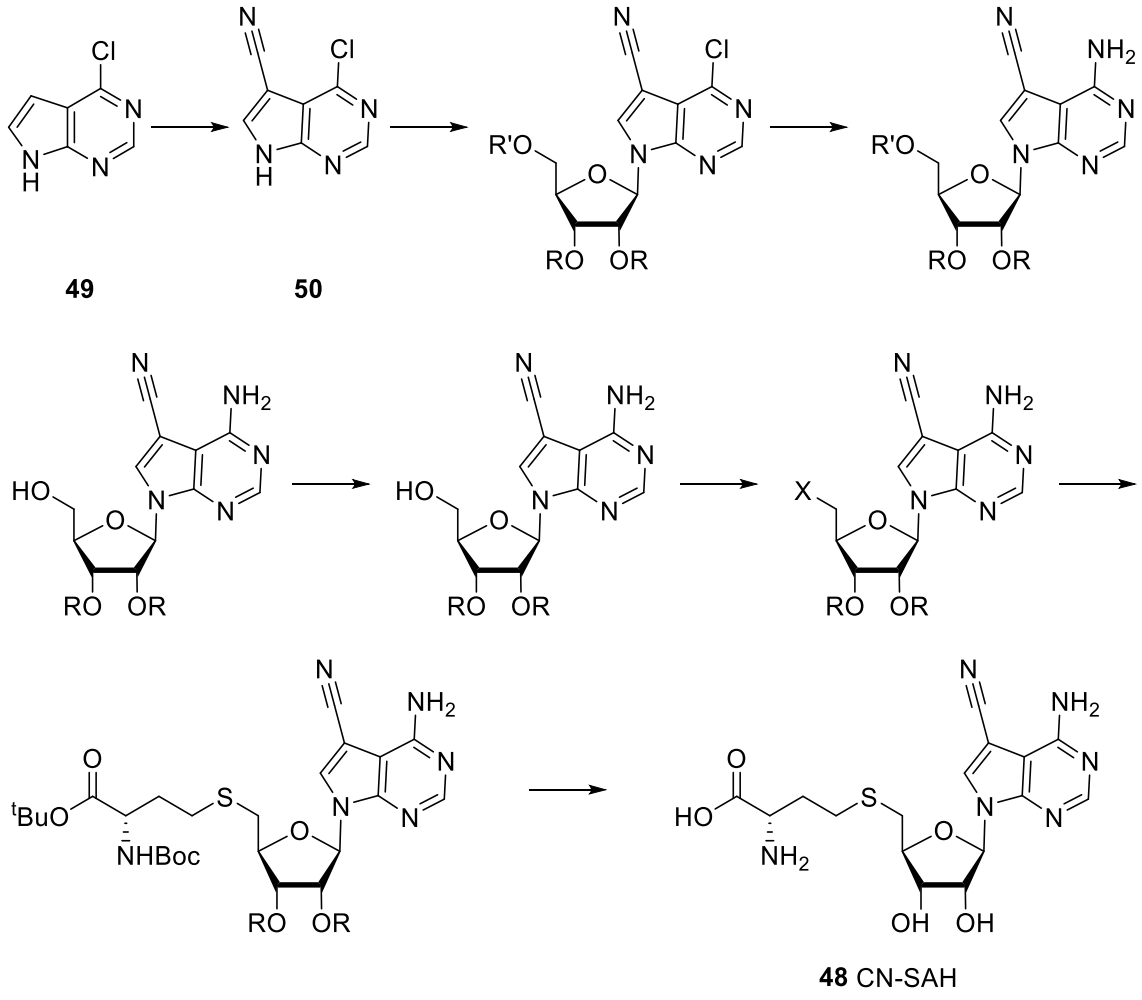


Figure 3.4 Replacement of the iodide in 5-ITC (**12**) for a nitrile, a non-traditional bioisosteric analogue, retained the potency against DOT1L and improved the LipE by 0.9 units. This halo-nitrile transformation could be translated into a biochemical tool.

CHAPTER 4
A biochemical tool: CN-SAH

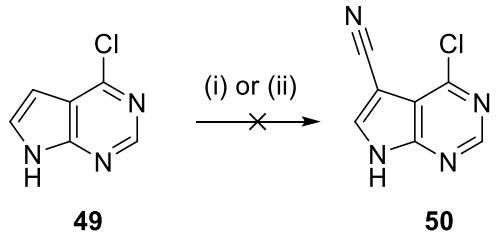
The synthetic route to obtain CN-SAH (**48**) was developed using previously published work on similar compounds and the synthetic route to Br-SAH (**15**)[55] (Scheme 4.1). There were two key differences from the synthetic route to Br-SAH (**15**). First, the 5-CN substituent was to be introduced at the beginning of the synthesis and carried through instead of the 5-Br. Second, poor yields were reported for the coupling of the nucleoside to the amino acid (7–13%),[55] it was hypothesised this could be improved through protection of both the sugar and amino acid.



Scheme 4.1 Overview of general synthetic route to CN-SAH (**48**) starting from the commercially available nucleobase **49**. R and R' = Suitable protecting groups. X = suitable leaving group.

4.1 Installation of 5-CN

Conditions to install the 5-CN were investigated first. An attractive approach was to install the nitrile via a one-step reaction starting from the commercially available nucleobase **49**. A publication by Ushijima *et al*. suggested direct cyanation could be achieved using a Vilsmeier-haack reaction, followed by treatment with molecular iodine in aqueous ammonia.[77] These conditions were applied to nucleobase **49**. However, the desired nitrile **50** was not obtained (Scheme 4.2, step (i)).

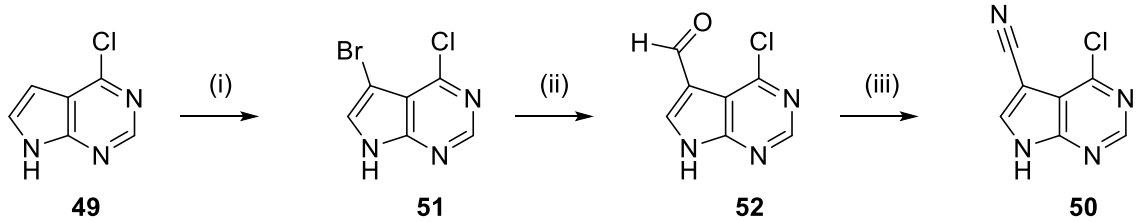


49 **50**

Scheme 4.2 Conditions for direct insertion of the nitrile to the commercially available nucleobase **49**. *Reagents and conditions*: (i) $POCl_3$, anhydrous DMF (8 eq), 0 °C, 3 hrs *then aq.* NH_3, I_2, rt, 3 hrs, no reaction; (ii) chlorosulphonyl isocyanate, DMF, - 20 °C to -10 °C, 1.5 hrs, no reaction observed.

In parallel, conditions published by Zhang *et al*. were attempted which used an alternative reagent chlorosulphonyl isocyanate to directly install the nitrile on nucleobase **49** (Scheme 4.2, step (ii)).[77, 78] However, nitrile **50** was not isolated. In both cases the literature used an indole template. Thus, it was hypothesised the reactions were not proceeding with the pyrrolopyrimidine template due to lack of activity of the pyrrole ring due to conjugation with the pyrimidine ring.

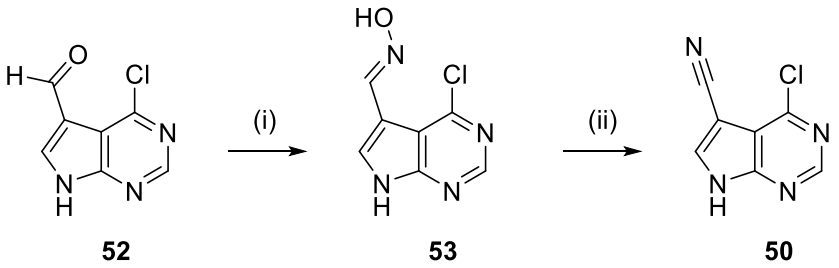
An alternative strategy was therefore employed using a 4-step synthetic route (Scheme 4.3). First, this required bromination of nucleobase **49** using *N*-bromosuccinimide resulting in **51**, this was followed by instalment of an aldehyde resulting in isolation of **52**; both reactions proceeded with moderate yields.[79] Reaction of aldehyde **52** with triflic acid and sodium azide to introduce the nitrile **50,** resulted in multiple products and did not go to completion, thus resulting in a low yield (16%).[80]



49 **51** **52** **50**

Scheme 4.3 Successful synthetic route to compound **50** (5-CN). *Reagents and conditions*: (i) NBS, DCM, rt, overnight (64%); (ii) n-BuLi, THF, -78 °C, 2 hrs *then* DMF, rt, 3 hrs (65%); (iii) Triflic acid, NaN_3, MeCN, rt, overnight (16%).

An alternative option was explored which introduced an oxime **53** from the aldehyde **52** using hydroxylamine hydrochloride followed by oxidation to the nitrile **50** using thionyl chloride (Scheme 4.4).[81] Both reactions yielded the product as a solid in good yields.

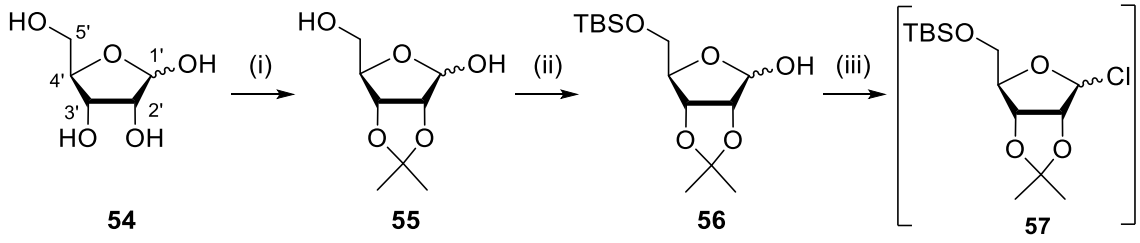


Scheme 4.4 Alternative conditions to oxidise 5-CHO (**52**) to 5-CN (**50**). *Reagents and conditions*: (i) 2M *aq.* NaOH, hydroxylamine hydrochloride, EtOH, 50 °C, overnight (84%); (ii) $SOCl_2$, DCM, rt, overnight (83%).

4.2 Glycosylation

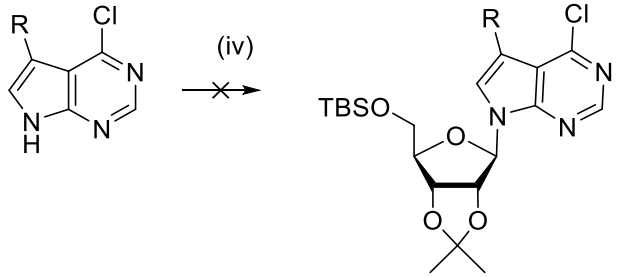
Having successfully installed the nitrile, attention was turned to glycosylation of **50**. The formation of a glycosidic linkage between a nucleophilic nucleobase and appropriately protected electrophilic sugar, has been extensively reported in the literature. The 'Salt method'[82] and the 'Vorbrüggen method' (Silyl-Hilbert-Johnson)[83] are the two most common routes. The methods differ by the mechanisms used to selectively achieve the desired anomeric configuration. The Salt method employs the anomeric effect to direct the formation of the β-anomer whereas the Vorbrüggen method utlises neighbouring group participation.

Formation of the glycosidic linkage was initially attempted using the Salt method. Orthogonal hydroxyl protecting groups for 2’ and 3’ versus 5’ were necessary to allow selective deprotection after glycosylation. Thus, ribose **54** was sequentially protected with acetonide and *tert*-butylsilyl protecting groups (Scheme 4.5).



Scheme 4.5 Synthetic route to activated ribofuranose **57**. *Reagents and conditions*: (i) H_2SO_4, Acetone, 1.5 hrs (73%); (ii) TBSCl, imidazole, DMF, 12 hrs (70%); (iii) tris(dimethylamino)phosphine, CCl_4, THF, -40 – 0 °C, 45 mins (α:β ratio 3:1).

The protected ribose **56**, was successfully chlorinated providing a solution of **57** in a 3:1 ratio of the α:β anomers respectively, observed through NMR analysis of the crude solution. Attempts to couple this solution *in-situ* to the nucleobase **50** to yield nucleoside **58** were unsuccessful (Scheme 4.6). A control reaction was carried out using nucleobase **49**, known to couple under these conditions, but only trace amounts of nucleoside **59** were isolated.[59]



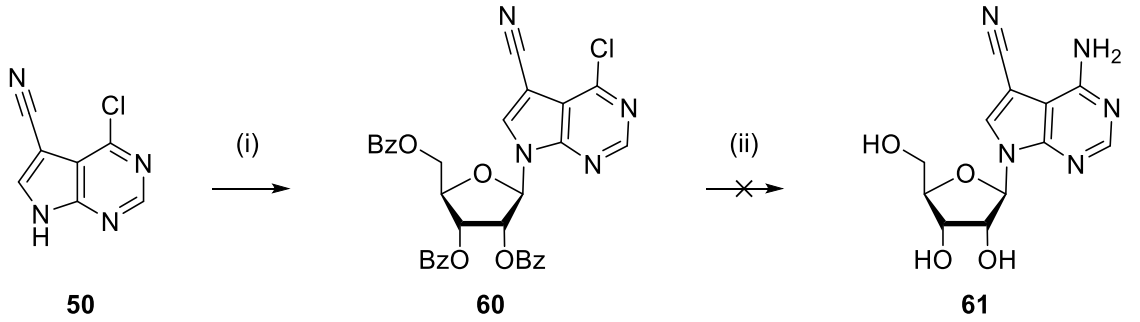
50: R = CN
49: R = H

58: R = CN
59: R = H

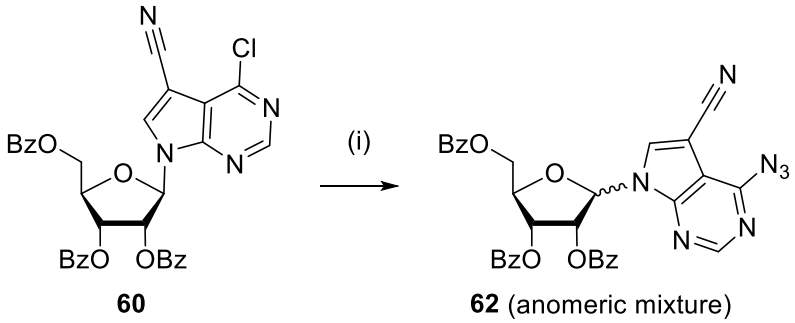
Scheme 4.6 Attempted synthetic route to isolate compound **58** or **59** using the salt method. *Reagents and conditions*: (iv) NaH, DMF, 0 °C *then* 24, 0 °C to rt, no reaction observed.

At this point it was unclear why the reaction was not proceeding. Several single point changes to the reaction conditions were investigated including using 15-crown-ether to increase the reactivity of NaH and switching to the more reactive KH, but the desired product was not isolated.

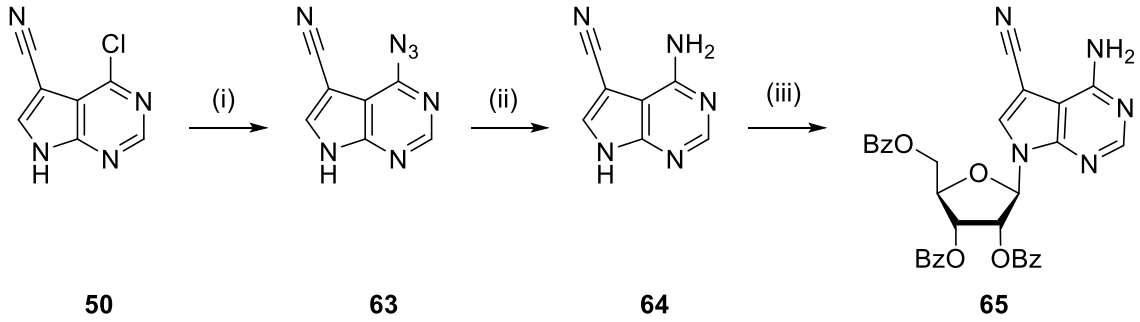
Switching to an alternative glycosylation method, Vorbrüggen reaction (Scheme 4.7, step (i)), resulted in the successful isolation of compound **60**. An attempt at global deprotection and concomitant installation of 4-NH_2 to obtain **61**, using the same conditions as outlined in the synthesis of Br-SAH (**15**), was unsuccessful (Scheme 4.7, step (ii)). These harsh conditions, which are required for direct amination, led to the destruction of the nitrile moiety. Milder conditions for indirect amination via the azide **62**, resulted in the epimerisation of the anomeric centre and the formation of an inseparable mixture of anomers in a 3:2 ratio (Scheme 4.8). Introduction of the 4-NH_2 prior to glycosylation (**63** and **64**) had a negative effect on the efficiency of the coupling due to an increase in side reactions and poor separation, resulting in isolation of compound **65** in a 7% yield (Scheme 4.9).



Scheme 4.7 Attempted synthetic route to compound **61**. *Reagents and conditions*: (i) *N,O-bis*(trimethylsilyl)acetamide, MeCN, rt, 10 mins *then* 1-*O*-Acetyl-2,3,5-tri-*O*-benzoyl-β-D-ribofuranose, TMSOTf, 3 hrs (45%); (ii) 7 N NH_3/MeOH, 120 °C, 2 hrs, no product isolated.



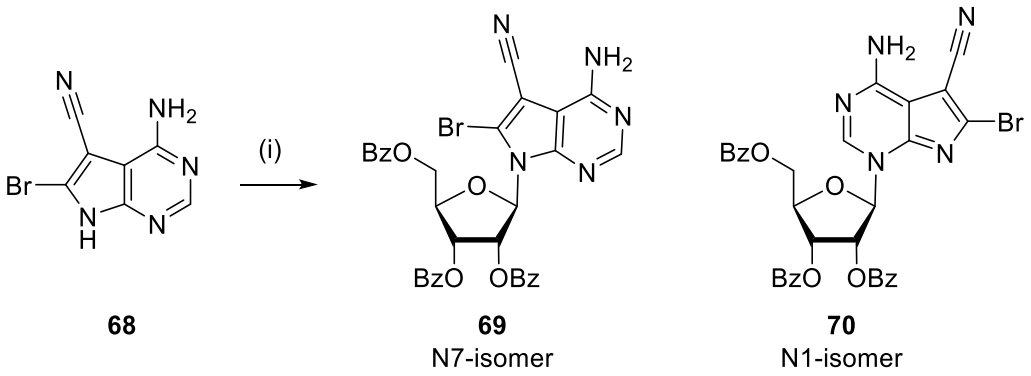
Scheme 4.8 Approach to install 4-N_3 resulted in epimerisation of the anomeric centre. *Reagents and conditions*: (i) NaN_3, DMF, 60 °C, 2 hrs (80%) (3:2 ratio of β:α anomers).



Scheme 4.9 Synthetic route to compound **65**. *Reagents and conditions*: (i) NaN_3, DMF, 60 °C, 2 hrs (29%); (ii) 10 mol% Pd/C, H_2, EtOH, rt, 12 hrs (33%); (iii) *N,O-bis*(trimethylsilyl)acetamide, MeCN, rt, 10 mins *then* 1-*O*-Acetyl-2,3,5-tri-*O*-benzoyl-β-D-ribofuranose, TMSOTf, 3 hrs (7%).

The difficulty in introducing the 4-NH_2, while maintaining an adequate yield for the glycosylation step and avoiding destruction of the nitrile moiety, was overcome using an alternative synthetic strategy. The glycosylation of the nucleobase is sensitive to slight electronic changes. It has been shown that incorporation of a 6-Br on the pyrrole ring **68** allowed successful glycosylation in a reasonable yield (~50%).[84, 85] Following the published glycosylation conditions as before – using the Lewis acid trimethylsilyl trifluoromethanesulphonate (TMSOTf) and *N,O-bis*(trimethylsilyl)acetamide – resulted

in the formation of the desired N7 regioisomer **69** and the N1 regioisomer **70** in a ratio of 1.5:1 respectively (Scheme 4.10). This ratio differed to literature reports which indicated a 2:1 formation of the N7:N1 isomers.[86] Furthermore, separation of the regioisomers proved problematic. However, the same paper did not report any difficulties in resolution of the regioisomers.[86]



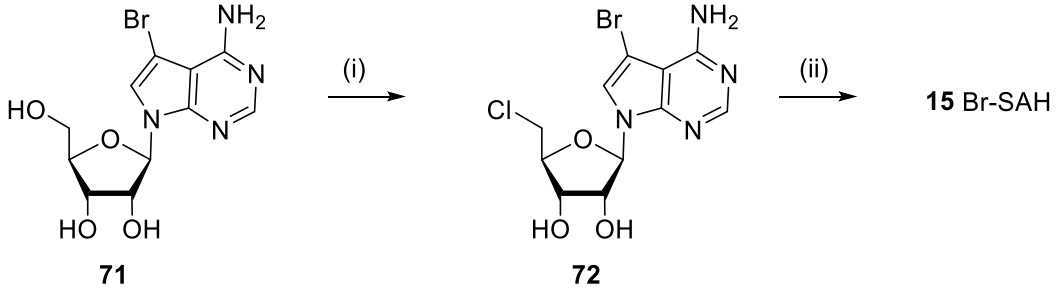
Scheme 4.10 Attempted synthetic route to compound **69.** *Reagents and conditions*: (i) *N,O-bis*(trimethylsilyl)acetamide, MeCN, rt, 10 mins *then* 1-*O*-Acetyl-2,3,5-tri-*O*-benzoyl-β-D-ribofuranose, TMSOTf, 3 hrs (**69**: 55% **70**: 35%).

At this point, Kim *et al*. published a thorough investigation into the regioselectivity of this reaction.[87] They found the regioselectivity of the reaction is heavily influenced by the reaction conditions. Through variation of the following four variables: the silylating agent (HDMS vs BSA); the protecting groups used for the sugar (OAc vs OBz); the equivalents of Lewis acid (TMSOTf) and the solvent (aromatic versus non aromatic), they identified conditions which gave greater control over the regioselectivity of the reaction, thus allowing selective isolation of the N1 and N7 regioisomers in good yields (>90% and 70% respectively). Adopting the conditions for selective isolation of the N7 regioisomer, which utilised an alternative silylating agent, hexamethyldisilazane (HMDS), the desired N7 isomer **69** was selectively formed and isolated in a good yield (49%) (Scheme 4.13, step (iii)). Most importantly, the N1 regioisomer was not formed thus greatly simplifying the purification step. This is in contrast to the lengthy purification required to remove the N1 regioisomer, when using the silylating agent *N,*O-*bis*(trimethylsilyl)acetamide.

4.3 Final steps

Having successfully selectively synthesised the N7 regioisomer **69**, attention was turned to reduction of the 6-Br and deprotection of the ribose to isolate the intermediate and natural product, toyocamycin (**29**). Both reactions proceeded in adequate yields (Scheme 4.13, (iv) and (v)).

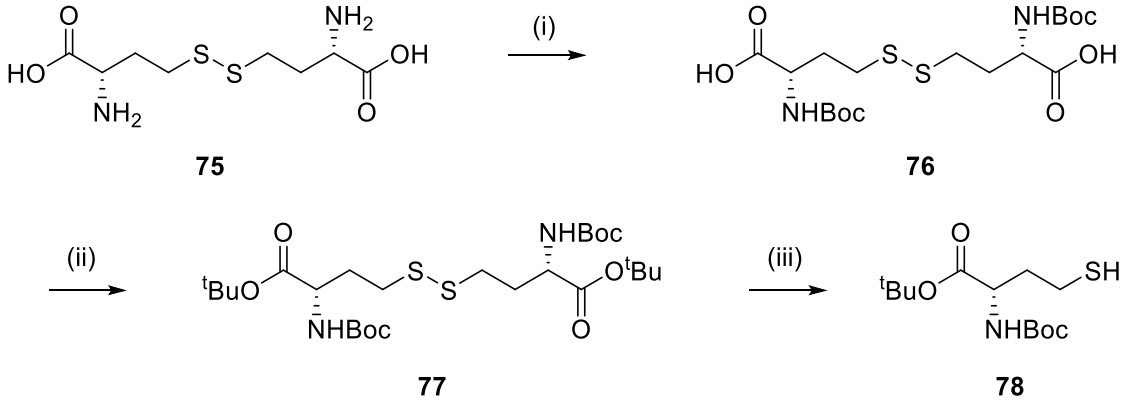
With a reliable route for the gram scale synthesis of toyocamycin (**29**) in hand, the final three steps of the synthesis were carried out following the synthetic procedure for synthesis of Br-SAH (**15**) (Scheme 4.11),[55] with the incorporation of two modifications (vide infra).



71 **72**

Scheme 4.11 Published conditions to synthesise Br-SAH (**15**). *Reagents and conditions*: (i) $SOCl_2$, HMPT, 0 °C, 18 hrs (61%); (ii) KI, K_2CO_3, L-homocysteine, 1 M aq. NaOH, 80 °C, 2 hrs (9%).

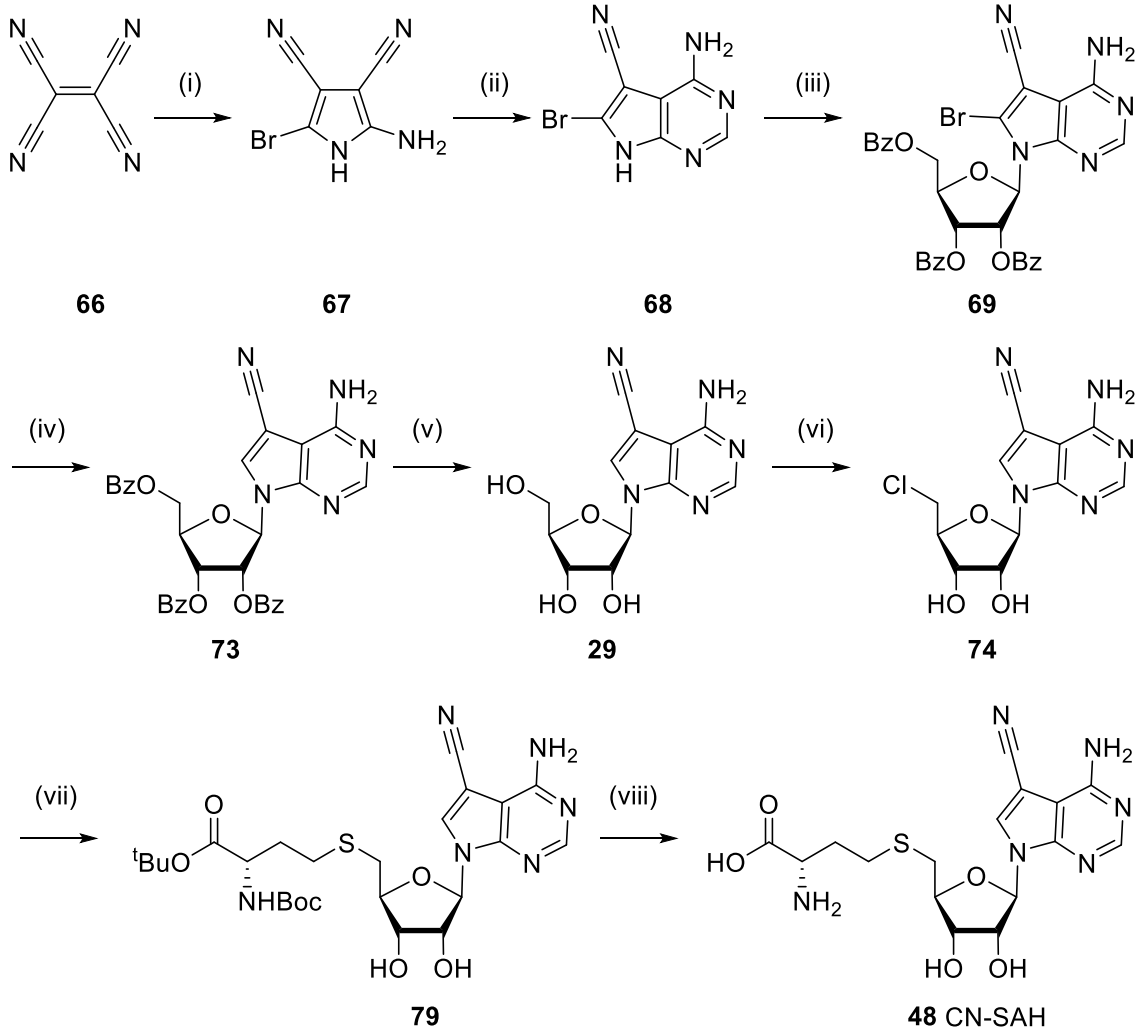
First, to activate the 2'-hydroxyl in the ribose unit, the carcinogenic solvent hexamethylphosphorous triamide (HMPT) (Scheme 4.11, step (i)) was exchanged for the more environmentally friendly reagent *N*,*N*′-dimethylpropylene urea (DMPU) (Scheme 4.13, step (vi)). This substitution had a slight negative effect on the yield, with 2'-chlororibose **74** isolated in a slightly lower though still moderate yield in comparison to the analogous step in the synthesis of Br-SAH (**15**) (50% vs. 61% respectively). Second, the yield of the S_N2 coupling step between homocysteine and **72** in the published route to Br-SAH (**15**) is poor (9%) (Scheme 4.11, step (ii)). It was hypothesised protection of the amine and acid of homocysteine could improve the yield. Thus, the acid and amine on the homocysteine chain **75** were sequentially protected using *tert*-butyl and Boc respectively to yield **78** (Scheme 4.12).



75 **76** **77** **78**

Scheme 4.12 Synthetic route to compound **78**. *Reagents and conditions*: (i) Boc_2O, 10% Na_2CO_3, dioxane, 0 °C to rt, 12 hrs (99%); (ii) *tert*-Butyl 2,2,2-trichloroacetimidate**,** DCM, rt, 12 hrs (78%); (iii) Bu_3P, H_2O, DMF, rt, 12 hrs (62%).

Coupling of the protected homocysteine **78** and 5'-chlororibose **74** proceeded with an excellent yield which was vastly improved in comparison to the analogous step in the synthesis of Br-SAH (**15**) (75% vs 9% respectively) (Scheme 4.13, step (vii)). Finally, deprotection using HCl yielded the biochemical tool CN-SAH (**48**) in a moderate yield (Scheme 4.10, step viii).



Scheme 4.13 Synthetic route to CN-SAH (**48**). *Reagents and conditions*: (i) 33% HBr in AcOH, acetone, EtOAc, 0 °C - rt, 3 hrs (60%); (ii) formamidine acetate, 2-ethoxyethanol, reflux, 24 hrs (36%); (iii) HMDS, MeCN, reflux, 18 hrs, *then* 1-*O*-Acetyl-2,3,5-tri-*O*-benzoyl-β-D-ribofuranose, TMSOTf, reflux, 3 hrs (49%); (iv) ammonium formate, EtOH, 5 mol% Pd/C, reflux, 3 hrs (65%); (v) 7 M NH_3 in MeOH, 0 °C - rt, 12 hrs (60%); (vi) $SOCl_2$, *N*,*N*′-dimethylpropylene urea (DMPU), 0 °C, 5 hrs (50%); (vii) KI, K_2CO_3, Boc-L-homocysteinate (**78**), DMF, 80 °C, 5 hrs (75%); (viii) 4 M HCl in dioxane, 0 °C, 7 hrs (67%).

4.4 Inhibition of DOT1L with CN-SAH

Pleasingly CN-SAH (**48**) was found to be a potent inhibitor of DOT1L with an IC_{50} value of 26 nM (Figure 4.1). This is a 3-fold increase in inhibition compared to Br-SAH (**15**) (IC_{50} 77 nM).

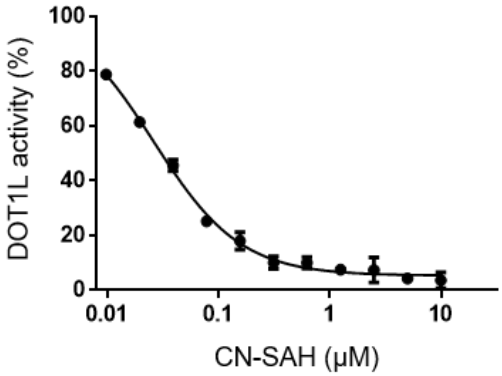


Figure 4.1 IC_{50} of CNSAH (**48**) was determined from an 11-point dose response IC_{50} curve (0.01–10 μM). Experiments were performed in triplicates with the mean presented as ± SD. Screening was carried out by members of the molecular biophysics team at the SGC (Toronto).

4.4.1 Crystal structure

A crystal structure was obtained of DOT1L in complex with CN-SAH (**48**) (PDB code: 5JUW) which verified CN-SAH (**48**) has the same binding mode as its analogue Br-SAH (**15**) (PDB code: 3SX0) and co-factor SAM (PDB code: 1NW3) (Figure 4.2). The 5-nitrile of CN-SAH (**48**) occupies the hydrophobic cavity, mirroring the 5-Br of Br-SAH (**15**). The activation loop of DOT1L is folded on the cofactor mimetic in a conformationally active state.

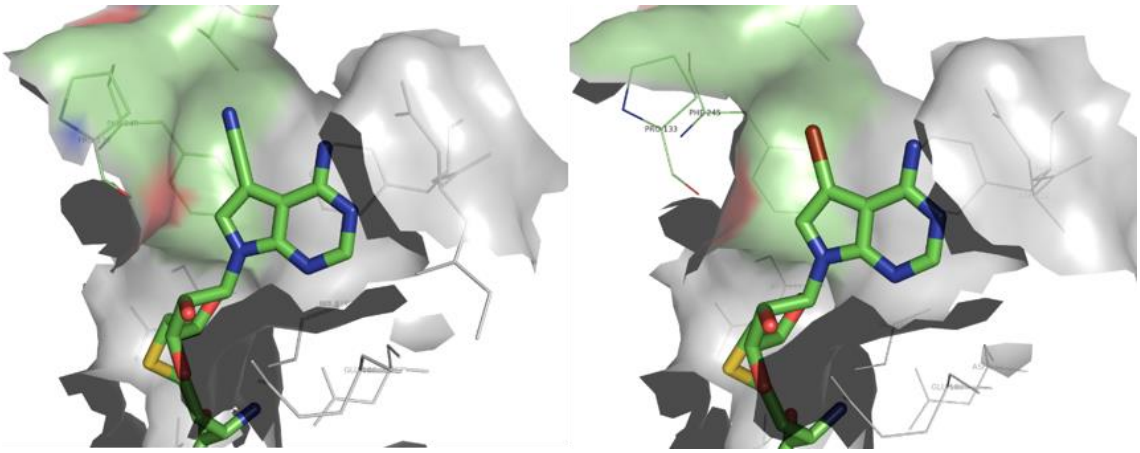


Figure 4.2 Co-crystal structures of (Left) CN-SAH (**48**) (PDB code: 5JUW, 2.4 Å) and (Right) Br-SAH (**15**) (PDB code: 3SX0, 2.3 Å). Obtained by the SGC (Toronto). Images generated using Pymol.

4.4.2 Biological profile

CN-SAH (**48**) was screened for selectivity against a representative panel of PKMTs, PRMTs and DNA methyltransferases (DNMTs) (Table 4.1). The screen was carried out by Eurofins Cerep (France) and confirmed CN-SAH (**48**) to be a potent inhibitor of DOT1L (IC_{50} = 13 nM). The methyltransferases (MTase) were chosen in line with previous selectivity studies of Br-SAH (**15**).[55]

Table 4.1 DOT1L inhibitory activity and MTase selectivity profile of CN-SAH (**48**) versus Br-SAH (**15**) and SAH.

Mtase	CN-SAH (IC_{50} µM)*	Br-SAH (IC_{50} µM)**	SAH (IC_{50} µM)
DOT1L	0.013	0.077	0.22
DNMT1	0.15	1.9	0.20
PRMT3	0.87	2.3	2.0
PRMT5	2.4	2.3	0.20
G9a	99	No inhibition	4.1
SETD7	No inhibition	No inhibition	47
SUV39H2	No inhibition	No inhibition	22
EZH2	No inhibition	No inhibition	65

*Inhibition data for CN-SAH (**48**) was obtained by Eurofins Cerep (France), SAH was used as a control throughout and data are presented for comparison.[76] Compounds were evaluated by 8-point dose-response curves with test concentrations from 0.03–100 µM, except against DOT1L which required a lower range, 0.1–1000 nM. Experiments were performed in duplicate with data presented as the mean. **Inhibition data for Br-SAH (**15**) was obtained by SGC (Toronto).[55]

As anticipated, CN-SAH (**48**) did not inhibit the selected PKMTs – SETD7, SUV39H2, EZH2 – at concentrations up to 100 µM, although weak activity (99 µM) was observed for the PKMT G9a. CN-SAH (**48**) exhibits inhibitory activity against PRMT3 (IC_{50} = 0.87 µM) and PRMT5 (IC_{50} = 2.4 µM) although with acceptable selectivity (66-fold and 185-fold respectively) this is modestly improved in comparison to Br-SAH (**15**) (both 30-fold). This is due to increased affinity for DOT1L rather than weaker activity for the PKMTs. However, CN-SAH (**48**) is a weak DNA methyltransferase 1 (DNMT1) inhibitor with only a modest selectivity (12-fold) compared to Br-SAH (**15**) (25-fold). The selectivity profile of CN-SAH (**48**) is consistent with Br-SAH (**15**) with some improvement in selectivity for PRMTs, although future development of DNMT selectivity is required.

In summary, the single point halo-nitrile transformation, **15** to **48**, as part of the nucleoside, has been shown to retain potency and selectivity for DOT1L through occupation of the hydrophobic DOT1L pocket, thereby supporting the proof of concept. It was hypothesised translation of this nitrile moiety into the sub-nanomolar DOT1L inhibitor SGC-946 (**16**), an epizyme-SGC hybrid, would allow further assessment of the potential physicochemical benefit through the resulting putative cell penetrant analogue CN-SGC946 (**80**) (Figure 4.3). However, synthetic efforts to synthesise CN-SGC946 (**80**) were made redundant due to the publication of an equivalent analogue bearing a 5-CN substituent, **81**. The synthetic approaches to obtain CN-SGC946 (**80**) and the structure and properties of the equivalent analogue **81** are discussed in the next section.

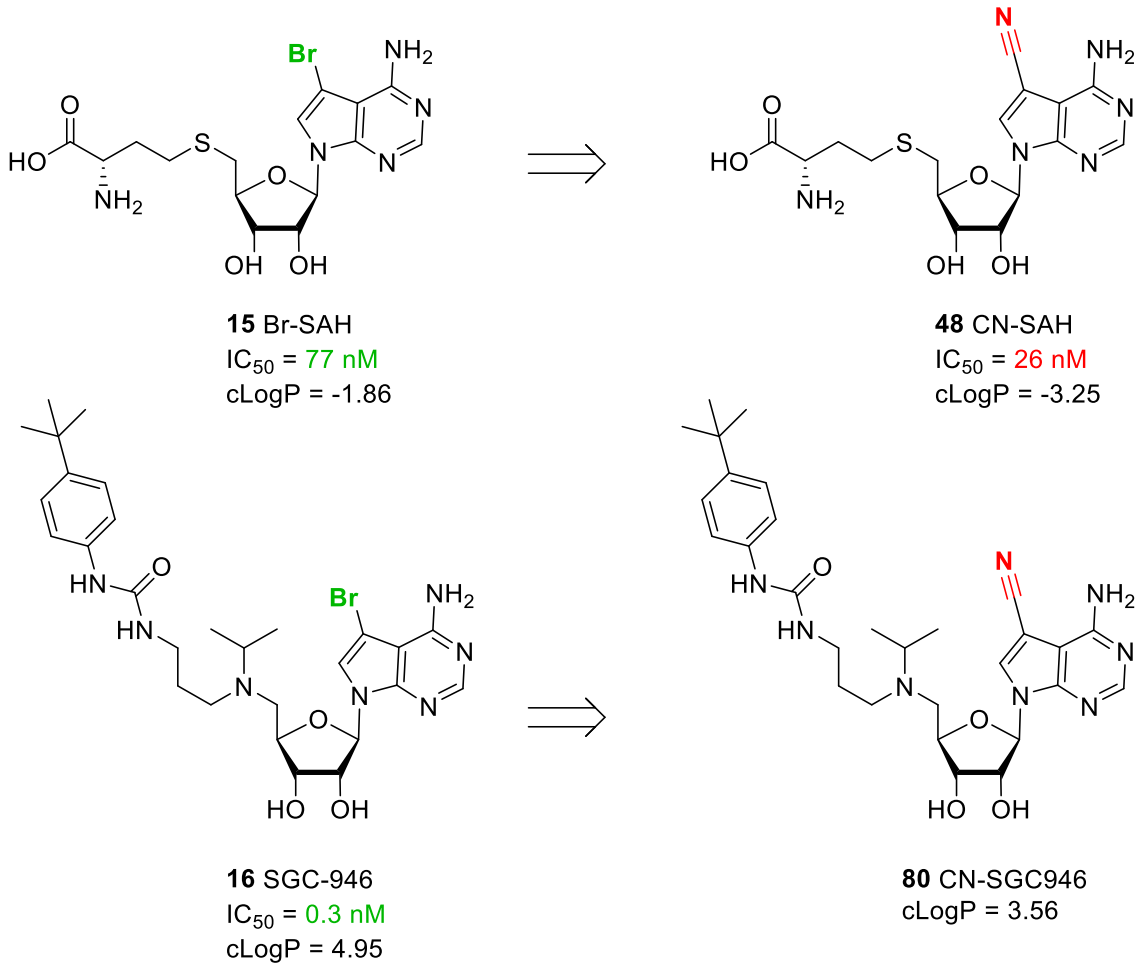
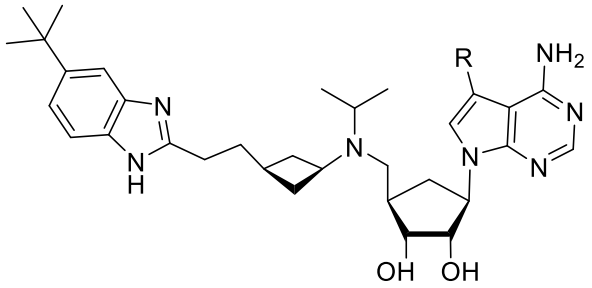


Figure 4.3 It was hypothesised this halo-nitrile transformation could be translated into a putative cell penetrant analogue, CN-SGC946 (**80**).

4.5 CN-SGC946: a putative cell penetrant analogue

The nitrile moiety was successfully transformed into the biochemical tool CN-SAH (**48**), which was shown to be a potent and selective co-factor competitive DOT1L inhibitor. Attention was turned to translating this moiety into a putative cell-penetrant analogue, CN-SGC946 (**80**), to assess the pharmacokinetic benefit *in vitro*. Initial efforts entailed three approaches:[††] (i) a modular approach, (ii) a step-wise approach and (iii) synthesis of SGC-946 (**16**) followed by Pd-catalysed insertion of 5-CN.

However, synthetic efforts to isolate CN-SGC946 (**80**) were rendered redundant due to the publication of compound **81** by Bradner *et al*., which exhibits a 5-CN moiety and also incorporates an O-deletion in the ribofuranose (Figure 4.4).[88] Thus, compound **81** tests the same hypothesis as compound **80**. The synthetic route published by Bradner *et al*. used to obtain compound **81** is summarised in Appendix 3. Of note, the nitrile substituent was introduced towards the end of the synthetic route via a palladium catalysed insertion using zinc cyanide.



81 R = CN
82: R = H

Figure 4.4 Structures of the published DOT1L inhibitors **81** and **82**. Compound **81** bears a 5-CN substitution.

Compound **81** and a 5-H analogue **82** were analysed in mouse *in vitro* studies. Compound **81** exhibited an improved half-life ($T_{½}$) of 12.6 hrs, in comparison to 2.9 hrs for EPZ-4777 (**13**) and 1.2 hrs for compound **17**. However, the 5-H analogue **82** proved to be more stable with a half-life of 26.3 hrs. This was off-set by the improved cellular activity of compound **81** versus its 5-CH analogue **82**. A western blot assay showed compound **81** reduces H3K27 trimethylation 'at a much lower dose' than both its 5-CH analogue **82** and EPZ-4777 (**13**). An anti-proliferative study using the MV4-11 cell line for AML showed the activity of compound **81** 'was greater than' the 5-H analogue **82**,

[††] Synthetic routes and reaction conditions are summarised in Appendix 2. Experimental data is not presented for these structures.

again confirming the importance of the 5-CN group. This data indicate the nitrile substituent is beneficial to both binding and the pharmacokinetic properties of this inhibitor template thus supporting the proof of concept hypothesised by this thesis.

4.6 Conclusions and future work

DOT1L is known to play a role in *MLL*-r leukaemia through its recruitment by *MLL* fusion proteins and subsequent hypermethylation of a characteristic set of genes known to drive leukaemogenesis. A new nucleoside inhibitor, CN-SAH (**48**), for the PKMT DOT1L has been designed through analysis of structural literature reviews of the DOT1L active site and published crystal structures of DOT1L inhibitors. Occupation of the DOT1L hydrophobic cavity has been demonstrated to provide selectivity and potency for DOT1L. A screen of synthetically tractable and commercially available compounds identified the natural product toyocamycin (**29**, 5-nitrile) as equipotent to micromolar DOT1L inhibitor 5-ITC (**14**, 5-iodo). More significantly, this single point iodo-nitrile transformation improves the LipE value by 1.6 units which indicates potential improvement to the pharmacokinetic properties of this template. A reduction in lipophilicity, which in this case is gained through loss of the lipophilic iodide for the polar nitrile, has been shown to improve the medicinal chemical properties of drug candidates. Additionally, it is moving away from the metabolically labile adenine core to a less recognisable structure.

This substitution was translated into a biochemical tool CN-SAH (**48**), an analogue of the DOT1L inhibitor Br-SAH (**15**). The synthesis of CN-SAH (**48**) was more challenging than anticipated but the desired product was isolated in an 8-step synthesis. Pleasingly CN-SAH (**48**) was 3-fold more potent than its analogue Br-SAH (**15**). CN-SAH (**48**) was also shown to be selective for DOT1L through a screen against other methyltransferases. Additionally, a crystal structure confirmed the nitrile substituent occupies the hydrophobic cavity. The retention of potency and biological properties from Br-SAH (**15**) to CN-SAH (**48**) supports the iodo-nitrile substitution as a promising bioisosteric transformation.

The next step was to assess the potential pharmacokinetic benefit *in vitro* of this iodo-nitrile transformation. The potent and selective cell penetrant DOT1L inhibitor SGC-946 (**16**) was identified as the template from which to translate the nitrile moiety. This substitution would have resulted in a putative cell penetrant analogue CN-SGC946 (**80**)**.** However, synthetic efforts to isolate **80** were rendered redundant due to disclosure of a cell penetrant analogue **81** which bears a 5-CN substituent and an O-

deletion in the ribofuranose ring. Compound **81** was shown to be active in mouse *in vivo* studies with an improved half-life and potency, underlining the importance of the 5-CN substitution to both activity and pharmacokinetic properties of the new inhibitor scaffold and thus supporting the hypothesis of this thesis.

Future work could involve identification of further, more dramatic structural alterations to move away from the metabolically labile adenosine core, while retaining potency and selectivity, to achieve a more drug-like template with oral bioavailability.

4.7 Epilogue

Since the point at which work on this project stopped (September 2015), there have been a number of publications detailing advancements towards small molecule inhibition of DOT1L, for both previously published inhibitors and novel scaffolds. A synopsis of these findings is given below.

A noteworthy development for previously published inhibitors is the completion of phase I clinical trials for treatment of patients with *MLL*-r leukaemia by the potent and selective DOT1L inhibitor EPZ-5676 (**4**, pinometostat), the first PMT to advance to the clinic. However, despite patient response and in some cases remission, observation of subsequent relapse in the adult trial treating MLL-r patients has led to investigations into potential mechanisms of EPZ-5676 (**4**) treatment emergent resistance (TER), which are on-going.[89]

Significant developments have been made in development of novel DOT1L inhibitors which are not derived from the metabolically labile nucleoside scaffold, which is responsible, in part, for the poor pharmacokinetics of previously published inhibitors, including EPZ-5676 (**4**). Several structurally novel DOT1L inhibitors have been reported and their structures are summarised in Figure 4.5.

Scheufler *et al.* reported a structurally novel pyrimidine series of DOT1L inhibitors, **83** and **84** which were identified via optimisation of a hit from a micromolar HTS using a scintillation proximity assay (SPA).[90] Both of these pyrimidine scaffolds display high potencies in biochemical assays (IC_{50} = 1.4 and 0.4 nM, and K_i = 0.36 and 0.08 nM respectively) and good selectivity against both PKMTs and PRMTs. Importantly both compounds exhibit good *in vitro* activity in cellular assays. In addition, **83** was evaluated in PK experiments in rats where it shows high total blood clearance, high volume distribution and a moderate half-life and oral bioavailability. Intriguingly, this series of inhibitors interact with a novel binding pocket induced adjacent to the co-factor binding site, as determined by X-ray crystallography of a related analogue **85** (PDB code: 5DRT). Despite not binding in the SAM binding site, the compounds are described as SAM-competitive as they interact with the lid loop of the co-factor binding pocket which disrupts SAM binding.

In a companion paper, Scheufler *et al*. reported the identification of a further novel series of DOT1L inhibitors, **86**, bearing a quinazoline scaffold. Compound **86** exhibited low nanomolar DOT1L inhibition (IC_{50} = 14 nM).[91] Compound **86** was obtained by

optimisation of a weak fragment-based screening hit (IC_{50} = 320 µM) through analysis of interactions observed in a co-crystal structure (PDB code: 5DTM). The crystal structure indicated this series also binds to DOT1L in a similarly induced pocket adjacent to the SAM binding site, again disrupting effective SAM binding through interaction with the lid loop.

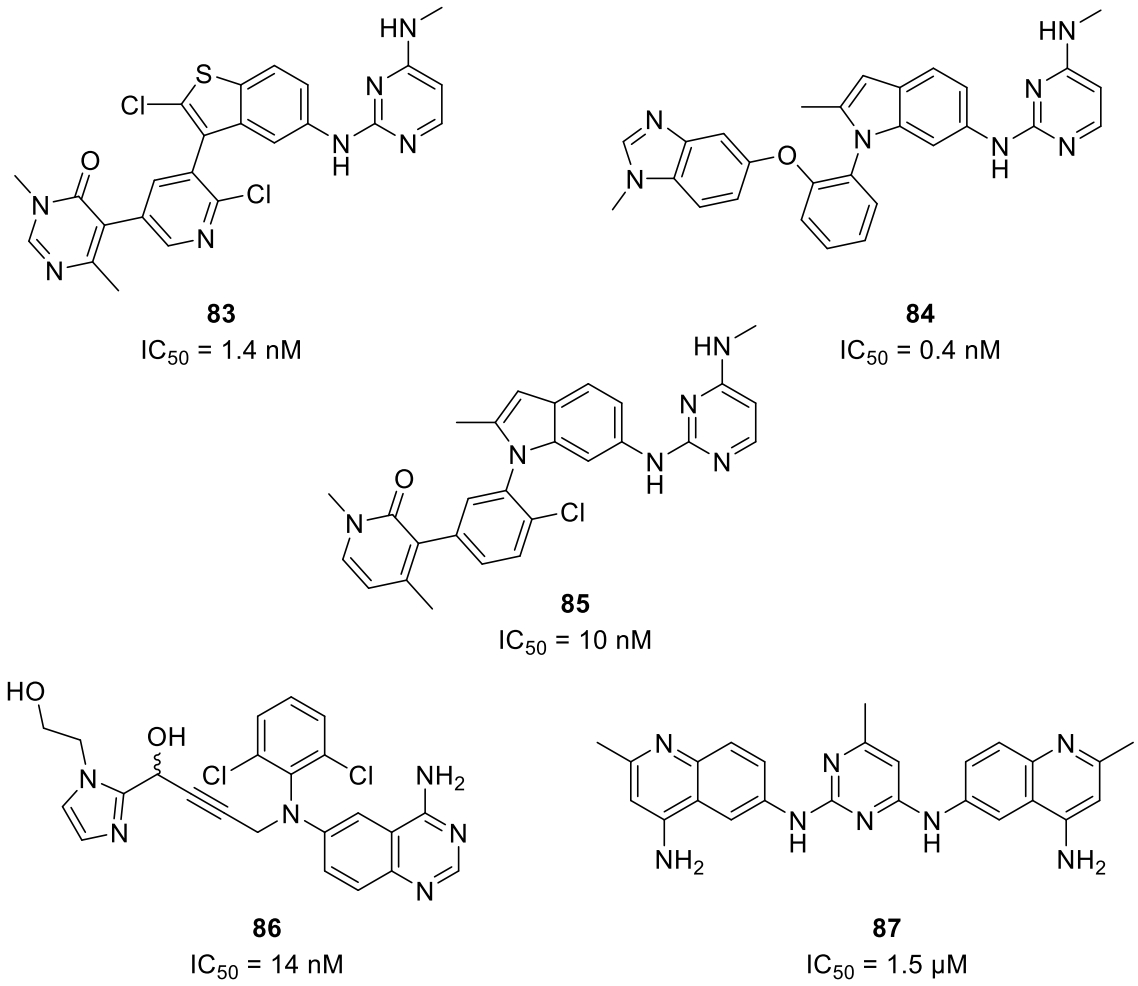


Figure 4.5 Structures of recently reported structurally novel DOT1L inhibitor scaffolds.

A further structurally novel micromolar DOT1L inhibitor, compound **87** (IC_{50} = 1.5 µM), was reported by Chen *et al*. using structure-based virtual screening in combination with biophysical assays.[92] Compound **87** selectively inhibits *MLL*-r cell proliferation (GI_{50} = 37.1 µM). Molecular modelling proposed the compound is SAM-competitive and binds in the co-factor binding site.

In summary, in response to the metabolically unstable pyrrolopyrimidine or adenosine nucleoside template used by previously published inhibitors, non-nucleoside DOT1L inhibitors have been identified through utilisation of HTS, fragment-based drug discovery and virtual-based screening methods including structures which have the

potential to be useful chemical tools for cellular and *in vivo* studies. Furthermore, two of the novel series of inhibitors interact with DOT1L via a novel induced binding site, demonstrating inhibition of DOT1L can be achieved without occupation of the co-factor binding site, the mechanism of inhibition employed by all previously published inhibitors.

Similar success could be obtained through continued optimisation of the nucleoside derivative. Single point changes can result in great improvement to PK properties as shown through compound **80** which incorporates a 5-CN in combination with a cyclopentane alternative to the ribofuranose, and exhibits improved *in vitro* and *in vivo* pharmacokinetic properties. Through combinations of these single point changes, an orally bioavailable, cell active and potent *in vivo* DOT1L inhibitor could be obtained. However, an important point to note is the observed relapse in patients from the clinical study of EPZ-5676 (**4**). This possibly indicates treatment of MLL via inhibition of DOT1L could be rendered redundant, or alternatively, DOT1L inhibitors would need to be administered in combination with other inhibitors which act through different inhibition mechanisms.

CHAPTER 5
Experimental

All oxygen and moisture sensitive reactions were carried out under argon, in oven dried glassware purged with argon prior to use.

All Nuclear Magnetic Resonance (NMR) spectra were recorded on a Bruker Avance 400 MHz or 500 MHz NMR spectrometer. Chemical shifts are reported in ppm (δ) with reference to the solvent and instrument. Data is reported in the following format; chemical shift, integration, multiplicity (s = singlet, t = triplet q = quartet, m = multiplet, br = broad), coupling (Hz), assignment.

All infra-red (IR) spectra were recorded on a PerkinElmer 100 Optica FT-IR spectrophotometer. Data are reported in the following format; type (neat, nujol etc.), cm^{-1}, absorption, intensity (s = strong, m = medium, w = weak), assignment.

All optical rotation values were recorded on a PolAAr 21 with reference to the temperature and cell length.

Melting points were performed on a Stuart SMP10 and are uncorrected.

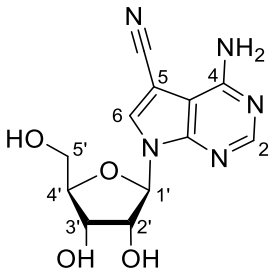
All normal phase purification was carried out by flash chromatography on silica gel 60 Å pore size, 230-400 mesh particle, 40-63 µM particle size. Analytical thin-layer chromatography (TLC) was performed on aluminium sheets with a silica gel layer 60 Å F_{254}. Visualisation was accomplished using UV light and staining by aqueous potassium permanganate ($KMnO_4$) followed by heating.

Low resolution mass spectrometry (LRMS) is performed on a Thermo LCQ Duo Mass Spectrometer operating under electrospray ionisation in positive and negative mode. Data is reported in the following format; method, mass, isotope identification if applicable, ion, charge, ion percentage. High resolution mass spectrometry (HRMS) was performed by the EPSRC UK National Mass Spectrometry Facility (NMSF) at Swansea University.

Purity was analysed on a Shimadzu LCMS-20/20 system using a Waters XTerra® C18 column 50 x 4.6 mm (particle size: 2.5 µm) at a flow rate of 1 mL/min. The eluent system and method were as follows; Eluent A H_2O + 0.1% FA; Eluent B MeCN + 0.1%

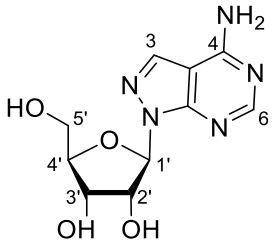
FA; initial fixed composition 10% B to 95% B over 8 min, held at 95% B for 2 min then returned to 10% B over 0.2 min and held for 2 min, the total run time was 12 min.

All reagents and laboratory grade solvents were purchased from commercial vendors and were used without further purification unless otherwise stated. Specifically: 4-Chloro-7*H*-pyrrolo[2,3-*d*]pyrimidine (CAS: 3680-69-1) was purchased from Alfar Aesar; D-(-)-ribose [50-69-1]; L-homocystine was purchased from Fluorochem (CAS: 626-72-2) and 1-*O*-acetyl-2,3,5-tri-*O*-benzoyl-β-D-ribofuranose (CAS: 6974-32-9) was purchased from Sigma-Aldrich. The compounds submitted for screening were purchased from the indicated commercial supplier (catalogue number) between January 2013 and March 2015: Compound **14** from Sigma-Aldrich (I100); Compound **27** from Sigma-Aldrich (A9251); Compound **28** from Sigma-Aldrich (T0642); Compound Compound **29** from Sigma-Aldrich (T3580); Compound **30** from Sigma-Aldrich (S5895); **31** from Sigma-Aldrich (T3830); Compound **32** from Carbosynth (ND06190); Compound **33** from Carbosynth (ND08351); Compound **34** from Sigma-Aldrich (D8296).

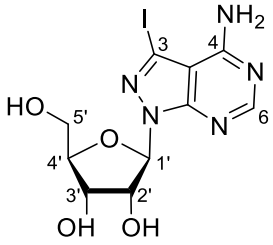


4-Amino-7-((2*R*,3*R*,4*S*,5*R*)-3,4-dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-7*H*-pyrrolo[2,3-*d*]pyrimidine-5-carbonitrile (Toyocamycin) (29). Compound **73** (0.53 g, 0.88 mmol) was suspended in cold 7 N methanolic ammonia (18 mL, 0.050 M) and stirred at 0 °C for 1 hr. The reaction mixture was allowed to warm to rt and stirred overnight. The solvent was concentrated under vacuum and the residue purified by flash column chromatography [EtOAc:MeOH:AcOH, 98:1:1–95:4:1]. Compound **29** was obtained as a white solid (0.22 g, 78%): R_f 0.22 [MeOH: $CHCl_3$, 1:9]; $[\alpha]_D^{25}$ +70 ° (c 0.5, MeOH); mp 208–210 °C dec (lit.[93] 243 °C); γ_{max} (neat)/cm^{-1} 3120 (br), 2231 (w), 1666 (s); ^{1}H NMR (400 MHz, DMSO-d_6) δ: 8.44 (1H, s, 2/6-C*H*); 8.23 (1H, s, 2/6-C*H*); 6.88 (2H, br s, 4-NH_2); 6.06 (1H, d, *J*= 5.6 Hz, 1'-C*H*); 5.43 (d, *J*= 5.6 Hz, 2'-OH); 5.18–5.14 (2H, m, 3'-OH, 5'-OH); 4.37 (1H, dd, *J*= 5.6, 5.6 Hz, 2'-C*H*); 4.12–4.09 (1H, m, 3'-C*H*); 3.93 (1H, ddd, *J*= 3.7, 3.7, 3.7 Hz, 4'-C*H*); 3.69–3.64 (1H, m, 5'-C*H*); 3.59–3.54 (1H, m, 5'-C*H*); ^{13}C NMR (100 MHz, DMSO-d_6) δ: 157.0 (*C*), 153.5 (*C*H-2), 150.1 (*C*), 132.3 (*C*H-6), 115.3 (*C*N), 101.2 (*C*), 87.8 (*C*H-1'), 85.4 (*C*), 82.9 (*C*H-4'), 74.2 (*C*H-2'),

70.2 (*C*H-3'), 61.2 (*C*H-5'); HRMS *m/z* (ESI$^+$) [Found (M+H)$^+$ 292.1040; $C_{12}H_{13}N_5O_4SH$ requires (M+H)$^+$ 292.1042]. The data are in good agreement with the literature values.[86]

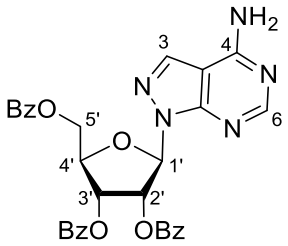


(2*R*,3*R*,4*S*,5*R*)-2-(4-Amino-1*H*-pyrazolo[3,4-*d*]pyrimidin-1-yl)-5-(hydroxymethyl)-tetrahydrofuran-3,4-diol (41). NaOMe (26 mg, 1.1 mmol) was added as one portion to a solution of compound **44** (0.50 g, 0.86 mmol) in MeOH (22 mL, 0.04 M). The solution was flushed with nitrogen and then stirred at rt for 1 hr and refluxed for a further 3 hrs. The reaction mixture was allowed to cool and the solvent was removed under vacuum. The crude residue was purified by column chromatography [$CHCl_3$:MeOH, 5:1] to provide compound **41** as a white solid (60 mg, 26%): R_f 0.34 [$CHCl_3$:MeOH, 9:1]; mp 248 °C dec (lit.[94] 247–249 °C); ^{1}H NMR (500 MHz, DMSO-d_6) δ: 8.19 (1H, s, 3/6-C*H*); 8.17 (1H, s, 3/6-C*H*); 7.90 (1H, br s, 4-NH_2); 7.69 (1H, br s, 4-NH_2); 6.08 (1H, d, *J*= 4.7 Hz, 1'-C*H*); 5.37 (1H, d, *J*= 4.1 Hz, 2'-O*H*); 5.14 (1H, d, *J*= 5.4 Hz, 3'-O*H*); 4.88 (1H, t, *J*= 5.8 Hz, 5'-O*H*); 4.61–4.58 (1H, m, 2'-C*H*); 4.19–4.21 (1H, m, 3'-C*H*); 3.91–3.88 (1H, m, 4'-C*H*); 3.59–3.55 (1H, m, 5'-C*H*); 3.42 (1H, m, 5'-C*H*); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 158.1 (*C*), 156.1 (*C*H-3/6), 154.0 (*C*), 133.3 (*C*H-3/6), 100.5 (*C*), 88.6 (*C*H-1'), 85.1 (*C*H-4'), 73.1 (*C*H-2'), 70.9 (*C*H-3'), 62.4 (*C*H-5'); LCMS RT 0.74 min; 87% content; *m/z* (ESI$^+$) 268 ([M+H]$^+$ 100%) 290 ([M+Na]$^+$ 72%). The data are in good agreement with the literature values.[95]

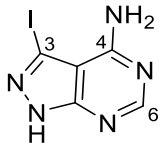


(2*R*,3*R*,4*S*,5*R*)-2-(6-Amino-7-iodo-1*H*-pyrazolo[3,4-*d*]pyrimidin-1-yl)-5-(hydroxymethyl)tetrahydrofuran-2,3-diol (42). NaOMe (4 mg, 0.18 mmol) was added as one portion to a solution of compound **47** (0.10 g, 0.14 mmol) in MeOH (3.5 mL, 0.04 M). The solution was flushed with nitrogen and then stirred at rt for 1 hr and refluxed for a further 3 hrs. The reaction mixture was allowed to cool and the solvent

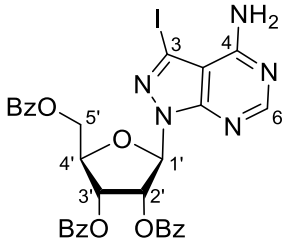
was removed under vacuum. The crude residue was purified by column chromatography [$CHCl_3$: MeOH, 6:1] and then re-crystallised from MeOH to provide compound **42** as a pale pink solid (30 mg, 54%): R_f 0.2 [$CHCl_3$:MeOH, 8:2]; mp 246 °C dec (lit.[64] >247 °C dec); ^{1}H NMR (500 MHz, DMSO-d_6) δ: 8.23 (1H, s, 6-C*H*); 8.03 (1H, br s, 4-NH_2); 6.75 (1H, br s, 4-NH_2); 6.03 (1H, d, *J*= 5.0 Hz, 1'-C*H*); 5.40 (1H, d, *J*= 6.0 Hz, 2'-O*H*); 5.17 (1H, d, *J*= 5.4 Hz, 3'-O*H*); 4.84 (1H, t, *J*= 5.8 Hz, 5'-*OH*); 4.58–4.55 (1H, m, 2'-C*H*); 4.17–4.15 (1H, m, 3'-C*H*); 3.90–3.87 (1H, m, 4'-C*H*); 3.57–5.53 (1H, m, 5'-C*H*); 3.45–3.40 (1H, m, 5'-C*H*); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 157.7 (*C*), 156.3 (*C*H-6), 154.4 (*C*), 103.5 (*C*), 91.6 (*C*), 88.3 (*C*H-1'), 85.3 (*C*H-4'), 72.9 (*C*H-2'), 70.7 (*C*H-3'), 62.2 (*C*H-5'); LCMS RT 0.85 min; >95% content; *m/*z (ESI$^-$) 392 ([M-H]$^-$ 100%). The data are in good agreement with the literature values.[64]



(2*R*,3*R*,4*R*,5*R*)-2-(4-Amino-1*H*-pyrazolo[3,4-*d*]pyrimidin-1-yl)-5-((benzoyloxy)-methyl)tetrahydrofuran-3,4-diyldibenzoate (44). 1-*O*-Acetyl-2,3,5-tri-*O*-benzoyl-β-D-ribofuranose (2.7 g, 5.4 mmol) was added to a suspension of 1*H*-pyrazolo[3,4-*d*]pyrimidin-4-amine (**45**) (0.50 g, 3.7 mmol) in anhydrous nitromethane (37 mL, 0.1 M). The reaction mixture was refluxed for 30 minutes, allowed to cool and boron trifluoride diethyletherate (0.66 mL, 5.4 mmol) was added as one portion. The resulting clear yellow solution was refluxed for 3 hrs, allowed to cool and the solvent removed under vacuum. The crude residue was purified by column chromatography [Hexane:EtOAc, 1:3] to yield compound **44** as a yellow foam (0.90 g, 42%): R_f 0.25 [Hexane:EtOAc, 2:3]; ^{1}H NMR (400 MHz, DMSO-d_6) δ: 8.27 (1H, s, 3/6-C*H*); 8.22 (1H, s, 3/6-C*H*); 8.01–7.99 (2H, m, Ar-C*H*); 7.94–7.9 (4H, m, Ar-C*H*); 7.70–7.63 (3H, m, Ar-C*H*); 7.54–7.44 (6H, m, Ar-C*H*); 6.69 (1H, d, *J*= 2.8 Hz, 1'-C*H*); 6.30–6.28 (1H, m, 2'-C*H*); 6.24–6.21 (1H, m, 3'-C*H*); 4.89–4.86 (1H, m, 4'-C*H*); 4.66–4.62 (1H, m, 5'-C*H*); 4.55–4.51 (1H, m, 5'-C*H*); *m/z* (ESI$^+$) 580 ([M+H]$^+$ 95%). The data are in good agreement with the literature values.[95]

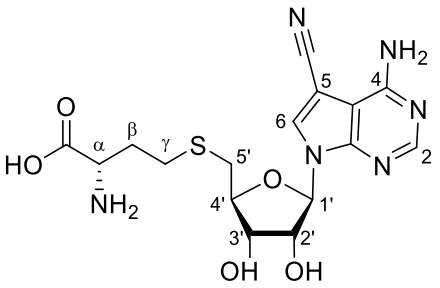


3-Iodo-1*H*-pyrazolo[3,4-*d*]pyrimidin-4-amine (46). *N*-Iodosuccinimde (0.25 g, 1.11 mmol) was added to a suspension of 1*H*-pyrazolo[3,4-*d*]pyrimidin-4-amine (**45**) (0.10 g, 0.74 mmol) in DMF (0.80 mL, 0.9 M). The solution was stirred at 80 °C overnight, allowed to cool and then half of the solvent was removed under vacuum. Hexane (2mL) was added to the reaction mixture with a few drops of MeOH and allowed to stir vigorously. The resulting precipitate was collected, washed with ether and dried under vacuum to afford compound **46** as a pale yellow solid (0.15 g, 78%): R_f 0.9 [Hexane:EtOAc, 3:7]; ^{1}H NMR (500 MHz, DMSO-d_6) δ: 13.79 (1H, br s, 1-N*H*); 8.16 (1H, s, 6-C*H*); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 157.5 (*C*), 156.0 (*C*H-6), 155.0 (*C*), 102.44 (*C*), 89.7 (*C*); *m/z* (ESI$^+$) 262 ([M+H]$^+$ 70%). The data are in good agreement with the literature values.[64]

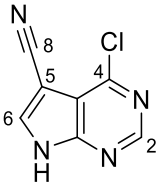


(2*R*,3*R*,4*R*,5*R*)-2-(4-Amino-3-iodo-1*H*-pyrazolo[3,4-*d*]pyrimidin-1-yl)-5-((benzoyloxy)methyl)tetrahydrofuran-3,4-diyl dibenzoate (47). 1-*O*-Acetyl-2,3,5-tri-*O*-benzoyl-β-D-ribofuranose (0.25 g, 0.50 mmol) was added to a suspension of compound **46** (90 mg, 0.34 mmol) in anhydrous $MeNO_2$ (37 mL, 0.1 M). The solution was stirred at rt for 15 minutes and then boron trifluoride diethyletherate (62 µL, 0.50 mmol) was added as one portion producing a clear yellow solution. After refluxing for 3 hrs, the resulting black solution was allowed to cool and the solvent removed under vacuum. The crude residue was purified by column chromatography [Hexane:EtOAc:MeOH, 50:49:1] to provide compound **47** as a brown solid (0.12 g, 44%): R_f 0.21 [Hexane:EtOAc, 1:3]; ^{1}H NMR (500 MHz, DMSO-d_6) δ: 8.25 (1H, s, 6-C*H*); 8.05–8.03 (2H, m, Ar-C*H*); 7.93–7.80 (4H, m, Ar-C*H*); 7.70–7.64 (3H, m, Ar-C*H*); 7.59–7.56 (2H, m, Ar-C*H*); 7.49–7.45 (4H, m, Ar-C*H*); 6.66 (1H, d, *J*= 3.8 Hz, 1'-C*H*); 6.31–6.30 (1H, m, 2'-C*H*); 6.13–6.11 (1H, m, 3'-C*H*); 4.89–4.86 (1H, m, 4'-C*H*); 4.61–4.60 (2H, m, 5'-C*H*); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 165.5 (*C*=O), 164.7 (*C*=O), 164.6 (*C*=O), 157.8 (*C*), 156.7 (*C*H-6), 154.6 (*C*), 134.0 (*C*H-Ar), 133.5 (*C*H-Ar), 129.4

(*C*H-Ar), 129.4 (*C*H-Ar,*C*H-Ar), 129.3 (*C*H-Ar), 128.9 (*C*H-Ar), 128.8 (*C*H-Ar), 128.8 (*C*H-Ar), 128.6 (*C*H-Ar), 128.4 (*C*H-Ar), 103.5 (*C*), 93.4 (*C*), 85.9 (*C*H-1'), 79.0 (*C*H-4'), 73.6 (*C*H-2'), 70.9 (*C*H-3'), 63.3 (*C*H-5'); *m/z* (ESI$^+$) 706 ([M+H]$^+$ 80%) 728 ([M+Na]$^+$ 100%)**.** The data are in good agreement with the literature values.[64]

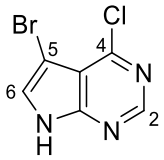


S-(((2*S*,3*S*,4*R*,5*R*)-5-(4-Amino-5-cyano-7*H*-pyrrolo[2,3-*d*]pyrimidin-7-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)-L-homocysteine CN-SAH (48). Compound **79** (77 mg, 0.14 mmol) was suspended in anhydrous DCM (0.46 mL, 0.30 M) and cooled to 0 °C, 4 M HCl dioxane (0.45 ml) was added dropwise and the reaction was stirred at 0 °C for 3 hrs. The reaction was diluted with DCM (5 mL) and extracted with H_2O (5 mL), the aqueous layer was freeze dried to yield CN-SAH (**48**) as an off white solid as the di hydrochloride salt (45 mg, 67%): $[\alpha]_D^{25}$ = +18 ° (c 0.5, MeOH); mp 132–134 °C; γ_{max} (neat)/cm^{-1} 2922 (br), 2232 (s), 1665 (s), 1518 (s), 1427 (s), 1047 (s); ^{1}H NMR (500 MHz, CD_3OD) δ: 8.51 (1H, s, 2/6-C*H*); 8.48 (1H, s, 2/6-C*H*); 6.28 (1H, d, *J*= 4.5 Hz, 1'-C*H*); 4.52 (1H, dd, *J*= 4.6, 4.5 Hz, 2'-C*H*); 4.28–4.23 (2H, m, 3'-C*H*,4'-C*H*); 4.15 (1H, t, *J*= 6.1 Hz, Cys-α-C*H*); 3.07–2.99 (m, 2H, 5'-CH_2); 2.81 (2H, t, *J*= 7.4 Hz, Cys-γ-CH_2); 2.32–2.25 (1H, m, Cys-β-C*H*); 2.20–2.13 (1H, m, Cys-β-C*H*); ^{13}C NMR (125 MHz, CD_3OD) δ: 171.4 (*C*=O), 152.5 (*C*), 149.7 (*C*), 145.8 (*C*H-2/6), 135.7 (*C*H-2/6), 113.9 (*C*N), 103.0 (*C*), 90.7 (*C*H-1'), 88.7 (*C*), 85.6 (*C*H-4'), 76.0 (*C*H-2'), 73.8 (*C*H-3'), 52.7 (Cys-α-*C*H), 35.0 (*C*H-5'), 31.6 (Cys-β-*C*H$_2$), 29.0 (Cys-γ-*C*H$_2$); LCMS RT 0.85 min; 92.5% content; *m/z* (ESI$^+$) 409.2 ([M+H]$^+$ 100%). HRMS *m/z* (ESI$^+$) [Found (M+H)$^+$ 409.1284; $C_{16}H_{20}N_6O_5SH$ requires (M+H)$^+$ 409.1289].

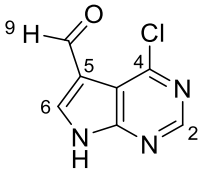


4-Chloro-7*H*-pyrrolo[2,3-*d*]pyrimidine-5-carbonitrile (50). Oxime **53** (4.3 g, 0.022 mol) was suspended in anhydrous DCM (145 mL, 0.15 M) in an oven dried flask purged with N_2. Thionyl chloride (16 mL, 0.22 mol) was added slowly and the resulting

reaction mixture was stirred at rt overnight. The solvent was concentrated under vacuum and the residue suspended in water (10 mL), the resulting solution was adjusted to pH 4 using sat. *aq.* $NaHCO_3$ (~5 mL). The solid precipitate was collected by filtration, rinsed with water and dried under vacuum to afford nitrile **50** (3.3 g, 83%): R_f 0.24 [Hexane:EtOAc, 1:1]; γ_{max} (neat)/cm^{-1} 3076 (br), 2240 (m), 1620 (s); ^{1}H NMR (400 MHz, DMSO-d_6) δ: 13.63 (1H, br s, 7-N*H*); 8.78 (1H, s, 2/6-C*H*); 8.68 (1H, s, 2/6-C*H*); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 152.4 (*C*H-2/6), 151.7 (*C*), 150.8 (*C*), 138.6 (*C*H-2/6), 115.1 (*C*N), 114.3 (*C*), 83.3 (*C*); *m/z* (ESI$^-$) 177 ([^{35}M-H]$^-$ 100%) 179 ([^{37}M-H]$^-$ 30%) 354 ([^{35}M-H]$^-$ 85%) 356 ([^{37}M-H]$^-$ 50%).[93] The data are in good agreement with the literature values.[93]

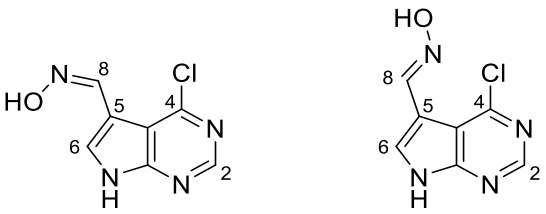


5-Bromo-4-chloro-7*H*-pyrrolo[2,3-*d*]pyrimidine (51). *N*-Bromosuccinimide (12 g, 0.065 mol) was added to a stirred suspension of 4-chloro-7*H*-pyrrolo[2,3-*d*]pyrimidine (**49**) (9.0 g, 0.059 mol) in DCM (0.29 L, 0.20 M). The reaction mixture was stirred at rt overnight with exclusion of light. The reaction was filtered and compound **51** was isolated as a yellow solid (8.7 g, 64%): R_f 0.36 [Hexane:EtOAc, 5:2]; ^{1}H NMR (500 MHz, DMSO-d_6) δ: 12.96 (1H, br s, 7-N*H*); 8.60 (1H, s, 2/6-C*H*); 7.92 (1H, s, 2/6-C*H*); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 150.9 (*C*), 150.8 (*C*H-2/6), 150.1 (*C*), 128.6 (*C*H-2/6), 113.6 (*C*), 85.8 (*C*); *m/*z (ESI$^-$) 230 ([^{79Br,35Cl}M-H]$^-$ 74%) 232 ([^{81Br,35Cl}M-H]$^-$[^{79Br,37Cl}M-H]$^-$ 100%) 234 ([^{81Br,37Cl}M-H]$^-$ 20%). The data are in good agreement with the literature values.[96]



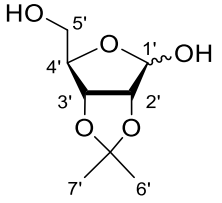
4-Chloro-7*H*-pyrrolo[2,3-*d*]pyrimidine-5-carbaldehyde (52). Compound **51** (8.1 g, 0.035 mol) was suspended in anhydrous THF (0.18 L, 0.2 M) and cooled to -78 °C. 2.5 M *n*-BuLi in hexane (31 ml, 0.080 mol) was added dropwise and the reaction mixture was stirred for 2 hrs at -78 °C. Anhydrous DMF (5.4 ml, 0.10 mol) was added dropwise and the reaction stirred for a further 3 hrs at rt. The reaction was diluted with water (10 mL) and *aq.* NH_4Cl (15 mL) was added until the formation of a precipitate, the reaction was filtered and the precipitate collected to yield compound **52** as an off white solid (4.1

g, 65%). The organic solvent was concentrated under vacuum and a second batch of off white precipitate was collected to yield compound **52** (1.1 g, 17%): R_f 0.32 [Hexane:EtOAc, 3:1]; ^{1}H NMR (500 MHz, DMSO-d_6) δ: 10.23 (1H, s, 9-C*H*); 8.73 (1H, s, 2/6-C*H*); 8.59 (1H, s, 2/6-C*H*); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 183.9 (*C*=O), 154.0 (C), 151.7 (*C*H-2/6), 151.2 (C), 137.6 (*C*H-2/6), 115.6 (C), 113.9 (C); *m/z* (ESI$^-$) 180 ([^{35Cl}M-H]$^-$ 100%) 182 ([^{37Cl}M-H]$^-$ 30%) 360 ([^{35Cl}2M-H]$^-$ 20%) 362 ([^{37Cl}2M-H]$^-$ 7%). The data are in good agreement with the literature values.[97]



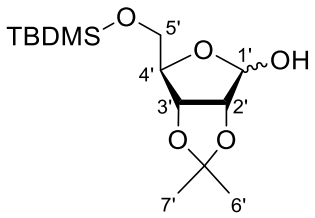
Z E

4-Chloro-7*H*-pyrrolo[2,3-*d*]pyrimidine-5-carbaldehyde oxime (53). A solution of 2M NaOH (18 mL, 0.041 mol) was added dropwise to a suspension of aldehyde **52** (4.9 g, 0.027 mol) and hydroxylamine hydrochloride (2.4 g, 0.034 mol) in EtOH (0.11 L, 0.25 M). The resulting reaction mixture was diluted with EtOH (~80 mL) and was stirred at 50 °C overnight. The solid was collected by filtration, washed with water (20 mL) and dried under vacuum to afford oxime **53** as two isomers (*E* and *Z*) in a 2:1 ratio (4.5 g, 84%): R_f 0.27, 0.13 [Hexane:EtOAc, 3:1]; ^{1}H NMR (400 MHz, DMSO-d_6) δ: [*E* or Z-isomer] 13.04 (1H, br s, 8-O*H*); 11.81 (1H, br s, 7-N*H*); 8.65 (1H, s, 2/6-C*H*); 8.57 (1H, s, 8-C*H*); 8.08 (1H, s, 2/6-C*H*); [*E* or *Z*-isomer] 12.95 (1H, br s, 8-O*H*); 11.09 (1H, br s, 7-N*H*); 8.61 (1H, s, 2/6-C*H*); 8.51 (1H, s, 8-C*H*); 8.00 (1H, s, 2/6-C*H*); ^{13}C NMR (125 MHz, DMSO-d_6) δ: [*E* or *Z*-isomer] 152.2 (*C*), 151.5 (*C*), 150.9 (*C*H-2/6), 136.4 (*C*H-2/6), 133.2 (*C*H-8), 113.4 (*C*), 104.8 (*C*), [E or *Z*-isomer] 152.2 (*C*), 150.6 (*C*), 150.3 (*C*H-2/6), 140.8 (*C*H-8), 126.0 (*C*H-2/6), 113.7 (*C*), 107.4 (*C*); *m/z* (ESI$^-$) 195 ([^{35Cl}M-H]$^-$ 100%) 197 ([^{37Cl}M-H]$^-$ 30%). The data are in good agreement with the literature values.[81]

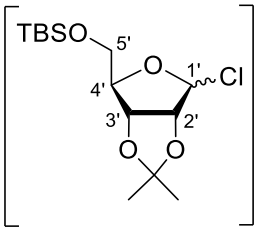


(3a*R*,6*R*,6a*R*)-6-(Hydroxymethyl)-2,2-dimethyltetrahydrofuro[3,4-*d*][1,3]dioxol-4-ol (55). H_2SO_4 (0.18 mL, 3.4 mmol) was added to a solution of D-(-)-ribose (**54**) (10 g,

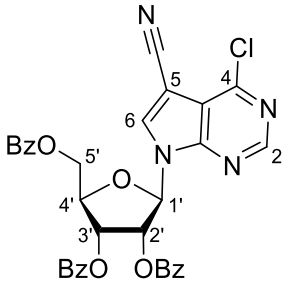
0.067 mol) in anhydrous acetone (0.13 L, 0.50 mol). The solution was stirred at rt for 3 hrs and then quenched with solid $NaHCO_3$ (spatula tip). The solvent was removed under vacuum and the crude residue was purified by column chromatography [Hexane:EtOAc, 1:1] to yield compound **55** as a clear colourless oil (9.3 g, 73%): R_f 0.3 [Hexane:EtOAc, 2:3]; ^{1}H NMR (400 MHz, $CDCl_3$) δ: 5.40 (1H, s, 1'-C*H*), 4.78 (1H, d, *J*= 5.6 Hz, 3'-C*H*); 4.56 (1H, d, *J*= 5.6 Hz, 2'-C*H*); 4.35 (3H, br s, 1'-O*H*); 3.75–3.66 (2H, m, 5'-C*H*); 1.39 (3H, s, CH_3); 1.27 (3H, s, CH_3); ^{13}C NMR (100 MHz, $CDCl_3$) δ: 113.2 (*C*), 104.0 (*C*H-1'), 88.1 (*C*H-4'), 87.5 (*C*H-2'), 83.6 (*C*H-3'), 62.0 (*C*H-5'), 27.0 (*C*H_3-6'/7'); *m/z* (ESI$^-$) 189 ([M-H]$^-$ 100%) 379 ([2M-H]$^-$ 90%). The data are in good agreement with the literature values.[56]



(3a*R*,6*R*,6a*R*)-6-(((*tert*-Butyldimethylsilyl)oxy)methyl)-2,2-dimethyltetrahydrofuro-[3,4-*d*][1,3]dioxol-4-ol (56). Imidazole (3.4 g, 0.049 mol) and TBSCl (3.9 g, 0.032 mol) were added to a suspension of compound **55** (4.0 g, 0.021 mol) in anhydrous DCM (88 mL, 0.24 M). The resulting solution was stirred at rt overnight. The organic solvent was removed under vacuum, suspended in water and extracted with EtOAc (3 x 100 mL). The combined organic layers were washed with brine (100 mL), dried over Na_2SO_4 and concentrated under vacuum. The crude residue was purified by column chromatography [Hexane:EtOAc, 9:1] to yield compound **56** as a white solid (2.8, 41%): R_f 0.26 [Hexane:EtOAc, 7:1]; ^{1}H NMR (500 MHz, CD_3OD) δ: 5.25 (1H, s, 1'-C*H*); 4.73–4.72 (1H, m, 3'-C*H*); 4.52 (1H, d, *J*= 6.0 Hz, 2'-C*H*); 4.12–4.10 (1H, m, 4'-C*H*); 3.73–3.65 (2H, m, 5'-CH_2); 1.44 (3H, s, 6'/7'-CH_3); 1.31 (3H, s, 6'/7'-CH_3); 0.93 (9H, s, CH_3); 0.11 (6H, s, CH_3); ^{13}C NMR (125 MHz, CD_3OD) δ: 113.4 (*C*), 104.2 (*C*H-1'), 88.4 (*C*H-4'), 87.7 (*C*H-2'), 83.6 (*C*H-3'), 65.9 (*C*H-5'), 27.0 (*C*H_3-6'/7'), 26.5 (*C*H_3), 25.2 (*C*H_3-6'/7'), 19.3 (*C*), -5.1 (*C*H_3), -5.2 (*C*H_3); *m/z* (ESI$^+$) 327 ([M+Na]$^+$ 100%). The data are in good agreement with the literature values.[56]

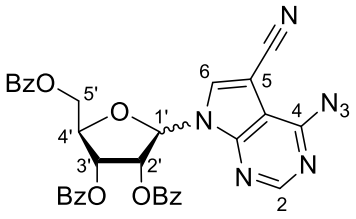


***tert*-Butyl(((3a*R*,4*R*,6a*R*)-6-chloro-2,2-dimethyltetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)methoxy)dimethylsilane (57).** CCl_4 (48 µL, 0.49 mmol) was added to a solution of compound **56** (100 mg, 0.33 mmol) in anhydrous THF (1.3 mL, 0.25 M) under N_2. The solution was cooled to - 70 °C (2-propanol/dry-ice) and tris(dimethylamino)phosphine (66 µL, 0.36 mmol) was added dropwise. The solution was warmed to - 40 °C over 45 minutes, prior to use the solution was warmed to 5 °C. The reaction was monitored by proton NMR by transferring an aliquot to an NMR tube (sealed with a septum and under nitrogen) and concentrating under vacuum to yield a colourless oil. After 2 hrs the reaction was observed to reach completion with formation of **57** as a solution of α and β anomers in a 3:2 ratio [diagnostic α peaks]: ^{1}H NMR (400 MHz, $CDCl_3$) δ: 6.15 (1H, d, *J*= 4 Hz, 1'-C*H)*; 4.60–4.74 (2H, m, 2'/3'/4'-C*H*); 4.40–4.37 (1H, m, 2'/3'/4'-C*H*); 3.81–3.61 (2H, d, *J*= 2.8 Hz, 5'-CH_2); 1.66 (3H, s, CH_3); 1.39 (3H, s, CH_3); 0.88 (9H, s, CH_3); 0.06 (3H, s, CH_3); 0.05 (3H, s, CH_3). NB: a diagnostic peak for the by-product hexamethylphosphoramide is observed at 2.60 (*J*= 9.3 Hz) and trace amounts of reagent tris(dimethylamino)phosphine at 2.48 (*J*= 9.3 Hz) which indicates the reaction is complete. The data are in good agreement with the literature values.[56]



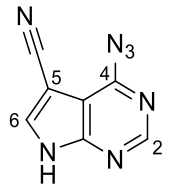
(2*R*,3*R*,4*R*,5*R*)-2-((Benzoyloxy)methyl)-5-(4-chloro-5-cyano-7*H*-pyrrolo[2,3-d]-pyrimidin-7-yl)tetrahydrofuran-3,4-diyldibenzoate (60). *N,O-bis*(trimethylsilyl)acetamide (75 µL, 0.31 mmol) was added to a suspension of compound **50** (50 mg, 0.28 mmol) in anhydrous MeCN (2.0 mL, 0.14 M) and the reaction was stirred at rt for 10 minutes. To the resulting clear green solution was added a solution of 1-*O*-acetyl-2,3,5-tri-*O*-benzoyl-β-D-ribofuranose (170 mg, 0.34 mmol) in anhydrous MeCN and TMSOTf (56 µL, 0.31 mmol). The solution was stirred at rt for a further 15 minutes and then heated to 80 °C for 2 hrs. The solution was

cooled, diluted with EtOAc (30 mL), quenched with sat. *aq.* $NaHCO_3$ (100 mL) and extracted with EtOAc (3 x 100 mL). The combined organic layers were washed with brine (100 mL), dried over Mg_2SO_4 and concentrated under vacuum. The crude residue was purified by column chromatography [Hexane:EtOAc, 4:1] to yield compound **60** as a pale yellow foam (83 mg, 47%): R_f 0.38 [Hexane:EtOAc, 3:5]; ^{1}H NMR (400 MHz, $CDCl_3$) δ: 8.66 (1H, s, 2/6-C*H*); 8.11–8.09 (2H, m, Ar-C*H*); 8.04–8.01 (2H, m, Ar-C*H*); 7.99 (1H, s, 2/6-C*H*); 7.92–7.90 (2H, m, Ar-C*H*); 7.66–7.36 (9H, m, Ar-C*H*); 6.61 (1H, d, *J*= 5.5 Hz, 1'-C*H*); 6.19 (1H, dd, *J*= 5.5, 5.5 Hz, 2'-C*H*); 6.12 (1H, dd, *J*= 5.5, 5.3 Hz, 3'-C*H*); 4.94 (1H, dd, *J*= 12.4, 3.4 Hz, 5'-C*H*); 4.86 (1H, ddd, *J*= 5.3, 3.7, 3.4 Hz, 4'-C*H*); 4.73 (1H, dd, *J*= 12.4, 3.7 Hz, 5'-C*H*); ^{13}C NMR (100 MHz, $CDCl_3$) δ: 166.0 (*C*=O), 165.3 (*C*=O), 165.1 (*C*=O), 153.5 (*C*), 153.0 (*C*H-2/6), 150.7 (*C*), 134.8 (*C*H-2/6), 134.0 (*C*), 133.9 (*C*) 133.8 (*C*), 129.8 (*C*H-Ar,*C*H-Ar), 129.6 (*C*H-Ar), 129.1 (*C*), 128.8 (*C*H-Ar), 128.6 (*C*H-Ar), 128.6 (*C*H-Ar), 128.2 (*C*), 116.7 (*C*N), 112.5 (*C*), 88.1 (*C*H-1'), 87.5 (*C*), 81.2 (*C*H-4'), 74.2 (*C*H-2'), 71.3 (*C*H-3'), 63.2 (*C*H-5'); *m/z* (ESI$^+$) 667 ([^{35Cl}M+2Na-H]$^+$ 100%) 669 ([^{37Cl}M+2Na-H]$^+$ 40%). The data are in good agreement with the literature values.[98]

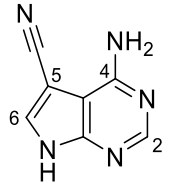


(3*R*,4*R*,5*R*)-2-(4-Azido-5-cyano-7*H*-pyrrolo[2,3-*d*]pyrimidin-7-yl)-5-((benzoyloxy)methyl)tetrahydrofuran-3,4-diyl dibenzoate (62) NaN_3 (0.14 g, 2.2 mmol) was added to a solution of compound **60** (0.34 g, 0.55 mmol) in anhydrous DMF (5.5 mL, 0.10 M). The reaction was heated to 60 °C for 2.5 hrs, allowed to cool and concentrated under vacuum. The residue was purified by column chromatography [$CHCl_3$: MeOH, 9:1] to yield compound **62** as an off white foam (274 mg, 80%) in a 3:2 mixture of inseparable diastereomers (β:α): R_f 0.24 [$CHCl_3$:MeOH, 4:1]; γ_{max} (neat)/cm^{-1} 3087 (w), 2923 (w), 2236 (w), 1720 (s), 1259 (s); ^{1}H NMR (400 MHz, $CDCl_3$) δ: [all peaks]: 9.23 (1H, s), 8.59 (1H, s); 8.11–8.08 (4H, m); 8.05 (1H, s); 8.05–8.00 (4H, m); 7.93–7.90 (4H, m); 7.84 (1H, s); 7.65– 7.35 (18H, m); 6.67 (1H, d, *J*= 5.1 Hz); 6.59 (1H, d, *J*= 5.1 Hz); 6.24–6.18 (2H, m); 6.14–6.09 (2H, m); 5.01–4.97 (1H, m); 4.94–4.90 (2H, m); 4.86–4.83 (1H, m); 4.77–4.70 (2H, m); [diagnostic peaks for the β-anomer]: 8.59 (1H, s, 2/6-C*H*); 7.84 (1H, s, 2/6-C*H*); 6.59 (1H, d, *J*= 5.1 Hz, 1'-C*H*); [diagnostic peaks for the α-anomer]: 9.23 (1H, s, 2/6-C*H*), 8.05 (1H, s, 2/6-C*H*); 6.67 (1H, d, *J*= 5.1 Hz, 1'-C*H*); ^{13}C NMR (100 MHz, $CDCl_3$) δ: 166.1, 166.0, 165.3, 165.3, 165.1, 165.0,

153.5, 151.5, 145.3, 140.6, 134.2, 134.1, 134.0, 133.9, 133.8, 133.7, 133.2, 132.4, 129.9, 129.8, 129.6, 129.6, 129.1, 128.9, 128.8, 128.7, 128.7, 128.6, 128.6, 128.5, 128.3, 128.1, 113.1 (CN), 112.1 (CN), 107.8, 150.0, 89.4, 89.0, 87.9, 86.5, 81.4, 81.0, 74.6, 74.2, 71.3, 71.2, 63.3, 63.0; HRMS *m/z* (ESI^+) [Found $(M+H)^+$ 673.1898; $C_{33}H_{25}N_{10}O_7$ requires $(M+H)^+$ 673.1902].

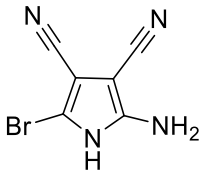


4-Azido-7*H*-pyrrolo[2,3-*d*]pyrimidine-5-carbonitrile (63). NaN_3 (1.78 g, 0.027 mol) was added to a suspension of compound **50** (2.4 g, 0.013 mol) in anhydrous DMF (55 mL, 0.25 M). The reaction was heated to 60 °C for 2 hrs. The excess DMF was concentrated under vacuum to give an oily residue, the remaining reaction mixture was diluted with EtOAc (100 mL) and water (100 mL) at which point the product partially precipitates. The precipitate was filtered and washed with water (20 mL) to yield compound **63** as an off white solid (0.91 g, 36%). The aqueous filtrate was combined with the remaining crude reaction mixture and extracted with EtOAc (3 x 100 mL). All organic extracts were combined and concentrated under vacuum to yield compound **63** as an off white solid (0.74 g, 29%): R_f 0.35 [Hexane:EtOAc, 1:3]; γ_{max} (neat)/cm^{-1} 2921 (br), 2230 (w), 1636 (s); ^1H NMR (500 MHz, DMSO-d_6) δ: 13.94 (1H, br s, 7-N*H*); 10.01 (1H, s, 2/6-C*H*); 8.58 (1H, s, 2/6-C*H*); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 145.3 (*C*), 142.3 (*C*), 135.6 (*C*H-2/6), 134.6 (*C*H-2/6), 114.5 (CN), 102.4 (*C*), 85.2 (*C*); *m/z* (ESI^-) 184 ($[M-H]^-$ 100%) 368 ($[2M-H]^-$ 30 %).

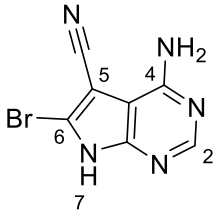


4-Amino-7*H*-pyrrolo[2,3-*d*]pyrimidine-5-carbonitrile (64). Azide **63** (1.63 g, 0.010 mol) was added to a solution of suspended 10 mol% Pd/C (0.94 mg) in anhydrous MeOH (44 mL). The reaction mixture was stirred at rt under a H_2 atmosphere (hydrogen balloon) for 12 hrs. The reaction was filtered through a pad of celite which was washed with EtOH (100 mL). The filtrate was concentrated under vacuum to yield **64** as a white solid (0.46 g, 33%): R_f 0.1 [Hexane:EtOAc, 1:3]; ^1H NMR (400 MHz, DMSO-d_6) δ: 12.65 (1H, br s, 7-N*H*); 8.18 (1H, s, 2/6-C*H*); 8.13 (1H, s, 2/6-C*H*); 6.68

(2H, br s, 4-N*H*$_2$); ^{13}C NMR (125 MHz, DMSO-*d*$_6$) δ: 156.9 (*C*), 153.5 (*C*H-2/6), 150.9 (*C*), 132.4 (*C*H-2/6), 116.0 (*C*N), 100.9 (*C*), 82.0 (*C*); *m/z* (ESI$^-$) 158 ([M-H]$^-$ 20%) 184 ([M+Na-2H]$^-$ 100%). The data are in good agreement with the literature values.[99]

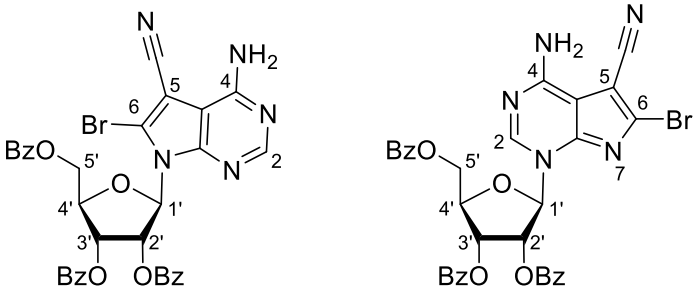


2-Amino-5-bromo-1*H*-pyrrole-3,4-dicarbonitrile (67). Tetracyanoethylene (**66**) (11 g, 0.084 mol) was suspended in acetone (56 mL) and EtOAc (117 mL) and cooled to 0 °C. Maintaining the internal temperature at 0 °C, 33% HBr in acetic acid (58 mL, 0.23 mol) was added dropwise to the suspension. After 3 hrs the resulting yellow precipitate was filtered and washed with EtOAc. The solid was then suspended in water (100 mL) and adjusted to pH 11 with 50% (w/w) *aq.* NaOH (~ 25 mL) and stirred until complete dissolution of the solid was achieved. The resulting red solution was adjusted to pH 5 using glacial AcOH (~ 40 mL) and the precipitate collected under vacuum filtration to yield compound **67** as a beige solid (11 g, 60%): R_f 0.3 [Hexane:EtOAc, 1:9]; ^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ: 12.43 (1H, br s, N*H*); 6.36 (2H, br s, NH$_2$); ^{13}C NMR (100 MHz, DMSO-*d*$_6$) δ: 149.9 (*C*), 115.0 (*C*N), 114.3 (*C*N), 103.8 (*C*), 92.2 (*C*), 70.8 (*C*); *m/z* (ESI$^-$) 209 ([^{79Br}M-H]$^-$ 55%) 211 ([^{81Br}M-H]$^-$ 52%) 418 ([^{79Br}2M-H]$^-$ 52%) 420 ([^{81Br}2M-H]$^-$ 100%). The data are in good agreement with the literature values.[87]



4-Amino-6-bromo-7*H*-pyrrolo[2,3-*d*]pyrimidine-5-carbonitrile (68). Formamidine acetate (4.9 g, 0.047 mol) was added to a solution of compound **67** (10 g, 0.047 mol) in 2-ethoxyethanol (91 mL, 0.52 M). The solution was purged with nitrogen then refluxed for 24 hrs and filtered while hot through celite. The celite cake was washed thoroughly with hot MeOH (150 mL) and the solvent concentrated under vacuum. The crude residue was recrystallised from DMF:H$_2$O (30 mL, 8:2) to yield compound **68** as a dark brown solid (4.1 g, 36%): R_f 0.1 [Hexane:EtOAc, 1:9]; ^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ: 13.72 (1H, br s, 7-N*H*); 8.22 (1H, s, 2-C*H*); 7.17 (2H, br s, 4-N*H*$_2$); ^{13}C NMR (125 MHz, DMSO-*d*$_6$) δ: 155.5 (*C*), 149.6 (*C*H-2), 149.0 (*C*), 124.5 (*C*), 115.4 (*C*N), 103.2 (*C*), 83.9

(*C*); *m/z* (ESI$^-$) 236 ([^{79Br}M-H]$^-$ 100%) 238 ([^{81Br}M-H]$^-$ 100%). The data are in good agreement with the literature values.[100]



69
N7-isomer

70
N1-isomer

(2*R*,3*R*,4*R*,5*R*)-2-(4-Amino-6-bromo-5-cyano-7*H*-pyrrolo[2,3-*d*]pyrimidin-7-yl)-5-((benzoyloxy)methyl)tetrahydrofuran-3,4-diyl dibenzoate (69).

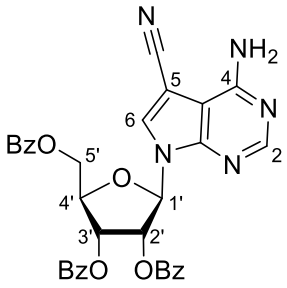
Method 1: *N,O-bis*(Trimethylsilyl)acetamide (3.4 mL, 14 mmol) was added to a stirred suspension of compound **68** (1.7 g, 6.9 mmol) in anhydrous MeCN (69 mL, 0.10 M) and the reaction mixture was stirred at rt for 10 mins. A solution of 1-*O*-Acetyl-2,3,5-tri-*O*-benzoyl-β-D-ribofuranose (3.8 g, 7.5 mmol) in anhydrous MeCN (~5 mL) was added followed by TMSOTf (3.8 ml, 21 mmol), at which point the solution turns colourless. The reaction mixture was stirred at rt for 15 mins and then heated at 80 °C for 3 hrs. The reaction was allowed to cool, quenched with sat. *aq.* $NaHCO_3$ (100 mL) and extracted with EtOAc (3 x 100 mL). The combined organic layers were washed with brine (100 mL), dried over Mg_2SO_4 and concentrated under vacuum. The residue was purified by column chromatography [CH_2Cl_2:EtOAc, 100:0 to 20:80] to elute the N1-regioisomer **70** as a yellow foam (1.6 g, 35%), followed by the N7-isomer **69** as a yellow foam (2.4 g, 51%).

(**69**): R_f 0.32 [$CHCl_3$: EtOAc, 1:1]; ^{1}H NMR (400 MHz, DMSO-d_6) δ: 8.14 (1H, s, 2-C*H*); 7.92–7.89 (6H, m, Ar-C*H*); 7.67–7.63 (3H, m, Ar-C*H*); 7.49–7.43 (6H, m, Ar-C*H*); 7.07 (2H, br s, 4-N*H*$_2$); 6.58 (1H, dd, *J*= 6.6, 3.3 Hz, 1'-C*H*); 6.41–6.38 (2H, m, 2'-C*H*,3'-C*H*); 4.89–4.85 (1H, m, 4'-C*H*); 4.78 (1H, dd, *J*= 12.4, 3.3 Hz, 5'-C*H*); 4.59 (1H, dd, *J*= 12.4, 4.3 Hz, 5'-C*H*); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 170.4 (*C*=O), 165.3 (*C*=O), 164.7 (*C*=O), 156.0 (*C*), 153.8 (*C*H-2'), 150.1 (*C*), 134.0 (*C*H-Ar), 133.9 (*C*H-Ar), 133.5 (*C*H-Ar), 129.4 (*C*H-Ar), 129.3 (*C*H-Ar), 129.2 (*C*H-Ar), 128.8 (*C*H-Ar), 128.7 (*C*H-Ar), 128.7 (*C*H-Ar), 128.5 (*C*H-Ar), 128.4 (*C*H-Ar), 121.4 (*C*), 114.0 (*C*N), 101.8 (*C*), 89.0 (*C*H-2'/3'), 88.0 (*C*), 78.9 (*C*H-4'), 73.0 (*C*H-1'), 69.6 (*C*H-2'/3'), 62.4 (*C*H-5'); *m/z* (ESI$^+$) 682

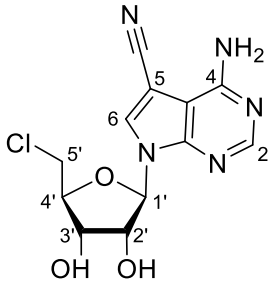
([^{79Br}M+H]$^+$ 48%) 684 ([^{81Br}M+H]$^+$ 48%) 704 ([^{79Br}M+Na]$^+$ 100%) 706 ([^{81Br}M+Na]$^+$ 90%). The data are in good agreement with the literature values.[87]

(2*R*,3*R*,4*R*,5*R*)-2-(4-Amino-6-bromo-5-cyano-1*H*-pyrrolo[2,3-*d*]pyrimidin-1-yl)-5-((benzoyloxy)methyl)tetrahydrofuran-3,4-diyl dibenzoate (**70**): R_f 0.37 [$CHCl_3$: EtOAc, 1:1]; ^{1}H NMR (400 MHz, DMSO-d_6) δ: 8.67 (1H, s, 2-C*H*); 7.95–7.89 (6H, m, Ar-C*H*); 7.67–7.63 (3H, m, Ar-C*H*); 7.49–7.43 (6H, m, Ar-C*H*); 6.61 (1H, d, *J*= 3.5 Hz, 1'-C*H*); 6.33–6.28 (2H, m, 2'-C*H*, 3'-C*H*); 4.91–4.88 (1H, m, 4'-CH_2); 4.80 (1H, dd, *J*= 12.3, 4.1 Hz, 5'-C*H*); 4.71 (1H, dd, *J*= 12.3, 5.4 Hz, 5'-C*H*). ^{13}C NMR (125 MHz, DMSO-d_6) δ: 165.4 (*C*=O), 164.6 (*C*=O,C=O), 156.0 (*C*), 144.5 (*C*H-1'), 144.5 (*C*), 134.0 (*C*H-Ar), 133.9 (*C*H-Ar), 133.5 (*C*H-Ar), 132.2 (*C*), 129.4 (*C*H-Ar,*C*H-Ar), 129.3 (*C*H-Ar,*C*H-Ar), 129.2 (*C*H-Ar,*C*H-Ar), 129.1 (*C*H-Ar), 128.7 (*C*H-Ar,*C*H-Ar,*C*H-Ar), 128.7 (*C*H-Ar,*C*H-Ar), 128.5 (*C*H-Ar), 128.4 (*C*H-Ar), 115.7 (*C*N), 104.3 (*C*), 92.2 (*C*H-1'), 83.8 (*C*), 79.6 (*C*H-4'), 73.6 (*C*H-2'/3'), 70.4 (*C*H-2'/3'), 63.3 (*C*H-5'); *m/z* (ESI$^+$) 682 ([^{79Br}M+H]$^+$ 20%) 684 ([^{81Br}M+H]$^+$ 20%) 704 ([^{79Br}M+Na]$^+$ 100%) 706 ([^{81Br}M+Na]$^+$ 100%). The data are in good agreement with the literature values.[87]

Method 2: To a stirred suspension of compound **68** (1.0 g, 4.2 mmol) in anhydrous hexamethyldisilazane (HMDS) (21 mL, 0.2 M) was added TMSCl (213 µL, 1.7 mmol). The reaction mixture was refluxed for 18 hrs, allowed to cool and the solvent was concentrated under vacuum and kept under an argon atmosphere. The dark brown residue was then suspended in anhydrous MeCN (42 mL, 0.10 M) and cooled to 0 °C. To this suspension was added 1-*O*-acetyl-2,3,5-tri-*O*-benzoyl-β-D-ribofuranose (2.3 g, 4.6 mmol) and TMSOTf (2.3 mL, 13 mmol). The reaction mixture was stirred at 0 °C for 30 minutes and then refluxed for 3 hrs, allowed to cool, quenched with sat. *aq.* $NaHCO_3$ (100 mL) and extracted with EtOAc (3 x 100 mL). The combined organic layers were washed with brine (100 mL), dried over Mg_2SO_4 and concentrated under vacuum. The residue was purified by column chromatography [CH_2Cl_2:EtOAc, 9:1] to selectively yield compound **69** as a pale yellow foam (1.5 g, 52%). The data are in good agreement with literature values and those obtained from method 1.

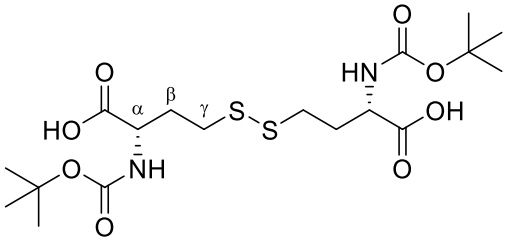


(2*R*,3*R*,4*R*,5*R*)-2-(4-Amino-5-cyano-7*H*-pyrrolo[2,3-*d*]pyrimidin-7-yl)-5-((benzoyloxy)methyl)tetrahydrofuran-3,4-diyl dibenzoate (73). 5 mol% Pd/C (232 mg, 0.34 mmol) and ammonium formate (1.1 g, 17 mmol) were added to a stirred solution of compound **69** (1.2 g, 1.7 mmol) in EtOH (43 ml, 0.040 M). The reaction mixture was refluxed for 5 hrs and filtered through a celite plug, washed thoroughly with hot EtOH (total 150 mL) and the filtrate was concentrated under vacuum. The residue was purified by flash column chromatography [Hexane:EtOAc, 1:1] to yield compound **73** as a white solid (0.86 g, 84%): R_f 0.2 [Hexane:EtOAc, 2:3]; ^{1}H NMR (400 MHz, CD_3OD) δ: 8.10 (2H, s, 2-C*H*,6-C*H*); 8.06–8.05 (2H, m, Ar-C*H*); 8.00–7.99 (2H, m, Ar-C*H*); 7.91–7.89 (2H, m, Ar-C*H*); 7.64–7.56 (3H, m, Ar-C*H*); 7.49–7.36 (6H, m, Ar-C*H*); 6.57 (1H, d, *J*= 5.1 Hz, 1'-C*H*); 6.32 (1H, dd, *J*= 5.5, 5.1 Hz, 2'-C*H*); 6.16 (1H, dd, *J*= 5.5, 5.4 Hz, 3'-C*H*); 4.89–4.84 (2H, m, 4'-C*H*, 5'-C*H*); 4.71 (1H, dd, *J*= 11.5, 3.3 Hz, 5'-C*H*); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 167.7 (*C*), 166.9 (*C*), 165.6 (*C*), 158.7 (*C*), 155.0 (*C*H-2/6), 151.6 (*C*), 135.1 (*C*H-Ar), 135.0 (*C*H-Ar), 134.8 (*C*H-Ar), 133.5 (*C*H-2/6), 131.0 (*C*), 131.0 (*C*H-Ar,*C*H-Ar), 131.0 (*C*H-Ar,*C*H-Ar), 130.9 (*C*H-Ar,*C*H-Ar), 130.4 (*C*), 130.1 (*C*), 129.9 (*C*H-Ar,*C*H-Ar), 129.8 (*C*H-Ar,*C*H-Ar), 129.8 (*C*H-Ar,*C*H-Ar), 115.8 (*C*N), 103.6 (*C*), 89.6 (*C*H-1'), 86.3 (*C*), 81.8 (*C*H-4'), 75.7 (*C*H-2'), 72.8 (*C*H-3'), 64.6 (*C*H-5'); *m/*z (ESI)$^+$ 604 ([M+H]$^+$ 20%) 648 ([M+2Na-H]$^+$ 100%). The data are in good agreement with the literature values.[86]

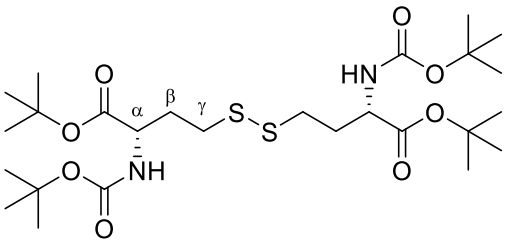


4-Amino-7-((2*R*,3*R*,4*S*,5*S*)-5-(chloromethyl)-3,4-dihydroxytetrahydrofuran-2-yl)-7*H*-pyrrolo[2,3-*d*]pyrimidine-5-carbonitrile (74). Compound **29** (120 mg, 0.41 mmol) was dissolved in *N*,*N*′-dimethylpropylene urea (DMPU) (2.1 mL, 0.20 M) and cooled to 0 °C, then $SOCl_2$ (150 µL, 2.1 mmol) was added dropwise to this solution. The reaction

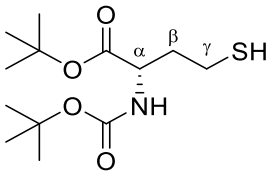
was stirred at 0 °C for 5 hrs then quenched with sat. *aq.* $NaHCO_3$ (20 mL) and extracted with EtOAc (2 x 20 mL). The organic fractions were washed with brine (40 mL), dried over Mg_2SO_4 and concentrated under vacuum. The crude residue was purified by column chromatography [$CHCl_3$:MeOH, 9:1] to afford **74** as an off white solid (85 mg, 67%): R_f 0.34 [$CHCl_3$:MeOH, 9:1]; $[\alpha]_D^{25}$ +50 ° (c 0.5, MeOH); mp 118–120 °C; ^{1}H NMR (500 MHz, DMSO-d_6) δ: 8.42 (1H, s, 6-C*H*); 8.24 (1H, s, 2-C*H*); 6.92 (2H, br, s, 4-NH_2); 6.12 (1H, d, *J*= 5.7 Hz, 1'-C*H*); 5.62 (1H, d, *J*= 5.5 Hz, 2-OH); 5.48 (1H, d, *J*= 4.7 Hz, 3-OH); 4.51 (1H, ddd, *J*= 5.7, 5.6, 5.5 Hz, 2'-C*H*); 4.14–4.06 (2H, m, 3'-C*H*, 4'-C*H*); 3.94–3.91 (1H, m, 5'-C*H*); 3.86–3.83 (1H, m, 5'-C*H*); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 157.0 (*C*), 153.7 (*C*H-2), 150.5 (*C*), 132.3 (*C*H-6), 115.2 (*C*N), 101.1 (*C*), 87.3 (*C*H-1'), 83.6 (*C*), 83.6 (*C*H-4'), 73.2 (*C*H-2'), 71.1 (*C*H-3'), 44.7 (*C*H-5'); HRMS *m/z* (ESI^+) [Found $(M+H)^+$ 310.0705; $C_{12}H_{12}ClN_5O_3H$ requires $(M+H)^+$ 310.0701]. The data are in good agreement with the literature values.[101]



(2*S*,2'*S*)-4,4'-Disulphanediyl*bis*(2-((*tert*-butoxycarbonyl)amino)butanoic acid) (76). 10% *aq.* Na_2CO_3 (19 mL, 0.1 M) was added to a suspension of L-homocystine (**75**) (0.50 g, 1.9 mmol) in dioxane (15 mL, 0.10 M). The solution was cooled to 0 °C and Boc_2O (0.90 g, 4.1 mmol) was added as one portion. The reaction mixture was stirred at rt overnight, the solution was adjusted to pH 4 using 10% *aq.* citric acid (10 mL) and the organic solvent was removed under vacuum. The remaining aqueous solution was extracted with EtOAc (30 mL) and the combined organic layers were washed with brine, dried over Na_2SO_4 and concentrated under vacuum. Compound **76** was isolated as a waxy white solid (0.87 g, 99%): ^{1}H NMR (400 MHz, CD_3OD) δ: 4.27–4.24 (2H, m, Cys-α-C*H*); 2.84–2.73 (4H, m, Cys-γ-C*H*); 2.31–2.22 (2H, m, Cys-β-C*H*); 2.07–1.97 (2H, m, β); 1.47 (18H, s, CH_3); ^{13}C NMR (125 MHz, CD_3OD) δ: 175.7 (C=O), 158.2 (C=O), 80.6 (*C*), 53.7 (*C*), 43.9 (Cys-α-*C*H), 35.7 (Cys-β-*C*H), 32.5 (Cys-γ-*C*H), 28.8 (*C*H_3); *m/z* (ESI^-) 467 ([M-H]$^-$ 100%) 934 ([2M-H]$^-$ 100%). The data are in good agreement with the literature values.[102]

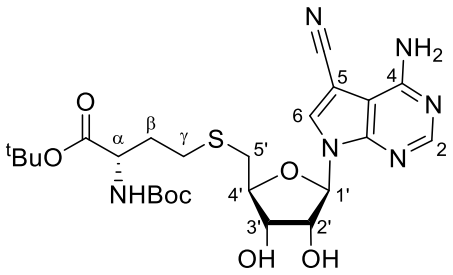


Di-*tert*-butyl-4,4'-disulphanediyl(2*S*,2'*S*)-*bis*(2-((*tert*-butoxycarbonyl)amino)-butanoate) (77). *tert*-Butyl-2,2,2-trichloroacetimidate (0.96 mL, 5.3 mmol) was added to a suspension of compound **76** (0.50 g, 1.1 mmol) in anhydrous DCM (5.3 mL, 0.2 M) and the reaction mixture was allowed to stir at rt overnight. The organic solvent was removed under vacuum and the crude residue purified by column chromatography [Hexane:EtOAc, 3:1]. Compound **77** was isolated as a waxy white solid (0.45 g, 78%): R_f 0.29 [EtOAc: Hexane: 1:7]; ^{1}H NMR (400 MHz, DMSO-d_6) δ: 7.19–7.17 (2H, br s, N*H*); 3.95–3.89 (2H, m, Cys-α-C*H*); 2.73–2.70 (4H, m, Cys-γ-C*H*); 2.04–1.97 (2H, m, Cys-β-C*H*); 1.94–1.85 (2H, m, Cys-β-C*H*); 1.40 (18H, s, *CH₃*); 1.38 (18H, s, *CH₃*); ^{13}C NMR (100 MHz, $CDCl_3$) δ: 171.2 (*C*=O), 155.3 (*C*=O), 82.3 (*C*), 79.8 (*C*), 53.2 (Cys-α-*C*H), 34.5 (Cys-γ-*C*H), 32.9 (Cys-β-*C*H), 28.3 (CH_3), 28.0 (CH_3); *m/z* (ESI^+) 603 ($[M+Na]^+$ 100%).[102]



***Tert*-butyl-(*tert*-butoxycarbonyl)-L-homocysteinate (78).** H_2O (0.37 mL, 1.4 M) and Bu_3P (142 µL, 5.3 mmol) were added to a solution of compound **77** (0.30 g, 0.52 mmol) in anhydrous DMF (4.3 mL, 0.12 M). The reaction mixture was stirred at rt overnight, quenched with excess water (50 mL) and extracted with EtOAc (3 x 50 mL). The combined organic layers were washed with brine (50 mL), dried over Na_2SO_4 and concentrated under vacuum. The crude residue was purified by column chromatography [Hexane:EtOAc, 20:1] to yield compound **78** as a waxy white solid (0.19 g, 62%): R_f 0.46 [Hexane:EtOAc, 7:1]; $[\alpha]\frac{25}{D} - 32\,°\,(c\ 0.5, \mathrm{MeOH})$; ^{1}H NMR (400 MHz, $CDCl_3$) δ: 5.06 (1H, br s, N*H*); 4.34–4.29 (1H, m, Cys-α-C*H*); 2.66–2.50 (2H, m, Cys-γ-C*H*); 2.13–2.05 (1H, m, Cys-β-C*H*); 1.95–1.85 (1H, m, Cys-β-C*H*); 1.60 (1H, m, S*H*); 1.48 (9H, s, CH_3); 1.45 (9H, s, CH_3); ^{13}C NMR (100 MHz, $CDCl_3$) δ: 171.4 (*C*=O), 155.4 (*C*=O), 82.2 (*C*), 79.9 (*C*), 53.0 (Cys-α-*C*H), 37.6 (Cys-β-*C*H), 28.3 (CH_3), 28.30

(CH_3), 20.8 (Cys-γ-*C*H); *m/z* (ESI$^+$) 314 ([M+Na]$^+$ 100%). The data are in good agreement with the literature values.[102]



***tert*-Butyl-S-(((2*S*,3*S*,4*R*,5*R*)-5-(4-amino-5-cyano-7*H*-pyrrolo[2,3-*d*]pyrimidin-7-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)-*N*-(*tert*-butoxycarbonyl)-L-homocysteinate (79).** Compound **78** (81 mg, 0.28 mmol) was added to a solution of compound **74** (64 mg, 0.21 mmol), KI (34 mg, 0.21 mmol) and K_2CO_3 (29 mg, 0.21 mmol) in DMF (1.4 mL, 0.15 M). The reaction was stirred at 80 °C for 3hrs, allowed to cool and diluted with EtOAc (20 mL). The organic layer was washed with brine (20 mL), dried over Mg_2SO_4 and concentrated under vacuum. The crude residue was purified by column chromatography [$CHCl_3$:MeOH, 9:1] to yield **79** as an off white solid (87 mg, 75%): R_f 0.21 [$CHCl_3$:MeOH, 9:1]; γ_{max} (neat)/cm^{-1} 3350 (br), 2976 (s), 2226 (s), 1702 (s), 1596 (s), 1367 (s), 1152 (s); ^{1}H NMR (CD_3OD) δ: 8.23 (1H, s, 2/6-C*H*); 8.18 (1H, s, 2/6-C*H*); 6.17 (1H, d, *J*= 4.9 Hz, 1'-C*H*); 4.52 (1H, dd, *J*= 5.0, 4.9 Hz, 2'-C*H*); 4.25 (1H, dd, *J*= 5.2, 5.0 Hz, 3'-C*H*); 4.18 (1H, ddd, *J*= 5.4, 5.4, 5.2 Hz, 4'-C*H*); 4.13–4.09 (1H, m, Cys-α-C*H*); 2.99–2.90 (2H, m, 5'-C*H*$_2$); 2.71 – 2.66 (1H, m, Cys-γ-C*H*); 2.62–2.57 (1H, m, Cys-γ-C*H*); 2.04–1.97 (1H, m, Cys-β-C*H*); 1.89–1.82 (1H, m, Cys-β-C*H*); 1.43 (18H, s, C*H*$_3$); ^{13}C NMR (125 MHz, CD_3OD) δ: 173.4 (*C*=O), 158.6 (*C*), 158.1 (*C*=O), 154.7 (*C*H-2/6), 151.6 (*C*), 132.9 (*C*H-2/6), 116.0 (*C*N), 103.4 (*C*), 90.3 (*C*H-1'), 85.7 (*C*), 85.5 (*C*H-4'), 82.7 (*C*), 80.6 (*C*), 75.6 (*C*H-2'), 73.9 (*C*H-3'), 54.7 (Cys-α-*C*H), 35.2 (*C*H-5'), 32.9 (Cys-β-*C*H$_2$), 30.2 (Cys-γ-*C*H$_2$), 28.7 (*C*H$_3$), 28.2 (*C*H$_3$); HRMS *m/z* (ESI$^+$) [Found (M+H)$^+$ 565.2428; $C_{25}H_{36}N_6O_7SH$ requires (M+H)$^+$ 565.2439].

PART 2

Design, Synthesis and Evaluation of Novel Small Molecule Inhibitors of the Ubiquitination Facilitator Keap1

CHAPTER 6
Introduction

The cells of the human body are continuously exposed to stresses caused by oxidative species (reactive oxygen species (ROS), reactive nitrogen species (RNS)) and electrophilic species from endogenous and exogenous sources. Exogenous sources include environmental and chemical carcinogens and radiation. Endogenous oxidative species can be produced as by-products from homeostatic processes associated with cell survival and function. For example aerobic respiration of mitochondria, inflammatory responses that produce oxidative conditions or molecules involved in cellular signalling.

Sustained oxidative stress can cause damage to biological macromolecules including DNA, lipids and proteins and has been associated with ageing as well as a range of diseases including cancer and neurodegenerative diseases such as Parkinson's or Alzheimer's disease.[103] To combat the damage caused by oxidative stress, cells have antioxidant and cytoprotective mechanisms which defend against and neutralise potential harm caused by oxidative and electrophilic species. Two defence mechanisms include antioxidants and antioxidant response enzymes.

Antioxidants are generally small molecules consumed as part of the diet or produced from endogenous sources. They are redox-active, short-lived and consumed or modified during the process, meaning they need to be constantly replenished. Examples include: glutathione (GSH), vitamin C (ascorbate), E (tocopherols) and K, lipoic acid and ubiquinol and other polyphenolic compounds.[104]

Antioxidant response enzymes are expressed either constitutively or are induced to maintain a healthy homeostasis. Compared to direct antioxidants they are a more efficient mechanism of detoxification as they have relatively long half-lives, are catalytic and thus not consumed in the process. Furthermore they catalyse a wide variety of chemical detoxification reactions including regeneration of antioxidants. Examples, which also include phase II metabolism enzymes, are: nicotinamide amine dinucleotide phosphate NAD(P)H, NAD(P)H quinone oxidoreductase 1 (NQO1), heme-oxygenase-1 (HO-1), γ-glutamyl-cysteine synthase, glutathione *S*-transferase (GST) and glutamate-cysteine ligase (GCL). The same transcription pathway controls the expression of many of these proteins. It involves the Kelch-like ECH-associated protein 1 (Keap1),

nuclear factor erythroid 2-related factor 2 (Nrf2) and antioxidant response elements (ARE) in the promoter regions of certain genes, i.e. the Keap1-Nrf2-ARE pathway.

6.1 Keap1-Nrf2-ARE pathway

As a master regulator of the antioxidant response, the Keap1-Nrf2-ARE system protects cellular proteins and DNA from oxidative damage caused by ROS and electrophiles.[105] Under conditions of stress Nrf2 activates the transcription of antioxidant response enzymes. Nrf2 is encoded by the gene *NFE2L2* and belongs to the Cap'n'Collar (CNC) subfamily of basic leucine zipper (bZIP) transcription factors. The Nrf2 protein in humans is 605 amino acids long and contains six highly conserved regions known as Nrf2-ECH homology (Neh) domains (Figure 6.1).[106]

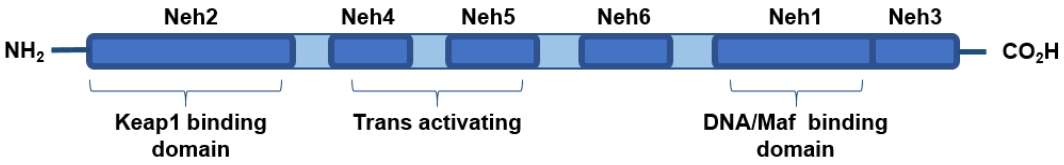


Figure 6.1 The organisation and domain structure of Nrf2. Reproduced from Abed *et al*.[107]

Under basal conditions Nrf2 has a very short half-life of ~20 minutes which is mediated by an actin-binding protein, Keap1.[108] Keap1 is a 67.9 kDa protein with 625 amino acid residues of which 27 are cysteine residues. It consists of five distinct domains: (i) the N-terminal region (NTR) (ii) the broad complex, tram-track and bric-a-brac (BTB) domain (iii) the intervening region (IVR) (iv) the Kelch domain and (v) the C-terminal region (CTR) (Figure 6.2).[109]

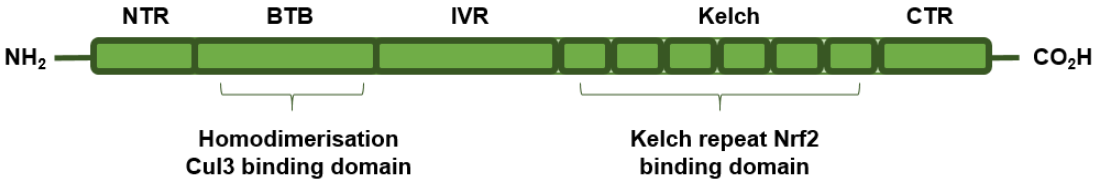


Figure 6.2 The organisation and domain structure of Keap1. Reproduced from Abed *et al.*[107]

Keap1 controls the steady state level of Nrf2 based on cellular redox conditions, through a series of redox sensitive cysteine residues distributed across the protein which are vulnerable to oxidation or covalent modification by electrophiles. It appears different cysteines are targeted by different reactive chemicals which has led to the concept of a 'cysteine code'.[110, 111] Two cysteine residues located in the IVR, Cys273 and Cys288, have both been shown to be essential for Keap1 control over Nrf2 under both basal and stress conditions.[112] Cys151, located in the BTB domain, is important mainly during stress conditions.[113] Whereas Cys226, Cys613 and Cys434 are required

for sensing specific toxins (for example heavy metals Cd^{2+}, Se^{2+} or RNS) and are located in the IVR, CTR and Kelch domain respectively.[114] Under basal conditions Keap1 sequesters Nrf2 into the cytoplasm, targeting it for proteasomal degradation and ubiquitination via a Cullin3-Rbx1 E3 ubiquitin ligase (Cul3-E3-ligase), bound to Keap1 via the BTB domain. It has been shown that the BTB domain is also responsible for forming a Keap1 homodimer which is essential for Keap1-Nrf2 association.[115] The BTB domain is an evolutionary conserved domain also found in actin-binding proteins and zinc finger transcription factors. The Kelch domain of Keap1 consists of six repeating Kelch motifs (KR1-KR6) that form a six-bladed β-propeller structure and binds directly to Nrf2.[116] The Keap1 Kelch domain binds to Nrf2 via two motifs in the Neh2 domain: the high affinity 'ETGE' motif (K_D ~5.3 nM, residues 78-82) and the lower affinity 'DLG' motif (K_D ~1000 μM, residues 24-31).[117] Each Nrf2 motif binds to a separate Kelch domain in the Keap1 homodimer in similar orientations (Figure 6.3).

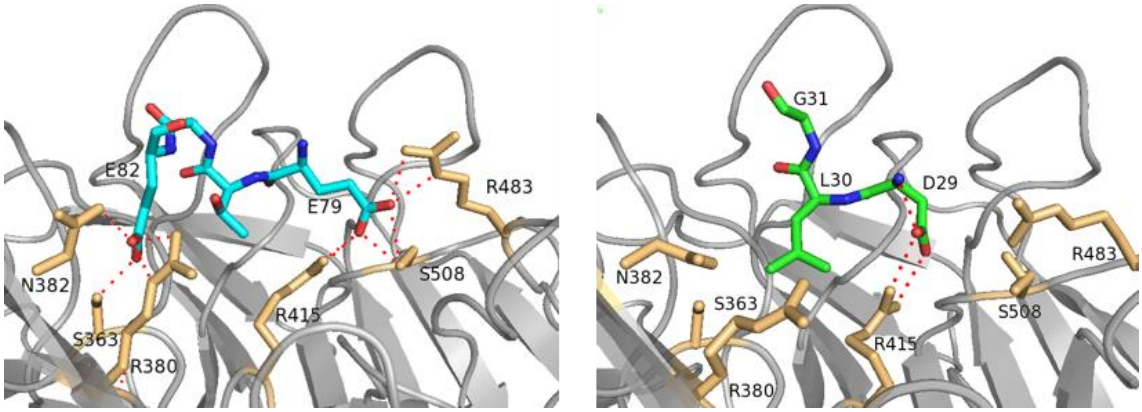


Figure 6.3 Binding orientation of (left) Nrf2 ETGE motif (PDB code: 2FLU, 1.6 Å) and (right) Nrf2 DLG motif (PDB code: 3WN7, 1.6 Å). Side chains of the Keap1 backbone involved in key interactions are shown in yellow. Key H-bond interactions have been highlighted with red dashes. Generated using PyMol.

A 'hinge and 'latch' model has been proposed to explain the mechanism of control of Nrf2 levels (Figure 6.4).[118] The higher affinity ETGE motif acts as the 'hinge', effectively 'fixing' Keap1 to Nrf2 through strong H-bond interactions, namely between Nrf2 Glu79 and Keap1 residues Arg415, Arg483 and Ser508 and Nrf2 Glu82 with Keap1 residues Ser363, Asn382 and Arg380.[119] The 100-fold lower affinity DLG motif acts a 'latch', binding in a similar orientation in the Keap1 Kelch domain to the ETGE motif. Under basal conditions the DLG motif is bound to Keap1, locking Nrf2 into a 'closed' orientation which positions the seven lysine residues located between the two motifs for ubiquitination via the Keap1-Cul3-ligase complex. This leads to subsequent proteasomal degradation of Nrf2.

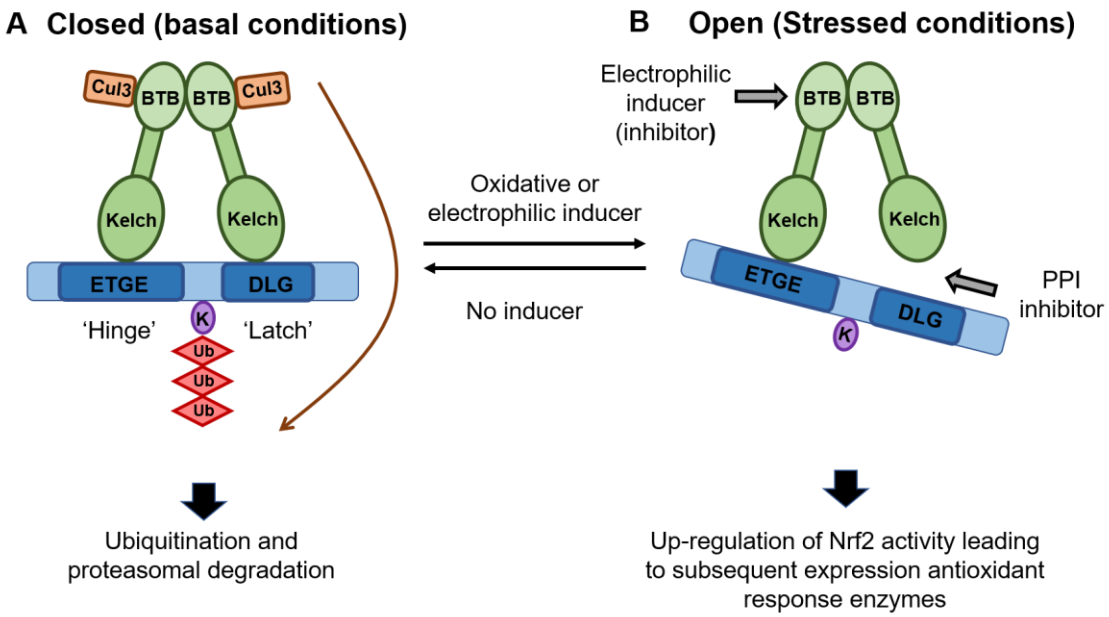


Figure 6.4 'Hinge and 'latch model' for the binding interaction between Keap1 and Nrf2. **A** The closed form (basal conditions) shows both the ETGE and DLG Nrf2 binding motifs from the Nrf2 Neh2 domain bound to both Keap1-Kelch domain in the Keap1 homodimer. Keap1 is bound to Cul3-E3-ligase (Cul3) which activates ubiquitination (Ub) of lysine residues (K) in Nrf2 leading to proteasomal degradation of Nrf2 **B** The open form (stressed conditions) shows only the ETGE Nrf2 motif bound to one of the Keap1-Kelch domains in the keap1 homodimer. Two arrows identify positions where the two forms of inhibitors, electrophilic inducers (inhibitors) and small molecule or protein PPI inhibitors, disrupt Keap1-Nrf2 binding. Electrophilic inducers disrupt the Keap1-Cul3-E3-liganse blocking ubiquitination whereas small molecule/protein PPI inhibitors hinder the ETGE/DLG-Kelch domain PPI interactions.

Under oxidative stress, cysteine residues in Keap1 become oxidised, resulting in conformational changes which disrupt the 'latch' mechanism, ultimately hindering the ubiquitination process, stabilising Nrf2 and increasing its half-life up to 200 minutes. Nrf2 subsequently accumulates and translocates to the nucleus where it forms a heterodimer with small Maf proteins and binds to DNA via its Neh1 domain. This activates transcription of genes containing antioxidant response elements (ARE) in their promoter regions. The Neh1 domain corresponds to a basic leucine zipper domain (bZip) and is found in many DNA binding eukaryotic proteins.

An alternative 'Keap1-Cul3 dissociation' model has also been suggested as a mechanism to control Nrf2 levels. In this model, it is proposed that, under induced conditions, covalent modification of cysteine residues in the BTB domain leads to disruption of the Keap1-Cul3 interaction. This hampers Keap1-Cul3-E3 ligase activity resulting in stabilisation of Nrf2 levels. Cys151 has been shown to be particularly important to achieve Nrf2 stabilisation, suggesting Cys151 is a principal factor for sensing oxidative conditions.

6.1.1 Inhibition

The Keap1-Nrf2-ARE pathway has been implicated in a wide range of disorders including cancer,[120] neurological disorders including Parkinson's disease (PD), Alzheimer's disease (AD)[121] and multiple sclerosis,[122] airway disorders,[123] cardiovascular diseases,[124] diabetes,[125] inflammatory bowel disease (IBD) and autoimmune diseases[107] as well as ageing,[126] It is therefore an important therapeutic target offering novel pharmacological opportunities.

Enhancement of Nrf2 activity has mainly been achieved through disruption of the interaction between Nrf2 and the Keap1 ubiquitination factor; a principle regulator in the degradation mechanism of Nrf2 and thus Nrf2 activity in cells.[110] There are two main methods which have been employed to disrupt Keap1: (i) electrophilic inducers and (ii) direct Nrf2 inducers.

6.1.1.1 Electrophilic inducers

Redox sensitive cysteine residues in the BTB domain and IVR region of Keap1 sense electrophilic and redox stress within cells. Covalent modification of these cysteine residues by electrophilic inducers, leads to obstruction of Nrf2 ubiquitination through apparent disruption of the Keap1-Cul3-E3-ligase complex, resulting in a non-functional form of the Keap1-Nrf2 complex (Figure 6.4).[127] Natural, semi-synthetic and synthetic electrophilic inducers have been reported and are discussed below (Figure 6.5).

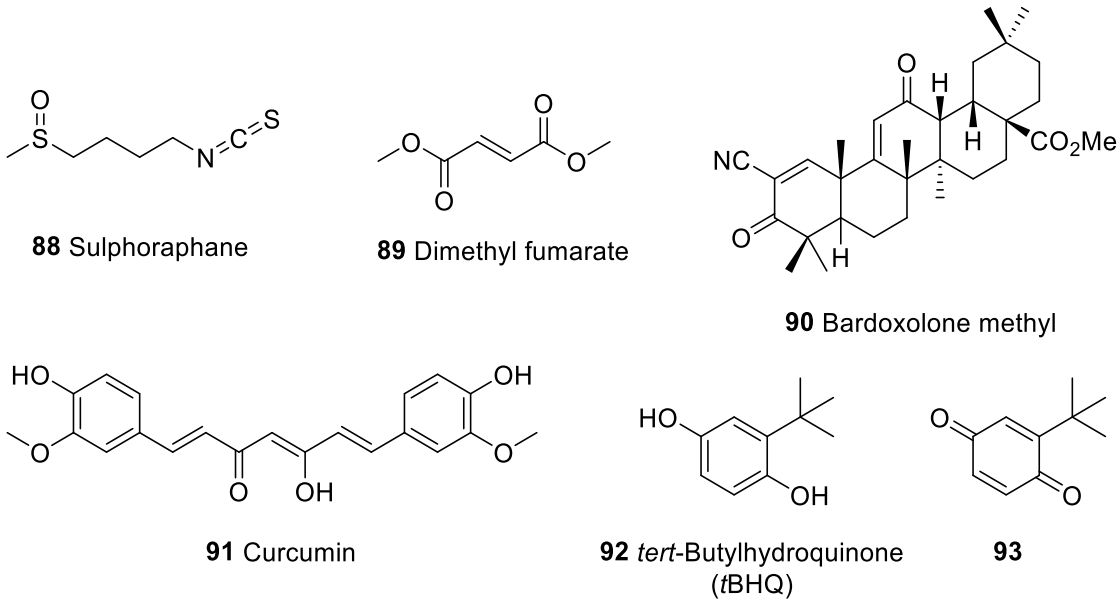


Figure 6.5 Summary of natural, semi-synthetic and synthetic electrophilic inducers of Nrf2 activity.

The isothiocyanate natural product sulphoraphane **88**, derived from glucoraphanin which is found in cruciferous vegetables, for example Brussels sprouts and broccoli, has been shown to induce Nrf2 accumulation and transcription of its associated genes at nanomolar concentrations.[128] It has been evaluated in animal models and clinical trials for several diseases including, but not limited to, cancer, diabetes and neurodegenerative conditions.

There are several examples (vide infra) of Michael acceptors containing α,β-unsaturated carbonyl functionalities which undergo 1,4-addition with reactive Keap1 cysteine thiols (Figure 6.6). Examples include, Dimethyl fumarate **89** (Tecfidera), which was recently approved by the FDA for treatment of multiple sclerosis[129] and the semi-synthetic triterpene bardoxolone methyl **90** which has been evaluated in clinical trials for the treatment of chronic kidney disease, pulmonary arterial hypertension and cancer, although the trials were terminated due to an associated increase in cardiovascular events [NCT01351675].[130, 131] Bardoxolone methyl **90** is one of the most potent Nrf2 inducers to date.[132] The polyphenolic *bis*-Michael acceptor curcumin **91** has displayed promising results in clinical trials for a variety of conditions despite its low bioavailability, diverse biological activities, and relatively low Nrf2-inducing potential.[133] *tert*-Butylhydroquinone **92** is an example of a *pro*-Michael acceptor; it is bio-activated to form *tert*-butyl-*p*-quinone **93** which reacts with cysteine thiols in Keap1. It has been shown to specifically interact with Cys151, as impairment of Nrf2-inducing activity was observed in cells with mutant C151S Keap1.[113]

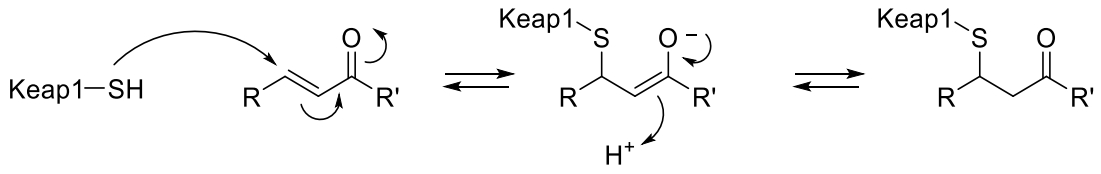


Figure 6.6 Mechanism of 1,4-addition of reactive cysteine thiols in Keap1 with an α,β-unsaturated carbonyl functionality.

The main limitation of electrophilic inducers are the associated off-target effects which occur due to their reactive nature. They have the potential to react with other proteins that contain accessible cysteine residues or with abundant cellular nucleophiles for example glutathione (GSH). An alternative, potentially more selective approach to enhance Nrf2 activity, is through direct disruption of the Keap1-Nrf2 interaction at the protein-protein interface.

6.1.1.2 Direct Nrf2 inducers

Inhibition of protein-protein interactions (PPI) with small molecules is generally regarded as a challenge due to the nature of the interfaces. Protein-protein interfaces are generally flat, large and often hydrophobic.[134] This contrasts quite sharply with the deep cavities which are normally targeted by small molecules in other targets such as enzymes. The Keap1-Nrf2 interface is relatively small in comparison with typical PPIs which suggests it may be tractable to inhibition by small molecule inhibitors. It involves the well-defined bowl-like C-terminal β-propeller Kelch domain from each subunit of Keap1 which is divided into 5 subpockets, P1-P5 and two sequences identified in the Neh2 domain of Nrf2, the DLG and ETGE motifs described earlier (Figure 6.3 and Figure 6.7).

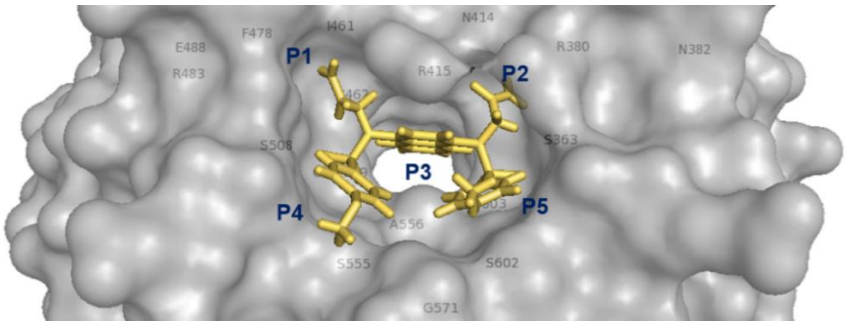


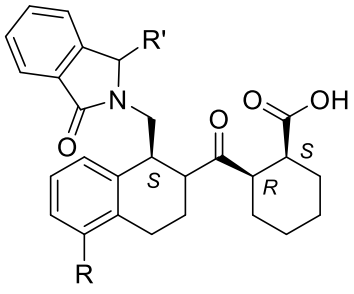
Figure 6.7 Subpockets P1-P5 of the Keap1-Nrf2 PPI interface (PDB code: 4XMB, 2.4 Å). Important residues for each pocket have been highlighted by their single letter code and associated number. The small molecule inhibitor **94** present in the co-crystal structure is shown in a stick representation. Image generated using PyMol.

Both the well-defined Keap1 Kelch domain binding pocket and small interacting regions in the Neh2 domain of Nrf2, suggest the discovery of potent PPI inhibitors is possible. Indeed, both peptide and small molecule inhibitors, identified mainly through HTS as well as fragment-based drug discovery (FBDD) and structure-based drug design (SBDD), have been shown to effectively inhibit the Keap1-Nrf2 PPI and are discussed below.

Lo *et al* demonstrated nanomolar binding affinity to Keap1 could be achieved through short peptide sequences (10 – 16 amino acid residues) which were based on the Nrf2 'ETGE' motif (the central 'DEETGE' motif was retained for each peptide).[135] Hancock *et al*. demonstrated shorter sequences were also able to effectively target the Keap1-Nrf2 PPI interaction.[136] They identified two residues, Glu78 and Phe83, in the central motif of the native peptide **95** ('D**E**ETGE**F**', IC_{50} = 5.4 μM) which were not critical to binding. Through substitution with proline and leucine residues, a large increase in affinity over the native peptide sequence was gained resulting in a high affinity 7-mer peptide **96**,

Ac-DPETGEL-OH (IC_{50} = 115 nM). Through computational methods, Lu *et al* identified a related 11-mer cyclic peptide inhibitor **97**, Ac-c[CLDPETGEYLC]-OH, with a similar sequence but higher affinity (IC_{50} = 9.5 nM).[137] Although potent, the cell permeability of these peptide inhibitor sequences was poor due to their high polarity. Efforts to increase cell entry were made through conjugation to cell permeable transactivating transcriptional activator (TAT) peptide sequences and fatty acids.[138,139] Although an up-regulation in Nrf2 was observed for both approaches, the activity remained in the mid-micromolar range.

Tetrahydroisoquinoline LH601A (**98**) was identified as the first small molecule Keap1 inhibitor (Figure 6.8).[140] LH601A (**98**) was identified from a HTS of the NIH Molecular Libraries Probe Production Centres Network (MLPCN). Of the 8 stereoisomers only one stereoisomer (*SRS*), assigned through crystal structures and confirmed through stereospecific synthesis, was found to be optimal for Keap1 binding. It was found to bind to the Keap1 Kelch domain with an IC_{50} of 3 μM and showed up-regulation of antioxidant response enzymes NQO1, HO-1 and TRX1 at mid-high micromolar concentrations in HEK293 cells, although had limited effect on glutathione-related genes.[141] A crystal structure of LH601A (**98**) with Keap1 elucidated the binding mode showing the quinoline located in the central P3-subpocket (PDB code: 4L7B), subsequent optimisation identified **99** which is a 3-fold more potent inhibitor than LH601A (**98**). It was suggested the increase in binding affinity was due to the methyl group acting as a 'lipophilic plug', providing a good shape and fit toward the pocket in the Keap1 Kelch domain.[142]



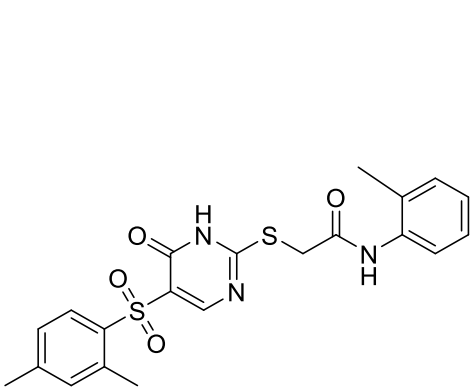
98: R = H, R = O
99: R = Me, R = H

Figure 6.8 Structure of the direct small molecule Keap1 inhibitors, LH601A (**98**) and **99.**

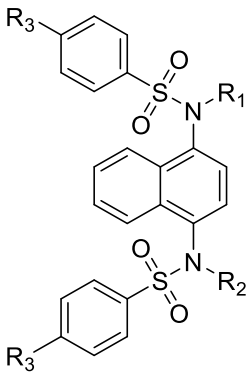
Another HTS identified compounds **100** and **101** using the Evotec Lead Discovery Library which was supplemented with ~2,000 compounds previously identified from virtual screening (Figure 6.9).[143] The benzenesulphonyl pyrimidone **100** and the symmetrical *bis*-sulphonamide **101** exhibited IC_{50} values of 118 μM and 3 μM

respectively. Compound **101** was shown to stabilise Nrf2 and up-regulate NQO1 expression at mid-low micromolar concentrations.

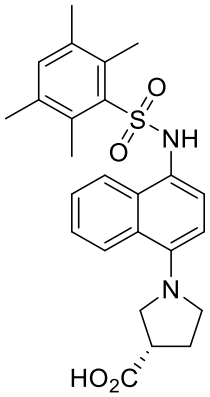
To develop and optimise compound **101**, Jiang *et al*. carried out molecular modelling studies and molecular binding determinant analysis between the Keap1 Kelch domain and the ETGE Nrf2 motif to determine crucial binding interactions in the Keap1 Kelch domain.[144] They found polar subpockets P1 and P2 significantly contribute to Keap1-Nrf2 binding through strong H-bond interactions. From this analysis, Jiang *et al*. reported the *bis*-acid analogue **102** (IC_{50} = 29 nM) which exhibits an improved Keap1 binding affinity in comparison to the micromolar analogue **101** (Figure 6.9).[144] The increase in activity was obtained through addition of two carboxylic acid side chains shown to occupy subpockets P1 and P2 as indicated by docking studies. The carboxyl groups mimic the Glu79 and Glu82 residues in the Nrf2 ETGE motif, forming H-bond interactions with Arg483 and Arg415 in subpockets P1 and P2 respectively. Despite the presence of two polar acid side chains, **102** was observed to induce Nrf2 activity in low micromolar concentrations in cells. Jiang *et al*. were able to improve the inhibition activity against Keap1 through substitution of the para-methoxy groups forming a *bis*-para-acetamido analogue **103** (IC_{50} = 14 nM)**;** importantly this substitution improved compound solubility resulting in enhanced Nrf2 inducing activity and *in vivo* efficacy.[145]



100



101: $R_1 = R_2 = H$, $R_3 = OMe$
102: $R_1 = R_2 = CH_2CO_2H$, $R_3 = OMe$
103: $R_1 = R_2 = CH_2COH$, $R_3 = NHCOMe$
94: $R_1 = R_2 = CH_2CONH_2$, $R_3 = OMe$
104: $R_1 = CH_2CO_2H$, $R_2 = H$, $R_3 = OMe$
105: $R_1 = CH_2CO_2Et$, $R_2 = H$, $R_3 = OMe$



106

Figure 6.9 Structures of direct inhibitors of Keap1 with a sulphonamide template.

Through comprehensive SAR investigations, Jain *et al*. found the polar interactions formed by the carboxyl groups did not need to be ionic, as shown through a *bis*-amide analogue **94** which retained activity (IC_{50} = 63 nM).[146] Furthermore they obtained a X-

ray crystal structure of the *bis*-amide analogue **94** with Keap1 which confirmed the proposed docking orientation of the *bis*-acid analogue **102** (PDB code: 4XMB, Figure 6.7). They also observed that both of the acid groups were not required and activity could be retained through a singly substituted analogue **104** (IC_{50} = 61 nM).

Jain *et al.* also observed the ester analogue **105** of the *mono*-acid **104,** induced HMOX1 and stabilised Nrf2 levels to the same extent as the *bis*-acid **102**, despite displaying a lower binding affinity *in vitro* (**105**: 85 nM *vs* **102**: 29 nM). They hypothesised the ester facilitated membrane traversal and was hydrolysed into the acid once inside the cell, thus further implying the carboxylic acids could be replaced for moieties with better pharmacokinetic properties. Further SAR within this study also indicated a significant decrease in affinity when the naphthalene core was replaced with a phenyl ring and a 1,4-substitution pattern was optimal.[146]

The *bis*-sulphonamide structure **101** was de-symmetrised resulting in compound **106** (Figure 6.9). Although **106** was found to exhibit nanomolar affinity (IC_{50} = 140 nM) with promising cell-based activity, *in vivo* evaluation was hampered due to oxidative metabolism.[147]

Additional compounds with heterocyclic moieties have also been identified as Keap1-Nrf2 inhibitors. Sun *et al*. employed a virtual screening method which identified a *bis*-acid urea derivative **107** (Figure 6.10).[148] Compound **107** was shown to successfully inhibit Keap1 with an IC_{50} of 9.8 µM. Satoh *et al*. disclosed a *mono*-acid oxadiazole compound **108** which was identified through molecular modelling and crystal structure studies (Figure 6.10).[149] Although its affinity for Keap1 was not described, it was shown to mimic the ETGE Nrf2 moiety. A related pyridine structure **109** was described by the same group with relatively poor Keap1-Nrf2 inhibition (>50 µM).[150] However, it was observed to exhibit *in vivo* efficacy in a model of non-alcoholic steatohepatitis. Finally, through an *in silico* fragment-based approach, Wells *et al*. developed a series of inhibitors based upon a 1,4-diphenyl-1,2,3-triazole scaffold.[151] The triazole scaffolds were found to inhibit the Keap1-Nrf2 interaction in Hepa1c1c7 cells leading to an increase in the expression of the Nrf2 target gene NQO1. Two structures from the series are presented (Figure 6.10).

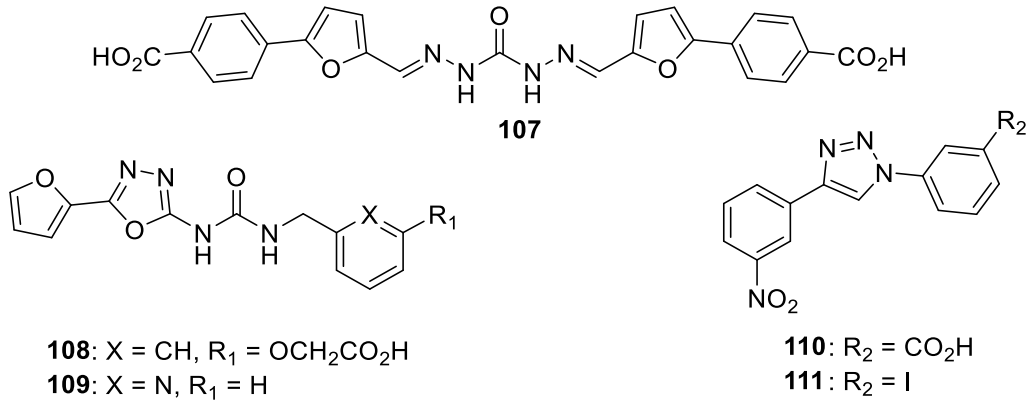


Figure 6.10 Direct Keap1 inhibitors with heterocyclic moieties, **107**–**111**

Through a fragment screen using X-ray crystallography as the primary assay, Davies *et al*. identified compound **112** as a potent Keap1-Nrf2 inhibitor with an *in vitro* EC_{50} of 12 nM (Figure 6.11).[152] The scaffold moves away from the *bis*-acid naphthalene structures exhibiting a *mono*-acid scaffold obtained through linking three fragments, followed by SAR-based optimisation. It has subsequently been shown to exert cytoprotective effects in an *in vivo* chronic obstructive pulmonary disease (COPD) model.

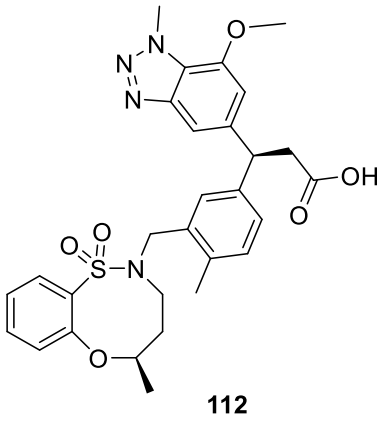


Figure 6.11 Structure of a Keap1 inhibitor **112**, identified through a fragment-based screen.

Direct inhibition of the Keap1-Nrf2 PPI by a small molecule has been shown to have a range of different therapeutic applications, as demonstrated through several *in vivo* disease model studies. Compound **103** has been shown to have potential therapeutic benefit towards inflammatory conditions, Crohn's disease, IBD and colorectal cancer whereas **109** and **112** have been shown to have potential therapeutic applications in non-alcoholic steatohepatitis and COPD respectively.[145, 150, 152]

Despite the success of both direct inhibition of the Keap1-Nrf2 PPI and indirect inhibition, there is still room for improvement in the physicochemical properties and observed off-target effects of the compounds reported to date, particularly for the

peptide inhibitors and electrophilic inducers. Success by GSK and Astex towards inhibition of the Keap1-Nrf2 PPI with a fragment-based approach is particularly interesting.[152] Fragment-based lead discovery is a well-established drug discovery technique used throughout the pharmaceutical industry which has had success in identifying small molecule inhibitors for challenging targets, including PPIs .[153]

6.2 Fragment-based drug discovery

There are many approaches currently adopted by the pharmaceutical industry to identify chemical leads including: use of an existing lead or drug, an existing natural product, traditional high-throughput screening (HTS) and more recent approaches including fragment-based drug discovery (FBDD).[154]

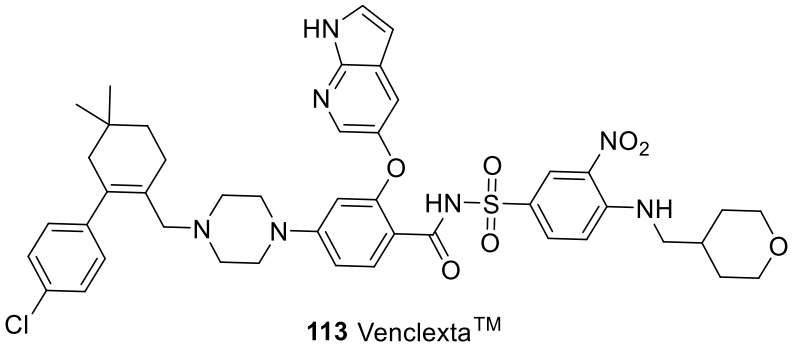
HTS aims to find a drug-sized starting point (typically up to 30 heavy atoms) through automated screening of 'drug-like' or 'lead-like' compounds.[155] A 'drug-like' compound generally obeys Lipinski's 'rule of 5' which establishes a guideline for orally bioavailable drugs and states that a drug is more likely to be orally bioavailable if it has some or all of the following properties: molecular weight ≤500 Da, calculated logP (lipophilicity) ≤5 and number of hydrogen bond donors and acceptors ≤5 and ≤10 respectively. Veber[156] and others[157] have enhanced the guideline to include: number of rotatable bonds ≤7 and polar surface area (PSA) 110–140 $Å^2$. The term 'lead-like' was introduced to describe hits from HTS which require, or are amenable to, further optimisation. Compounds with 'lead-like' properties are scaled-down with respect to Lipinski's guidelines to accommodate the increased molecular weight and lipophilicity gained from optimisation.[158] Although multiple lead compounds have been and still are identified using the HTS strategy,[155] it is limited due to (i) issues with false positives, particularly from compound aggregates,[159] (ii) difficulties screening against challenging or new targets which tend to yield few to no hits[160] and (iii) the limited success rate of bioactive hits during subsequent development.[161]

Fragment-based drug discovery (FBDD) has emerged in recent years as an alternative to HTS.[154] It has become a widely-utilised and well-established strategy for identifying potential chemical leads in both academia and industry with a number of drugs having entered the clinic.[162] The development of sensitive biophysical assays to enable the detection of weakly binding ligands helped to drive the success of the FBDD approach. Traditional fragment-based screening techniques include nuclear magnetic resonance (NMR), X-ray crystallography, surface plasmon resonance (SPR) and more recently differential scanning fluorimetry (DSF) and weak affinity chromatography.[162]

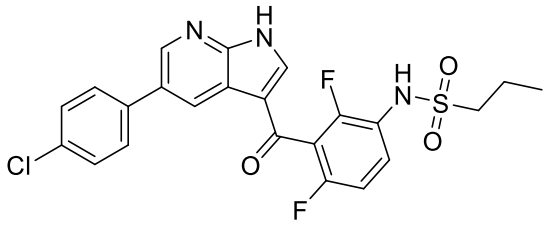
In contrast to HTS, FBDD screens smaller compounds (up to 15-20 heavy atoms) with the aim of identifying small fragments hits which can be subsequently optimised through structure-based drug design (SBDD). In some instances, the fragments can be linked if they are observed to bind in proximal binding sites. The fragment libraries are generally 'rule of 3' compliant,[163] a similar guideline to the 'rule of 5' but applicable to fragments where: cLogP ≤3, molecular weight <300 Da, number of H-bond acceptors and H-bond donors are each ≤ 3 and number of rotatable bonds ≤3, although as with the 'rule of 5' there are exceptions.[164]

FBDD is advantageous in comparison to HTS for a number of reasons, for example, due to the small size of fragments it is possible to cover more of the chemical space with a relatively small library. Furthermore, the identified fragments are more efficient binders, i.e. they have a higher ligand efficiency (LE) than most hits identified through HTS, which can result in highly LE drugs through careful subsequent SBDD. It is also cheaper as the fragment libraries tend to be smaller and synthetically tractable, with isolation of fragments in a small number of synthetic steps. Also, although fragments tend to have fewer interactions, they often have higher hit rates due to a lower molecular complexity in comparison to the larger compounds found in HTS libraries, i.e. they can bind to a greater number of sites on a greater number of targets due to fewer unfavourable interactions.[165] The observed higher hit rates enable the possibility of finding hits for the more challenging targets such as protein-protein interactions (PPIs).

PPIs are historically difficult targets for identifying small molecule inhibitors due to the generally large (1,500-3,000 Å^2), flat and featureless surface of the PPI binding area.[134] Facilitated by the recognition of 'hot spots' which are areas on the protein surface integral to binding, and the development of screening techniques and computational tools, the 'druggability' of PPIs with small molecules became more tangible.[153] Indeed, there are now multiple fragment-derived drug candidates for PPIs which have reached clinical trials; for example the FDA approved BCL-2 inhibitor ABT-199 **113** (Venclexta™) which is indicated in chronic lymphoid leukaemia (Figure 6.12).[166] The speed with which fragment-based approaches can develop inhibitors was highlighted by the drug Vemurafenib **114** (Zelboraf™), developed as an inhibitor of BRAF-mutant cancer (Figure 6.12).[167] It was the first fragment-derived drug to be approved, taking only 6 years from project initiation.



113 Venclexta™



114 Zelboraf™

Figure 6.12 Structures of PPI inhibitors derived from a fragment-based approach

Despite the many positives there are a number of disadvantages associated with the FBDD approach:[158] (i) although the fragment libraries are smaller than those for HTS, the fragments may still need to be synthesised and purified before use which is time-consuming and costly (ii) fragments have lower affinities thus higher assay concentrations are required as well as greater assay sensitivity (iii) it can also be challenging to link fragments because of the requirement to identify optimal linking groups and finally (iv) careful subsequent SBDD is required to maintain the LE and desired physicochemical properties.

6.2.1 Target-guided synthesis

The templated approach to drug design was developed during the rise of FBDD methodology. The concept of templated reactions was initially recognised in nature; for example, the transcription of DNA into RNA is a well-documented natural templated reaction.[168] In the last three decades, synthetic templated fragment-based approaches, also termed target-guided synthesis (TGS), have emerged as novel means of lead discovery where the biological target actively assembles its own inhibitory compound.[169] The newly formed inhibitors usually display much higher affinities for their biological targets than the individual fragments since they simultaneously engage in multiple binding interactions.[170] TGS shows great promise in lead discovery applications by combining the synthesis and screening of libraries of low molecular weight compounds in a single step.[169] Rideout *et al*. reported the first observation of TGS where they noted a synergism between the cytotoxic effects of decanal and *N*-

aminoguanidines.[171] They attributed this to the selective self-assembly of cytotoxic hydrazones *in situ*. Since then TGS has developed significantly and can be divided into two main categories: dynamic combinatorial chemistry (DCC) and kinetic target-guided synthesis (kTGS).

6.2.1.1 Dynamic combinatorial chemistry

The DCC approach uses thermodynamically controlled reversible reactions for *in situ* target-guided ligand assembly (Figure 6.13).[172] A fragment library is synthesised with building blocks exhibiting complementary reactive functional groups X and Y. The fragments react via a reversible reaction to make an equilibrium mixture of all possible products also known as a 'dynamic combinatorial library' (DCL). Subsequent incubation of the DCL with the protein or biological target skews the equilibrium towards the ligand which has the greater affinity for the target. 'Freezing' of the products, for example reduction of imines to amines, is often required to enable analysis of the reaction mixture and identify potential inhibitors of the target.[173]

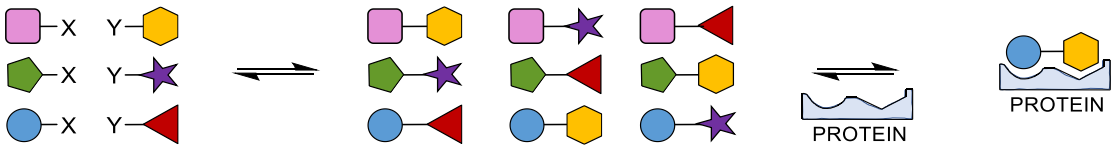


Figure 6.13 Cartoon representation of principles for *in situ* dynamic combinatorial chemistry.

The reversible reactions are selected based on synthetic tractability of the functionalised building blocks in tandem with the conditions required to generate the DCL (for example pH, temperature etc); ideal reactions must be biorthogonal and compatible with physiological conditions.[173] Despite these restrictions there is an expanding set of reversible covalent reactions which have been used to generate DCLs; three reactions often employed are: imine, hydrazone and disulphide bond formation (Figure 6.14).[173]

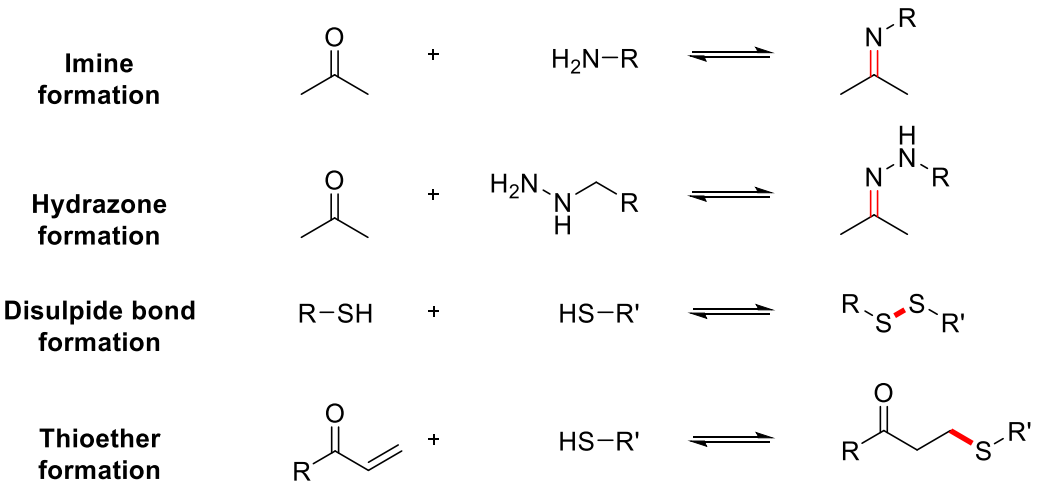


Figure 6.14 Common reversible reactions often employed for generating DCLs: imine, hydrazone and disulphide bond formation. New bonds formed are highlighted in red.

The approach has been successfully applied to many protein targets including, but not limited to, human carbonic anhydrase I and II using the formation of imines as the reversible reaction and glutathione *S*-transferase (GST) where hydrazones and thioethers have been formed in the reversible reactions.[174]

6.2.1.2 Kinetic target-guided synthesis

Conversely, the kTGS approach uses irreversible, kinetically controlled reactions for target-guided ligand assembly (Figure 6.15). A fragment library is synthesised with building blocks exhibiting complementary reactive groups A and B. The biological target is incubated with the fragments and accelerates productive interactions with the building blocks and stabilises the transition state.

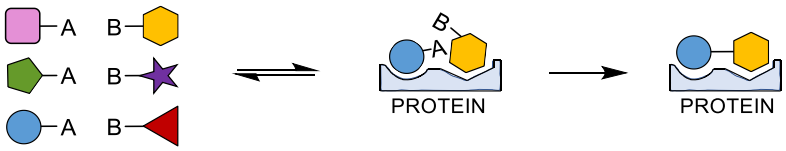


Figure 6.15 Cartoon representation of principles for *in situ* kinetic target-guided synthesis.

The irreversible reaction which links the fragments is required to be: (i) bioorthogonal, (ii) compatible with physiological conditions (iii) slow background rate of reaction due to a high kinetic barrier (iv) result in a stable covalent bond (v) have no or minor side products and (vi) be stable in aqueous media.[175] Despite these relatively stringent criteria a number of irreversible reactions have been employed to identify inhibitor scaffolds for a diverse range of biological targets and disease associations.[176] A few examples of the irreversible reactions commonly used in kTGS and their application to particular targets are discussed (vide infra) (Figure 6.16).

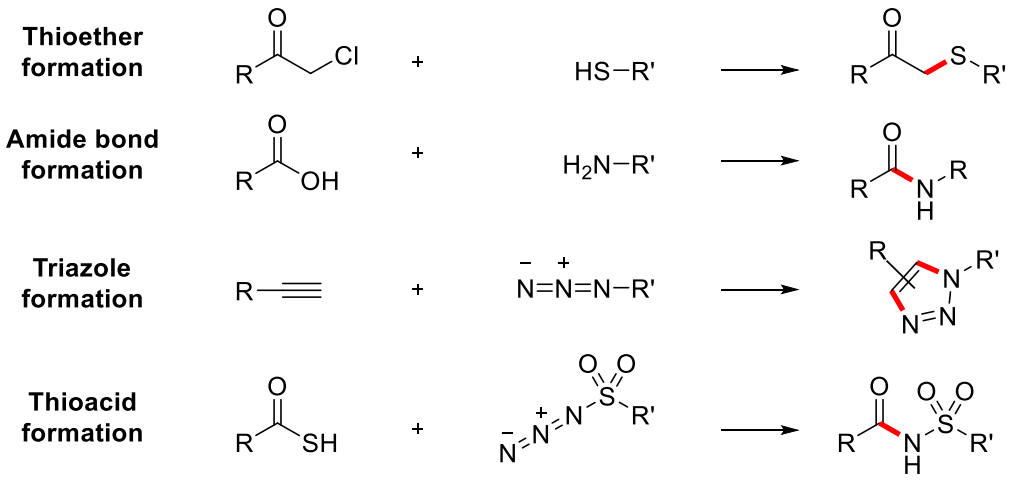


Figure 6.16 A selection of the irreversible reactions used in kTGS. New bonds formed are highlighted in red.

Nguyen *et al*. successfully demonstrated kTGS could be used to identify nanomolar inhibitors of bovine carbonic anhydrase (CA II), a Zn^{II} metalloenzyme, using an irreversible reaction between thiols and alkyl chlorides.[177] They observed that the most potent compound from their small series of sulphonamide compounds was selectively formed in the presence of CA II from a mixture of fragments.

Gelin *et al*. screened adenosine scaffolds against NAD kinase with *in situ* formation of amide bonds from an amine and carboxylic acid without prior functionalisation.[178] From this study they successfully identified a dinucleoside scaffold with micromolar *in vitro* activity. It was noted a His223 residue is required for the catalytic successful amidation, demonstrated through a single point mutation (H223E) which abolishes the reaction. This observation also demonstrated that the protein is catalysing the reaction.

Of the published reactions, the *in situ* click chemistry approach is the most widely employed. This was developed by Sharpless and colleagues using the Huisgen azide alkyne 1,3-dipolar cycloaddition (AAC) with the enzyme acetylcholinesterase (AChE) as the biological target.[179] They showed that the enzyme assembled a potent triazole inhibitor from a small library of fragments bearing complementary azide and alkyne functionalities. Previous work by Mock *et al*. had demonstrated proof of concept by establishing the AAC is accelerated by 4 to 5 orders of magnitude in the presence of a host (cucurbituril).[180] The reaction has been found to be particularly compatible with the identification of ligands for various biomolecules including enzymes, RNA and DNA, protein-protein interactions (PPIs) and antibody-like protein capture agents.[161] An alternative sulpho-click reaction via *in situ* generated thioacids has also been

developed and used in kTGS experiments with the protein target BcL-X_L and was the first example of kTGS used to investigate a PPI.[175, 181, 182]

kTGS has the potential to explore previously unknown conformations of proteins and novel binding sites, crucially without having prior knowledge of the protein structure. This presents an exciting opportunity to explore new targets which have so far been out of reach due to lack of structural data. Furthermore, application of kTGS to well characterised targets could allow the identification of novel scaffolds. This was observed by Willand *et al*. in identification of inhibitors for multi-drug resistant tuberculosis through evaluation of inhibition of the transcriptional repressor EthR.[183]

Advances in the methodology and sensitivity of the product detection technique vastly improved the kTGS approach, making it more widely applicable to research groups in both academia and industry. The sensitivity of the analytical technique to detect the product composition from kTGS studies was greatly improved through the introduction of the LC/MS-SIM (Liquid Chromatography Mass Spectrometry Single Ion Monitoring) which separates the crude reaction mixture prior to analysis using standard LC/MS with SIM for greater sensitivity.[184] This increase in sensitivity allowed for the introduction of multicomponent reactions by Manetsch *et al*.[184] Multicomponent mixtures of fragments, where a pool of ligands are screened at once rather than one ligand at time, increases the screening throughput and also reduces the quantity of protein required.

Although significant advances in the methodology have been made, kTGS still faces limitations. The main limitation of the kTGS approach is the occurrence of false negatives. These occur due to varying and often interrelated reasons including: (i) slow turnover due to higher affinity of the templated products resulting in poisoning of the biocatalyst by the templated products themselves (ii) low binding affinity of fragments despite their synergistic effects (iii) long incubation times due to slow reaction rates resulting in destruction or alteration of the protein tertiary structure and (iv) despite the recent progress, detection limits and sensitivity of the analytical techniques employed to identify hits.[161] Some of these issues have been addressed through multidisciplinary approaches. For example, mutant proteins have been designed which increase the dissociation constant of the protein-ligand complex promoting multiple rounds of biocatalysis and reducing catalyst poisoning from the ligands themselves.[185] Mutant proteins have also featured a covalent ‘tether’ for one of the two reactive fragments so as to create an artificial anchor which promotes the reaction to detectable quantities and also facilitates the use of lower affinity fragments.[186]

In summary, kTGS is an exciting fragment-based drug discovery technique which uses a novel approach where the target itself templates its own inhibitors. It has been shown to be a versatile and efficient tool for identifying ligands for a diverse range of targets, including the more difficult targets such as PPI and targets with little to no structural information. Likewise, novel scaffolds can be identified for targets which have been thoroughly characterised. The approach is continually advancing because the main limitation of kTGS, false negatives, is being addressed with a variety of multidisciplinary tactics and furthermore the repertoire of reactions is rapidly expanding.[161]

6.3 Aim

Despite the success of both direct inhibition and indirect inhibition of the Keap1-Nrf2 PPI, there is still room for improvement; off-target effects are the main limitation of electrophilic inducers and many of the small molecule and peptide scaffolds have poor cellular permeability due to the presence of acid groups. It was hypothesised an approach based on kTGS could be applied to identify new classes of Keap1-Nrf2 PPI inhibitors, as well as provide an opportunity to identify different binding modes through exploration of the binding pocket. A proof of concept was developed using a biased approach composed of three components (Figure 6.17):

(i) identification of biased fragments – initially, a biased fragment will be designed based on scaffolds known to bind to the Keap1 Kelch domain and evaluated for affinity to the Keap1 Kelch domain in biophysical assays

(ii) kTGS experiments – this biased fragment will then be used as an anchor, to screen a focused library of fragments (bearing a complementary reactive function group) in the presence of the Keap1 Kelch domain, the formation of the bidentate products will be evaluated

(iii) hit validation – a short list of scaffolds will be re-synthesised for validation in biophysical assays.

The new triazole lead compounds can then be further optimised and refined using structure-activity relationship (SAR) data, crystal structures and other medicinal chemistry techniques.

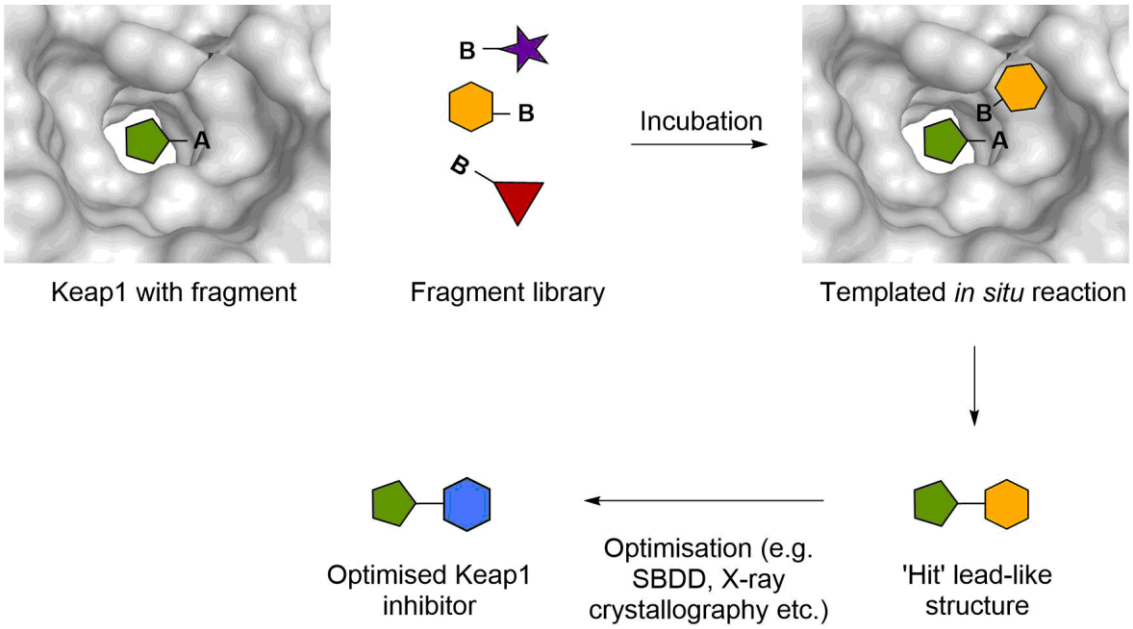


Figure 6.17 Schematic representation of the application of kTGS approach in identifying novel Keap1-Nrf2 PPI inhibitors.

CHAPTER 7
Biased fragment identification

In order to validate the use of kTGS as an approach for the development of novel Keap1 inhibitors, an approach based on the use of biased fragments was employed. This required the development of an appropriately labelled fragment to act as a handle within the Keap1 binding site and selection of an appropriate biocompatible *in situ* coupling reaction. Incubation of this fragment with a diverse library of fragments containing a complementary reactive group would allow rapid exploration of the available binding landscape and the identification of novel Keap1 inhibitor scaffolds.

7.1 *In situ* coupling reaction

As previously indicated in section 6.2.1.2, several reaction types have been investigated which are compatible with *in situ* fragment linking. For the purposes of this proof-of-concept study, the *in situ* click 1,3-dipolar Huisgen cycloaddition of azides and alkynes was selected due to its wide and successful application in kTGS (Figure 7.1).[161] The target-guided nature of the experiment can mean either, or both, of the 1,4- or 1,5-regioisomers can be formed and is dictated by the protein environment.

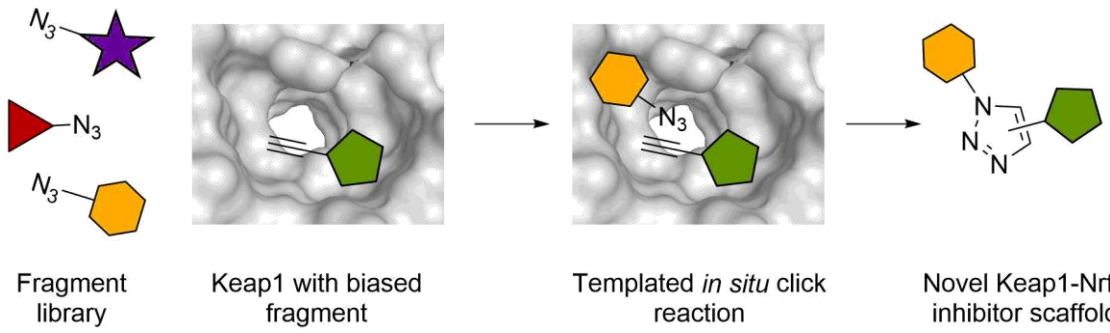


Figure 7.1 Schematic representation of an *in situ* 1,3-dipolar cycloaddition templated by Keap1 which links the two fragments together to create a novel triazole structure. Both regioisomers are represented in the product.

7.2 Biased fragment design

A biased method was proposed to validate kTGS as a suitable approach to identify novel inhibitors of the Keap1-Nrf2 PPI. This strategy is often used in proof-of-concept studies to probe the viability of the application through utilisation of a fragment which is known to bind to the target.[176] The structure of the biased fragment was derived from a published direct low nanomolar inhibitor **102** from the naphthalene series (Figure

7.2A).[144] Inhibitor **102** is a useful template to explore the Keap1-Nrf2 interaction as the binding conformation and interactions between **102** and Keap1 and a related *bis*-amide analogue **94**, have been thoroughly characterised using a range of biophysical techniques.[145] Furthermore, there is scope for optimisation of this series as the carboxylic acid groups increase the polarity, which limits the membrane permeability and activity in cell-based assays.[144] The binding conformation of this class of compound was confirmed through co-crystallisation of the *bis*-amide analogue **94** with the Keap1 Kelch domain (Figure 7.2A,B).[146] The core naphthalene ring occupies subpocket P3 of the Keap1-Nrf2 binding site. The symmetrical side chains, comprising a *bis*-amide and sulphonamide, occupy subpockets P1/P2 and P4/P5 respectively.

Guided by the X-ray crystal structure the biased fragment was designed to contain two components: a binding moiety to anchor the fragment to the target and a reactive functional group (azide or alkyne) which allows *in situ* coupling with fragments containing a complementary reactive functional group.

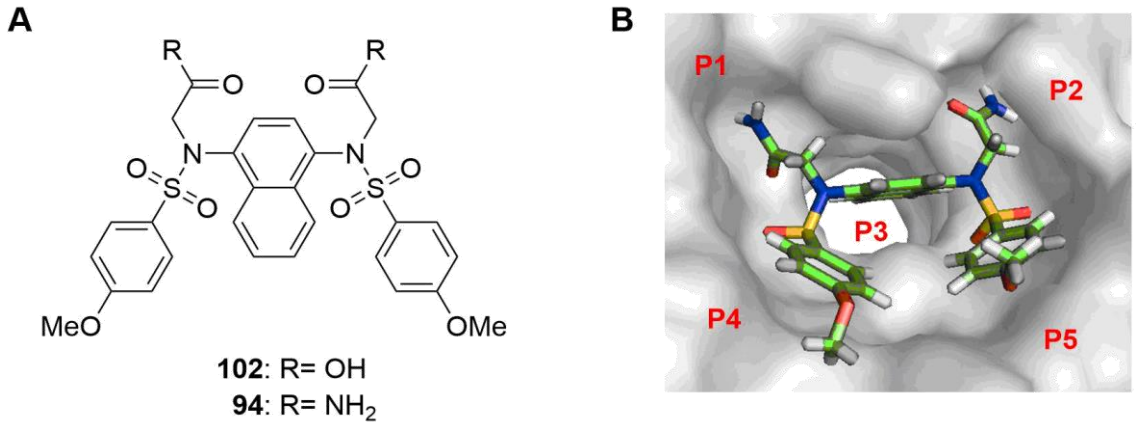


Figure 7.2 A Structures of Keap1 inhibitors **102** and **94**. **B** Crystal structure of Keap1 bound to Keap1 inhibitor **94** (PDB code 4XMB, 2.4 Å). Generated using PyMol.

As previously discussed in section 6.1.1.2, the core naphthalene ring and at least one, but not necessarily both, of the acetic acid side chains significantly contribute to the binding interactions for this series.[146] Thus, the core naphthalene ring and one of the mirrored side chains were identified as the anchor portion of the biased fragment. It was proposed that the reactive functional group (R) could be incorporated through replacement of the second side chain (Figure 7.3). It was hypothesised that the biased fragment would retain sufficient affinity and bind in the same orientation as the published inhibitor **102** and its *bis*-amide analogue **94** due to retention of several key interactions.

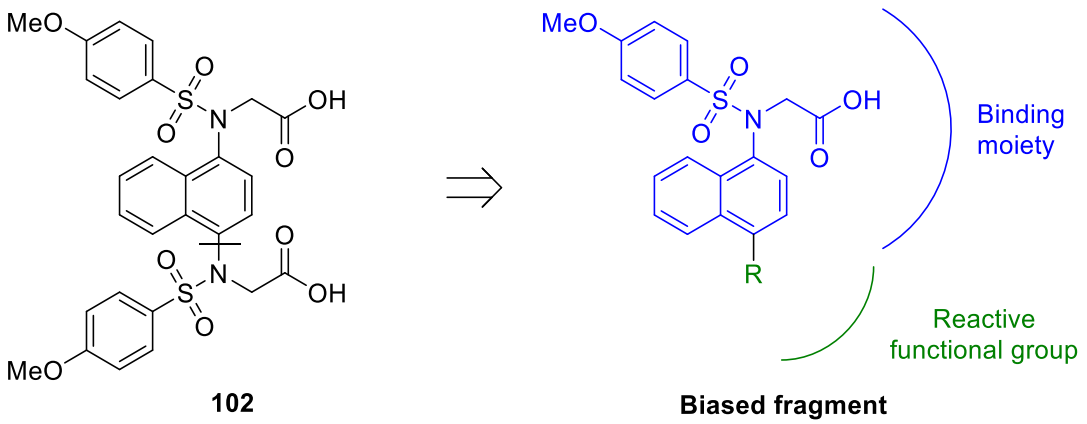
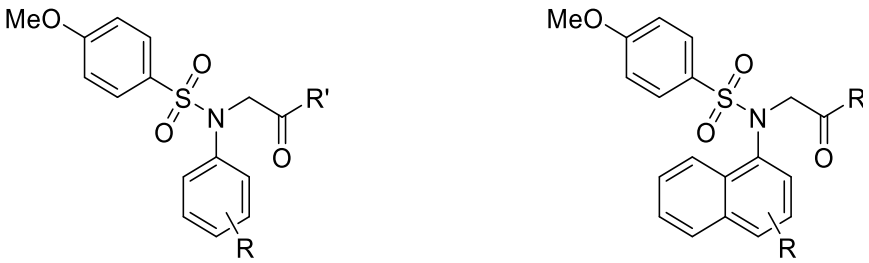


Figure 7.3 Replacement of one of the mirrored side chains in **102** with the reactive functionality, R (green). The remaining portion of the biased fragment acts as the binding function (blue).

Fragments are typically weak binders with low micromolar affinities.[162] However, fragments of varying affinities, micromolar to nanomolar, have been successfully employed as scaffolds for kTGS;[176] for example, a high affinity (nanomolar) fragment was used to identify a tacrine based AChE inhibitor and a low affinity (micromolar) fragment (>100 μM) was used to identify HIV-1 protease inhibitors.[170, 187] Thus, potent inhibitors of the target can be identified provided the fragments have sufficient affinity to anchor to the target protein.[184] It was hypothesised a second biased fragment with a lower affinity could be screened alongside the naphthalene fragment to assess the advantages, if any, that a higher affinity scaffold may offer. The lower affinity fragment could also identify a different set of scaffolds because the fragments might adopt different binding conformations or occupy different binding subpockets. Thus, a second analogous biased fragment with a predicted lower affinity for Keap1 was also designed through removal of one of the core naphthalene rings resulting in a phenyl scaffold.

A small subset of analogues, based on both the naphthalene and phenyl scaffolds, were designed and synthesised. They were subsequently screened to enable selection of a biased fragment within each series for use in *in situ* optimisation to investigate potential interactions in the Keap1-Nrf2 pocket. The subset variables were: (i) azides versus alkynes (R) (ii) acid versus ester (R') (iii) substitution at position 3- versus 4- on the aromatic core and (iv) linker lengths (Figure 7.4). It was envisaged that investigation of both the linker length and different substitution positions on the aromatic core could provide useful information on the topographical limitations of the pocket.



Phenyl biased fragment | **Naphthalene biased fragment**

Figure 7.4 Two fragment scaffolds based on a known potent Keap1 inhibitor, using a phenyl scaffold (left) and naphthalene scaffold (right). Where: R = N_3, or C≡CH or $OCH_2C{\equiv}CH$ at positions -3 or -4 on the aromatic ring; R' = OH or OEt.

7.3 Synthesis of biased fragment scaffolds

The synthesised analogues comprise six analogues of the phenyl scaffold (**115**-**124**) and one analogue of the naphthalene scaffold (**125**) (Figure 7.5).

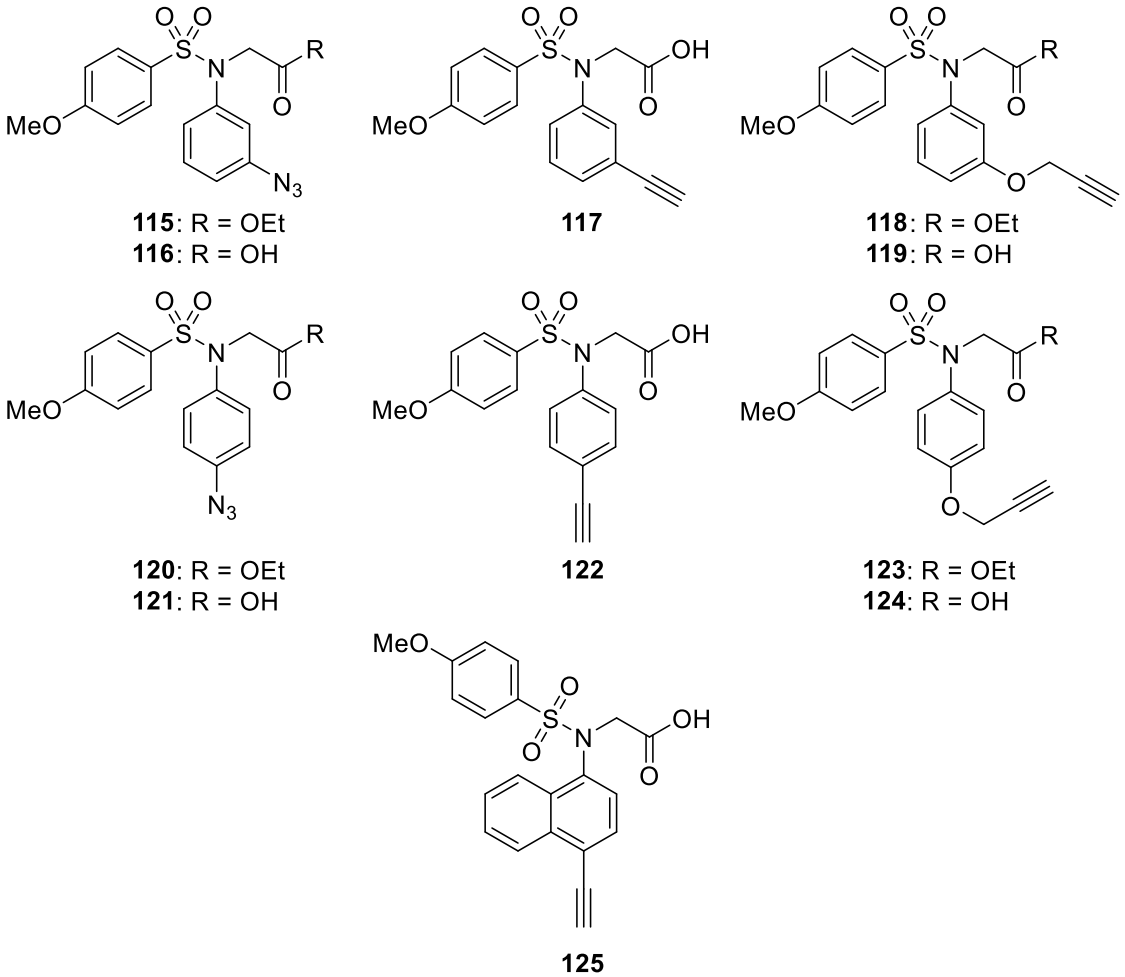
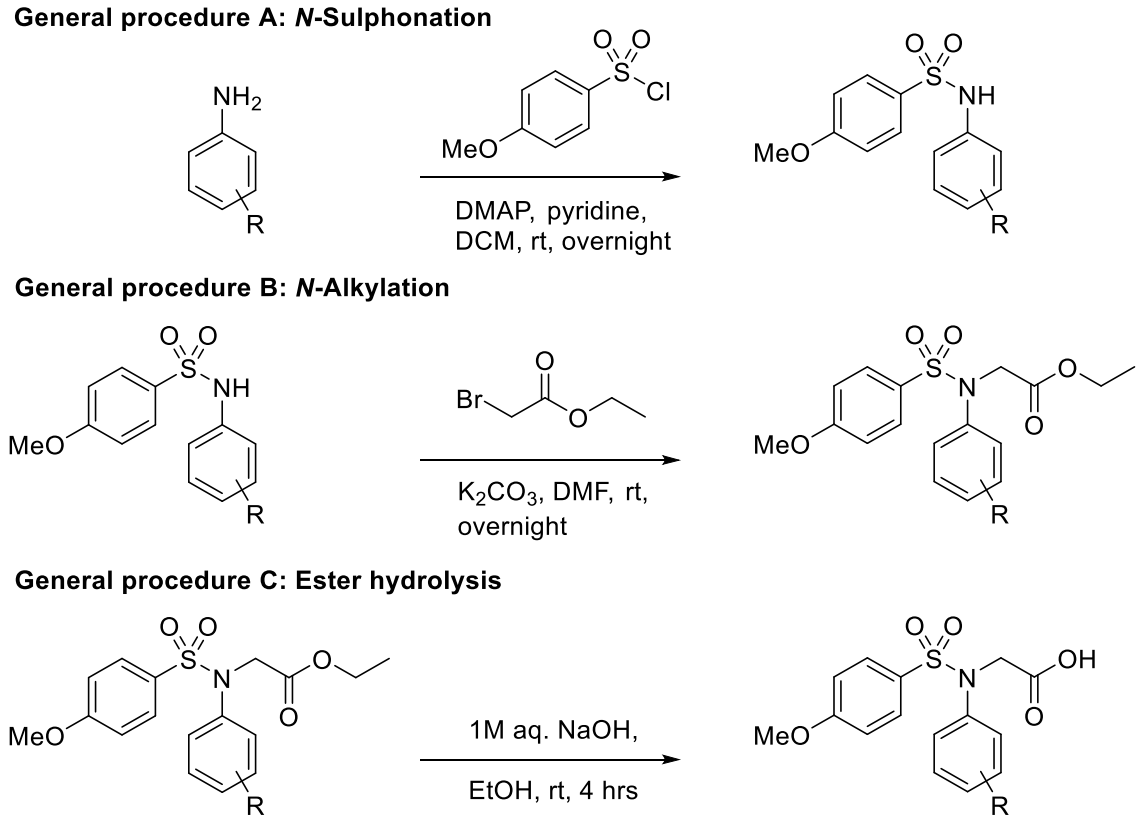


Figure 7.5 Summary of synthesised biased fragments for both scaffolds **115**-**125**

Compounds based on the naphthalene scaffold were not as easily attainable due to the high cost and limited availability of suitable building blocks. Individual synthetic schemes for each analogue with yields and reagents are discussed below.

7.3.1 General procedures A-C

The amino aryl starting materials were treated in a similar manner to install the sulphonamide by treatment with 4-methoxybenzenesulphonyl chloride in the presence of DMAP and pyridine. The sulphonamides were *N*-alkylated with ethyl bromoacetate and the ester functionality hydrolysed with sodium hydroxide in ethanol (Scheme 7.1, procedures A-C).[188]

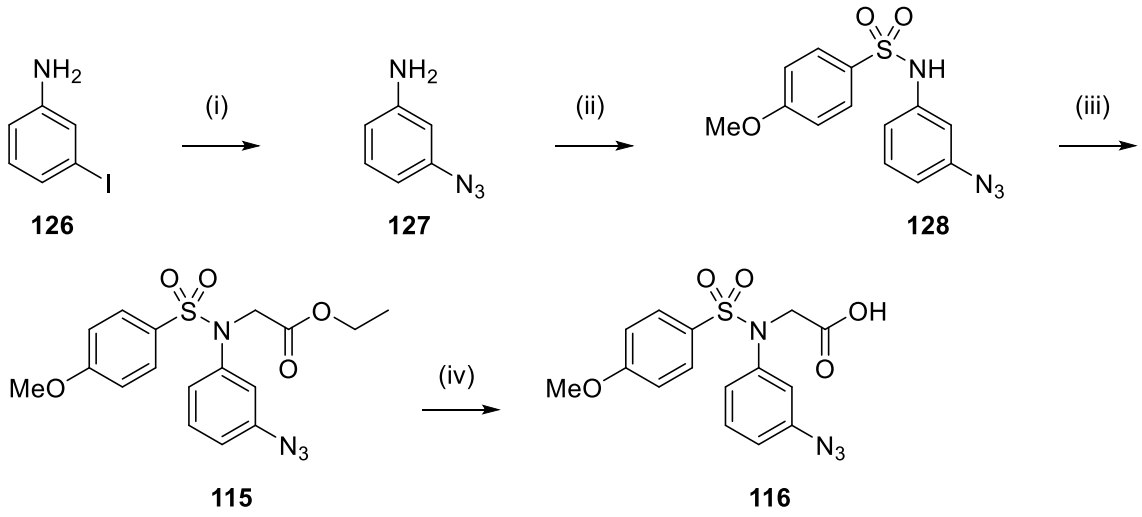


Scheme 7.1 Conditions for general procedures A, B and C.[188] Where reactive functional group: R= N_3, CΞCH or with a longer linker $OCH_2CΞCH$ at substitution positions -3 or -4 on the aromatic ring.

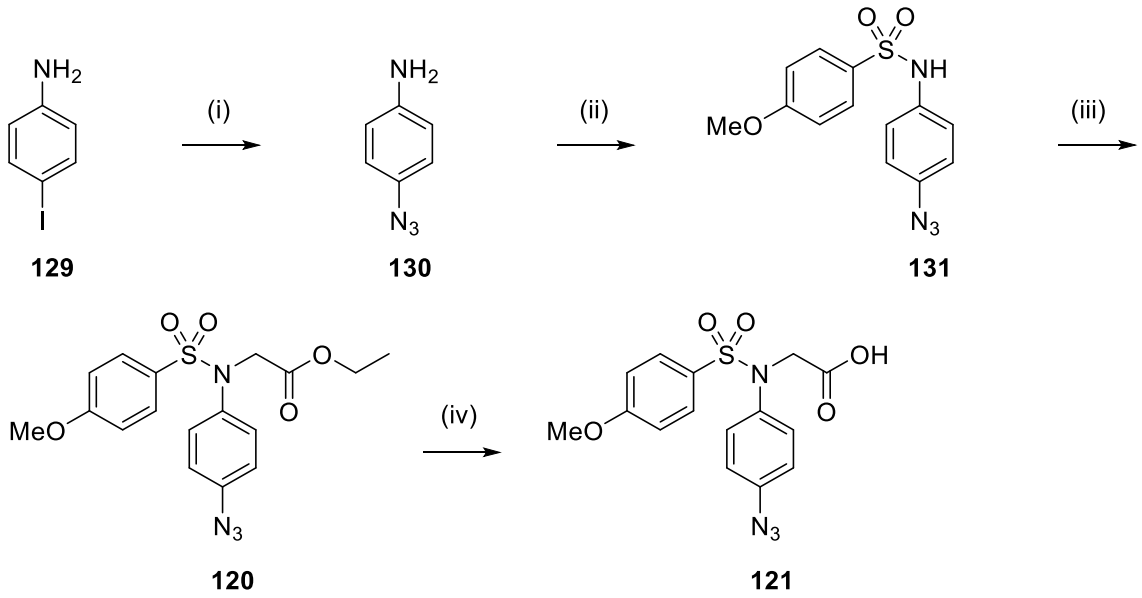
7.3.2 Synthesis of compounds 116, 117, 121 and 122

The four compounds with the shortest linker i.e. where the functional group is directly attached to the phenyl ring, were synthesised first. Azide compounds, **116** (R-3) and its analogue **121** (R-4), were each synthesised over four steps (Scheme 7.2 and Scheme 7.3). First, the azide was installed through a copper catalysed substitution of the aryl iodide with sodium azide. The sulphonamide and alkyl ester were installed following general procedures A and B (Scheme 7.2 and Scheme 7.3, (step (ii) and step (iii)).

Finally, the alkyl ester was hydrolysed to the acid to yield the final compounds **116** and **121**.



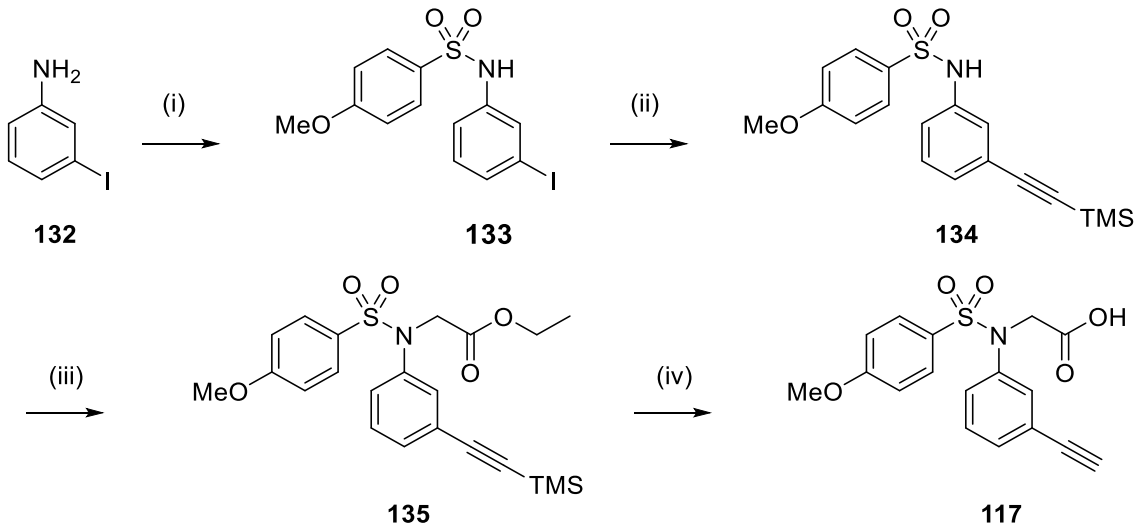
Scheme 7.2 Synthetic route to compound **116**. *Reagents and conditions*: (i) NaN_3, sodium ascorbate, CuI, N^1,N^2-dimethylethane-1,2-diamine, DMSO, H_2O, room temperature, 1 hr (82%); (ii) General procedure A (94%); (iii) General procedure B (97%) (iv) General procedure C (77%).



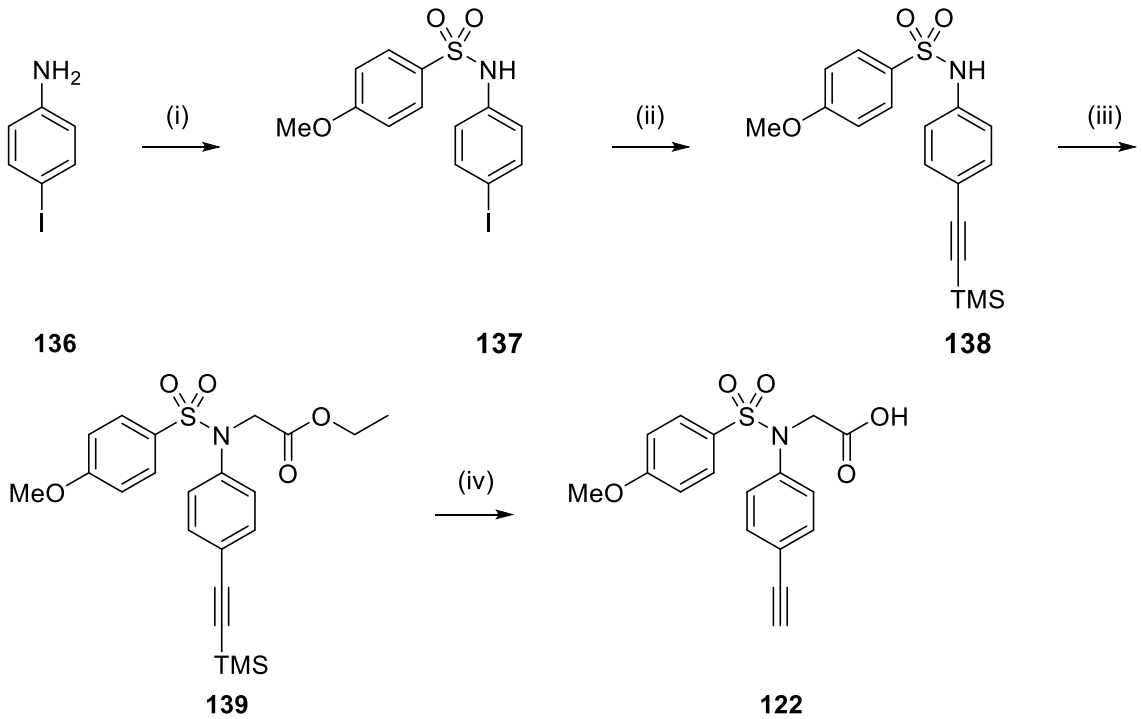
Scheme 7.3 Synthetic route to compound **121**. *Reagents and conditions*: (i) NaN_3, sodium ascorbate, CuI, N^1,N^2-dimethylethane-1,2-diamine, DMSO, H_2O, room temperature, 1 hr (91%); (ii) General procedure A (94%); (iii) General procedure B (92%); (iv) General procedure C (90%).

The corresponding alkyne analogues **117** (R-3) and **122** (R-4) were each synthesised over four steps (Scheme 7.4 and Scheme 7.5). First, the sulphonamide arm on the amine was installed onto the commercially available iodoaniline using general procedure A. The alkyne was introduced through a Sonogashira coupling in which the aryl iodide was cross-coupled with TMS-acetylene catalysed by CuI and

$Pd[(PPh_3)_2Cl_2]$.[189] Step (iii) introduced the ester through a second substitution onto the amine using general procedure B. Finally, a one-step TMS-deprotection and base catalysed ester hydrolysis following general procedure C, yielded final compounds **117** and **122**.



Scheme 7.4 Synthetic route to compound **117**. *Reagents and conditions:* (i) General procedure A (80%); (ii) CuI, $Pd[(PPh_3)_2Cl_2]$, Et_3N, TMS-acetylene, dry THF, room temperature, 12 hr (83%); (iii) General procedure B, taken forward as crude mixture; (iv) General procedure C (21%).

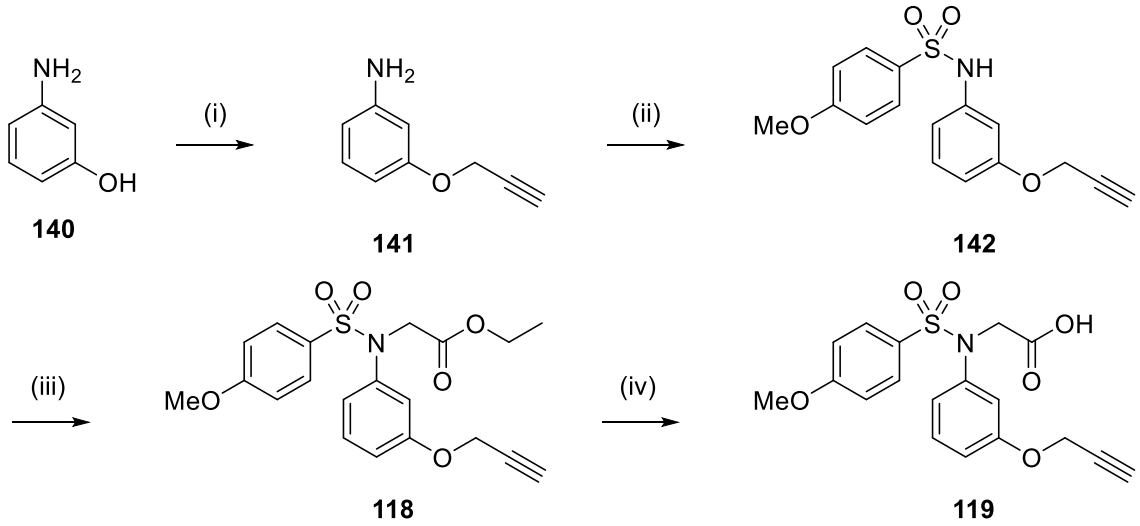


Scheme 7.5 Synthetic route to compound **122**. *Reagents and conditions*: (i) General procedure A (91%); (ii) CuI, $Pd[(PPh_3)_2Cl_2]$, Et_3N, TMS-acetylene, dry THF, room temperature, 12 hr (66%); (iii) General procedure B (88%); (iv) General procedure C (33%).

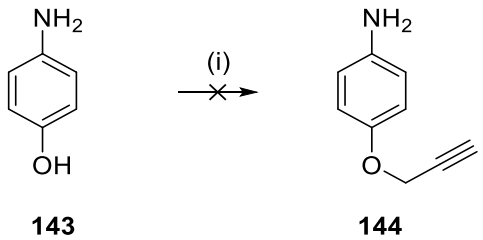
7.3.3 Synthesis of compounds 119 and 124

Compound **119**, an analogue of alkyne **117** with a longer linker (R = $OCH_2C\Xi CH$), was synthesised over four steps (Scheme 7.6). First, the extended linker was introduced via nucleophilic substitution of 3-hydroxyaniline using propargyl bromide. Step (ii) and (iii) installed the *N*-sulphonamide and *N*-alkyl ester using general procedures A and B respectively. Finally, a base catalysed ester hydrolysis following general procedure C, yielded the final compound **119**.

Synthesis of compound **124,** an analogue of compound **119** with substitution at R-4, was initially unsuccessful. Isolation of the intermediate compound **144** from step (i) using the same conditions as its R-3 analogue **141**, was unsuccessful (Scheme 7.7). Repetition of the same conditions but at room temperature and again at 0 °C, did not yield the desired product **144**. It was hypothesised these conditions could result in some oxidation of **143** to form the unstable 4-iminocyclohexa-2,5-dien-1-one.

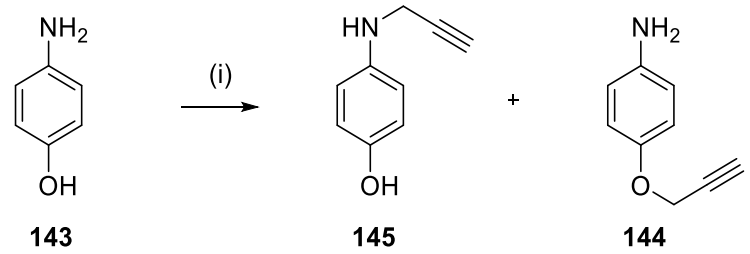


Scheme 7.6 Synthetic route to compound **119**. *Reagents and conditions:* (i) 3-hydroxyaniline, NaOH, MeOH, 40 °C, 1 hr then propargyl bromide in toluene (80% v/v), dry MeCN, room temperature, 12 hr (74%); (ii) General procedure A (91%); (iii) General procedure B (quant.); (iv) General procedure C (72%).



Scheme 7.7 Attempted synthesis of intermediate compound **144**. *Reagents and conditions*: (i) 4-hydroxyaniline, NaOH, MeOH, 40 °C, 1 hr *then* propargyl bromide in toluene (80% v/v), dry MeCN, room temperature, 12 hrs. Reaction unsuccessful.

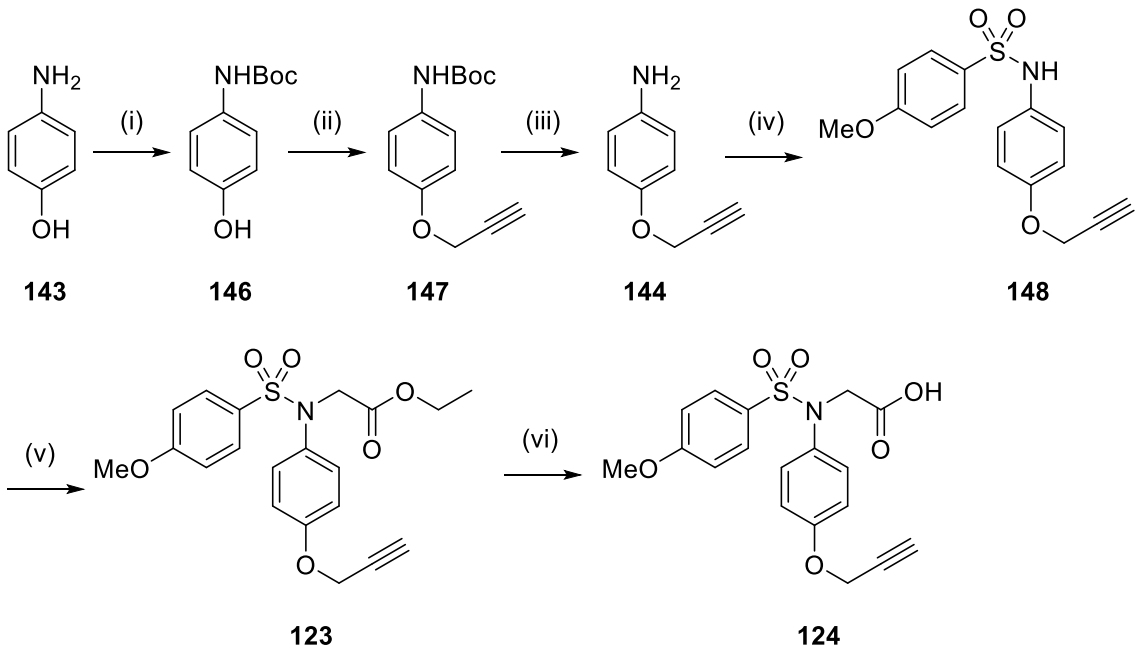
An alternative method was explored following the conditions published by Zhang et *al.* which utilised potassium carbonate as the base (Scheme 7.8).[190] However, these conditions resulted in the formation of **145** and the desired product **144** as a by-product.



143 **145** **144**

Scheme 7.8 Alternative conditions to isolate compound **144** resulted in *N*-propargylation of **143**. *Reagents and conditions:* (i) 4-hydroxyaniline, K_2CO_3, propargyl bromide in toluene (80% v/v), DMF, room temperature, 12 hr.

To overcome this, the amine was *N*-Boc protected prior to propargylation, allowing isolation of compound **147** in an 80% yield using the conditions set out by Zhang *et al* (Scheme 7.9, steps (i)-(iii)).[190] The amine was deprotected using 4 M HCl in dioxane and the desired scaffold **124** was subsequently obtained using general procedures A-C as previously described (Scheme 7.9).



143 **146** **147** **144** **148**

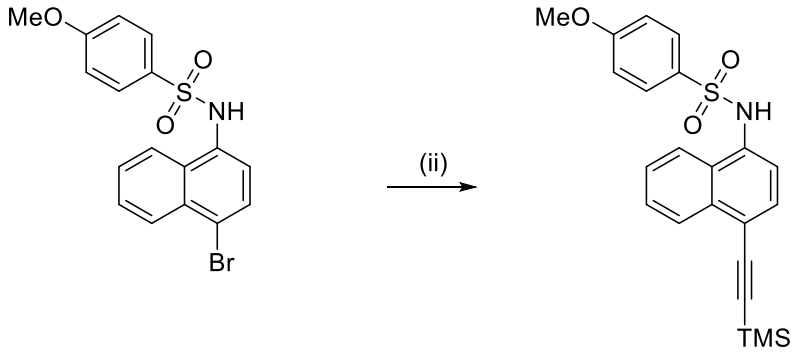
123 **124**

Scheme 7.9 Synthetic route to compound **124**. *Reagents and conditions*: (i) Boc_2O, 4-hydroxyaniline, THF, room temperature, 18 hr (*quant.*); (ii) K_2CO_3, propargyl bromide in toluene (80% v/v), DMF, room temperature, 12 hr (80%); (iii) 4M HCl in dioxane, MeOH, room temperature 2 hr (95%); (iv) General procedure A (95%); (v) General procedure B (*quant*.); (vi) General procedure C (82%).

7.3.4 Synthesis of compound 125

Synthesis of compound **125** was attempted following the same synthetic route as outlined for the corresponding phenyl analogue **122**. Employing the Sonogashira conditions, previously outlined for the synthesis of the phenyl analogue **150**, to synthesise naphthalene intermediate **149,** was unsuccessful (Table 7.1, entry 1). This is potentially due to the less labile C-Br bond in compound **149**, in comparison to the C-I bond in the phenyl analogues. Through optimisation of the reaction variables (temperature, solvent, heat source, reagent and catalyst equivalents) intermediate **149** was isolated in an 80% yield (Table 7.1, entry 5). Subsequent steps were performed using previously outlined conditions affording **125** (Scheme 7.10).

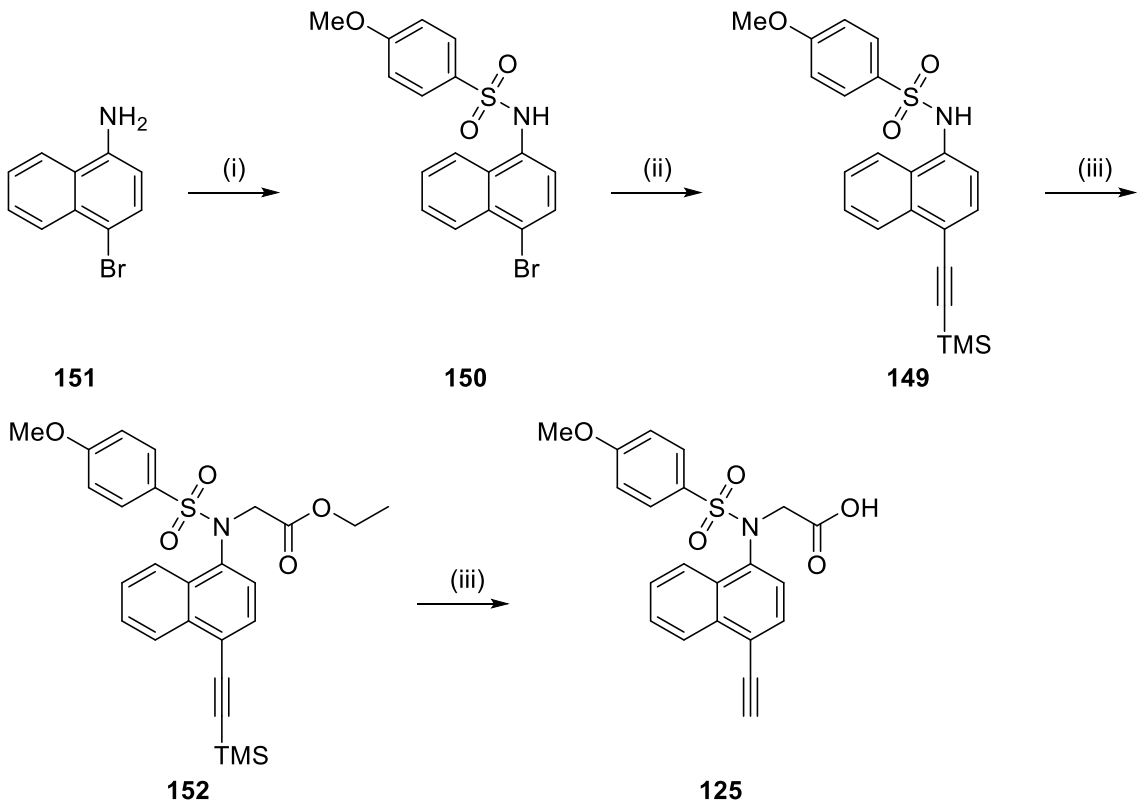
Table 7.1 Optimisation of the conditions for the Sonogashira cross-coupling between the aryl iodide and TMS-acetylene.



150 **149**

Entry	Pd mol%	Cu mol%	TMS-acetylene	Solvent	Time (hr)	T (°C)	150:149
1	5	10	1.1	THF	12	r.t.	-
2	5	10	1.1	THF	12	50	-
3	5	10	1.1	DMF	1.5	*120	2:1
4	10	20	2.0	DMF	1.5	*120	1:2
5	30	30	3.0	DMF	1.5	*120	0:1

Reaction progress was monitored by LCMS and 1H NMR. *microwave heat source.



Scheme 7.10 Synthetic route to compound **125**. *Reagents and conditions*: (i) General procedure A (83%); (ii) CuI, Pd[$(PPh_3)_2Cl_2$], Et_3N, TMS-acetylene, DMF, microwave irradiation 130 °C, 30 minutes (80%); (iii) General procedure B (70%); (iv) General procedure C (71%).

7.4 Keap1 over-expression and purification

To enable biophysical evaluation of the compounds synthesised in section 7.3, the human Keap1 Kelch domain was expressed from glycerol stocks of competent BL21(DE3) *Escherichia coli* (*E. Coli*) containing the expression vector pET15bhKelch (residues 321–609) with a hexahistidine tag fused at the *N*-terminus.[136] The protein was purified using immobilised metal ion affinity chromatography. Peak fractions, identified to contain protein by SDS PAGE and Bradford assay, were pooled and buffer exchanged using PD-10 columns and supplemented with glycerol prior to storage at -80 °C. Protein concentration and purity for each fraction was determined using the Bradford assay and SDS PAGE respectively, a representative gel is shown in Figure 7.6. MALDI-TOF was used to determine the protein molecular weight which was compared to the expected protein molecular weight calculated using UniProt and ExPASy. Detailed protocols are described in section 10.1.

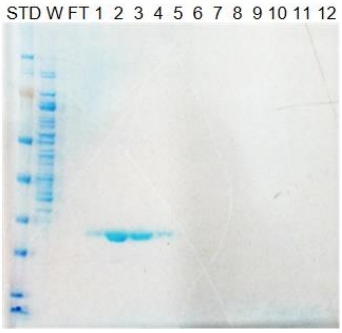


Figure 7.6 Example of SDS PAGE to determine protein purity where; STD = SeeBlue® Plus2 Pre-Stained Standard, FT= flow through, W= wash, 1 – 12= elution fractions.

7.5 Biophysical evaluation of biased fragments

With purified protein in hand, the biased scaffolds' relative affinity for Keap1 was evaluated using two biophysical assays, differential scanning fluorimetry (DSF) and fluorescence polarisation (FP).

7.5.1 Differential scanning fluorimetry

7.5.1.1 Biophysical basis of DSF assay

The stable tertiary structures formed by proteins are integral to their function as well as their interaction with drugs. The formation of these structures are controlled by various forces, the most important of which are hydrophobic interactions and H-bonding.[191] Under thermal conditions where the temperature is increased, these forces are disrupted and the protein denatures or unfolds. The 'melting' temperature or melt point (T_m) of a protein is defined as the temperature at which the concentrations of folded (F) and unfolded (UF) protein are equal.[192] The stability of a protein can be affected by non-specific interactions such as buffers, salts and detergents. Ligands can also affect protein stability through interactions at a specific site. Differential scanning fluorimetry (DSF), also known as thermal shift, is a technique used to assess the relative binding of a ligand to a protein. A fluorescent dye is used to monitor the protein unfolding as the temperature is increased and is typically measured using a real-time quantitative polymerase chain reaction (qPCR) machine.[193] Solvatochromatic dyes (for example SyproOrange®) are quenched in aqueous environments, such as, a buffer containing stable, folded protein with few exposed hydrophobic surfaces. As the protein unfolds, hydrophobic sites are exposed which the dye binds to resulting in increasing fluorescence. A plot of fluorescence intensity as a function of temperature generates a sigmoidal curve (Figure 7.7).

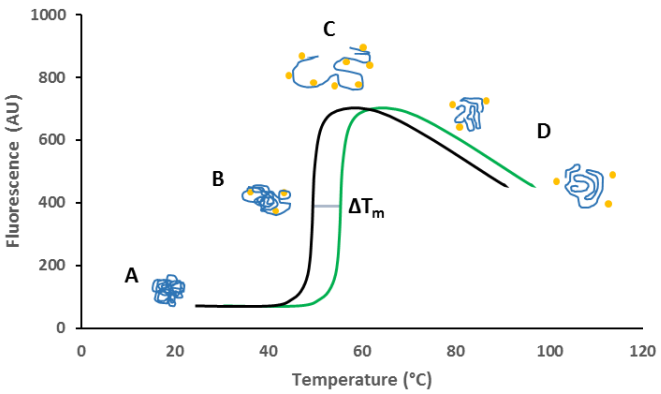


Figure 7.7 Representative sigmoidal curve of thermal denaturation of a protein (blue) with dye (yellow). A= melting protein, B= Dye binding to protein, C= Fluorescence peak, D= Protein aggregation and dye dissociation. Protein control (black sigmoidal curve) and Ligand + protein (green sigmoidal curve). ΔT_m= thermal shift.

The protein melt point (T_m) is indicated by the inflection point of the curve and is calculated using a Boltzman fitting function. A positive thermal shift (ΔT_m) indicates a ligand is binding to the protein resulting in a stabilising effect. Fragments are typically weak binders and induce relatively small positive ΔT_m in the range of 0.5–2.0 °C.[194] Negative, or destabilising ΔT_m are also observed in DSF. In some cases they can indicate optical interference due to ligand precipitation, usually at higher ligand concentrations; this can be confirmed through visual inspection. Alternatively, a negative ΔT_m can indicate a ligand is binding to a partially unfolded or pseudo-unfolded protein species to which the dye also binds, resulting in a negative ΔT_m. This can sometimes lead to false negatives if the ligand binds to both unfolded and folded protein species resulting in a negative thermal shift which masks the specific binding event. The ΔT_m of a ligand has been shown to be dependent on the affinity and concentration of the ligand.[192]

7.5.1.2 DSF assay optimisation

Prior to screening the ligands, DSF conditions were optimised for each experimental variable: buffer, dye concentration, protein concentration, DMSO content and ligand concentration. All experiments were run on an Applied Biosystems® 7500 Real-Time PCR system using Life Technologies 7500 software v.2.0.0 (T: 25–95 °C, step and hold for 30 seconds measuring at 1 °C intervals, 5x SyproOrange®).

First, a selection of buffers falling in a range between pH 7.4 – pH 7.6 were screened to determine the effect on the measured T_m of the protein in comparison to the storage buffer Tris pH 7.4. A review of the literature suggested the use of Tris in DSF assays should be avoided due to potential interference from the buffer in the assay due to pH

changes at different temperatures.[162] Dulbecco's phosphate buffer solution (DBPS) was selected as it did not exhibit a statistically different T_m versus the storage buffer Tris and because phosphate buffers are less prone to pH instability as the temperature changes (Figure 7.8A).[193]

Five protein concentrations between 2–20 µM were evaluated in the DSF assay (Figure 7.8B). A concentration of 5 µM was selected as it exhibited an optimal fluorescence response (20,000 AU) which would allow robust evaluation of changes in protein melting temperatures while minimising protein consumption.

A review of the literature suggested that the total DMSO content should be ≤ 2% *v/v* due to protein instability at higher DMSO percentages. However, this needs to be balanced against ligand solubility. DMSO content was assessed against protein stability and as expected a destabilising effect was observed with increasing DMSO content (Figure 7.8C). From the values obtained, 5% *v/v* DMSO was selected based on a balance between minimal destabilisation of the protein while maximising ligand solubility.

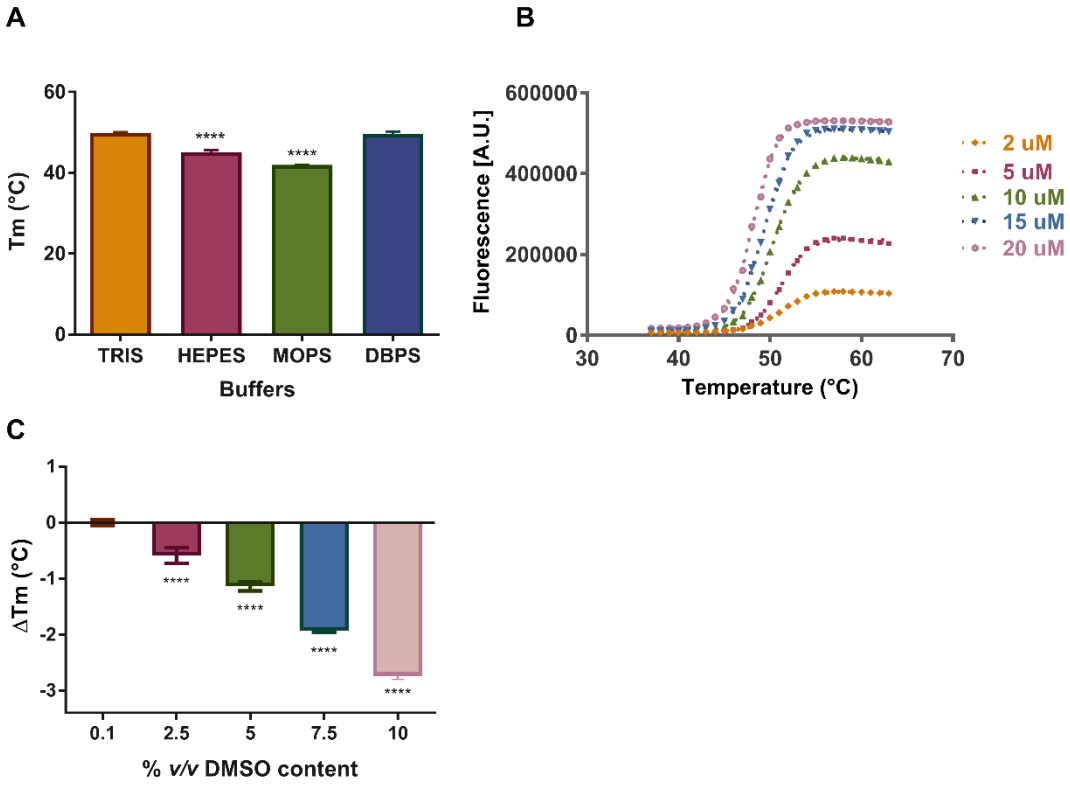


Figure 7.8 Optimisation of DSF conditions: buffer, protein concentration and DMSO content. All experiments were conducted in triplicate. Significance of changes evaluated by one-way ANOVA with Dunnetts correction; *P<0.05; **P<0.01; ***P<0.001; ****P<0.001 versus blank control (K1). **A** Bar graph of the thermal melt temperatures (T_m) of Keap1 in different buffers **B** Fluorescence versus temperature graph of melt temperatures at different protein concentrations **C** Thermal shift (ΔT_m, °C) with respect to DMSO content.

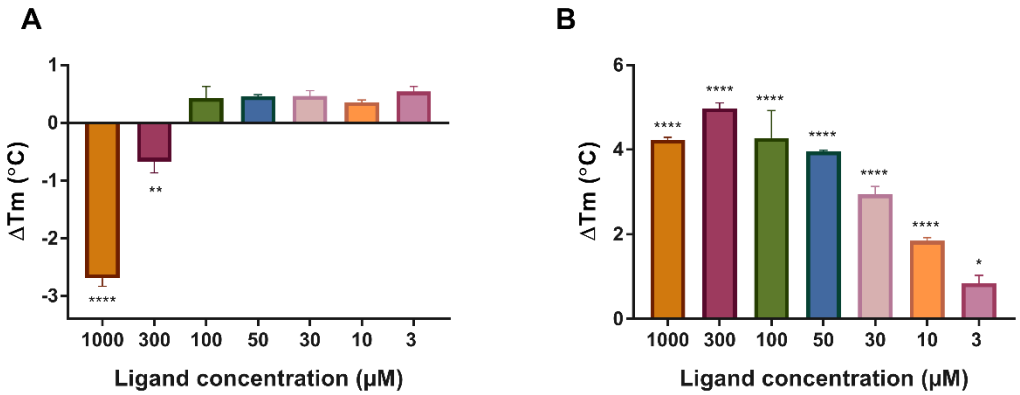


Figure 7.9 Optimisation of DSF conditions: ligand concentration. All experiments were conducted in triplicate. Significance of changes evaluated by one-way ANOVA with Dunnetts correction; *P<0.05; **P<0.01; ***P<0.001; ****P<0.001 versus blank control (K1). **A** Thermal shifts of **122** at different concentrations to determine optimal single dose concentration (5% DMSO, DBPS, 5 μM protein) **B** Screen of **125** at different concentrations to determine optimal single dose concentration (5% DMSO, DBPS, 5 μM protein).

Subsequently, with optimal buffer composition, protein concentration and % *v/v* DMSO determined, one ligand from each scaffold type, **122** and **125**, was selected to identify the optimal concentration for single dose screening. Six concentrations between the range 3–1,000 μM were evaluated (Figure 7.9A,B). The negative value observed for **122** suggests the compound is insoluble at concentrations above 100 μM. Of the concentrations screened, 50 μM (i.e. a ligand:protein ratio of 10:1) was selected, as a measurable thermal shift was observed. A 10:1 ligand:protein ratio has been reported as the optimum proportion which further supports the use of these concentrations.[193]

Finally, the DMSO content was re-evaluated with the inclusion of a ligand in the assay to measure ligand solubility at lower DMSO percentages. Three DMSO percentages were compared, 1%, 2% and 5%; a significant difference was not observed between the thermal shift values for either ligand with no obvious signs of insolubility (Figure 7.10A,B). This suggested the ligands were sufficiently soluble at the lower DMSO concentration. Thus to minimise the effects on protein stability, 1% *v/v* DMSO content was employed in all subsequent experiments. A summary of the optimised conditions is presented in Table 7.2.

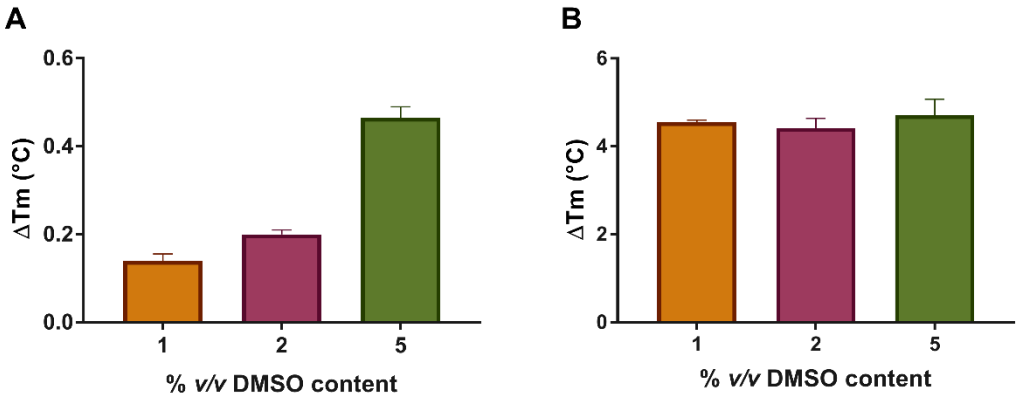


Figure 7.10 Optimisation of DSF conditions: re-evaluation of ligand concentration. All experiments were conducted in triplicate. Significance of changes evaluated by one-way ANOVA with Dunnetts correction; *P<0.05; **P<0.01; ***P<0.001; ****P<0.001 versus blank control (K1). **A** Re-evaluation of DMSO content for **122** (DBPS, 5 µM protein, 50 µM ligand) **B** Re-evaluation of DMSO content for **125** (DBPS, 5 µM protein, 50 µM ligand).

Table 7.2 Optimised thermal shift conditions. Well volume, 40 µL; Temperature range 25–95 °C measuring at 1 °C intervals; SyproOrange® was used as the protein stain.

Variable	Optimised condition
Buffer	DBPS
Dye concentration	SyproOrange®, 5x
Protein concentration	5 µM
Ligand concentration	50 µM
DMSO content	1%

7.5.1.3 Evaluation of biased fragments using DSF

The synthesised compounds **115**-**125** were screened using the optimised conditions described above. A summary of the T_m and ΔT_m of all compounds tested are shown in Table 7.3 and Figure 7.11. As expected there is a clear difference in binding affinities between the two series, demonstrated by comparing the thermal shifts of the naphthalene fragment **125** (4.55 °C) versus its phenyl analogue **122** (0.12 °C). This exemplifies the importance of the naphthalene vs. benzene ring for binding to the protein.

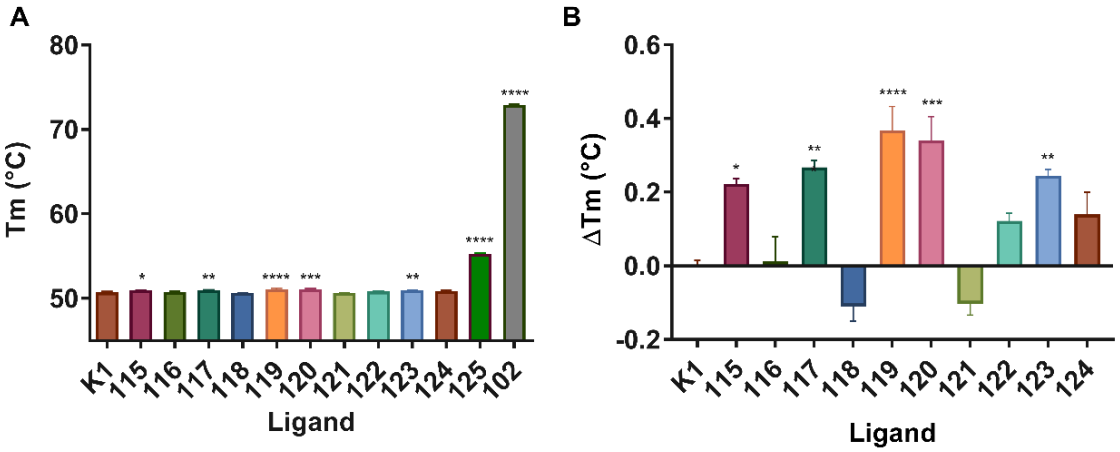


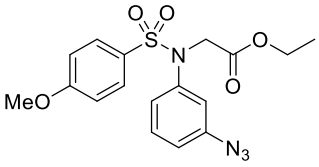
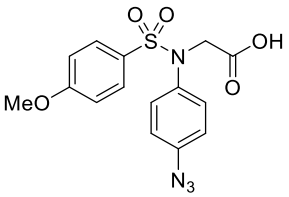
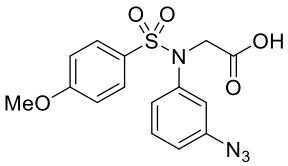
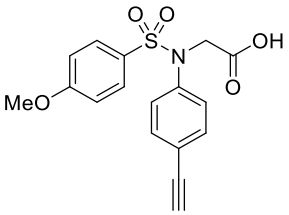
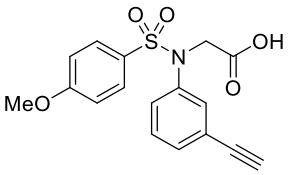
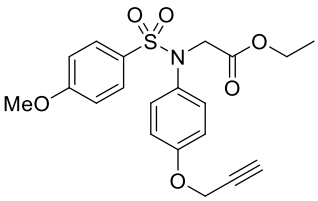
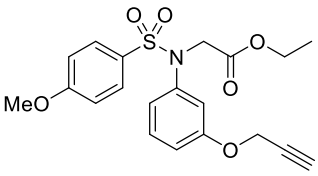
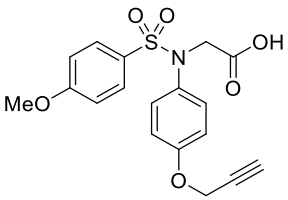
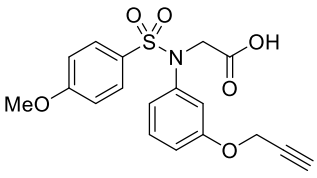
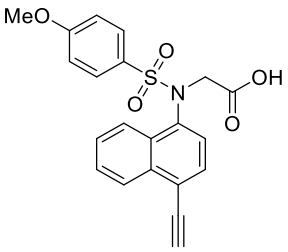
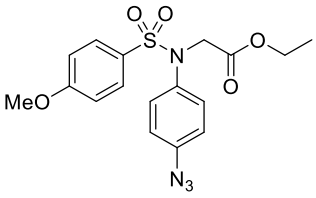
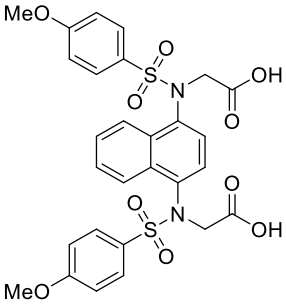
Figure 7.11 Evaluation of biased fragments using DSF. All experiments were conducted in triplicate with inhibitor **102** as the positive control. Significance of changes evaluated by one-way ANOVA with Dunnetts correction; *P<0.05; **P<0.01; ***P<0.001; ****P<0.001 versus protein (K1). Compound **102** was synthesised by Dr Nikolaos Dimitrios-Georgakopoulos **A** Thermal melting temperature (T_m, °C) of each ligand with respect to the protein (P). **B** Thermal shift (ΔT_m, °C) values with respect to the protein (P) for the phenyl series.

In comparison to ΔT_m values observed for typical fragments, the observed ΔT_m for the phenyl scaffolds were relatively small[185] ranging from -0.11 °C for **118** to the largest shift at 0.37 °C for **119**, whereas the thermal shift for the naphthalene scaffold was relatively high. A structure-activity relationship (SAR) analysis of the phenyl series indicated possible trends. However, due to the small library size, further validation is necessary. The negative ΔT_m values for compounds **118** and **121** suggest that the compounds may be insoluble or that the ligand is binding to a pseudo-unfolded form of the protein.

(i) Azide versus alkyne

A statistically significant difference in ΔT_m values between alkynes **117** and **122** (0.27 and 0.12 °C respectively) versus their azide analogues **116** and **121** (0.01 and -0.1 °C respectively) was observed. This suggests an alkyne is preferred at position 3- and 4- when R' = OH and the functional groups are directly attached to the aromatic core (i.e. no OCH_2 linker). Analogues of **116** and **121** with extended linkers were not synthesised and so analysis versus their respective alkyne ligands was not possible.

Table 7.3 Thermal shift values (ΔT_m) for all compounds with respect to the melt temperature of Keap1.

#	Structure	ΔT_M	#	Structure	ΔT_M
115		0.22	**121**		-0.10
116		0.01	**122**		0.12
117		0.27	**123**		0.25
118		-0.11	**124**		0.14
119		0.37	**125**		4.55
120		0.34	**102**		22.19

Values are highlighted using a heat map; green (high) to red (low).

(ii) Acid versus ester

The ester derivatives of the azide analogues **115** and **120** (0.22 and 0.34 °C respectively) exhibited a statistically significant difference in ΔT_m in comparison to the corresponding acids **116** and **121** (0.01 and -0.1 °C respectively). This may be due to the limited solubility of the acid analogues. In the alkyne series no clear preference for either acid or ester was observed.

(iii) Increasing linker length

No significant difference in ΔT_m was observed with variation of linker length, as exemplified by comparison of **117** (0.27 °C) to its analogue **119** (0.37 °C) as well as **122** (0.12 °C) to its analogue **124** (0.14 °C).

(iv) R-3 versus R-4

The observed ΔT_m of fragment **119** (0.37 °C) with an acid at R' and an alkyne substituted at R-3 with an extended linker (R = $OCH_2C\Xi CH$), is significantly larger than the thermal shift of its analogue where the alkyne is substituted at R-4, **124** (0.14 °C)**.** The corresponding analogues, with the shorter linker (R = $C\Xi CH$) **117** (0.27 °C) and **122** (0.12 °C), reflect the same trend. However, the observed difference in ΔT_m is not statistically significant. Intriguingly, the opposite observation is indicated where R' is an ester (R-4 > R-3, R=alkyne). Where R is an azide, there is not a significant difference between the ΔT_m values at different substitution positions.

In summary, this small study suggests an azide is preferred at position R-4 on the aromatic core in combination with an ester functionality **120**, possibly due to solubility issues for the acid analogues. An alkyne is preferred at R-3 on the aromatic core in combination with an acid functionality (Figure 7.12). This is in contrast to the findings published by Jain *et al*. who found a 1,4-substitution to be optimal.[146] The ΔT_m values for the short (**117**) and longer (**119**) linker lengths for the alkyne analogues were statistically comparable, thus both alkynes are represented below.

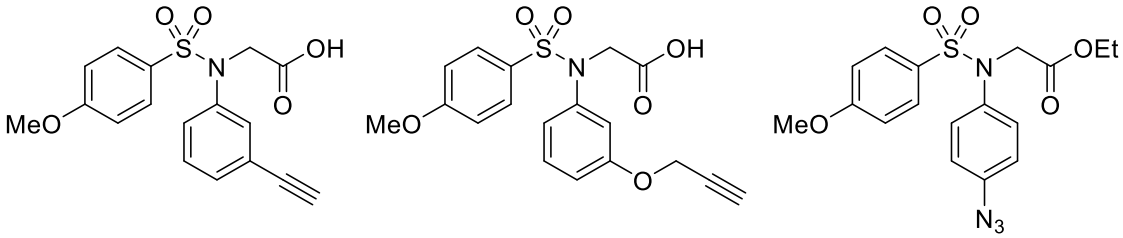


Figure 7.12 Phenyl scaffolds **117**, **119** and **120** identified from the thermal shift screen**.**

7.5.2 Fluorescence polarisation

7.5.2.1 Biophysical basis of Fluorescence polarisation

Fluorescence polarisation (FP) is a competitive binding assay using a fluorescently labelled inhibitor to measure binding affinity.[195] When a fluorescent molecule is excited with plane polarised light, the emitted light is in a fixed plane relative to the absorption plane. The fluorescence polarisation observed after irradiating a fluorescent solute is proportional to its rotational relaxation time. The rotational relaxation time is the time taken to rotate through an angle of 68.5° and is related to the solvent viscosity, temperature, the effective molecular volume of the fluorophore and the gas constant. Assuming constant temperature and viscosity, large molecules or complexes rotate more slowly resulting in emission of polarised light, for example a fluorophore-labelled peptide bound to a protein (Figure 7.13A). Conversely, smaller molecules rotate much faster and emit depolarised light, for example a free fluorophore (Figure 7.13B).

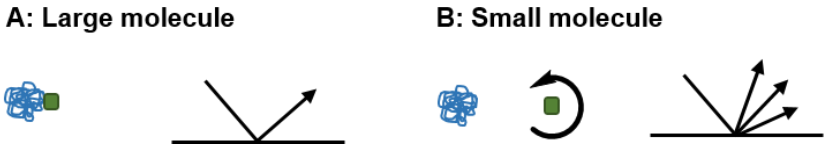
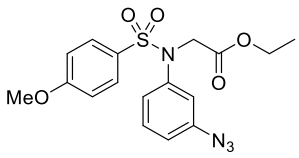
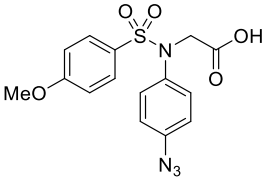
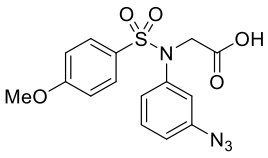
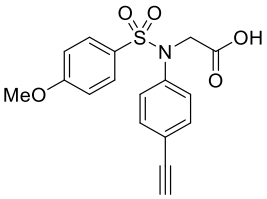
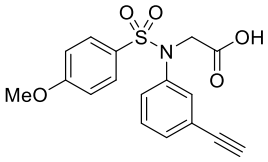
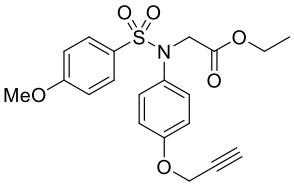
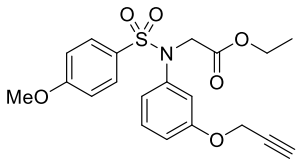
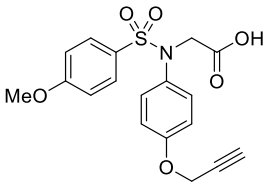
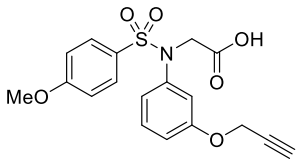
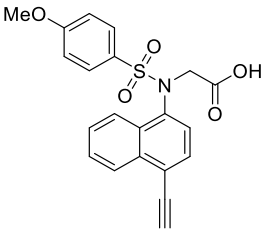
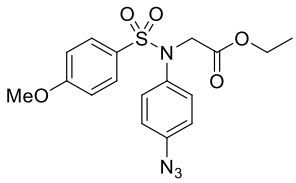


Figure 7.13 Schematic representation of florescence polarisation assay. **A** Polarised light from large molecules **B** Depolarised light from small molecules. The protein is represented in blue and the fluorophore is represented in green. Black straight lines represent angles of light absorption and emission. Curly black arrow represents free rotation.

7.5.2.2 Evaluation of biased fragments using FP

Conditions for the FP binding assay of the Keap1-Nrf2 interaction were followed as published by Hancock et *al*.[136] Each fragment was screened at two fixed dose concentrations, 10 and 100 µM, and the results are displayed in Table 7.4.

Table 7.4 Fixed dose percentage inhibition of the Keap1-Nrf2 interaction at two concentrations, 10 and 100 µM.

#	Structure	% inhibition 10 µM	% inhibition 100 µM	#	Structure	% inhibition 10 µM	% inhibition 100 µM
115		n/a	*a*	**121**		1.69	n/a
116		0.92	4.75	**122**		n/a	*a*
117		0.41	a	**123**		3.09	7.46
118		n/a	n/a	**124**		1.56	7.15
119		3.26	14.9	**125**		38.7	68.2
120		n/a	*a*				

a = Fluorescence interference. n/a = no inhibition. Values are highlighted using a heat map; green (high) to red (low).

The naphthalene series appeared to be more active than the phenyl series in the FP assay as demonstrated by the observed difference in the percentage inhibition between analogues **125** (38.7% @ 10 µM) and **122** (no inhibition @ 10 µM). This concurs with the DSF assay results.

Fragments from the phenyl series **116**, **119**, **123** and **124** exhibited an increase in the percentage inhibition between the two test concentrations (10 µM and 100 µM) which indicates binding to the protein. However, the difference is not statistically significant; experiments at higher concentrations would provide more information on the relative binding of each fragment. As observed with the DSF assay, **119** is the most active compound of the phenyl series displaying 3.26% inhibition at 10 µM.

Although the SARs observed in the DSF assay for the alkyne series (R = alkyne) are reflected in the FP assay, (for example, fragments with an acid at R' and an alkyne substituted at R-3, **117** and **119** (0.41% and 3.26% @ 10 µM) display a higher percentage inhibition than the corresponding analogues in which the alkyne is at R-4, **122** and **124** (no inhibition and 0.14% @ 10 µM respectively)) this difference was not statistically significant.

Intriguingly, the FP assay found one of the azide compounds **116** where R' is an acid (4.75% @ 100 µM) to be more active than its ester analogue **115** (no inhibition @ 100 µM). This is the opposite trend to that observed in the DSF assay. One possible reason for this discrepancy may be due to the increased solubility of the compounds under the FP assay conditions; a higher DMSO content is used for the FP assay in comparison to the DSF assay (11% versus 1%) which would help to solubilise the more polar acid analogues thus allowing detection of their activity. Fluorescence interference was observed for **115**, **117, 120** and **122** at 100 µM which suggests these compounds may be insoluble and form aggregates at higher concentrations causing optical interference.

The two most active ligands from each series, **119** and **125**, were further evaluated in dose-response assays to obtain IC_{50} values. Unfortunately, it was not possible to obtain a full dose-response curve for **119** due to compound insolubility at higher concentrations, which resulted in fluorescence interference. The partial curve is represented in Figure 7.14A. A full dose response was obtained for the more active naphthalene fragment **125** and is plotted alongside the published inhibitor **102** for comparison (Figure 7.14B). The IC_{50} of **125** was calculated to be 3.16 ± 0.89 µM versus the parent analogue inhibitor **102** with an IC_{50} of 14.7 ± 2.2 nM (published IC_{50} = 30 nM).[144] As anticipated the binding affinities of the scaffolds are significantly different but both sit within the range of affinities used in published kTGS experiments.[176]

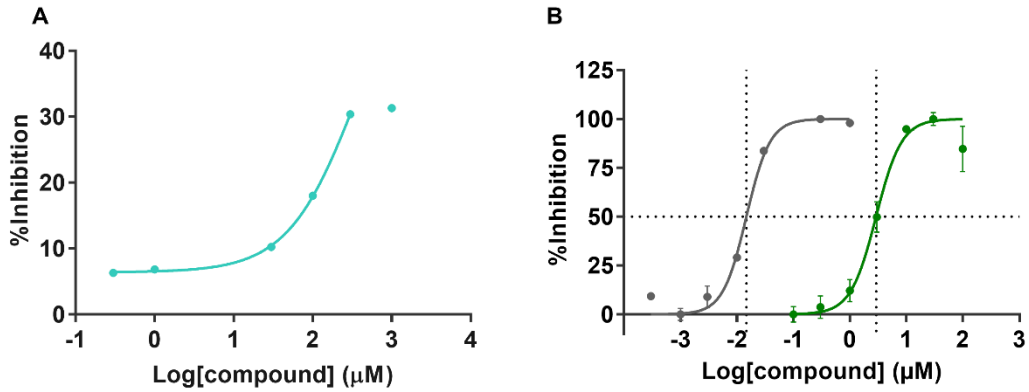


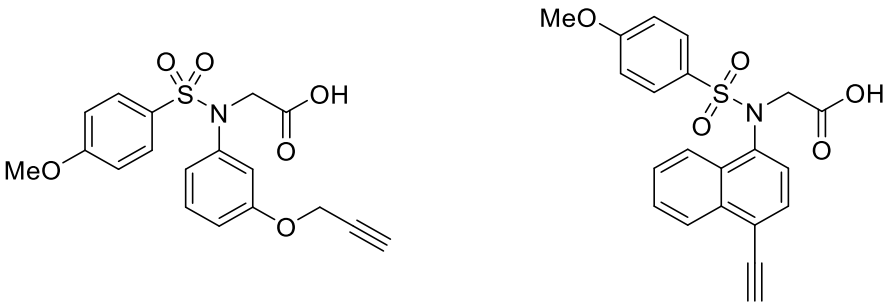
Figure 7.14 FP-based binding assay using a 7-mer Nrf2 ETGE motif derived peptide, FITC-DEETGEF-OH as the competitive inhibitor **A** Partial dose-response curve for **119 B** Dose-response competitive binding curves for **102** (Black); **125** (Green). Values were measured in triplicate. Binding curves were fitted using the non-linear regression functions in Prism. Compound **102** was synthesised by Dr Nikolaos Dimitrios-Georgakopoulos.

7.6 Conclusions

Synthesis of a small library of biased fragments based on two scaffolds derived from the published inhibitor **102**, a naphthalene scaffold and a phenyl scaffold, was achieved. One analogue of the naphthalene scaffold and six analogues of the phenyl scaffolds were obtained. Relative binding affinities of the biased fragments were evaluated using two complementary biophysical assays; FP and DSF. The small library of analogues indicated some SARs for phenyl type ligands, particularly that an alkyne is preferred at R-3 on the aromatic core in combination with an acid functionality (R'). Insolubility was a factor for the azide series as exemplified through the observed switch in relative affinities of the ester and acid functionalities (R') between the FP and DSF assays. The DSF assay uses a lower DMSO content which could impair solubility of the compounds, resulting in masking of their activity. This would need to be a consideration for further analysis within this series.

Compound **119** was identified as the most suitable ligand from the phenyl series to use in the kTGS assay as it behaved consistently in both biophysical assays with statistically relevant results in comparison to the controls. Compound **119** displayed an IC_{50} >300 μM in the FP competitive binding assay in comparison to the naphthalene fragment **125** which has a low micromolar Keap1-Nrf2 inhibition with an IC_{50} of 3.16 ± 0.89 μM (Figure 7.15). Both fragment affinities are within the range observed in the literature for fragments used in kTGS.

The following chapter presents the analysis and evaluation of the screening of both biased fragments against a focused library of azides in the presence of the Keap1 Kelch domain.



119

IC_{50} >300 µM
ΔT_m 0.4 °C

125

IC_{50} ~ 3 ± 1 µM
ΔT_m 4.6 °C

Figure 7.15 The structures of the selected biased fragments, **119** and **125**, from each series with their corresponding thermal shift (ΔT_m) and IC_{50} values calculated by FP.

CHAPTER 8
Kinetic target-guided synthesis screen

With the biased fragments in hand, **119** and **125**, the next steps towards evaluation of the biased kTGS approach in the development of novel inhibitors of Keap1, involved incubation of these biased fragments in the presence of Keap1 together with a diverse library of fragments containing a complementary reactive functional group. This process leads to the *in situ* generation of 1,3-dipolar cycloaddition products or 'hits' which have the potential to be Keap1 inhibitors. This chapter presents the analysis and evaluation of these hits, supported by a molecular modelling study, with subsequent selection of a shortlist for synthesis and validation.

8.1 Evaluation of biased fragments in a kTGS screen

Both biased fragments selected in section 7.6, **119** and **125**, were incubated in the presence of Keap1 against a focused library of 102 azide fragments. Multicomponent mixtures of azides were used to increase the screening throughput and also reduce the quantity of protein required.[184] Thus, the azide library was divided into groups of five, avoiding combinations where the masses of products were similar to circumvent complications arising in product detection. Two matched control samples were prepared using either Bovine serum albumin (BSA) or Dulbecco's Phosphate-Buffered Saline (DPBS) (vehicle) buffer in place of the Keap1 protein with equivalent ligand and library concentrations and DMSO content.

To support hit identification, the retention time of hits were compared to a synthetically prepared 1,4-disubstituted triazole. The mixtures were analysed on an Agilent Triple Quad LC-MS/MS system using both an MS2 scan and single ion monitoring (SIM) scan. A triazole scaffold was defined as a hit if the expected triazole mass was either not observed in the presence of the controls or there was an observable product amplification in the presence of protein over each of the controls. Product amplification was calculated by comparing the percent area of the product peak to the total ion count (TIC) and evaluating the difference between the protein and the controls.

The hit triazoles identified for the phenyl and naphthalene scaffolds are presented by their observed percentage product amplification in the presence of protein over each control (DPBS buffer and BSA) in Figure 8.1 and Figure 8.2 respectively.

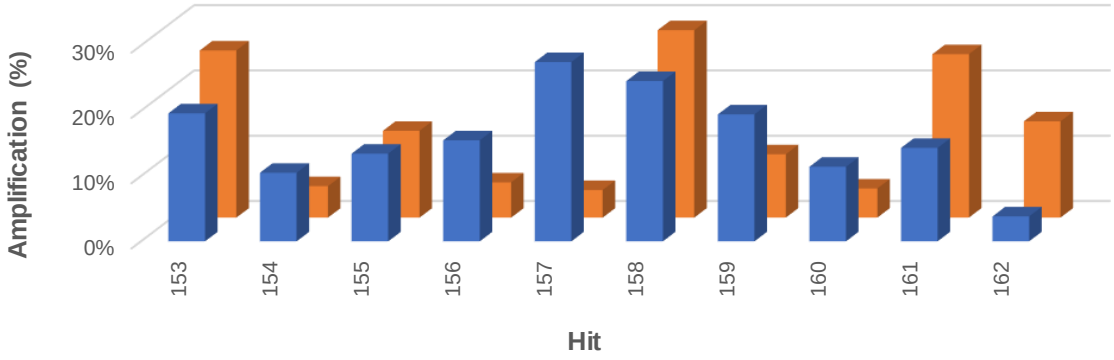
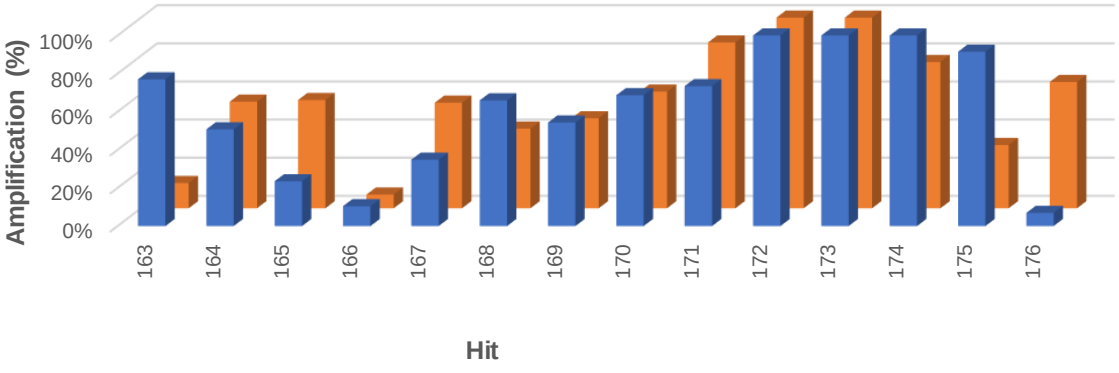


Figure 8.1 Summary of hits identified from the kTGS screen of the phenyl scaffold **119** versus a focused library of 102 azides. Amplified triazole formation in the presence of protein is expressed as a percentage versus the two controls, buffer (blue) and Bovine serum albumin (BSA) (orange), for each expected triazole.



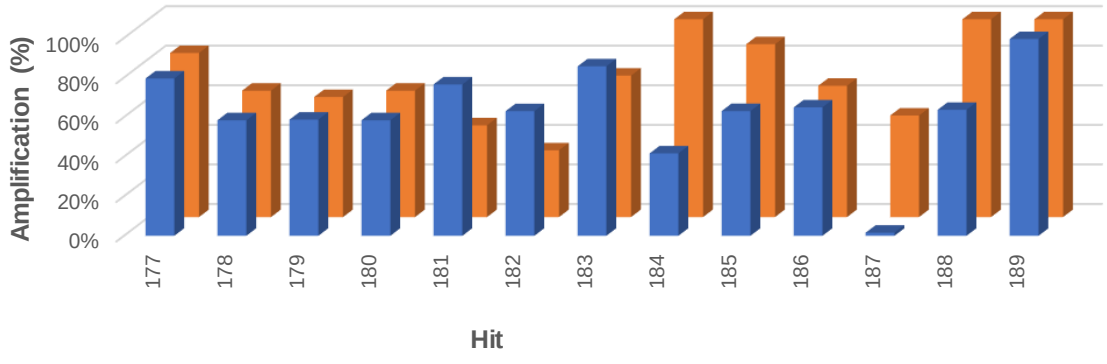


Figure 8.2 Summary of hits identified from the kTGS screen of the naphthalene scaffold **125** versus a focused library of 102 azides. Amplified triazole formation in the presence of protein is expressed as a percentage versus the two controls, buffer (blue) and Bovine serum albumin (BSA) (orange), for each expected triazole.

Table 8.1 Summary of the structures of the azide fragments which formed a 'hit' triazole with phenyl scaffold **119** in the presence of Keap1. The triazole structures are indicated by # and signify both 1,4 and 1,5 regioisomers unless specified. The azide fragments are indicated by their azide library reference by 'Azide ref'. The percent triazole formation (%TF) in the presence of protein is presented versus the two controls buffer and Bovine serum albumin (BSA). The relative magnitude of %TF values are assessed by a purple data bar; full cell (highest value) to empty cell (lowest value). The highest %TF value is highlighted in blue. The observed product amplification (PA), i.e. the difference between the percentage triazole formation in the presence of Keap1 versus the two controls, is expressed as a percentage increase (%PA) relative to each control and colour coded by a heat map using green (highest) to red (lowest).

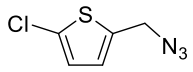
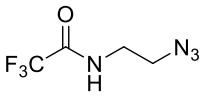
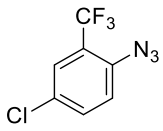
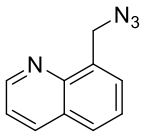
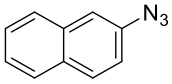
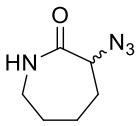
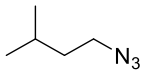
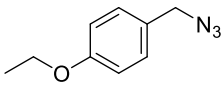
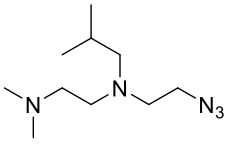
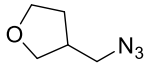
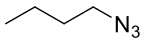
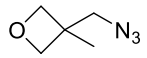
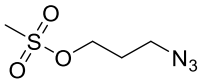
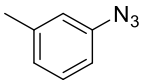
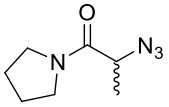
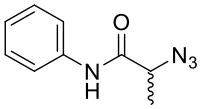
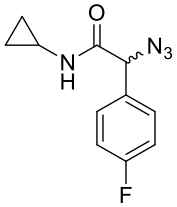
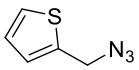
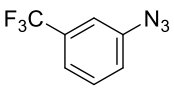
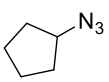
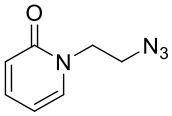
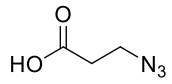
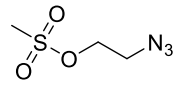
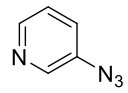
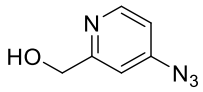
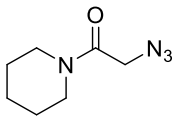
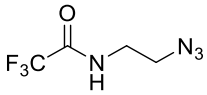
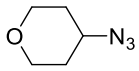
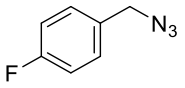
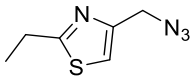
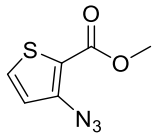
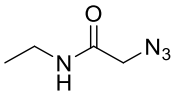
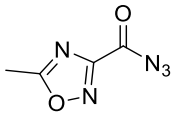
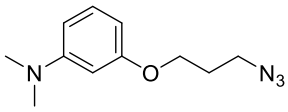
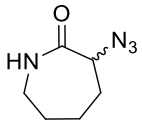
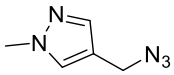
#	Azide Ref	Structure	%TF protein	%TF buffer	%TF BSA	%PA buffer	%PA BSA
153	AzD5-2		0.075%	0.060%	0.056%	20%	26%
154	AzB4-2		0.371%	0.332%	0.353%	11%	5%
155	AzD7-2		0.0594%	0.0514%	0.0515%	13%	13%
156	AzF3-2		0.0091%	0.0077%	0.0086%	15%	5%
157	AzE2-2		0.0162%	0.0118%	0.0155%	27%	4%
158	AzA5-2		0.033%	0.025%	0.023%	25%	29%
159	AzG1		0.0018%	0.0015%	0.0016%	19%	10%
160	AzA2		0.0042%	0.0037%	0.0040%	11%	4%
161	AzB4		0.026%	0.022%	0.019%	14%	25%
162	AzB1		0.015%	0.014%	0.013%	4%	15%

Table 8.2 Summary of the structures of the azide fragments which formed a 'hit' triazole with naphthalene scaffold **125** in the presence of Keap1. The triazole hits are indicated by # and signify both 1,4 and 1,5 regioisomers unless specified. The azide fragments are indicated by their azide library reference by 'Azide ref'. The percent triazole formation (%TF) in the presence of protein is presented versus the two controls buffer and Bovine serum albumin (BSA). The relative magnitude of %TF values are assessed by a purple data bar; full cell (highest value) to empty cell (lowest value). The highest %TF value is highlighted in blue. The observed product amplification (PA), i.e. the difference between the percentage triazole formation in the presence of Keap1 versus the two controls, is expressed as a percentage increase (%PA) relative to each control and colour coded by a heat map using green (highest) to red (lowest).

#	Azide ref	Structure	%TF protein	%TF buffer	%TF BSA	%PA buffer	%PA BSA
163	Az-5		0.10%	0.02%	0.08%	77%	13%
164	Az-3		0.20%	0.10%	0.09%	51%	56%
165	AzE3		0.11%	0.08%	0.05%	24%	57%
166	AzG3		0.02%	0.02%	0.02%	10%	7%
167	Az-10		0.29%	0.19%	0.13%	35%	55%
168	AzH1		3.48%	1.18%	2.03%	66%	42%
169	AzE1		1.12%	0.51%	0.59%	54%	47%
170	AzA3		1.03%	0.32%	0.40%	69%	61%
171	Az-16		0.14%	0.04%	0.02%	73%	87%
172	AzH3-2		0.004%	0.000%	0.000%	100%	100%
173	AzA3-2		0.01%	0.00%	0.00%	100%	100%
174	AzD2-2		0.76%	0.00%	0.18%	100%	77%

175	AzH5-2		32.06%	2.73%	21.42%	91%	33%
176	AzC7-2		0.14%	0.13%	0.05%	7%	66%
177	AzF1-2		0.21%	0.04%	0.04%	80%	83%
178	AzA1-2		0.74%	0.31%	0.27%	58%	64%
179	AzE5-2		1.43%	0.59%	0.56%	59%	61%
180	AzB4-2		0.74%	0.31%	0.27%	58%	64%
181	AzH4-2		0.31%	0.07%	0.17%	77%	46%
182	AzF2-2		0.20%	0.07%	0.13%	63%	34%
183	AzE1-2		0.19%	0.03%	0.05%	86%	72%
184	AzG1-2		0.06%	0.04%	0.00%	42%	100%
185	AzG3-2		0.23%	0.08%	0.03%	63%	87%
186	AzD6-2		0.32%	0.11%	0.11%	65%	66%
187	AzB1-2		1.43%	1.40%	0.69%	2%	51%
188	AzA5-2		0.03%	0.01%	0.00%	64%	100%
189	AzE6-2		1.62%	0.01%	0.00%	99%	100%

A cut-off, i.e. criteria, for either percentage triazole formation in the presence of protein or product amplification, was not employed to select the hits. This was because of the uncertainty of the relative importance of these values as factors in selection of hits; this is discussed in more detail later in this chapter. The hit rate for the naphthalene scaffold was higher than the hit rate observed for the phenyl scaffold. It is difficult to compare the hit rates versus published screens as the cut off criteria are not defined and so the hit rate is often variable.[176] Of the 102 azides screened, 27 formed triazoles which exhibited amplified product formation in the presence of the Keap1 Kelch domain relative to the controls, giving a 26% hit rate. Using the phenyl scaffold, 10 triazoles were identified as hits giving a 10% hit rate. The observed difference in hit rate between the two scaffolds likely stems from the difference in affinities of the two biased fragments, **119** and **125**. In line with this, the average percentage product amplification and average percentage triazole formation for the naphthalene scaffold, are both greater than that observed for the phenyl scaffold. The identification of twice as many hits presented a clear advantage for the higher affinity naphthalene scaffold over the phenyl scaffold. A greater number of hits allows a better evaluation of structural trends and a greater chance of finding a suitable lead. A higher percentage triazole formation is also advantageous as it enables a more robust detection of hit compounds.

The structures of the azide fragment of the hits identified for the phenyl and naphthalene scaffolds are summarised in Table 8.1 and Table 8.2 respectively. Alongside each fragment the percentage triazole formation in the presence of Keap1 and in the presence of the controls are presented as well as the respective percentage product amplification in the presence of Keap1 over each control.

The hits were evaluated and compared within and between each series considering the following factors: (i) structure-activity relationships (SARs) (ii) percentage triazole formation and (iii) 1,4 or 1,5 regioselectivity, with the aim of generating a short list of hits for subsequent validation.

(i) Structure-Activity Relationships

A diverse range of structures including ethers, amides, cyclic alkanes, aromatic groups and sulphonates were identified as hits for both scaffolds. This diversity likely stems from the nature of the kTGS approach. As previously, discussed the kTGS approach has the potential to probe the macromolecular complexity and dynamic nature of the protein presenting opportunities to ‘capture’ the protein in its lesser known conformations.[176] This could result in, as is reflected in this study, the identification of a

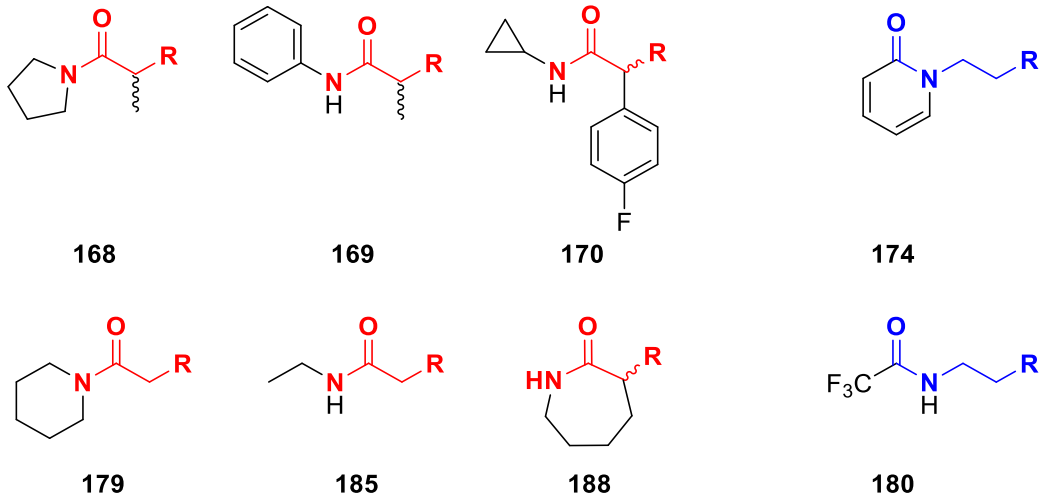
diverse set of potential inhibitor scaffolds, limited only by the diversity of the fragment library. Despite the diversity there are recurrent features within the triazole hits, particularly aromatic and amide functional groups which are discussed in more detail in the following paragraphs.

For both the naphthalene and phenyl scaffolds over half of the fragments identified contain an aromatic functional group which is unsurprising given the nature of the binding site. There are multiple arginine and tyrosine residues which could interact with the aromatic groups through cation-π and π-π interactions respectively.[143] Several of the aromatic azides identified as kTGS hits for the naphthalene series indicate a preference for a substitution at R-3 relative to the azide (Figure 8.3). This is not indicated in the phenyl series. However, as a smaller pool of hits were identified, such trends may not be as easily observed.

167 **172** **177** **178**

Figure 8.3 Structures of the azide fragments with aromatic scaffolds identified in the naphthalene series. The scaffolds indicate a preference for an R-3 substitution. R represents the naphthalene fragment **125** connected by a triazole linker.

An amide bond is another recurrent feature which was identified for both biased scaffolds. Eight fragments containing amide bonds with different *N*-substitutions were identified for the naphthalene scaffold in the *in situ* experiments (Figure 8.4).



168 **169** **170** **174**

179 **185** **188** **180**

Figure 8.4 Trends identified in the hits from the naphthalene scaffold **125** centred around an amide bond, methylene linker (red) versus ethylene linker (blue). R represents the naphthalene fragment **125** connected by a triazole linker.

Triazole hits, **168**, **169**, **170, 179, 185** and **188,** are analogues containing an amide backbone connected to the triazole via a methylene linker and various *N*-substitutions. Analogous hits **174** and **180** are also amides but with the azide connected via an *N*-ethylene bridge**.** Two of these azide fragments were also identified as hits in the *in situ* experiment with the phenyl scaffold, **154** and **158**. This indicates an amide scaffold is a potentially important functional group to be included as part of a novel Keap1-Nrf2 inhibitor structure.

Previously reported examples of kTGS for ligand optimisation have intentionally biased the fragment library to contain features of known ligands. When the kTGS experiments are run, these biased ligands were identified as hits which consequently provided a proof of concept and thus validation of the approach.[181] This azide library was not spiked with fragments which resemble the published inhibitor from which the biased fragments were derived. However, analysis of the hits found that structures which resemble the published inhibitor were observed for both biased scaffolds**.** One of the hits, **175**, identified in the *in situ* experiments with the naphthalene scaffold **125**, contained an ethylene acid structure which bears a similarity to the acid side chain found in inhibitor **102** (Figure 8.5). This structure could consequently form similar H-bond or salt bridge interactions to its parent analogue through the triazole and carboxylate motifs. For the phenyl fragment **119,** a hit with a *p*-ethoxybenzyl ether azide fragment **160,** was identified which bears a similarity to the *p*-methoxysulphonamide side chain in the published inhibitor **102** (Figure 8.5). 'Selection' of these fragments as hits by the kTGS experiments substantiates the methodology and provides, in part, a proof of concept that the kTGS approach can be applied to Keap1.

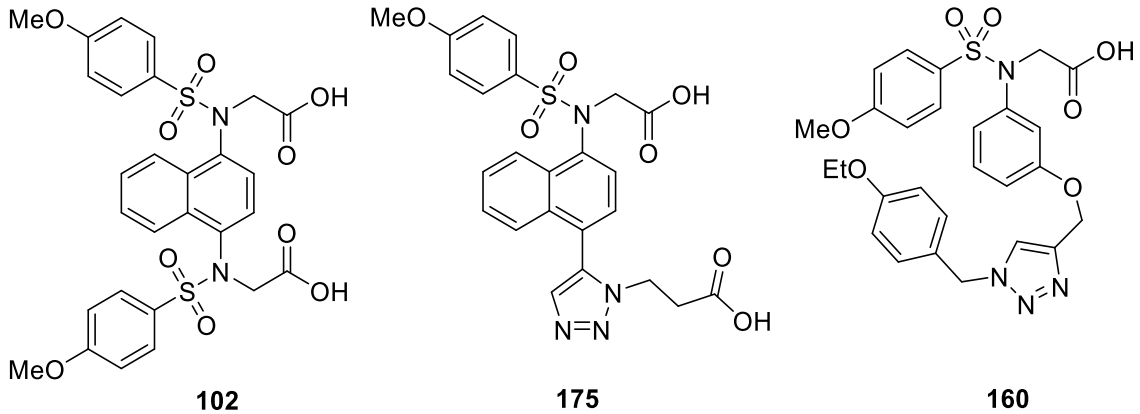


Figure 8.5 Structurally similar hit compared to the published inhibitor **102,** identified for the naphthalene scaffold, **175** and identified for the phenyl scaffold, **160**.

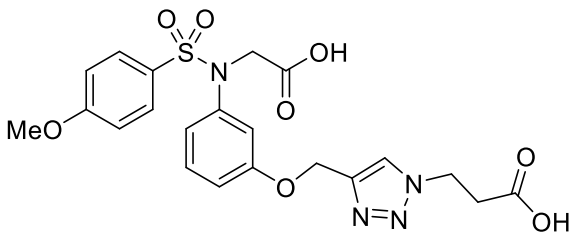
It was hypothesised in section 7.2 that the difference in affinities between the two scaffolds could result in identification of different inhibitor structures because the biased fragments may adopt different binding orientations. Although the same azide fragments were identified in both series (for example **154** and **180** (AzB4-2) and **158** and **188** (AzA5-2)) and some different azide fragments but with common features were identified in both series (for example **155** (AzD7-2) identified as a hit for the phenyl scaffold shares similar substitution patterns with naphthalene hits **172** (AzH3-2) and **182** (AzF2-2)), without structural data it is difficult to assess whether this is because the fragments are binding in similar or different conformations.

(ii) Percentage triazole formation

Grimster *et al*. reported a correlation between the amount of product formed and its affinity.[196] The results from this investigation were analysed to assess whether this observation held true for this system. The naphthalene fragment **125**, which has a higher affinity for Keap1, showed a greater average percentage triazole formation compared to the lower affinity phenyl fragment **119.** This indicates that a higher affinity fragment results in a higher percentage triazole formation which correlates with Grimster's findings. However, analysis of the percentage triazole formation of structurally related hits – fragments which mimic the published inhibitor and the SARs discussed previously – do not conclusively indicate a potential relationship between percentage triazole formation and affinity. This analysis is discussed in more detail in the following paragraph.

The hits which mimic the published inhibitor, **175** and **160** are discussed first. For **175** (a hit identified for the naphthalene scaffold) the percentage triazole formation in the presence of protein was 10-fold larger than the average for the series, highlighted in blue in Table 8.2. This large increase, associated with a structure which resembles a known inhibitor, indicates a potential relationship between percentage triazole formation and affinity. Yet, this higher percentage triazole formation was not observed for the corresponding hit identified for the phenyl scaffold, **160,** where the percentage triazole formation in the presence of Keap1 was below the series average at 0.0042%. The difference in contributions to binding, between the two azide fragments, could account for this; the acid side chain (i.e. azide fragment AzH5-2) has been shown to be much more important to binding compared to the sulphonamide side chain (i.e. azide fragment AzA2-2).[144] The equivalent acid mimic for the phenyl scaffold, **190**, was not observed in the phenyl hits. This is potentially due to the result of a false negative.

Additional steps can be introduced in the method to help reduce the occurrence of false negatives and is discussed further in the chapter, in relation to this specific example.



190

Figure 8.6 Equivalent acid mimic for the phenyl scaffold, **190** (AzH5-2), wasn't observed.

The hit with the largest percentage triazole formation in the presence of Keap1 for the phenyl scaffold was **154**, a trifluoromethyl amide, and is highlighted in blue in Table 8.1. However, the background percentage triazole formations, i.e. the percentage triazole formation of the controls, were also comparably high. This results in a low percentage product amplification which indicates a low catalytic contribution towards formation of the product from the protein. This structure is also a hit for the naphthalene scaffold but exhibits an average, rather than an elevated percentage triazole formation for the series and an average percentage product amplification. Despite the low percentage product amplification observed for **154**, this is an intriguing structure as it could be a more metabolically stable replacement for the carboxylic acid side chain.

Finally, to further analyse the potential relationship between percentage triazole formation and inhibitor affinity, the percentage triazole formation exhibited by the SAR scaffolds identified previously were analysed (aromatic scaffolds in Figure 8.3 and amide scaffolds in Figure 8.4). The percentage triazole formation observed for the 8 naphthalene scaffolds with an amide functionality are all above average for the series with the exception of **188** which is below average. This indicates a potential correlation between percentage product formation and affinity. However, the first SAR which contained hits with an aromatic functionality are evenly distributed across the observed range for percentage triazole formation for both scaffolds. This observation indicates there is not a correlation between percentage product formation and affinity. However, this relationship cannot be confirmed until the hits are validated and their affinity assessed.

As such, although there are several positive indicators to suggest a relationship between percentage triazole formation and affinity, it is not observed universally across the results. Thus, this analysis does not fully support a correlation between the two

measurements. To fully assess this relationship, an investigation through re-synthesis and bio-evaluation of a selection of hits is required. If this relationship was validated it could lead to an early indicator of high affinity scaffolds. This relationship was not actively explored any further as part of this thesis but could be investigated as part of the future work.

(iii) Regioisomer selectivity

It has been observed in published studies that the biological target in *in situ* experiments can govern the absolute regioselectivity of the Huisgen cycloaddition reaction, i.e. formation of either 1,4 or 1,5-disubstituted triazoles. For example, Manetsch *et al*. noted the hits formed in the presence of acetylcholinesterase (AChE) were exclusively 1,4-regiosiomers.[184] To assess the regiospecificity of Keap1 in this kTGS study, the hits from both scaffolds were analysed through comparison of the retention time against the 1,4-synthetic standard. The results obtained were inconclusive. It was not possible to establish the regioisomer assignment of the product formed in the presence of Keap1 due to slight fluctuations in the retention time of the product formed in the presence of Keap1 versus the 1,4-synthetic control. The differences in retention time could indicate the formation of the 1,5-regioisomer or alternatively, could be due to lack of reproducibility of the method. Future experiments could include: (i) a 1,5-synthetic standard to provide a more thorough assessment of the regiospecificity of the *in situ* coupling and (ii) further LC/MS method optimisation to ensure robust and reproducible results (to assess the extent of resolution of the regioisomers, with, if necessary, subsequent optimisation to ensure resolution). The overall impact on the total experimental time would need to be considered, as a small increase in an individual run has a significant impact on the overall screen time required for a library. This series of optimisations would refine hit identification and allow determination of regioisomer preference in the presence of protein which could potentially reduce unnecessary synthesis of regioisomers.

This analysis also highlighted three additional areas of method development: (i) use of BSA as a control protein (ii) method sensitivity and (iii) occurrence of false negatives due to multicomponent reactions, which are discussed below.

(i) Use of BSA as a control protein

A further area which requires optimisation, that has an impact on hit identification, was identified in relation to the unrelated protein control, BSA. BSA was used as a control to

identify non-specific catalysis as it is cheap, stable and unrelated to the protein of interest, Keap1. Ideally, little to no background reaction should be observed. However, BSA is capable of templating reactions in its own right.[197] There are several examples for both scaffolds where a high percentage product formation is observed in the presence of BSA. For example, **163** (a naphthalene scaffold hit) exhibits a 0.08% product formation in the presence of BSA relative to the protein and buffer at 0.10% and 0.02% respectively, and **157** (a phenyl scaffold hit) exhibits a 0.0155% product formation in the presence of BSA relative to the protein and buffer at 0.0162% and 0.0118% respectively. Furthermore, recent work within the Wells group involving biomimetic HPLC, found naphthalene structures have a high binding affinity for albumins (unpublished results). This result further indicates BSA may not be an optimal control protein for use with these scaffolds. However, it might be a useful assay for screening out potential hits with high BSA binding. Future work could involve bio-evaluation of the fragment library with respect to BSA (for example in a DSF assay) to assess its use as an unrelated protein and consequently its use as a control.

Other control experiments could also be investigated. For example, proteins containing a point mutation known to abrogate ligand binding have also been used as a control in kTGS experiments; specifically, in efforts to identify inhibitors of Bcl-X_L, two proteins, in which amino acids essential for binding were mutated and incubated with the biased fragments. If the templated products were not observed to form in the presence of the mutant controls, it confirmed the fragments were binding in the desired binding site.[181]

Alternatively, a control experiment could be incorporated which does not rely on the use of an unrelated protein. For example, towards the identification of inhibitors for carbonic anhydrase, a control experiment was used where a known inhibitor (ethoxazolimide) was also incubated with the library fragments, biased fragment and biological target.[169] If the percentage triazole formation was observed to decrease relative to the control which did not contain the known inhibitor, it indicated the scaffold was binding to the biological target thus validating the scaffold as a hit.

(ii) Method sensitivity

It is important to note that the robust detection of hits and differentiation of product amplification due to the presence of protein is reliant on the sensitivity of the analytical techniques. Ultimately, the analytical techniques employed in the kTGS approach need to be able to accurately detect the presence of small quantities of new molecules in the presence of large amounts of protein and ligands. The sensitivity of the analytical

technique was greatly improved through development of the LC/MS-SIM method by Manetsch *et al*.[184] However, with such small quantities of product formation, particularly for the phenyl series, where a low percentage triazole formation and small product amplifications were observed, it is important to establish the lower detection limit and sensitivity of the method to allow robust identification of hits. Future work could involve method development through a calibration curve which measures the relationship between concentration and relative response of analyte using the ion count. This would give an estimate of the signal to noise ratio and thus give a measure of the method sensitivity. Ultimately this may help to provide a clearer definition of a hit and consequently improve hit identification.

(iii) False negatives

False negatives can be attributed to multiple and sometimes related reasons as discussed in section 6.2.1.2 and highlighted earlier by phenyl scaffold **190**. Additional steps can be introduced in the procedure used to identify hits to help reduce the occurrence of false negatives. For instance, although multicomponent reactions speed up the hit identification process and reduce the amount of protein required, competitive binding from fragments in the same well can result in false negatives due to hindered triazole formation, resulting in undetectable amounts of product. Running parallel kTGS experiments, where the library is split into orthogonal combinations, mitigates the identification of false negatives while retaining the economy of multicomponent reactions.[198] Future work could involve optimisation of the methodology to incorporate this step for use in subsequent experiments.

Despite the areas highlighted above which require method development, the identification of fragments which mimic the published inhibitor provides, in part, a proof of concept that the kTGS approach can be applied to Keap1. However, re-synthesis and bio-evaluation of a shortlist of hits is required to validate the approach. To support the kTGS investigation and aid selection of a shortlist of hits for validation, a parallel molecular modelling study was carried out using the same biased scaffolds and fragment library.

Molecular modelling is a frequently used structure-based drug discovery (SBDD) technique because of its ability to accurately predict the conformation adopted by ligands within a protein binding site.[199] There are published examples where modelling studies have been successfully utilised to identify novel inhibitors of the Keap1-Nrf2 PPI.[149] It was envisaged the modelling would provide a complementary ranking of the

predicted binding affinities, SARs of the triazole structures and an insight into the binding conformations adopted by each triazole to help with a short list of hits for validation. Comparison of the results obtained in both studies could also provide some insights to optimise and improve the kTGS methodology.

8.2 Molecular modelling

AutoClickChem software[200] was employed to generate molecular models of all possible triazole products, including stereoisomers (*S* and *R*) and regioisomers (1,4 or 1,5), from a combination of each biased fragment **119** and **125** using the same focused azide library used in the kTGS study. The crystal structure of the published inhibitor **101** bound to Keap1 (PDB code: 4IQK) was used to dock the triazoles. Full experimental details are discussed in section 10.4.

(i) Complementary predicted binding affinities and SARs

The compounds were docked into the binding pocket of the Keap1 Kelch domain using Autodock Vina with default settings.[201] The binding pocket was defined by a 24 x 20 x 20 Å grid centred on the Nrf2 binding site. The predicted binding affinities generated for each of the possible ~230 triazole products were plotted for both scaffolds with the ligands ranked in order of binding affinity (Figure 8.7). As illustrated by the plots of the two scaffolds, the naphthalene series is predicted to have a greater affinity than the phenyl series; this corresponds with the behaviour of the biased fragments **119** and **125** in the biophysical assays (section 7.5).

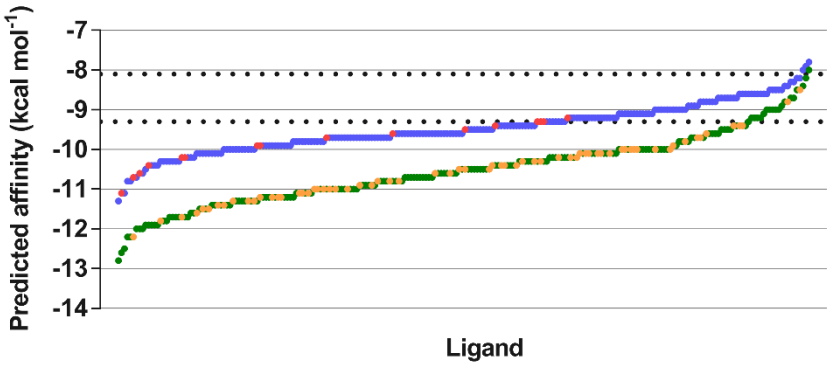
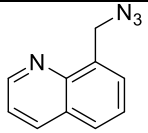
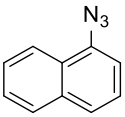
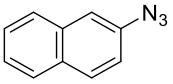
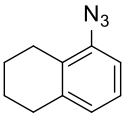
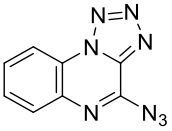
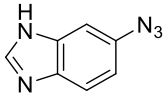
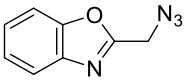
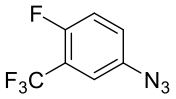
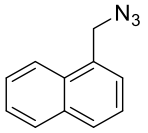
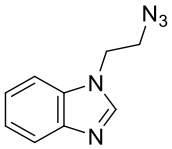


Figure 8.7 Plot of the predicted binding affinities for each triazole generated from the molecular modelling study. Naphthalene scaffold (green) and phenyl scaffold (blue). The kTGS hits are also highlighted for each series; kTGS hits identified for the naphthalene scaffold (yellow) and kTGS hits identified for the phenyl scaffold (red). The two horizontal dotted lines indicate the predicted binding affinities of the biased fragments **119** (8.1 kcal mol^{-1}) and **125** (9.3 kcal mol^{-1}).

The top hits from the molecular modelling study were defined by the point of inflexion of the curve, equating to 14 hits for both biased fragments. The structure of the azide fragment for each hit and regioisomer preference for the triazoles are summarised in Table 8.3 and Table 8.4 for the phenyl and naphthalene scaffolds respectively.

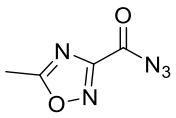
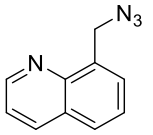
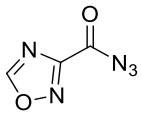
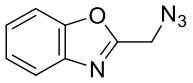
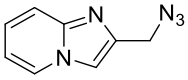
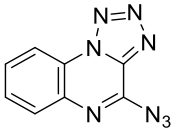
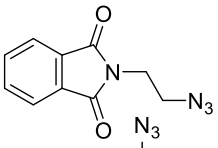
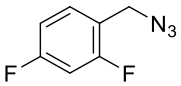
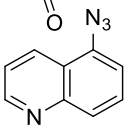
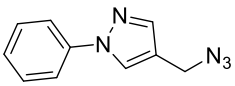
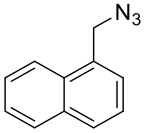
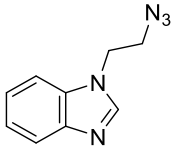
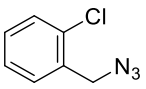
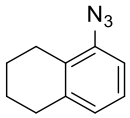
Table 8.3 Top molecular modelling hits for the phenyl scaffold **119**.

#	Azide ref	Structure	Isomer	#	Azide ref	Structure	Isomer
156	AzF3-2		Both	**194**	AZA04		1,4
157	AzE2-2		Both	**195**	AzB02		1,4
191	AZC03		Both	**196**	AzE4-2		1,5
192	AzB6-2		1,5	**197**	AzB5-2		1,5
193	AZH02		Both	**198**	AzC6-2		1,4

The triazole structures are indicated by # and the regioisomer preference is specified under 'isomer'. 'Both' indicates both 1,4 and 1,5-regioisomers were identified in the top hits. The azide fragments are indicated by their azide library reference, 'Azide ref'.

Four of the top 14 hits for the phenyl fragment **119** included both the 1,4 and 1,5-regioisomers resulting in 10 different scaffolds. Analysis of the remaining triazole scaffolds for this series found an even distribution of 1,4 and 1,5 regioisomers (Figure 8.8). All 14 of the scaffolds for the naphthalene fragment **125** were from different azide fragments and intriguingly all 14 hits were observed to be 1,5-regioisomers. Furthermore, this trend extends to the top half of the entire naphthalene library indicating, for this scaffold, a 1,5-regiosiomer is preferred for binding in the Keap1-Nrf2 binding surface (Figure 8.8). This information was not discerned in the kTGS study highlighting the requirement for future LC-method optimisation and inclusion of a 1,5-synthetic standard to improve hit identification. This predicted regio-preference can be validated through comparison of the affinity of 1,4 vs 1,5 regioisomers through subsequent re-synthesis and bio-evaluation.

Table 8.4 Top molecular modelling hits for the naphthalene scaffold **125**.

#	Azide ref	Structure	Isomer	#	Azide ref	Structure	Isomer
186	AzD6-2		1,5	**205**	AzF3-2		1,5
199	AzH2-2		1,5	**206**	AzB6-2		1,5
200	AzC4-2		1,5	**207**	AZC03		1,5
201	AzB2-2		1,5	**208**	AzC1-2		1,5
202	Az-11		1,5	**209**	AzE3-2		1,5
203	AzH02		1,5	**210**	AzC6-2		1,5
204	Az-19		1,5	**211**	AzB02		1,5

The triazole structures are indicated by # and the regioisomer preference is specified under ‘isomer’. The azide fragments are indicated by their azide library reference, ‘Azide ref’.

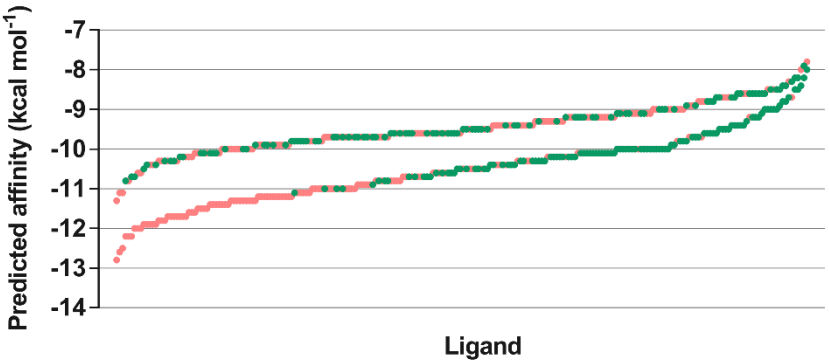


Figure 8.8 Analysis of the regioisomer preference governed by Keap1 for the phenyl scaffold (top) and naphthalene scaffold (bottom); 1,5-regioisomer (pink) versus 1,4-regioisomer (green).

Analysis of the structures of the top triazole hits from the molecular modelling study found that, in contrast to the kTGS study, there was little diversity observed between the azide scaffolds for either of the biased fragments. Predominantly mono or bicyclic aromatic or heterocyclic aromatic structures were identified for both series. This is foreseeable as these groups would form favourable interactions with the multiple arginine and tyrosine residues contained within the pocket, as discussed earlier in the chapter. This similarity in features between hits was unsurprising as, in this instance, only one protein conformation of Keap1 was used in the modelling study. To capture the dynamic nature of the protein in its different conformations, amino acid side chain flexibility can be considered in the docking calculations or, protein structures derived from several different crystal structures, can be used. However, these approaches are more complex, require longer computational time and may be more difficult to analyse. This highlights an advantage of the kTGS approach as it allows assessment of the protein in a dynamic form as previously discussed.

The triazole hits identified from the kTGS study are highlighted on the plot of relative binding affinities generated by the molecular modelling study (Figure 8.8). Both the 1,4 and 1,5 regioisomers are highlighted because it was not possible to clearly distinguish which regioisomer was formed in the presence of Keap1 from the kTGS experiment. It was anticipated there would be a degree of overlap in the hits between the two studies because the biased fragments were predisposed towards the crystal structure and so would influence the binding orientation, particularly for the higher affinity naphthalene fragment. Thus, identification of similar scaffolds by both studies would validate the method used to identify the hits for the kTGS approach. However, as can be seen from the plot and tables, only one azide fragment **186** (AzD6-2) was identified by both studies in the naphthalene series and two azide fragments **157** (AzE2-2) and **156** (AzF3-2) (as both the 1,4 and 1,5-regioisomers) were identified by both studies for the phenyl scaffold. Multiple peaks were not observed in the LC/MS-SIM traces from the kTGS experiments. However, this could be due to lack of resolution between the two isomers or one isomer being favoured. Alternatively, these bicyclic scaffolds may have a slow off-rate resulting in poisoning of the protein; this could reduce triazole formation and potentially result in formation of undetectable amounts of product.[161]

Further analysis of the scaffolds identified by the naphthalene series found that, although individual fragments do not concur, there are similar structures with similar substitution patterns across both studies (Figure 8.9).

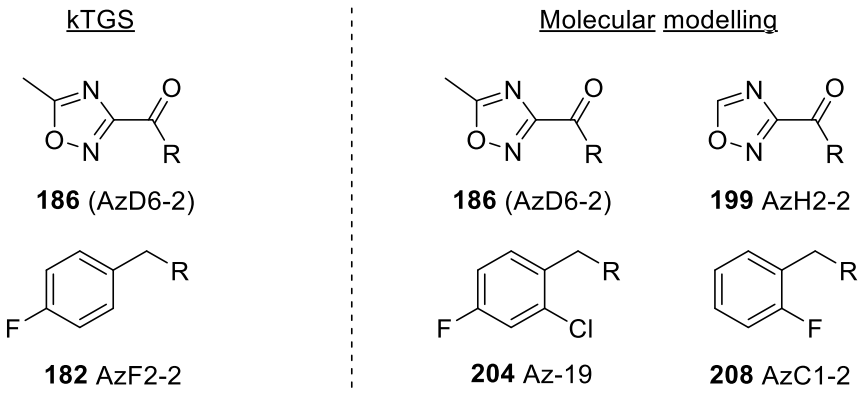


Figure 8.9 Similar monocyclic structures identified by the molecular modelling study and the kTGS study for the naphthalene scaffold **125.**

This indicates there is more agreement between the two studies than was initially obvious. For example, although the azide fragment **182** (AzF2-2) is not part of a scaffold in the top triazole hits of the molecular modelling study, two structures with similar substitution patterns **204** (Az-19) and **208** (AzC1-2) are identified.

This was a much smaller agreement than anticipated for the higher affinity naphthalene scaffold with only 15% of hits overlapping, compared to 20% for the lower affinity phenyl series. The structures missing from the kTGS study which were identified by the molecular modelling study were mainly bicyclic aromatic structures. These missing structures could be due to over emphasis of π-stacking interactions in the binding pocket. Alternatively, the missing structures could arise because the kTGS experiments are run as competitive, multicomponent assays. This differs to the molecular modelling study in which all potential products are ranked. Thus, as discussed, competitive binding from fragments in the same well could result in hindered triazole formation consequently resulting in undetectable amounts of product, and thus false negatives.

(ii) Binding conformations

The biased fragments **119** and **125** are predicted to adopt the same orientation with the carboxylic acid side chain occupying subpocket P2 (Figure 8.10A,B and D,E). Intriguingly, although P1 has been indicated by published studies as the more important pocket to occupy, the acid side chain is predicted to occupy polar subpocket P2. Comparison of the binding orientations of the top 14 hits between the two series indicates the two biased fragments **119** and **125** are adopting different poses when they are part of the full triazole structures. The lower affinity phenyl fragment flips and adopts an orientation where the acid side chain occupies polar subpocket P1 (Figure 8.10F). The naphthalene series remain orientated with the acid occupying subpocket

P2 and the triazole pointing into subpocket P1 (Figure 8.10C). It is unclear why the two series adopt different orientations, though this could be driven by steric factors with the naphthalene core locking the fragment into a more rigid pose with less flexibility than the phenyl core. This could also be a contributing factor for the lack of concurrence between the kTGS study and the molecular modelling study, because in the kTGS study the protein is dynamic and able to adopt less common conformations and so can accommodate the naphthalene fragment in a less sterically strained orientation.

The main role of the triazole is to act as a linker which forms as a product of the reaction between the azide and alkyne but it is also capable of forming interactions in its own right.[176] Occupation of an important polar subpocket P1, raises the question as to how much the triazole functional group is influencing the binding conformation of the products. It could be biasing the identification of potential hits for the naphthalene and phenyl scaffolds within the molecular modelling study.

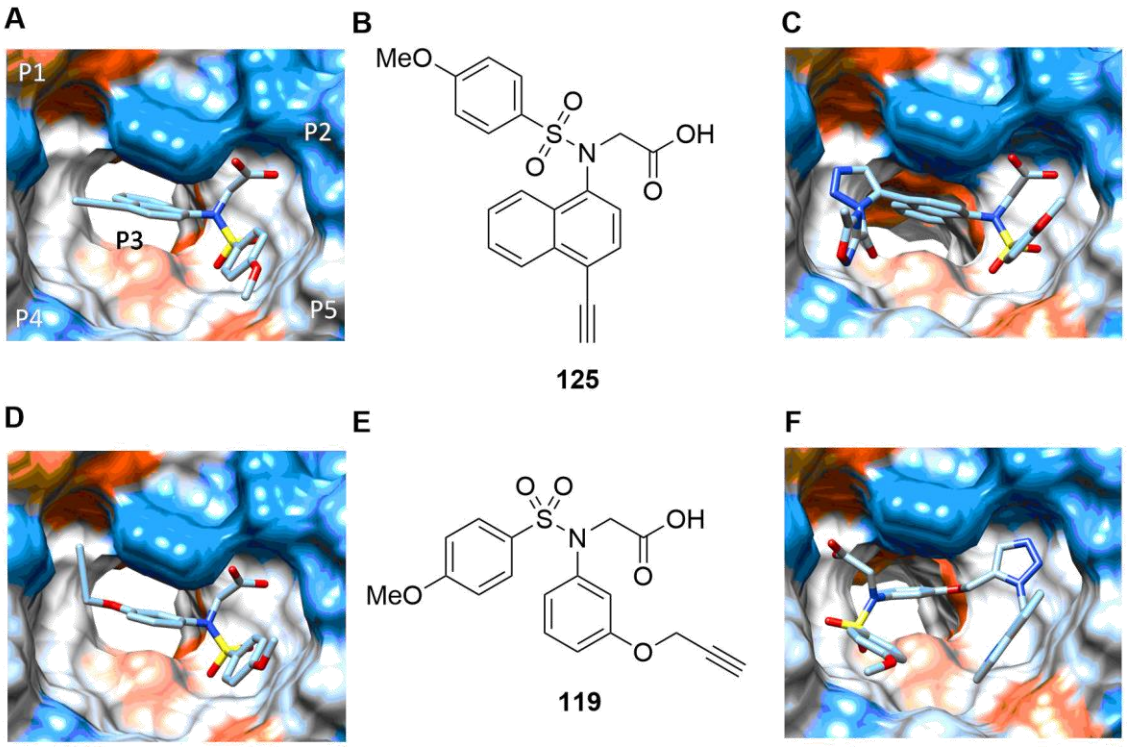


Figure 8.10 Structures of (**A/B**) the naphthalene and (**D/E**) phenyl biased fragments in their respective binding conformations. Binding conformation of triazoles for (**C**) the naphthalene scaffold, 1,5-regioisomer of **186** (AzD6-2) and (**F**) phenyl scaffold, 1,5-regioisomer of **156** (AzF3-2).

8.3 Summary and conclusions

The aim of this chapter was to identify novel small molecule indirect inhibitors of the Keap1-Nrf2 PPI interaction using the kTGS approach. In summary, the kTGS experiments identified a diverse set of 27 triazole scaffolds for the naphthalene series

and 10 triazole scaffolds for the phenyl series as potential novel Keap1-Nrf2 PPI inhibitors. Within these hits, two SARs were observed across both scaffolds; over half of the fragments exhibited aromatic functional groups, for example **177** and **178** and multiple fragments exhibited amide bonds, for example **168** and **169**. The higher affinity naphthalene scaffold appeared to present an advantage over the lower affinity phenyl scaffold, as twice as many hits were identified. This allows a better evaluation of the structural trends and so a greater chance of identifying promising lead structures. Furthermore, a clear difference in percentage triazole formation and product amplification was also evident with greater values observed for the naphthalene scaffold. This is also advantageous as it affords a greater chance of detection of hits.

The results were analysed to evaluate the potential relationship between percentage triazole formation and affinity as hypothesised by Grimster *et al*. and other groups.[196] Although there are indicators to suggest a correlation between the two factors, it was not a universal observation and thus there is not enough evidence within this study to fully support the hypothesis. This relationship was not further evaluated as part of this thesis. Further investigation and analysis could be incorporated into future work.

The regiospecificity governed by Keap1 for these scaffolds, if present, also remains unclear. This was due to slight changes in the retention time between the observed peak from the triazole generated in the presence of Keap1 and the 1,4-synthetic standard and the lack of a 1,5-synthetic standard. To refine the hit identification process, future work is required to optimise the LC-method to deliver reproducible retention times and thus distinguish between regioisomers. A 1,5-synthetic plate is also required to validate the presence of 1,5-regioisomers. Three further aspects of the method were also highlighted for development: (i) evaluation of the use of BSA as a control (and thus investigation of alternative control experiments) (ii) method sensitivity and (iii) false negatives occurring due to the use of multicomponent reactions.

A parallel molecular modelling study was also carried out to support the kTGS investigation. The outcomes from the molecular modelling study and their impact on the kTGS study are discussed below. The molecular modelling study, carried out to support the kTGS approach, highlighted the advantage of kTGS which can result in identification of a wider range of scaffolds. For the molecular modelling study both scaffolds identified mainly aromatic groups in the top triazole hits, particularly multicyclic aromatic groups. Surprisingly less agreement was observed between the kTGS study and molecular modelling study for the naphthalene scaffold than was anticipated. The molecular modelling study showed a clear preference for the 1,5-

disubstituted triazole regioisomer of the naphthalene scaffold. This, regiospecificity was not discerned in the kTGS study highlighting the need for optimisation of the LC method and inclusion of a 1,5-synthetic control as previously discussed in analysis of the kTGS results.

From the results of this analysis, a selection of hits were identified for synthesis and subsequent validation in two biophysical assays, in support of proof of concept (Figure 8.11 and Figure 8.12) and is detailed below.

The smaller number of hits identified by the kTGS study for the phenyl scaffold, alongside a better concurrence with the molecular modelling study, presents a concise group of compounds with which to assess the kTGS method and approach. A total of nine triazoles were selected for the phenyl scaffold; four triazole hits were selected from the kTGS study covering a diverse range of scaffolds (**153**, **154**, **155**, and **158**) and three from the molecular modelling study to represent the bicyclic aromatic structures (**192**, **196** and **198**) (Figure 8.11). Two further hits were selected which were identified by both studies, **156** and **157** (Figure 8.11).

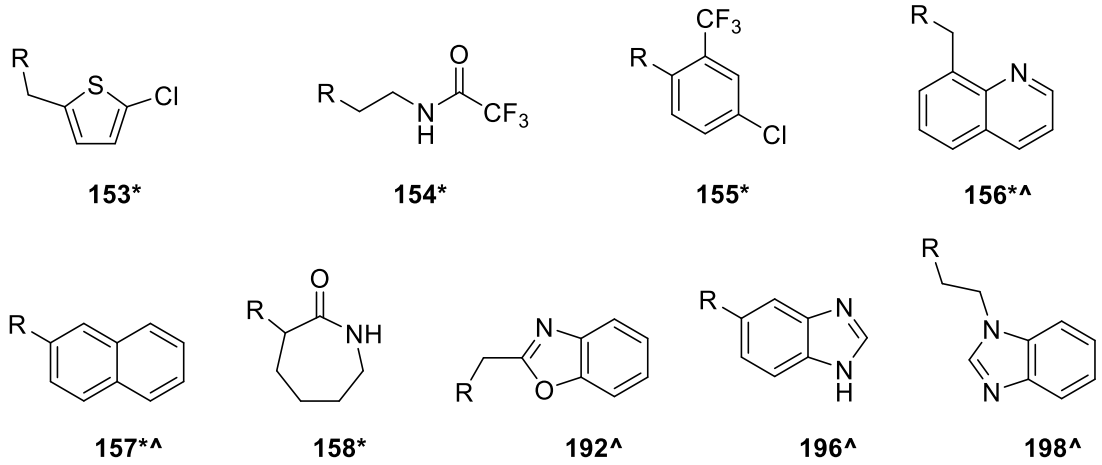
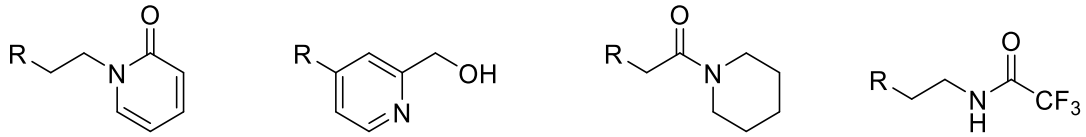


Figure 8.11 Summary of triazole hits identified through kTGS (*) and *in silico* (^) screens for the phenyl scaffold **119**. Both regioisomer will be synthesised and tested. R represents the phenyl fragment **119** and triazole linker.

For the naphthalene scaffold, a smaller group of four triazole hits from the kTGS study were selected. They were chosen based on the structural trends observed as previously discussed; two triazole hits with an amide bond were selected (**174** and **179)** and one triazole hit containing an aromatic group (**178**) was selected. Triazole hit **180** was also selected to represent the overlap between the two scaffolds.

To assess regioisomer selectivity it was envisaged both the 1,4- and 1,5-disubstituted triazoles would be synthesised for all of the selected alkyne and azide combinations.

Synthesis of the remaining triazole hits from both scaffolds could be carried out in future work.



174 **178** **179** **180**

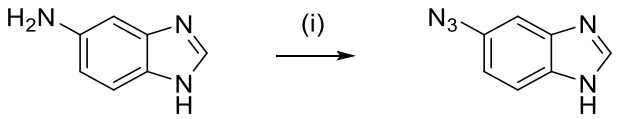
Figure 8.12 Summary of triazole hits identified through the kTGS screen for the naphthalene scaffold **125**. Both regioisomer will be synthesised and tested. R represents the naphthalene fragment **125** and triazole linker.

CHAPTER 9
Hit Validation

In the kTGS experiments, 27 and 10 triazole scaffolds were identified as hits for the naphthalene and phenyl scaffolds respectively. Supported by a parallel molecular modelling study, a shortlist of hits were selected for re-synthesis and subsequent bio-evaluation to enable validation of the kTGS approach. The results are presented in this chapter and are analysed to evaluate the approach of kTGS as a tool to identify novel Keap1 inhibitors.

9.1 Synthesis of azide fragments

All azide fragments, with the exception of **AzE4-2**, were obtained from a commercial source (Enamine). Azide fragment **AzE4-2** was successfully synthesised by diazo transfer, following published conditions, from 5-aminobenzoimidazole **212** in a moderate yield (Scheme 9.1).[202]



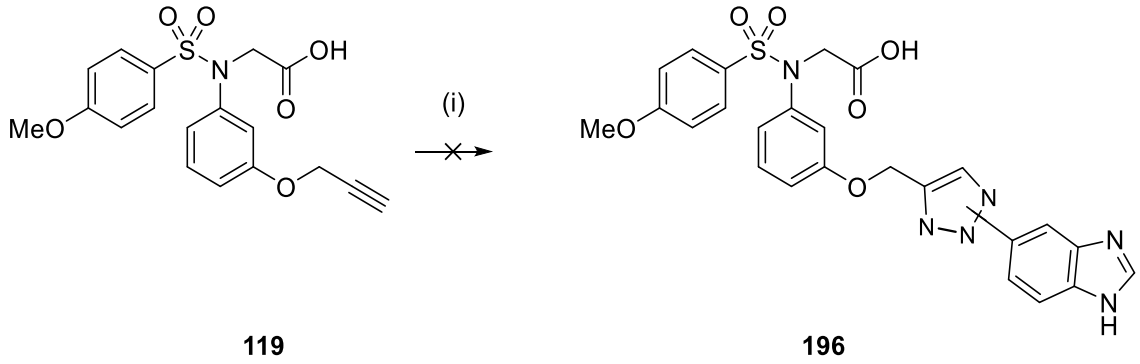
212 **AzE4-2**

Scheme 9.1 Synthesis of **AzE4-2**. *Reagents and conditions*: (i) 9 M HCl, 0 °C *then* $NaNO_2$ in H_2O, 0 °C, 1 hr *then* NaN_3 in H_2O, 0 °C, 1 hr (43%).

9.2 Synthesis of 1,4-disubstituted triazoles

The thermal Huisgen 1,3-dipolar Azide-Alkyne cycloaddition (AAC) is a metal-free catalysed reaction which leads to the formation of 1,2,3-triazoles.[203] Although popular. due to the lack of a metal catalyst, conditions can involve elevated temperatures and result in the formation of both 1,4-disubstituted and 1,5-disubstituted triazoles with low yields. The lack of regioselectivity could be beneficial in this instance as it potentially allows isolation of both regioisomers from one reaction. These conditions were initially attempted on a model system but only trace amounts of product were observed after heating for 24 hours at reflux (Scheme 9.2). Furthermore, it was not evident by TLC or LCMS if both regioisomer products were formed. However it is possible the two isomers were not resolved. Due to the lengthy time-scale, prolonged elevated

temperature and potential issues with resolution of the regioisomers, these conditions were not taken forward.



119 **196**

Scheme 9.2 Thermal Huisgen 1,3-dipolar azide-alkyne cycloaddition. *Reagents and conditions*: (i) **AzE4-2**, DMF, reflux, 24 hrs**.** Compound **196** was not isolated.

The copper catalysed azide-alkyne cycloaddition (CuAAC) is an alternative method to the thermal Huisgen conditions described above. It is advantageous over thermal conditions for the following reasons:[204, 205] (i) it is regiospecific favouring 1,4-disubstituted triazoles (ii) uses lower temperatures and (iii) is compatible with aqueous conditions. The active catalyst Cu(I) can be obtained from Cu(I) salts or generated *in situ* using a Cu(II) salt and sodium ascorbate as a reducing agent. The proposed mechanism is depicted in Figure 9.1. The azide coordinates to a σ-bound copper acetylide with an additional π-bond to a second copper atom. The σ-bound copper acts as a stabilising donor ligand while a ring contraction takes place forming a triazoyl copper derivative. Protonolysis releases the desired 1,4-disubstituted triazole and regenerates Cu(I).

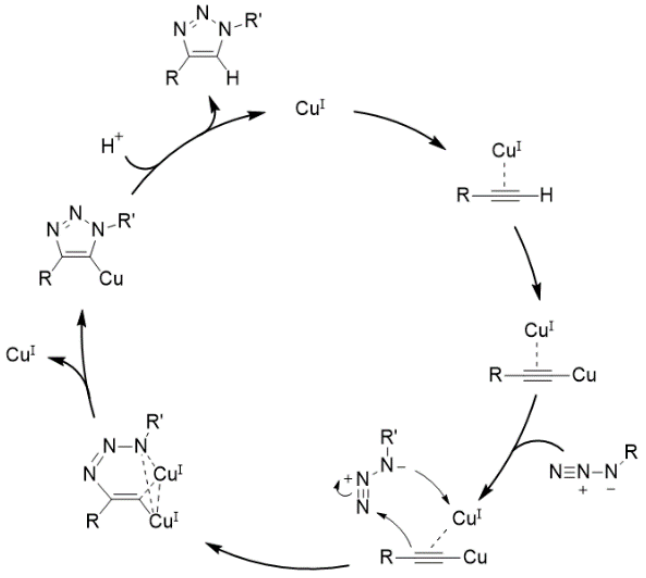
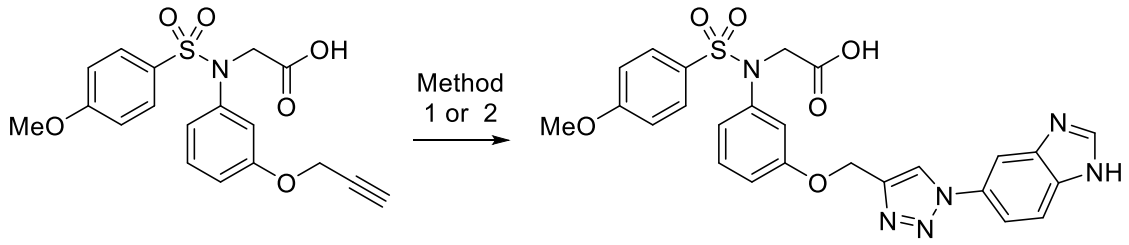


Figure 9.1 Catalytic Cu cycle for 1,4-triazole formation. Reproduced from Worrell *et al*.[206]

The same model system used to investigate the Huisgen conditions was employed to compare the Cu(I) salt method versus the Cu(II) salt method with *in situ* generation of Cu(I). Copper chloride was chosen as the Cu(I) salt (method 1) and Cu_2SO_4 was employed as the Cu(II) salt using sodium ascorbate for *in situ* generation of Cu(I) (method 2). Conditions were identified though investigation of a series of changes to temperature and catalyst equivalents which provided sufficient quantity of compound to enable bio-evaluation (Table 9.1). The final two columns of the table indicate the progress of the reaction for each method presented as percentage conversion to the product as observed by LCMS.

Table 9.1 Investigation of the conditions for the CuAAC using alkyne **119** and azide **AzE4-2** as a model system.



119 **196a**

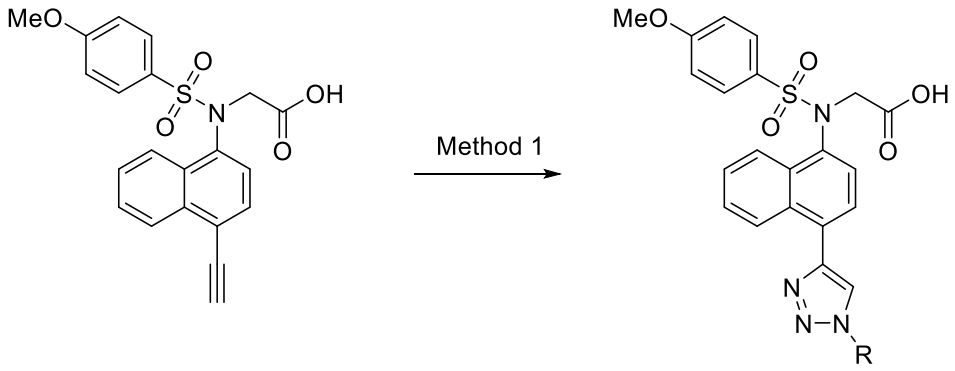
Entry	Cu mol%	Sodium ascorbate (eq)$^+$	Time (hr)	T (°C)	%Conversion	
					Method 1	Method 2
1$^+$	2	0.1	12	rt	-	n/a
2$^+$	2	0.1	36	100	-	30%
3	3	0.1	1	*130	30%	30%
4	50	1.3	0.5	*130	100%	100%

Reaction progress was monitored by LCMS. *Reagents and conditions*: Method 1: CuCl, ^{t}BuOH, H_2O. Method 2: Cu_2SO_4, sodium ascorbate, ^{t}BuOH, H_2O. $^+Cu_2SO_4$ only. *microwave heat source. The reaction was monitored by LCMS and the progress presented as the percentage conversion.

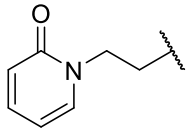
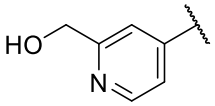
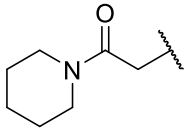
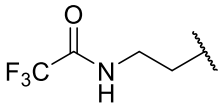
Initially a low mol% of catalyst at room temperature was employed. However, no reaction was observed after 12 hours stirring under these conditions (Table 9.1, entry 1). An increase in temperature resulted in the formation of product. However, the reaction did not reach completion after 36 hours of heating (Table 9.1, entry 2). Changing the heat source to microwave irradiation rapidly increased the reaction progress but the reaction did not reach completion with 30% conversion observed for both methods (Table 9.1, entry 3). Due to time constraints, a large increase in catalyst equivalents was employed to drive the reaction to completion (Table 9.1, entry 4). Using conditions from entry 4, both methods yielded the desired 1,4-regioisomer **196a**

in equivalent, although low, isolated yields with 30% for method 1 and 26% for method 2. As similar yields were observed for both methods, method 1 was chosen to synthesise the remaining hit scaffolds for both series as fewer reagents were required. Percentage yields are summarised in Table 9.2 and Table 9.3 for the naphthalene and phenyl series respectively.

Table 9.2 Yields for synthesis of 1,4-disubstituted triazoles from the naphthalene series.



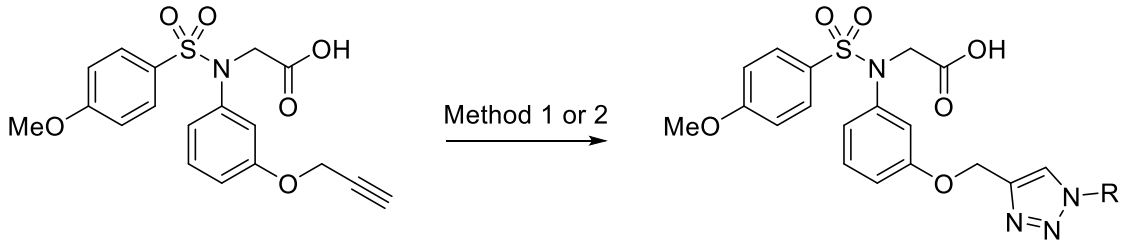
125

Compound	R	Method 1 yield (%)
174a		12%
178a		0%
179a		18%
180a		37%

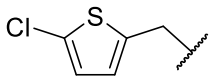
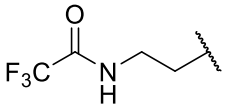
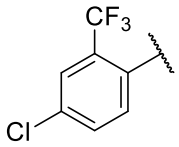
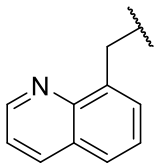
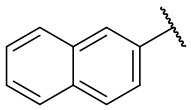
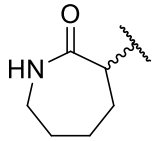
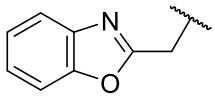
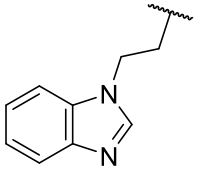
Reagent and conditions: Method 1: CuCl, ^{t}BuOH and H_2O.

Compound **178a** was not isolated using the conditions from method 1; it was hypothesised the pyridine nitrogen was coordinating with the copper disrupting the catalytic cycle. The reaction was repeated with 1.3 eq of CuCl. However, compound **178a** was not obtained. Due to time constraints, no further attempts to optimise the synthesis of compound **178a** were made. Future work could explore the use of the thermal cyclisation which does not use a metal catalyst.

Table 9.3 Yields for synthesis of 1,4-disubstituted triazoles from the phenyl series.



119

Compound	R	Method 1 yield (%)	Method 2 yield (%)
153a		55%	
154a		22%	51%
155a		84%	
156a		24%	
157a		38%	
158a		19%	67%
192a		32%	
198a		n/a	10%

Reagents and conditions: Method 1: CuCl, ^{t}BuOH, H_2O. Method 2: Cu_2SO_4, sodium ascorbate, ^{t}BuOH, H_2O.

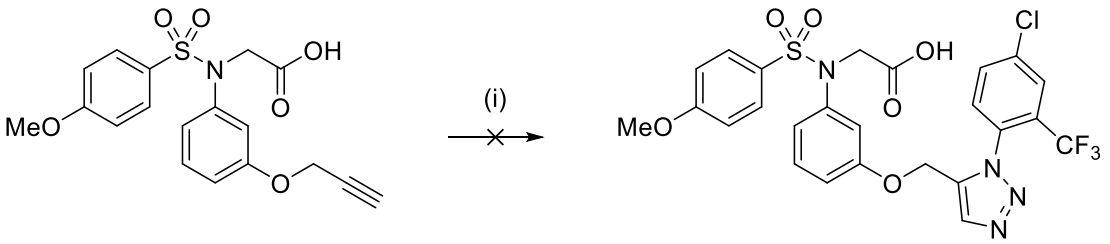
For the phenyl series not enough material was obtained to sufficiently analyse and test three compounds, **154a, 158a** and **198a** due to a combination of small reaction scales and low yields. Employment of method 2 resulted in isolation of **154a** and **158a** in increased yields over method 1. However, a low yield was observed again for compound **198a**.

Moderate to low yields were observed for both methods for the synthesis of the 1,4-triazoles of the selected scaffolds with the exception of **155a** which was isolated in a good yield. With samples of all 1,4-regioisomers of triazole hits in hand, with the exception of **178a**, attention was turned to the synthesis of the 1,5-regioisomer analogues.

9.3 Synthesis of 1,5-disubstituted triazoles

The sister process to CuAAC, ruthenium catalysed azide-alkyne cycloadditions (RuAAC), leads to regioselective formation of 1,5-disubstituted triazoles.[207, 208] However, the catalytic activity and regioselectivity of ruthenium complexes have been shown to vary with different ligands and does not always result in the formation of 1,5-disubstitued triazoles. For example, Boren *et al.* showed that both $Ru(OAc)_2(PPh_3)_2]$ and $RuCl_2(PPh_3)_3$ complexes lead to the sole formation of the 1,4-disubstituted triazole products in a 46% yield or a mix of both regioisomers in trace yields respectively.[207] They also observed ruthenium complexes with η_5-pentamethylcyclopenta-dienyl [Cp*RuCl] are both active and selective in promoting the 1,5-reaction. Guided by this study, two complexes with the pentamethylcyclopentadiene (Cp*) ligand were selected based on commercial availability and stability: *bis*(triphenylphosphine) [$Cp^*RuCl(PPh_3)_2$] and cyclooctadiene [Cp*RuCl(COD)]. A model system using the phenyl scaffold was employed to identify optimal conditions for the synthesis of 1,5-disubstituted triazoles using the Cp*RuCl complexes (Table 9.4). Attempts to obtain the desired 1,5-disubstitued triazole **155b** were unsuccessful irrespective of the ruthenium complex employed. It was hypothesised the carboxylic acid side chain could be coordinating to the ruthenium, hindering the reaction. Further conditions using ruthenium complexes were not investigated at this point due to time constraints. Future work could investigate use of the ester analogue in place of the acid or carrying out the triazole synthesis with a simpler propargyl derivative first with subsequent attachment to the rest of the target compound.

Table 9.4 Conditions used to investigate two ruthenium complexes to obtain a 1,5-disubstituted triazole **155b.**

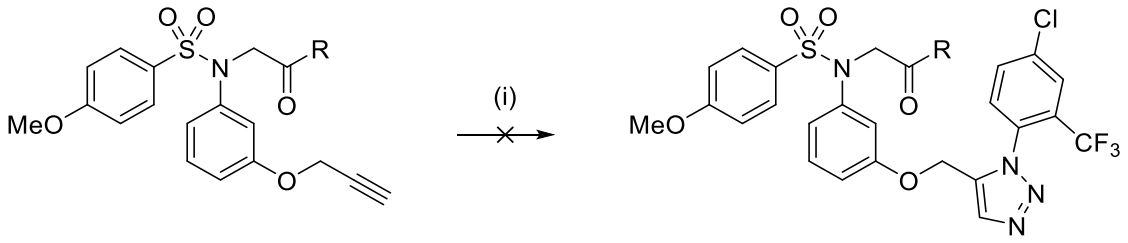


119 **155b**

Entry	Ru mol%	Time (hr)	T (°C)	%Conversion	
				Method 1	Method 2
1	1	12	rt	0%	0%
2	1	0.5	100	0%	0%
3	1	0.5	*100	0%	0%

Reagents and conditions: Method 1: Cp*RuCl(PPh_3)$_2$, AzD7-2, anhydrous toluene, anhydrous DMF Method 2: Cp*RuCl(COD), **AzD7-2**, anhydrous toluene, anhydrous DMF. *microwave heat source. All reactions were carried out under an inert atmosphere. Percentage conversion to the product is indicated in the final two columns and was monitored by LCMS.

In parallel, an alternative method was explored following the conditions published by Smith *et al.* which utilised Et_2Zn as catalyst (Scheme 9.3)**.**[209] Treatment of **119** with Et_2Zn and azide **AzD7-2** did not yield the desired compound **155b.** Again, it was hypothesised the carboxylic acid side chain could be coordinating to the zinc, hampering the reaction. The conditions were repeated using the ester analogue **213**. However, the desired compound **214** was not obtained. No further attempts to isolate the 1,5-disubstituted triazoles were made using Et_2Zn. However there is scope for further investigation. Future work could encompass evaluation of different reaction parameters such as temperature, reagent equivalents and solvent choice.



119: R = OH
213: R = OEt

155b: R = OH
214: R = OEt

Scheme 9.3 Attempted synthesis of 1,5-disubstituted triazole using Et_2Zn as catalyst. *Reagents and conditions*: (i) AzD7-2, 1 M Et_2Zn in hexane, 1-methylimidazole, anhydrous THF, room temperature, 12 hrs. Compounds **155b** and **214** were not obtained. All reactions were carried out under argon.

Further investigation into conditions to isolate 1,5-disubstituted triazoles was put on hold in favour of testing the 1,4-analogues. Future investigation could encompass further evaluation of the ruthenium complexes and zinc metal as neither investigation was exhausted. Other avenues of investigation could also include alternative reagents, for example utilisation of the base tetramethylammonium hydroxide[210] or a Grignard reagent, for example bromomagnesium acetylides.[211]

9.4 Biophysical evaluation of triazole hits

The synthesised 1,4-disubstituted triazoles were screened to determine their binding affinity to Keap1 in two biophysical assays, differential scanning fluorimetry (DSF) and fluorescence polarisation (FP). To validate the ligands as Keap1-Nrf2 inhibitors, requires an observed affinity to Keap1 in both assays. Furthermore, to provide a proof of concept that kTGS can be applied to Keap1, the triazole also needs to exhibit an *increased* affinity in comparison to its biased fragment. Thus, the affinity of each triazole was compared to their respective biased fragment **119** or **125**. The findings for each scaffold are discussed in the following paragraphs.

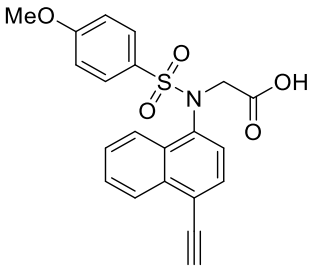
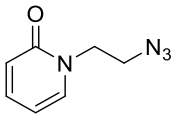
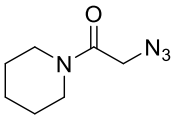
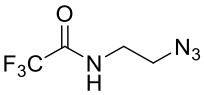
The thermal shift and FP data indicate all three of the naphthalene 1,4-disubstituted triazoles **174**, **179** and **180** are inactive relative to the biased fragment **125** (Figure 9.2B and Table 9.5). The complete loss of affinity suggests the key interactions formed by the naphthalene core and side chains have been disrupted and so the compounds are not binding to Keap1 (Figure 9.2A,B).

As discussed in section 8.1, it was unclear from the kTGS study which regioisomers were formed in the presence of Keap1 due to lack of resolution in the LC method. However, the molecular modelling study indicated that the 1,5-disubstituted triazole is favoured for the naphthalene series. This regiospecificity could account for the lack of affinity observed for the 1,4-disubstituted triazoles. The 1,5-regiosiomers were not isolated and so this observation cannot be validated at this point. These observations further highlight the requirement for LC-method optimisation to establish the regioisomer formed in the presence of Keap1.

It is also possible that despite the high ligand concentrations used in the assay, the incubation times were not sufficient to reach equilibrium in the biophysical assays. When evaluating the naphthalene scaffolds in the FP assay, the Wells group has observed that ligand binding plateaus after relatively long incubation times (1-3 hrs) in

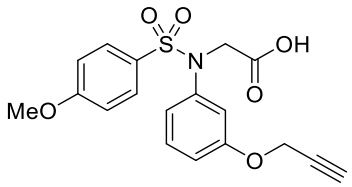
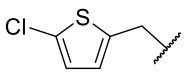
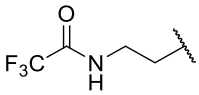
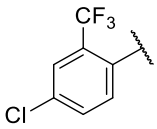
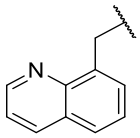
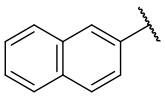
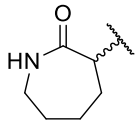
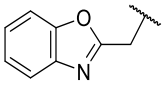
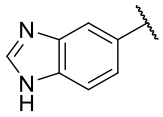
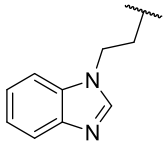
some cases. Thus, longer ligand-protein incubation times could be evaluated as part of future method optimisation.

Table 9.5 Summarised thermal shift and FP data for the naphthalene triazoles.

Ref	Structure	Study	ΔTm (°C)*	%inhibition @ 50 μM^
125		-	**3.79**	**71%**
174a		kTGS	0.09	n/a
179a		kTGS	0.17	n/a
180a		kTGS	0.04	n/a

All experiments were conducted in triplicate *Thermal shift (ΔT_m, °C) values with respect to the protein (P). ^Fixed dose percentage inhibition of the Keap1-Nrf2 interaction at ligand concentration 50 μM. A heat map highlights the results for both the FP and thermal shift data; Significantly increased activity (green), moderate to no change in activity (orange), no inhibition (n/a) (red) relative to compound **125**.

Table 9.6 Summarised thermal shift and FP data for the phenyl triazoles.

Ref	Structure	Study	Isomer‡	ΔTm*	%inhibition @ 50 μM^
119		n/a	n/a	0.19	13%
153a		kTGS	not assigned	0.34	14%
154a		kTGS	not assigned	0.30	18%
155a		kTGS	not assigned	-0.11	n/a
156a		Both	Both	1.81	32%
157a		Both	Both	-1.90	10%
158a		kTGS	not assigned	0.10	8%
192a		MM	1,5	-0.12	n/a
196a		MM	1,5	0.45	18%
198a		MM	1,4	1.10	40%

All experiments were conducted in triplicate. The study which identified the triazole as a hit is specified alongside the structure as well as the (‡) regioisomer preference as indicated in the molecular modelling study (MM). Regioisomer preference from the kTGS study was inconclusive and so is presented as unassigned (n/a). *Thermal shift (ΔT_m, °C) values with respect to the protein (P). ^Fixed dose percentage inhibition of the Keap1-Nrf2 interaction at ligand concentration 50 μM. [a]Fluorescence interference. A heat map highlights the results for both the FP and thermal shift data; Significantly increased activity (green), moderate to no change in activity (orange), no inhibition (n/a) (red) relative to compound **119**

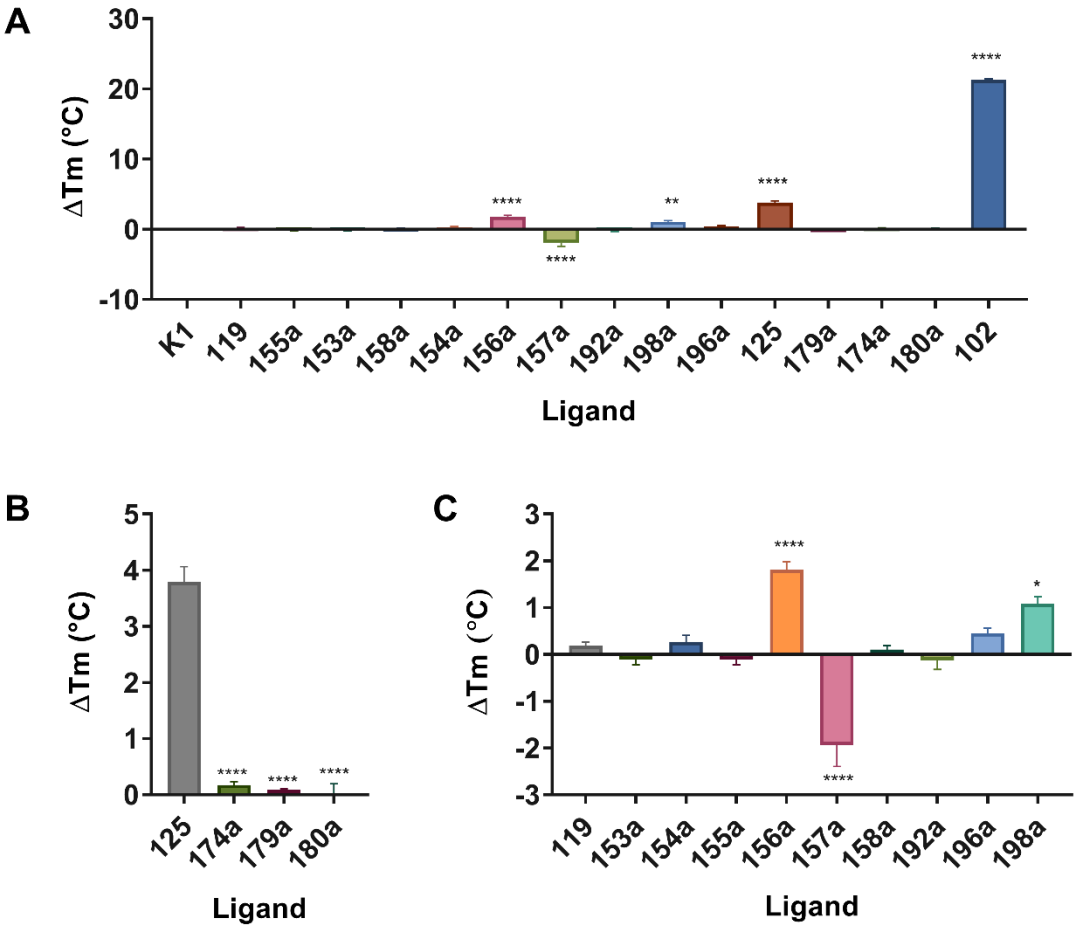


Figure 9.2 Evaluation of triazoles in a DSF assay. All experiments were conducted in triplicate with inhibitor **102** as the positive control. Significance of changes evaluated by one-way ANOVA with Dunnetts correction; *P<0.05; **P<0.01; ***P<0.001; ****P<0.001 versus protein (K1) or the biased inhibitor **119** or **125**. Compound **102** was synthesised by Dr Nikolaos Dimitrios-Georgakopoulos **A** Thermal shift (ΔT_m, °C) values for each triazole with respect to the protein (K1). **B** ΔT_m values for each triazole for the naphthalene series with respect to biased fragment **125** **C** ΔT_m values for each triazole from the phenyl series with respect to biased fragment **119**.

For the phenyl scaffold, the thermal shift and FP data indicate a more mixed response in comparison to the naphthalene series (Figure 9.2A,C and Table 9.6). The structures of each hit are summarised alongside their biophysical results. The shortlist of hits for the phenyl series included scaffolds identified by both the molecular modelling study and the kTGS study. The table indicates which study the hit was identified by as well as the regioisomer preference.

A significant increase in affinity is observed relative to the phenyl fragment **119** for two of the nine synthesised phenyl triazoles compounds, **156a** and **198a**. The thermal shift assay found compound **156a** and **198a** increased the stability of the protein by 1.61 °C and 0.91 °C respectively. This observation was reinforced by the single dose FP assay which also found both triazoles **156a** (32% inhibition @ 50 μM) and **198a** (40%

inhibition @ 50 μM) had a significantly increased affinity in comparison to the biased fragment **119** (13% inhibition @ 50 μM). Triazoles **156a** and **198a** were further evaluated in dose-response experiments using the FP assay to obtain IC_{50} values and were found to be comparable micromolar Keap1 inhibitors (IC_{50} = 257 μM ± 53 μM and IC_{50} = 321 μM ± 82 μM respectively) (Figure 9.3).

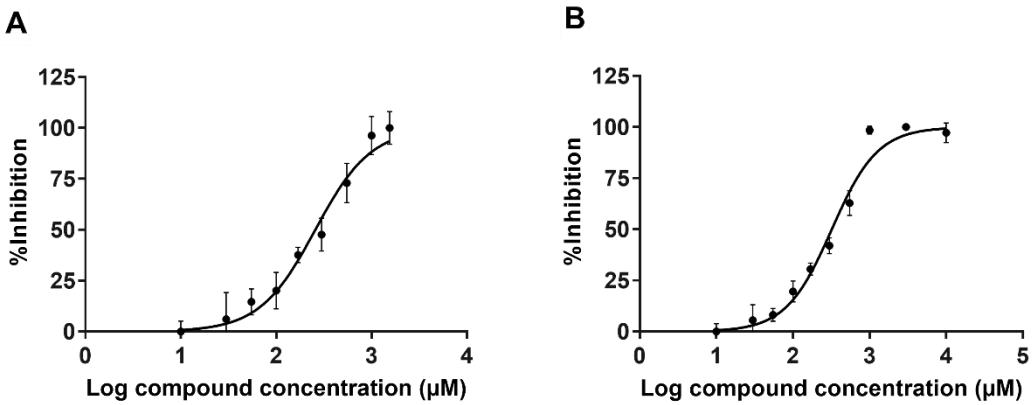


Figure 9.3 Two 9-point dose-response competitive binding curves using an FP-based binding assay with a 7-mer Nrf2 ETGE motif derived peptide, FITC-DEETGEF-OH as the competitive inhibitor. Values were measured in triplicate. Binding curves were fitted using the non-linear regression functions in Prism 7.0. **A** Compound **156a B** Compound **198a**.

Although the binding affinities are relatively low in comparison to the nanomolar IC_{50} exhibited by the published inhibitor **102**, they are both moving to a more drug-like structure through a lower molecular weight and removal of the *bis*-carboxylic acid scaffold which limits cell permeability. These scaffolds can be further optimised and refined through structure-activity relationship (SAR) guided drug design and other medicinal chemistry techniques to improve the Keap1 binding affinity and optimise their physicochemical properties. This positive result provides the proof of concept that the kTGS approach can be applied to Keap1 to identify novel inhibitors and paves the way, alongside method development, for future work to evaluate an unbiased kTGS approach to identifying novel Keap1 inhibitors.

It is important to note, compound **156a** was identified by both the kTGS study and molecular modelling study whereas compound **198a** was only identified as a hit triazole by the molecular modelling study. Analysis of the LC/MS-SIM traces indicates the triazole **198a** formed in the presence of Keap1. However, it did not have an observable product amplification over the control with DBPS buffer and so was dismissed as a 'non-hit'. As it has subsequently been shown to be active, this is an example of a false negative, a recognised disadvantage of the kTGS approach. False negatives often occur due to issues with the methodology, for example slow off-rates leading to

undetectable quantities of product. This result highlights an area of method development and has raised further questions regarding the methodology, in particular how hits are identified and defined for this protein. In section 8.1 several areas of method optimisation were discussed including: LC-method optimisation, assessment of the method sensitivity and the use of BSA as an unrelated control protein. Further areas of investigation and optimisation could include assessment of the off-rate of the triazole product; a slow off-rate can result in hindered triazole formation due to catalyst poisoning. This has been addressed by Tieu *et al.* using a mutant form of the target protein which has a lower affinity for the product.[185]

The remaining seven 1,4-disubstituted triazoles displayed either comparable activity to the phenyl fragment **119** (**153a**, **154a**, **158a** and **196a**) or were inactive under the conditions tested (**155a**, **157a** and **192a**) (Table 9.6). The molecular modelling study also identified three of the seven scaffolds as hits (**196a**, **157a** and **192a**) and intriguingly, were indicated as 1,5-disubstituted triazoles. This regiospecificity governed by Keap1 could account for the lack of affinity exhibited by the synthesised 1,4-disubstituted analogues. Thus, evaluation of the 1,5-disubstituted triazole analogues is required. The 1,5-disubstituted triazole analogues have not been isolated and so the hypothesis cannot be tested at this point. The large negative value observed for **155a** could indicate the compound is insoluble causing fluorescence interference in the assay.

Hit triazole **158a** has one chiral centre and was synthesised as a racemic mixture. It is inactive relative to the phenyl fragment **119** with a thermal shift (ΔT_m) of 0.1 °C with respect to the protein and 8% inhibition @ 50 µM in the FP assay. All chiral scaffolds within the fragment library are racemic mixtures. Some activity would still be expected for a racemic mixture if only one enantiomer of the 1,4-triazole is active. This may indicate that both of the 1,4-disubstituted enantiomers of **158a** are inactive and so to fully assess the activity of this scaffold a racemic mixture of the 1,5-triazole **158b** needs to be synthesised and evaluated.

Published studies have noted a relationship between percentage triazole formation and affinity. As discussed in section 8.1, the analysis of the results from the kTGS study did not find a clear relationship between the two. This relationship was not investigated further as part of this thesis, i.e. the hit triazoles nominated for synthesis were not selected to investigate this relationship. However, through comparison of the percentage triazole formation with the observed affinity, an initial impression can be gained. This analysis indicates that there is no correlation between activity and catalytic

effect or percentage triazole formation. In fact, for the most active hit, **156a**, both the percentage triazole formation (0.0091%) and catalytic effect (15% *c.f.* buffer and 5% *c.f* BSA controls) were below the series average**.** Although without affinity data for the 1,5-disubstituted regioisomer analogues for all the shortlisted hits, the relationship cannot be fully evaluated and further investigation is still required.

9.5 Summary

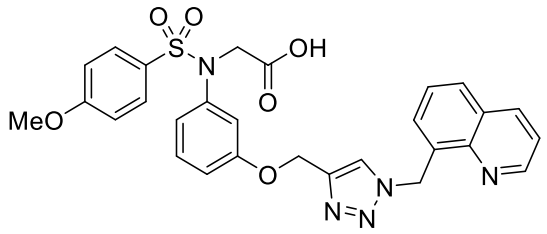
Synthesis of the 1,4-regioisomers of the shortlisted hits for the naphthalene and phenyl scaffolds was successful with the exception of naphthalene scaffold **178a** which was not isolated. It was hypothesised the pyridine was coordinating to the Cu-catalyst hindering the reaction. Future synthetic efforts could utilise non-metal catalysed reactions, for example thermal Huisgen 1,3-dipolar cycloaddition. The 1,5-regioisomers of the shortlisted hits were not isolated using either zinc or ruthenium catalysed reactions. Although neither investigation was exhausted, attention was turned to screening the 1,4-regioisomers.

All three naphthalene 1,4-regiosiomers were found to be inactive in both biophysical assays. This could be attributed to the 1,5-regiospecificity exhibited by Keap1 as indicated in the molecular modelling study. The biophysical evaluation of the phenyl scaffold was more mixed. Of the nine scaffolds screened, seven were found to be inactive or display no increased affinity versus the biased fragment **119**. Synthesis and evaluation of their 1,5-analogues will allow greater assessment of these results and their implications on potential method development. However, two triazole scaffolds, **156a** and **198a**, were observed to have an improved affinity to Keap1 in comparison to the biased fragment **119** in both biophysical assays. Evaluation by an FP dose-response experiment found both compounds to be micromolar inhibitors of Keap1. Although they have comparably lower affinity to the published inhibitor **102**, they both move towards a more drug-like template with one less acid group (shown to limit cell permeability) and lower molecular weights. Both scaffolds have the potential for further optimisation using medicinal chemistry techniques including SAR-based drug design approaches and confirmation of binding orientation through crystallography.

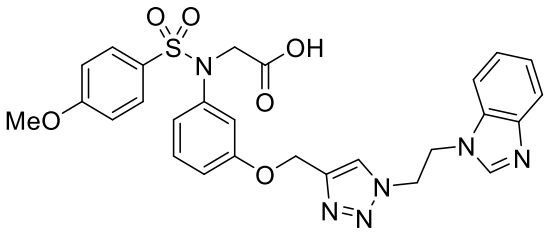
Identification of a scaffold (**156a**) with an improved affinity over the biased fragment provides proof of concept that the kTGS approach can be used to identify novel Keap1-Nrf2 inhibitors. However, the second inhibitor scaffold (**198a**) was only identified as a hit by the molecular modelling study and dismissed as a ‘non-hit’ by the kTGS. Although false negatives are a known limitation of the kTGS approach, this, alongside

the less positive results for the naphthalene and remaining phenyl triazoles, raises important questions regarding the methodology, focusing on how hits are identified and defined for Keap1. Central areas of method development highlighted for future work include: (i) LC-method optimisation (ii) assessment of the method sensitivity (iii) the use of BSA as an unrelated control protein and (iv) off-rates of *in situ*-generated triazole products and associated catalyst poisoning.

These scaffolds provide a tool with which to explore the Keap1-Nrf2 binding interaction and pave the way for evaluating the unbiased kTGS approach for identifying novel Keap1 inhibitors.



156a
32% inhibition @ 50 μM
ΔTm 1.81 °C



198a
40% inhibition @ 50 μM
ΔTm 1.10 °C

Figure 9.4 Structures of the top hits and their respective percentage inhibition value at single dose concentration (50 μM) and thermal shift value (ΔTm, °C) with respect to biased fragment **119.** Triazole **156a** was identified as a hit by both the kTGS study and molecular modelling study, **198a** was identified as a hit by the molecular modelling study only.

9.6 Conclusions and future work

As a master regulator of the antioxidant response the Keap1-Nrf2-ARE pathway protects cellular proteins and DNA from oxidative damage. This pathway has been implicated in several areas of disease including cancer, neurological diseases (AD and PD), inflammatory diseases as well as ageing. Thus, it has been identified and targeted as a potential therapeutic pathway. Enhancement of Nrf2 activity has mainly been achieved through disruption of the Keap1 ubiquitination factor via two mechanisms: electrophilic inducers and direct small molecule or peptide inhibitors. Although there has been success, with many inhibitors reaching clinical trials and the recent approval of dimethyl fumarate **89** by the FDA, there is room for improvement; off-target effects are the main limitation of electrophilic inducers and many of the small molecule and peptide scaffolds have poor cellular permeability.

It was hypothesised that application of the relatively novel fragment-based kTGS approach, where the biological target templates the formation of its own inhibitors,

would allow exploration of the topography of the pocket and identification of novel Keap1-inhibitor scaffolds. This required development of an appropriately labelled fragment to act as a handle within the Keap1 binding site and selection of a biocompatible *in situ* coupling reaction. For the purposes of this proof-of-concept study, the biased kTGS approach was applied using the widely successful 1,3-dipolar Huisgen azide alkyne cycloaddition as the *in situ* coupling reaction. Two biased scaffolds with naphthalene and phenyl cores, designed from the *bis*-acid inhibitor **102**, were incubated with a diverse fragment library**.** From these kTGS experiments, a novel Keap1-Nrf2 inhibitor **156a** was identified from the phenyl series. Importantly, it has an improved affinity in comparison to the biased scaffold **119** which demonstrates the biased kTGS approach can be applied to Keap1 to identify novel inhibitor scaffolds. The kTGS approach was supported by a molecular modelling study which identified a second micromolar inhibitor **198a**, which was not identified as a hit in the kTGS experiment. Both scaffolds move towards a more drug-like template; lower molecular weights and fewer acid groups which could possibly improve cell permeability. However, there is potential for further optimisation using medicinal chemistry techniques including SAR-based drug design approaches and confirmation of binding orientation through crystallography.

Key areas of method development were highlighted for future work to improve hit identification and develop a more robust method. These included: (i) LC-method optimisation (ii) assessment of method sensitivity (iii) the use of BSA as an unrelated control protein and (iv) off-rates of *in situ* triazole products and associated catalyst poisoning.

In summary, this work demonstrates the biased kTGS approach can be applied to Keap1 to identify novel inhibitor scaffolds. The novel scaffolds provide a tool with which to explore the Keap1-Nrf2 binding interaction and paves the way for evaluating the unbiased kTGS approach, which does not template against a known inhibitor for identifying novel Keap1 inhibitors. This will allow further exploration of the topography of the pocket and the opportunity to move away from the acid scaffold to a more drug-like structure.

CHAPTER 10
Experimental

10.1 Biology

All chemical reagents were purchased from commercial vendors unless otherwise stated. Sodium dodecyl sulphate (SDS) polyacrylamide gel electrophoresis (PAGE) was carried out using NOVEX™ precast gels and buffers from Invitrogen Ltd in an XCell Surelock™ Mini-Cell electrophoresis system. All solutions for growth were autoclaved using a Priorclave Front Loading Autoclave machine and stored at room temperature unless otherwise stated. Centrifugation was carried out on a Beckman Coulter Avanti J-E Centrifuge. A Jenway 6305 UV spectrophotometer was used to record optical density and calculate protein concentration. All sterile work was carried out using aseptic technique.

10.1.1 Keap1 over-expression

10.1.1.1 Stock solutions and mediums

Table 10.1 Stock solutions and mediums for expression of Keap1.

Stock	Description	Storage
20 g/L LB Broth	Luria-Bertani (LB) broth mix (25 g) was dissolved in d-H_2O (1 L).	Room temperature
0.3 M Ampicillin	Ampicillin (500 mg) was dissolved in d-H_2O (5 mL), sterile filtered (Whatman 0.45 µm) and aliquoted (5 x 1 mL) into sterile Eppendorf tubes.	Aliquots were stored at -20 °C.
1 M IPTG	IPTG (476 mg) was dissolved d-H_2O (2 mL), sterile filtered (Whatman 0.45 µm) and aliquoted (2 x 1 mL) into sterile Eppendorf tubes.	Aliquots were stored at -20 °C.

10.1.1.2 Procedure[212]

Human Kelch protein was heterologously expressed as a His-tagged polypeptide from glycerol stocks of transformed Rosetta 2 BL21(DE3) *Escherichia coli* (*E. Coli*) cells containing the expression vector pET15bhKelch. 1 mL of glycerol stock was added to 1 L of LB medium supplemented with ampicillin (final concentration 100 µg/mL). The inoculated medium was incubated at 37 °C whilst shaking at 220 rpm for 4-5 hours at which point the cells had reached the exponential growth phase (OD_{597nm} = 0.4–0.6). The 1 L bacterial culture was incubated at 21 °C whilst shaking at 220 rpm for a further 30 minutes. The culture was induced with IPTG (final concentration 1 mM) and incubated for 24 hours at 21 °C. The cells were pelleted (3,600 x g at 4 °C for 30 minutes) and stored at -80 °C.

10.1.2 Keap1 extraction and purification

10.1.2.1 Stock solutions and Purification buffers

Table 10.2 Stock solutions for extraction and purification of keap1.

Stock	Description	Storage
1 M Tris pH 7.4	Tris(hydroxymethyl)aminomethane (Tris) (121 g) was dissolved in d-H_2O (800 mL), the pH was adjusted to 7.4 using conc. HCl and the volume made up to 1 L with d-H_2O.	Stock solutions were stored at 4 °C.
5 M Sodium chloride	NaCl (292 g) was dissolved in d-H_2O (1 L).	Stock solutions were stored at 4 °C.
1 M Imidazole pH 7.4	Imidazole (68 g) was dissolved in d-H_2O (1 L) and the pH adjusted to 7.4 using conc. HCl.	Stock solutions were stored at 4 °C
1 M DTT	Dithiothreitol (DTT) (3.1 g) was dissolved in d-H_2O (20 mL).	Stock solutions were stored at 4 °C

Table 10.3 Buffers for the extraction and purification of Keap1.

Buffer	Description	Storage
Lysis buffer	0.5 mL of 1 M Tris (pH 7.4) to give a final concentration of 20 mM 0.75 mL of 5 M NaCl to give a final concentration of 150 mM 0.75 mL of 1 M Imidazole to give a final concentration of 30 mM 12.5 µL of 1 M DTT to give a final concentration of 0.5 mM 1.25 g of 100% glycerol to give a final concentration of 5% glycerol 50 µL of protease inhibitor cocktail per 1 g/cell pellet The final volume was adjusted to 25 mL with d-H_2O.	25 mL was prepared on day of use
Purification buffer (PB)	4 mL of 1 M Tris (pH 7.4) to give a final concentration of 20 mM 20 mL of 5 M NaCl to give a final concentration of 0.5 M 100 µL of 1 M DTT to give a final concentration of 0.5 mM The final volume was adjusted to 25 mL with d-H_2O.	200 mL was prepared on day of use
30 mM Imidazole in PB	1 M imidazole (3 mL) was added to Purification buffer (97 mL).	100 mL was prepared on day of use
300 mM Imidazole in PB	1 M imidazole (5 mL) was added to Purification buffer (20 mL).	25 mL was prepared on day of use
Storage buffer	2 mL of 1 M Tris (pH 7.4) to give a final concentration of 20 mM 500 µL of 1 M DTT to give a final concentration of 5 mM. The final volume was adjusted to 25 mL with d-H_2O.	25 mL was prepared on day of use

10.1.2.2 Procedure

Bacterial pellets were thawed on ice in 25 mL of extraction buffer. The cells were lysed by sonication (5 x 1 minute sonication with 30 seconds of rest between each burst). Insoluble proteins and cell debris were pelleted (27,000 x g, 40 minutes) and the supernatant filtered (Whatman 0.22 µm). The supernatant was loaded onto Ni-NTA agarose gel (1-2 mL) in a 10 mL polypropylene column which was pre-equilibrated with 30 mM imidazole in purification buffer. The flow through was collected and kept on ice. The resin was washed with 50 mL of 30 mM imidazole in purification buffer to remove impurities and the flow through was collected and kept on ice. The protein was eluted with 25 mL of purification buffer containing 300 mM imidazole and the flow through was collected in 25 x 1 mL fractions and kept on ice. The protein concentration and purity of each fraction was determined using Bradford assay and SDS PAGE respectively. Peak fractions were pooled and the buffer exchanged using PD-10 columns; the protein was eluted with final buffer and collected in 1 mL fractions. Protein concentration and purity for each fraction was determined using Bradford assay and SDS PAGE respectively. The fractions were supplemented with 10% glycerol, aliquoted (4 x 250 µL) and stored at -80 °C. A sample was taken and analysed by MALDI-TOF to confirm isolation of the correct protein.

10.1.3 Protein analysis

10.1.3.1 SDS PAGE

Protein purity was determined by SDS PAGE. The sample to be analysed (7 µL) was mixed with NuPAGE® Novex® sample buffer (5 µL), NuPAGE® reducing agent (2 µL) and heated to 72 °C on a heat block for 10 minutes. The sample was loaded onto a NuPAGE® Novex® Bis-Tris gradient gel (4-12%) (Invitrogen) alongside SeeBlue® Plus2 Pre-Stained Standard. The gel was run for 30-45 minutes at 200 V and 150 mA in SDS PAGE running buffer (30 mL NuPAGE® Novex® MOPS SDS Running buffer (20x) (Invitrogen) and 570 mL H_2O). The gel was removed from the plastic case with a metal spatula, washed with H_2O and visualised with Instant Blue (Expedeon) for 30 minutes before de-staining with H_2O.

10.1.3.2 Bradford Assay

The protein concentration of the solution was determined by measuring the optical density. The sample (2 µL) was diluted with Quick Start™ 1X Bradford dye (BioRad) (1 mL) in a 1 mL plastic cuvette. The optical density was recorded on a Jenway 6305

spectrophotometer at 595 nM and used to calculate the protein concentration as follows:

$$C = A \times \left(\frac{16}{V}\right)$$

Where: C = Concentration (mg/mL); A = Absorbance at 595 nM; V = Volume of sample (2 uL)

The concentration units were converted to µM/mL using the mass of protein calculated using two programs, UniProt and ExPASy.

10.1.3.3 MALDI-TOF

The molecular mass of the protein was determined by matrix assisted laser desorption ionisation (MALDI) time of flight (TOF) mass spectrometry on an Applied Biosystem Voyager-DE Pro. The protein sample was prepared for MALDI-TOF analysis using a ZipTip® (C18) pipette tip and the solutions in Table 10.4 using the following procedure: The ZipTip® was hydrated with 2 x 10 µL of solution A, followed by washing with 2 x 10 µL of solution B, the solutions were expelled into waste between each loading. The tip was then loaded with 10 µL of the sample and expelled back into the sample tube, this was repeated 6 times. The tip was then washed with 10 x 10 µL of solution B, expelling the solution into waste between each wash. The sample was eluted by washing the tip with 2 x 1.3 µL of solution C into an empty sample tube.

The analytical results were compared to the expected protein mass which was calculated using two programs, UniProt and ExPASy.

Table 10.4 ZipTip® preparation solutions

Solution	Description
Solution A	50:50 MeCN:H_2O + 0.1% TFA
Solution B	H_2O + 0.1% TFA
Solution C	75:25 MeCN:H_2O + 0.1% TFA

10.2 Biophysical assays

10.2.1 Differential scanning fluorimetry protocol

All experiments were run on an Applied Biosystems® 7500 Real-Time PCR system using Life Technologies 7500 software v.2.0.0. Output data was analysed using Protein Thermal Shift solution software, the mid-temperature (T_m) of the melt curves were analysed using Excel calculations and fitted using the Boltzman equation on GraphPad Prism. Thermal shift (ΔT_m) values were calculated by subtracting the T_m of the control and from that of each ligand. Experiments were run in Applied Biosystems® MicroAmp® 96-well reaction plates and sealed with Applied Biosystems® MicroAmp® Optical Adhesive Film. Optimised conditions are summarised as follows; Protein concentration, 5 µM; Sypro® Orange, 5x; Ligand concentration for fixed dose, 50 µM; Buffer, Dulbecco's Phosphate-Buffered Saline (DBPS); DMSO content, 1% v/v; Total well volume, 40 µL; Temperature range, 25–95 °C measuring at 1 °C intervals with a step and hold for 30 seconds. All experiments were run in triplicate with a positive control.

10.2.2 Fluorescence Polarisation protocol

All experiments were run on a BMG LABTECH PHERAstar®. Inhibition curves were fitted using non-linear regression curves in GraphPad Prism from which the IC_{50} values were determined. Experiments were run in Corning® 96-well black flat bottom reaction plates and excluded from light. Optimised conditions were used as described by Hancock *et al* and are summarised as follows;[136] Protein concentration, 200 nM; Fluorescent probe concentration 1 nM; Ligand concentration for fixed dose, 10 µM and 100 µM; Buffer, Dulbecco's Phosphate Buffer (DBPS); DMSO content, 11% v/v and Total well volume, 100 µL. The Nrf2 ETGE motif derived 7-mer peptide, FITC-DEETGEF-OH was utilised as the fluorescent probe as published. Plates were read at two time points after mixing of the components, 30 min and 120 min. All experiments were performed in triplicate with a positive control.

10.3 Kinetic target-guided synthesis protocol

Conditions used for the kTGS plate were by developed and optimised by Carrie-Anne Molyneaux[213] using conditions published by Manetsch et *al*.[184] Amendments were made to protein and ligand concentration to accommodate solubility of fragments **119** and **125**. The azide library was divided into groups of five, ensuring an even distribution of molecular weight in each group. A stock solution of each mixture was prepared in

Dulbecco's phosphate-buffered saline (DPBS) buffer with final azide concentration at 10 mM and DMSO content ≤1%. Purified Keap1 kelch domain was prepared at 12.5 µM in DPBS buffer. The biased fragment was prepared at 7.5 mM in DBPS with 5% DMSO content. In a 96-well plate the stocks were mixed as follows; 10 µL of biased ligand; 80 µL of protein; 10 µL of azide mixture. Two matched control samples were prepared using either Bovine serum albumin (BSA) or DPBS (vehicle) buffer in place fo the KEap1 protein with equivalent ligand and library concentrations and DMSO content. Plates were sealed and incubated at 37 °C for three days. To support hit identification the retention time of hits were compared to a synthetically prepared 1,4-disubstituted triazole. Conditions for the synthesis plate were used as published by Deprez-Poulain et *al* and are summarised as follows;[198] 3 µL of 25.5 mM biased ligand in DMSO; 3 µL of 5.1 mM azide mixture in DMSO; 2 µL of 20 mM aq. $CuSO_4$; 296 µL of a 1:1 solution of *t*-BuOH:H_2O. The plates were shaken at room temperature for three days.

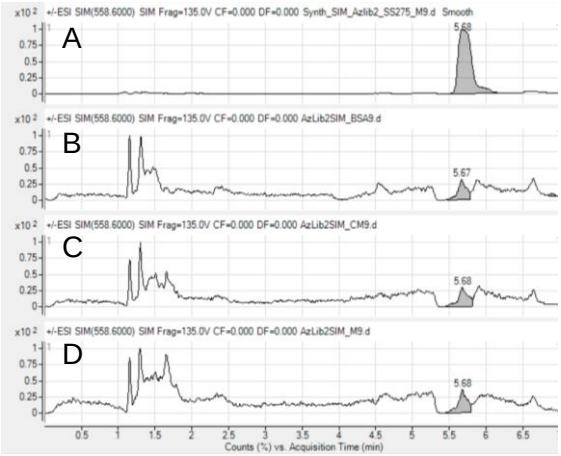


Figure 10.1 Representative 'hit' LC/MS-SIM trace of **A** the synthetic control (1,4-regioisomer) **B** BSA control **C** buffer control and **D** protein (Keap1), for an expected triazole mass.

Analysis of the mixtures was performed on an Agilent Triple Quad LC-MS/MS system using an AquaSil C18 column 50 x 4.6 mm (5 µM particle size) with an MS2 scan and a single ion monitor (SIM) scan. Identified triazole masses were validated against the synthesis plate and measured against the controls for evaluation as a potential hit. An example LC/MS-SIM trace is shown in Figure 10.1. A triazole was defined as a hit if the expected mass was either not evident in the control wells or there was an observable product amplification in the presence of protein over the controls. Product amplification was calculated by comparing the percent area of the peak relative to the total ion count (TIC) and evaluating the difference between the protein and the controls.

10.4 Molecular modelling protocol

The docking experiments were performed by Dr Geoff Wells employing the protocol described as follows; the Keap1 Kelch domain protein structure in complex with *bis*-sulphonamide **101** was obtained from the Protein Data Bank (PDB code: 4IQK).[143] The protein was processed using a text editor to prepare pdb files for the protein, ligand and protein-ligand complex for use in subsequent analyses. The protein pdb file was opened in Autodock Tools (ADT),[214] the solvent and ions were removed and the resulting structure was saved as a pdbqt file for use in Autodock Vina.[201] The small molecule ligands were prepared by constructing the 2D structures in ChemBioDraw Ultra 14.0 which were saved in sdf format. The 2D representations were converted into 3D structures using an in-house script that used Balloon 1.6.1.1220[215] to generate an initial 3D conformer, followed by Openbabel 2.3.2[216] to add hydrogens appropriate for pH 7.4 and to perform a weighted conformer search (nconf = 1000, weighted, gasteiger partial charges). Energy minimisation was performed on the lowest energy conformer using obmimimize (n = 10000, c = 1e-30) followed by MOPAC2012[217] using the semi-empirical PM7 parameters (keywords: PM7, MMOK, T=30000, PDBOUT). The compounds were saved in a mol2 file format and converted to pdb, pdbqt and mol2 files formats (the latter two containing partial charge information) using Openbabel 2.3.2. The properties of the compounds were determined using Openbabel 2.3.2 (molecular weight, number of non-hydrogen atoms, cLogP). The alkyne and azide fragment libraries were parameterised similarly from the corresponding sd files, with visual inspection of the output to check for anomalies.

Autodock Vina was run using an Exhaustiveness setting of 9 and a binding site box centred on the ligand binding pocket that was 24 x 20 x 20 Å in size. All other parameters were set to their default values. The lowest energy conformation of each docked ligand was extracted from the results file using a custom script and analysed in Chimera 1.11.2.[218]

10.5 Chemistry

All oxygen and moisture sensitive reactions were carried out under argon, in oven dried glassware purged with argon prior to use.

All Nuclear Magnetic Resonance (NMR) spectra were recorded on a Bruker Avance 400 MHz or 500 MHz NMR spectrometer. Chemical shifts are reported in ppm (δ) with reference to the solvent and instrument. Data are reported in the following format;

chemical shift, integration, multiplicity (s = singlet, t = triplet q = quartet, m = multiplet, br = broad), coupling (Hz), assignment.

All infra-red (IR) spectra were recorded on a PerkinElmer 100 Optica FT-IR spectrophotometer. Data are reported in the following format; type (neat, nujol etc.), cm^{-1}, absorption, intensity (s = strong, m = medium, w = weak), assignment.

All normal phase purification was carried out by flash chromatography on silica gel 60 Å pore size, 230-400 mesh particle, 40-63 µM particle size. Analytical thin-layer chromatography (TLC) was performed on aluminium sheets with a silica gel layer 60 Å F_{254}. Visualisation was accomplished using UV light and staining by aqueous potassium permanganate ($KMnO_4$) followed by heating. All reverse phase (RP) column chromatography was carried out on a Biotage Isolera™ using a SNAP C18 12 g cartridge.

Low resolution mass spectrometry (LRMS) was performed on a Shimadzu LCMS-20/20 system using a Waters XTerra® C18 column 50 x 4.6 mm (particle size: 2.5 µm) at a flow rate of 1 mL/min. The eluent system and method were as follows; Eluent A H_2O + 0.1% FA; Eluent B MeCN + 0.1% FA; initial fixed composition 10% B to 95% B over 4 min, held at 95% B for 1 min then returned to 5% B over 1 min and held for 1 min, the total run time was 7 min. Data are reported in the following format; method, mass, isotope identification if applicable, ion, charge, percentage purity, retention time. High resolution mass spectrometry (HRMS) was performed by the EPSRC UK National Mass Spectrometry Facility (NMSF) at Swansea University.

Analytical reverse phase high performance liquid chromatography (HPLC) was carried out on an Agilent technologies 1200 series HPLC system using a Waters XSelect® CSH C18 Column 50 x 6 mm (particle size: 2.5 µm) at a flow rate of 1 mL/min. The eluent system and method were as follows; Eluent A H_2O + 0.2% TFA; Eluent B MeCN + 0.2% TFA; initial fixed composition 5% B to 95% B over 20 min, held at 95% B for 4 min then returned to 5% B over 1 min, the total run time was 25 min. Data are reported in the following format; percentage purity, retention time.

General procedures are described when reaction conditions are repeated for the synthesis of analogues. Specific reagent quantities, yield and characterisation data is subsequently reported for each compound.

All reagents and laboratory grade solvents were purchased from commercial vendors and were used without further purification unless otherwise stated. The following azide

fragments were purchased from Enamine (catalogue number) in September 2016: **AzA1-2** (EN300-120046), **AzB4-2** (EN300-131361), **AzD2-2** (EN300-116577), **AzE5-2** (EN300-115709), **AzA5-2** (EN300-115516), **AzD5-2** (EN300-115516), **AzD7-2** (EN300-65322), **AzE2-2** (EN300-78371), **AzF3-2** (EN300-68750), **AzB6-2** (EN300-117186), **AzC6-2** (EN300-120703).

10.5.1 General procedures A-D

General procedure A: *N*-sulphonation[188]

DMAP (0.1 eq.) and *p*-methoxybenzenesulphonyl chloride (1.2 eq.) were added to a solution of aromatic amine (1 eq.) in DCM (0.2 M). Pyridine (2 eq.) was added to the reaction mixture and the reaction was stirred overnight at room temperature and monitored by TLC. Upon completion, the reaction was diluted with DCM and washed with H_2O followed by brine. The organic layer was dried over Na_2SO_4, filtered and the solvent removed under vacuum to yield the crude product. Reagent quantities are specified for each compound along with the yield and purification method.

General procedure B: *N*-alkylation[188]

Ethyl bromoacetate (1.5 eq.) and K_2CO_3 (1.5 eq.) were added to a solution of sulphonamide (1 eq.) in DMF (0.2 M), the reaction was stirred overnight at room temperature and monitored by TLC. Upon completion, the resulting solution was neutralised with 0.05 M aq. HCl, diluted with EtOAc and washed with H_2O and brine. The organic layer was dried over Mg_2SO_4, filtered and the solvent removed under vacuum to yield the crude product.Reagent quantities are specified for each compound along with the yield and purification method.

General procedure C: ester hydrolysis[188]

1M aq. NaOH (2.0 eq.) was added to a stirred solution of ester in EtOH (0.2 M). The reaction was stirred at room temperature for 1-4 hour and monitored by TLC. Upon completion, 10 v/v% aq. HCl was added dropwise to the reaction mixture to achieve ~ pH 2, at which point the product precipitated and was collected by filtration. The precipitate was washed with small quantities of water followed by diethyl ether and dried under vacuum. When precipitate was not observed, or an oily residue precipitated, the crude reaction mixture was diluted with EtOAc and washed with H_2O and brine. The organic layer was dried over Na_2SO_4, filtered and the solvent removed

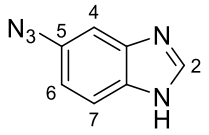
under vacuum to yield the crude product.Reagent quantities are specified for each compound along with the yield and purification method where applicable.

General procedure D: 1,4-regioisomer triazole formation

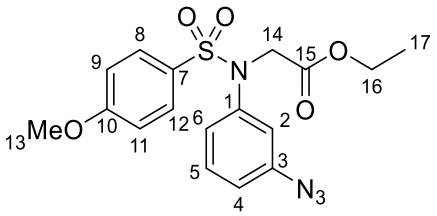
Method 1: Alkyne (1 eq.), azide (1.1 eq.) and CuCl (0.5 eq.) were suspended in a 1:1 solution of ^{t}BuOH: H_2O in a microwave vial and heated to 130 °C for 30 minutes unless otherwise indicated.

Method 2: Alkyne (1 eq.), azide (1.1 eq.), $CuSO_4.5H_2O$ (0.5 eq.) and sodium ascorbate (1.25 eq.) were suspended in a 1:1 solution of ^{t}BuOH: H_2O in a microwave vial and heated to 130 °C for 30 minutes unless otherwise indicated.

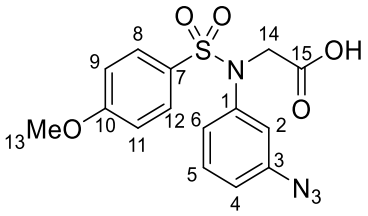
In both cases the reaction mixture was filtered, diluted with EtOAc and washed with H_2O and brine. The organic layer was dried over Na_2SO_4, filtered and the solvent removed under vacuum to yield the crude product. Reagent quantities are specified for each compound along with the yield and purification method.



5-Azido-1*H*-benzoimidazole (AzE4-2). A solution of $NaNO_2$ (0.32 g, 4.7 mmol) in H_2O (~1 mL), was added dropwise to a solution of 5-amino-1*H*-benzoimidazole **212** (0.50 g, 3.8 mmol) in 9 M HCl (15 mL, 0.25 M) cooled to 0 °C. The reaction was stirred at 0 °C for 1 hr at which point a solution of NaN_3 (0.12 g, 1.9 mmol) in H_2O (~ 0.5 mL) was added dropwise. The reaction was stirred for 1 hour at 0 °C, neutralised with solid K_2CO_3 and extracted with EtOAc (1 x 50 mL). The organic layer was washed with H_2O and brine and dried over Na_2SO_4. The solvent was removed under vacuum to yield **AzE4-2** as a pale brown solid (0.26 g, 43%): R_f 0.40 [EtOAc:MeOH:AcOH, 94:5:1]; γ_{max} (neat)/cm^{-1} 2530 (br), 2103 (s), 1629 (w), 1474 (w); ^{1}H NMR (500 MHz, DMSO-d_6) δ: 12.52 (1H, br s, 1-N*H*); 8.24 (1H, s, 2-C*H*); 7.62 (1H, d, *J*= 8.5 Hz, 7-C*H*); 7.31 (1H, d, *J*= 2.1 Hz, 4-C*H*); 6.95 (1H, dd, *J*= 8.5, 2.1 Hz, 6-C*H*); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 143.2 (*C*H-2), 133.4 (*C*), 116.9 (*C*H-7), 113.8 (*C*H-6), 105.1 (*C*H-4), HPLC RT 0.775 min; 75% content; HRMS *m/z* (ESI$^+$) [Found (M+H)$^+$ 160.0615; $C_7H_6N_5H$ requires (M+H)$^+$ 160.0618]. The data are in good agreement with the literature values.[202]

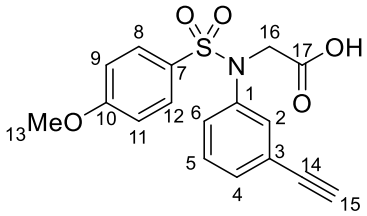


Ethyl-*N*-(3-azidophenyl)-*N*-((4-methoxyphenyl)sulphonyl)glycinate (115). Following general procedure B: Ethyl bromoacetate (0.14 mL, 1.2 mmol), K_2CO_3 (0.17 g, 1.2 mmol), compound **128** (0.25 g, 0.82 mmol), DMF (4.1 mL, 0.2 M). Compound **115** was isolated without purification as a peach solid (0.31 g, 97%): R_f 0.93 [EtOAc:MeOH:AcOH, 94:5:1]; γ_{max} (neat)/cm^{-1} 2100 (s), 1724 (s), 1590 (s), 1480 (s), 1254 (s), 1150 (s); ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.66–7.64 (2H, m, 8-C*H*,12-C*H*); 7.29–7.26 (1H, m, 5-C*H*); 6.99–6.92 (4H, m, 4-C*H*,6-C*H*,9-C*H*,11-C*H*); 6.90 (1H, dd, *J*= 2.2, 2.2 Hz, 2-C*H*); 4.39 (2H, s, 14-C*H*); 4.17 (2H, q, *J*= 7.3 Hz, 16-CH_2); 3.87 (3H, s, 13-CH_3); 1.24 (3H, t, *J*= 7.3 Hz, 17-CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 168.6 (*C*), 163.2 (*C*), 141.5 (*C*), 141.0 (*C*), 130.2 (*C*H-5), 130.2 (*C*), 130.0 (*C*H-8,*C*H-12), 124.6 (*C*H-4/6), 119.3 (*C*H-2), 118.6 (*C*H-4/6), 114.0 (*C*H-9,*C*H-11), 61.6 (*C*H_2-16), 55.6 (*C*H_3-13), 52.5 (*C*H_2-14), 14.1 (*C*H_3-17); HPLC RT 12.8 min; >95% content; HRMS *m/z* (ASAP) [Found $(M+H)^+$ 391.1070; $C_{17}H_{19}N_4O_5SH$ requires $(M+H)^+$ 391.1071].

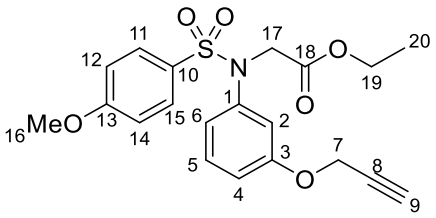


***N*-(3-Azidophenyl)-*N*-((4-methoxyphenyl)sulphonyl)glycine (116).** Following general procedure C: compound **115** (0.30 g, 0.83 mmol), 1 M aq. NaOH (1.7 mL, 1.7 mmol), EtOH (4.2 mL, 0.2 M). Compound **116** was isolated upon precipitation as a colourless powder (0.23 g, 77%): R_f 0.60 [EtOAc:MeOH:AcOH, 94:5:1]; γ_{max} (neat)/cm^{-1} 2964 (br), 2111 (s), 1717 (s), 1587 (s), 1260 (s), 1161 (s); ^{1}H NMR (500 MHz, DMSO-d_6) δ: 12.91 (1H, br s, OH); 7.76–7.60 (2H, m, 8-C*H*,12-C*H*); 7.37 (1H, dd, *J*= 8.0, 8.0 Hz, 5-C*H*); 7.11–7.08 (2H, m, 9-C*H*,11-C*H*); 7.06 (1H, m, 4/6-C*H*); 7.00 (1H, m, 4/6-C*H*); 6.87 (1H, dd, *J*= 2.2, 2.2 Hz, 2-C*H*); 4.42 (2H, s, 14-CH_2); 3.84 (3H, s, 13-CH_3); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 169.9 (*C*); 162.8 (*C*); 141.3 (*C*); 139.8 (*C*), 130.3 (*C*), 129.8 (*C*-5); 129.6 (*C*H-8,*C*H-12); 124.0 (*C*H-4/6); 118.2 (*C*H-2); 118.1 (*C*H-4/6), 114.4

(*C*H-9,*C*H-11); 55.7 (*C*H_3-13), 51.8 (*C*H_2-14); HPLC RT 10.4 min; >95% content; HRMS *m/z* (ESI^-) [Found $(M-H)^-$ 361.0603; $C_{15}H_{12}N_4O_5S$ requires $(M-H)^-$ 361.0612].

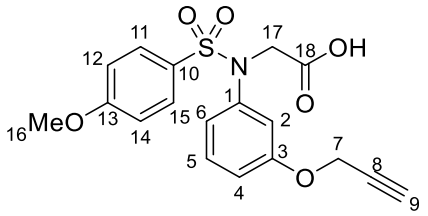


***N*-(3-Ethynylphenyl)-*N*-((4-methoxyphenyl)sulphonyl)glycine (117).** Following general procedure C: compound **135** (0.30 g, 0.67 mmol), 1M aq. NaOH (1.4 mL, 1.4 mmol), EtOH (3.4 mL, 0.2 M). Compound **117** was isolated upon extraction as peach crystals (49 mg, 21%.): R_f 0.70 [EtOAc:MeOH:AcOH, 94:5:1]; ^{1}H NMR (500 MHz, DMSO-d_6) δ: 6.71–6.68 (2H, m, 8-C*H*,12-C*H*); 6.52–6.43 (2H, m, 5-C*H*,4-C*H*); 6.39 (1H, dd, *J*= 2.0, 2.0 Hz, 2-C*H*); 6.31 (1H, ddd, *J*= 8.0, 1.7, 1.7 Hz, 6-C*H*); 6.22–6.18 (2H, m, 9-C*H*,11-C*H*); 3.52 (2H, s, 16-CH_2); 3.36 (1H, s, 15-C*H*); 2.95 (3H, s, 13-CH_3); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 170.0 (*C*), 162.9 (*C*), 140.1 (*C*), 131.0 (*C*H-2), 130.8 (*C*H-4/5), 129.8 (*C*), 129.6 (*C*H-8,*C*H-12), 129.5 (*C*H-4/5), 128.4 (*C*H-6), 122.3 (*C*), 114.5 (*C*H-9,*C*H-11), 82.6 (*C*), 81.6 (*C*H-15), 55.8 (*C*H_3-13), 51.8 (*C*H_2-16); HPLC RT 10.1 min; >95% content; HRMS *m/z* (ESI^-) [Found $(M-H)^-$ 344.0587; $C_{17}H_{14}NO_5S$ requires $(M-H)^-$ 344.0598].

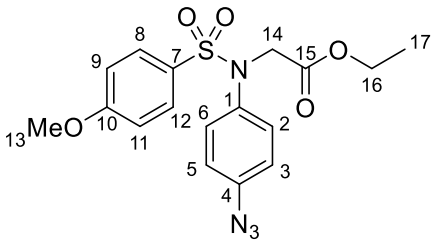


Ethyl-*N*-((4-methoxyphenyl)sulphonyl)-*N*-(3-(prop-2-yn-1-yloxy)phenyl)-glycinate (118). Following general procedure B: ethyl bromoacetate (0.26 mL, 2.4 mmol), K_2CO_3 (0.11 g, 2.4 mmol), compound **142** (0.50 g, 1.6 mmol), DMF (7.9 mL, 0.2 M). Compound **118** was isolated as a yellow oil without purification (0.64 g, *quant*.): R_f 0.30 [Hexane:EtOAc, 7:3]; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.66–7.64 (2H, m, 11-C*H*,15-C*H*); 7.20 (1H, dd, *J*= 8.2, 8.2 Hz, 5-C*H*); 6.92–6.89 (3H, m, 4-C*H*,12-C*H*,14-C*H*); 6.88 (1H, dd, *J*= 2.2, 2.2 Hz, 2-C*H*); 6.83 (1H, m, 6-C*H*); 4.62 (2H, d, *J*= 2.5 Hz, 7-CH_2); 4.39 (2H, s, 17-CH_2); 4.16 (2H, q, *J*= 7.1 Hz, 19-CH_2); 3.86 (3H, s, 13-CH_3); 2.51 (1H, t, *J*= 2.5 Hz, 9-C*H*); 1.24 (3H, t, *J*= 7.1 Hz, 20-CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 168.8 (*C*), 163.0 (*C*), 160.0 (*C*), 141.1 (*C*), 130.4 (*C*), 130.0 (*C*H-11,*C*H-15), 129.7 (*C*H-5), 121.3 (*C*H-6), 115.2 (*C*H-4), 114.8 (*C*H-2), 113.9 (*C*H-12,*C*H-14), 78.1(*C*-8), 75.7 (*C*H-9),

61.4 (*C*H_2-19), 55.9 (*C*H_2-7), 55.6 (*C*H_3-16), 52.7 (*C*H_2-17), 14.1 (*C*H_3-20); HPLC RT 12.2 min; >95% content; HRMS *m/z* (ASAP) [Found $(M+H)^+$ 404.1261; $C_{20}H_{22}NO_6SH$ requires $(M+H)^+$ 404.1168].

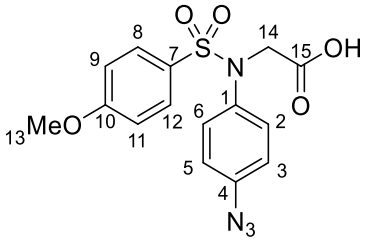


***N*-((4-Methoxyphenyl)sulphonyl)-*N*-(3-(prop-2-yn-1-yloxy)phenyl)glycine (119).** Following general procedure C: compound **118** (0.50 g, 1.3 mmol), 1 M aq. NaOH (2.6 mL, 2.6 mmol), EtOH (6.4 mL, 0.2 M). Compound **119** was isolated upon precipitation as a beige solid (0.33 g, 72%): R_f 0.68 [EtOAc:MeOH:AcOH, 94:5:1];^{1}H NMR (500 MHz, DMSO-d_6) δ: 12.9 (1H, br s, 18-O*H*); 7.62–7.60 (2H, m, 11-C*H*,15-C*H*); 7.24 (1H, dd, *J*= 8.5, 8.5 Hz, 5-C*H*); 7.09–7.07 (2H, m, 12-C*H*,14-C*H*); 6.92–6.89 (1H, m, 4-C*H*); 6.80–6.78 (2H, m, 2-C*H*,6-C*H*); 4.73 (2H, d, *J*= 2.5 Hz, 7-C*H*); 4.38 (2H, s, 17-CH_2); 3.83 (3H, s, 13-CH_3); 3.57 (1H, t, *J*= 2.5 Hz, 9-C*H*); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 170.0 (*C*); 162.7 (*C*); 157.3 (*C*); 141.0 (*C*), 130.1 (*C*), 129.6 (*C*H-5), 129.5 (*C*H-11,*C*H-15), 120.3 (*C*H-6), 114.5 (*C*H-2), 114.4 (*C*H-12,*C*H-14), 113.4 (*C*H-4), 78.9 (*C*-8), 78.5 (*C*H-9), 55.7 (*C*H_3-16), 55.6 (*C*H_2-7), 51.9 (*C*H_2-17); HPLC RT 9.99 min; >95% content; HRMS *m/z* (ESI^-) [Found $(M-H)^-$ 374.0695; $C_{15}H_{13}N_4O_5S$ requires $(M-H)^-$ 374.0704].

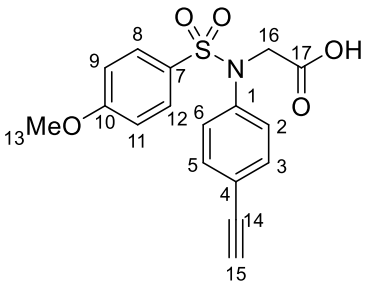


Ethyl-*N*-(4-azidophenyl)-*N*-((4-methoxyphenyl)sulphonyl)glycinate (120). Following general procedure B: Ethyl bromoacetate (0.44 mL, 3.9 mmol), K_2CO_3 (0.44 g, 3.2 mmol), compound **131** (0.80 g, 2.6 mmol), DMF (13 mL, 0.2 M). The crude residue was purified by column chromatography [Hexane:EtOAc, 19:1–9:1] to yield compound **120** as a bright yellow oil (0.95 g, 92%): R_f 0.26 [Hexane:EtOAc, 7:3]; γ_{max} (neat)/cm^{-1} 2121 (s), 2094 (s), 1749 (s), 1595 (s), 1498 (s), 1258 (s), 1153 (s); ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.62–7.60 (2H, m, 8-C*H*,12-C*H*); 7.21–7.18 (2H, m, 3-C*H*,5-C*H*); 6.96–6.91 (4H, m, 2-C*H*,6-CH,9-C*H*,11-C*H*); 4.38 (2H, s, 14-C*H*); 4.16 (2H, q, *J*= 7.2 Hz, 16-CH_2); 3.87 (3H, s, 13-CH_3); 1.24 (3H, t, *J*= 7.2 Hz, 17-CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ:

168.7 (*C*), 163.1 (*C*), 140.0 (*C*), 136.4 (*C*), 130.4 (*C*H-3,*C*H-5), 130.2 (*C*), 130.0 (*C*H-8,*C*H-12), 119.6 (*C*H-2,*C*H-6), 113.9 (*C*H-9,*C*H-11), 61.5 (CH_2-16), 55.6 (CH_3-13), 52.7 (CH_2-14), 14.1 (*C*-$H_3$17); HPLC RT 12.8 min; >95% content; HRMS *m/z* (ASAP) [Found $(M+H)^+$ 391.1066; $C_{17}H_{19}N_4O_5SH$ requires $(M+H)^+$ 391.1076].

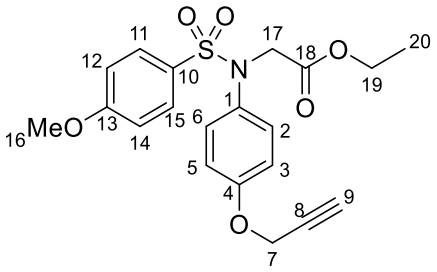


***N*-(4-Azidophenyl)-*N*-((4-methoxyphenyl)sulphonyl)glycine (121).** Following general procedure C: compound **120** (0.62 g, 1.7 mmol), 1 M aq. NaOH (3.4 mL, 3.4 mmol), EtOH (8.6 mL, 0.2 M). Compound **121** was isolated upon precipitation as a beige powder (0.56 g, 90%): R_f 0.60 [EtOAc:MeOH:AcOH, 94:5:1]; γ_{max} (neat)/cm^{-1} 2986 (br), 2124 (s), 1708 (s), 1498 (s), 1348 (s), 1229 (s), 1162 (s); ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.59 (2H, d, *J*= 8.8 Hz, 8-C*H*,12-C*H*); 7.18 (2H, d, *J*= 8.8 Hz, 3-C*H*,5-C*H*); 6.97–6.92 (4H, m, 2-C*H*,6-C*H*,9-C*H*,11-C*H*); 4.42 (2H, s, 14-CH_2); 3.88 (3H, s, 13-CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 171.9 (*C*); 163.3 (*C*); 140.3 (*C*); 136.2 (*C*), 130.4 (*C*H-3,*C*H-5), 130.0 (*C*H-8,*C*H-12); 129.9 (*C*); 119.8 (*C*H-2,*C*H-6); 114.1 (*C*H-9,*C*H-11); 55.6 (CH_3-13), 52.2 (CH_2-14); HPLC RT 10.5 min; >95% content; HRMS *m/z* (ESI^-) [Found $(M-H)^-$ 361.0606; $C_{15}H_{12}N_4O_5S$ requires $(M-H)^-$ 361.0612].

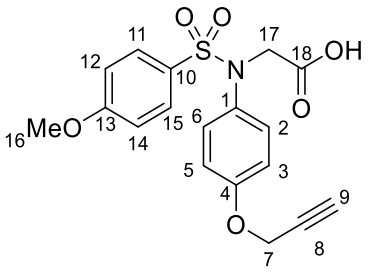


***N*-(4-Ethynylphenyl)-*N*-((4-methoxyphenyl)sulphonyl)glycine (122).** Following general procedure C: compound **139** (45 mg, 0.12 mmol), 1M aq. NaOH (0.24 mL, 0.24 mmol), EtOH (0.6 mL), 0.2 M). Compound **122** was isolated upon precipitation as a yellow solid (15 mg, 33%.): R_f 0.70 [EtOAc:MeOH:AcOH, 94:5:1]; ^{1}H NMR (500 MHz, DMSO-d_6) δ: 7.58 (2H, d, *J*= 9.0 Hz, 8-C*H*,12-C*H*); 7.42 (2H, d, *J*= 8.7 Hz, 3-C*H*,5-C*H*); 7.18 (2H, d, *J*= 8.7 Hz, 2-C*H*,6-C*H*); 7.07 (2H, d, *J*= 9.0 Hz, 9-C*H*,11-C*H*); 4.39 (2H, s, 16-CH_2); 4.23 (1H, s, 15-C*H*); 3.83 (3H, s, 13-CH_3); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 169.9 (*C*), 162.8 (*C*), 140.4 (*C*), 132.3 (*C*H-3,*C*H-5); 129.8 (*C*), 129.6 (*C*H-8,*C*H-12),

127.3 (*C*H-2,*C*H-6), 120.4 (*C*), 114.5 (*C*H-9,*C*H-11), 82.8 (*C*), 81.6 (*C*H-15), 55.8 (CH_3-13), 51.7 (CH_2-16); HPLC RT 10.3 min; >95% content; HRMS *m/z* (ESI^-) [Found $(M-H)^-$ 344.0588; $C_{17}H_{14}NO_5S$ requires $(M-H)^-$ 344.0598].

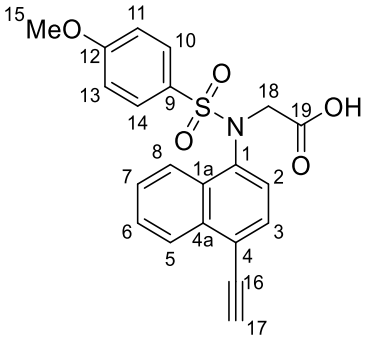


Ethyl-*N*-((4-methoxyphenyl)sulphonyl)-*N*-(4-(prop-2-yn-1-yloxy)phenyl)-glycinate (123). Following general procedure B: ethyl bromoacetate (0.21 mL, 1.9 mmol), K_2CO_3 (0.26 g, 1.9 mmol), compound **148** (0.40 g, 1.3 mmol), DMF (6.3 mL, 0.2 M). Compound **123** was isolated as a yellow oil without purification (0.64 g, *quant*.): R_f 0.30 [Hexane:EtOAc, 7:3]; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.63–7.60 (2H, m, 11-C*H*,15-C*H*); 7.15–7.12 (2H, m, 2-C*H*,6-C*H*); 6.93–6.86 (4H, m, 3-C*H*,5-C*H*,11-C*H*,15-C*H*); 4.67 (2H, d, *J*= 2.5 Hz, 7-C*H*); 4.36 (2H, s, 17-C*H*); 4.16 (2H, q, *J*= 7.1 Hz, 19-C*H*); 3.87 (3H, s, 16-C*H*); 2.54 (1H, t, *J*= 2.5 Hz, 9-C*H*); 1.24 (3H, t, *J*= 7.1 Hz, 20-C*H*); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 168.9 (*C*), 163.0 (*C*), 157.2 (*C*), 133.2 (*C*), 130.6 (*C*H-2,*C*H-6); 130.0 (*C*H-11,*C*H-15), 115.2 (*C*H-12,*C*H-14), 113.8 (*C*H-3,*C*H-5), 78.1 (*C*), 75.8 (*C*H-9), 61.4 (CH_2-19), 55.9 (CH_2-7), 55.6 (CH_3-16), 53.0 (CH_2-17); 14.1 (CH_3-20); HPLC RT 12.2 min; >95% content; HRMS *m/z* (ASAP) [Found $(M+H)^+$ 404.1171; $C_{20}H_{22}NO_6SH$ requires $(M+H)^+$ 404.1168].

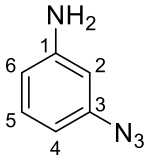


***N*-((4-Methoxyphenyl)sulphonyl)-*N*-(4-(prop-2-yn-1-yloxy)phenyl)glycine (124).** Following general procedure C: compound **123** (0.50 g, 1.2 mmol), 1 M aq. NaOH (2.5 mL, 2.5 mmol), EtOH (6.2 mL, 0.2 M). Compound **124** was isolated upon extraction as a beige powder (0.38 g, 82%): R_f 0.68 [EtOAc:MeOH:AcOH, 94:5:1]; ^{1}H NMR (500 MHz, DMSO-d_6) δ: 12.80 (1H, br s, 18-O*H*); 7.56 (2H, d, *J*= 8.8 Hz, 11-C*H*,15-C*H*); 7.09–7.06 (4H, m, 2-C*H*,6-C*H*,12-C*H*,14-C*H*); 6.91 (2H, d, *J*= 8.8 Hz, 3-C*H*,5-C*H*); 4.78 (2H, d, *J*= 2.4 Hz, 7-CH_2); 4.32 (2H, s, 17-CH_2); 3.83 (3H, s, 16-CH_3); 3.58 (1H, t, *J*=

2.4 Hz, 9-C*H*); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 170.1 (*C*), 162.6 (*C*), 156.4 (*C*), 132.9 (*C*), 130.2 (*C*), 129.8 (*C*H-2,*C*H-6), 129.5 (*C*H-11,*C*H-15), 114.9 (*C*H-3,*C*H-5), 114.3 (*C*H-12,*C*H-14), 79.1 (*C*-8), 78.4 (*C*H-9), 55.7 (*C*H$_3$-16), 55.5 (*C*H$_2$-7), 52.4 (*C*H$_2$-17); HPLC RT 9.99 min; >95% content; HRMS *m/z* (ESI$^-$) [Found (M-H)$^-$ 374.0700; $C_{18}H_{15}NO_6S$ requires (M-H)$^-$ 374.704].

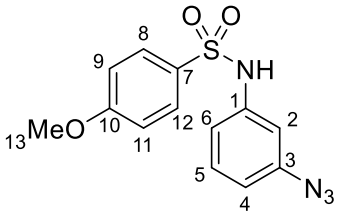


***N*-(4-Ethynylnaphthalen-1-yl)-*N*-((4-methoxyphenyl)sulphonyl)glycine (125).** Following general procedure C: compound **152** (0.42 g, 0.85 mmol), 1 M aq. NaOH (1.7 mL, 1.7 mmol), EtOH (4.3 mL, 0.2 M). Compound **125** was isolated upon precipitation as a colourless powder (0.24 g, 71%): R_f 0.75 [EtOAc:MeOH:AcOH, 94:5:1]; ^{1}H NMR (500 MHz, DMSO-d_6) δ: 12.83 (1H, br s, 19-O*H*); 8.25–8.22 (2H, m, 5-C*H*,8-C*H*); 7.70–7.66 (2H, m, 3-C*H*,6-C*H*); 7.63–7.58 (3H, m, 7-C*H*,10-C*H*,14-C*H*); 7.16 (1H, d, *J*= 7.6 Hz, 2-C*H*); 7.10–7.07 (2H, m, 11-C*H*,13-C*H*); 4.71 (1H, s, 17-C*H*); 4.45 (1H, d, *J*= 18 Hz, 18-C*H*$_2$); 4.40 (1H, d, *J*= 18 Hz, 18-C*H*$_2$); 3.85 (3H, s, 15-C*H*$_3$); ^{13}C NMR (125 MHz, CD_3OD) δ: 169.9 (*C*), 162.9 (*C*), 137.5 (*C*), 133.7 (*C*), 131.8 (*C*), 130.4 (*C*H-3/6), 130.0 (*C*H-10,*C*H-14), 129.3 (*C*), 127.7 (*C*H-3/6), 127.0 (H*C*-7), 126.7 (*C*H-2), 125.4 (*C*H-5/8), 125.1 (*C*H-5/8), 120.1 (*C*), 114.4 (*C*H-11,*C*H-13), 87.0 (*C*-16), 80.8 (*C*H-17), 55.8 (*C*H$_3$-15), 53.0 (*C*H$_2$-18); HPLC RT 11.5 min; >90% content; HRMS *m/z* (ESI$^-$) [Found (M-H)$^-$ 394.0744; $C_{21}H_{15}NO_5S$ requires (M-H)$^-$ 394.0755].

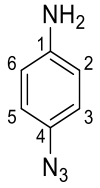


3-Azidoaniline (127). 3-Iodoaniline **126** (0.50 g, 2.3 mmol), NaN_3 (0.30 g, 4.6 mmol), sodium ascorbate (23 mg, 0.11 mmol), CuI (43 mg, 0.23 mmol) were suspended in a 5:1 solution of DMSO:H_2O (6:1.6 mL). The reaction mixture was placed under an argon atmosphere and *N*^{1},*N*^{2}-dimethylethane-1,2-diamine (37 µL, 0.34 mmol) was added. The reaction was allowed to stir at room temperature for 1 hr and monitored by LCMS.

Upon completion, the reaction was diluted with EtOAc and washed with H_2O and brine. The organic layer was dried over Na_2SO_4, filtered and the solvent removed under vacuum to yield the crude product. The crude residue was purified by column chromatography [Hexane:Acetone, 9:1–7:3] to yield compound **127** as a yellow waxy solid (0.25 g, 82%): R_f 0.24 [Hexane: Acetone, 8:2]; γ_{max} (neat)/cm^{-1} 3355 (w), 2103 (s), 1586 (s), 1490 (s); ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.13 (1H, dd, *J*= 8.0, 8.0 Hz, 5-C*H*); 6.46 (2H, m, 4-C*H*,6-C*H*); 6.33 (1H, m, 2-C*H*); 3.75 (2H, br s, 1-NH_2); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 148.1 (*C*-1), 141.2 (*C*-3), 130.7 (*C*H-5), 112.0 (*C*H-4/6), 109.3 (*C*H-4/6), 105.6 (*C*H-2); LCMS RT 2.76 min; >99% content; *m/z* 176.1 ([M+MeCN+H]$^+$ 100%). The data are in good agreement with the literature values.[219]

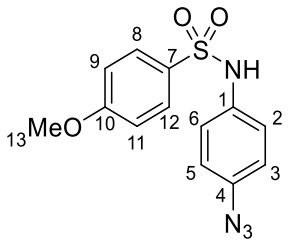


***N*-(3-Azidophenyl)-4-methoxybenzenesulphonamide (128).** Following general procedure A: compound **127** (0.20 g, 0.96 mmol), DMAP (12 mg 0.10 mmol), *p*-methoxybenzenesulphonyl chloride (0.24 g, 1.1 mmol), pyridine (0.15 mL, 1.9 mmol), DCM (4.0 mL, 0.23 M). The crude residue was purified by column chromatography [Hexane:EtOAc, 95:5 - 80:20] to yield compound **128** as a pale yellow oil (0.27 g, 94%): R_f 0.32 [Hexane:EtOAc, 7:3]; γ_{max} (neat)/cm^{-1} 3697 (w), 2107 (s), 1496 (w), 1330 (s), 1246 (s), 1150 (s), 1041 (s); ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.77–7.75 (2H, m, 8-C*H,*12-C*H*); 7.23–7.19 (1H, m, 5-C*H*); 6.96 (1H, br s, 1-N*H*); 6.94–6.91 (2H, m, 9-C*H,*11-C*H*); 6.85–6.82 (1H, m, 2-C*H*); 6.78–6.76 (2H, m, 4-C*H*,6-C*H*); 3.85 (3H, s, 13-CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 163.3 (*C*), 141.2 (*C*), 136.2 (*C*), 130.5 (*C*H-5), 130.2 (*C*), 129.5 (*C*H-8,*C*H-12), 117.1 (*C*H-2), 115.4 (*C*H-4/6), 114.3 (*C*H-9,*C*H-11), 111.3 (*C*H-4/6), 55.6 (*C*H_3-13); HPLC RT 11.1 min; >95% content; HRMS *m/z* (ESI$^+$) [Found (M+NH_4)$^+$ 322.0969; $C_{13}H_{16}N_5O_3SNH_4$ requires (M+NH_4)$^+$ 322.0968].

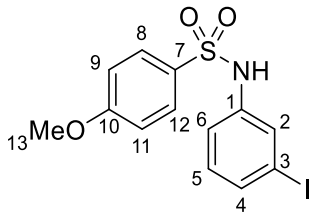


4-Azidoaniline (130). 4-Iodoaniline **129** (1.0 g, 4.6 mmol), NaN_3 (0.59 g, 9.1 mmol), sodium ascorbate (45 mg, 0.23 mmol), CuI (87 mg, 0.46 mmol) were suspended in a

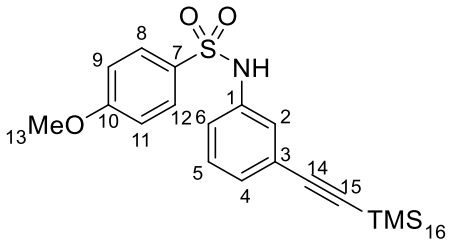
5:1 solution of DMSO:H_2O (12:3.2 mL). The reaction mixture was placed under an argon atmosphere and N^1,N^2-dimethylethane-1,2-diamine (74 µL, 0.68 mmol) was added. The reaction was allowed to stir at room temperature for 1 hr and monitored by LCMS. Upon completion, the reaction was diluted with EtOAc and washed with H_2O and brine. The organic layer was dried over Na_2SO_4, filtered and the solvent removed under vacuum to yield the crude product. The crude residue was purified by column chromatography [Hexane:Acetone, 9:1–8:2] to yield compound **130** as red crystalline needles (0.55 g, 91%): R_f 0.24 [Hexane:Acetone, 8:2]; γ_{max} (neat)/cm^{-1} 3395 (w), 2108 (s); 1586 (s), 1490 (s); ^{1}H NMR (500 MHz, $CDCl_3$) δ: 6.86–6.83 (2H, m, 3-C*H*,5-C*H*); 6.70–6.67 (2H, m, 2-C*H*,6-C*H*); 3.64 (2H, br s, 1-NH_2); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 143.7 (*C*-1), 130.2 (*C*-4), 120.0 (*C*H-3,*C*H-5), 116.3 (*C*H-2,*C*H-6); LCMS RT 1.52 min; >95% content; *m/z* 176.1 ([M+MeCN+H]$^+$ 100%). The data are in good agreement with the literature values.[219]



***N*-(4-Azidophenyl)-4-methoxybenzenesulphonamide (131).** Following general procedure A: compound **130** (0.45 g, 3.4 mmol), DMAP (41 mg 0.34 mmol), *p*-methoxybenzenesulphonyl chloride (0.83 g, 4.0 mmol), pyridine (0.54 mL, 6.7 mmol), DCM (15 mL, 0.23 M). The crude residue was purified by column chromatography [Hexane:EtOAc, 8:2–6:4] to yield compound **131** as a pale red solid (0.96 g, 94%): R_f 0.1 [Hexane:EtOAc, 8:2]; γ_{max} (neat)/cm^{-1} 3707 (w), 3244 (w), 2108 (s), 1496 (s), 1150 (s), 1327 (s), 1266 (s), 1032 (s); ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.70–7.69 (2H, m, 8-C*H*,12-C*H*); 7.09–7.07 (2H, m, 3-C*H*,5-C*H*); 6.93–6.91 (4H, m, 2-C*H*,6-C*H*, 9-C*H*,11-C*H*); 6.73 (1H, s, 1-N*H*); 3.86 (3H, s, 13-CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 162.9 (*C*); 137.3 (*C*); 133.0 (*C*); 130.0 (C); 129.2 (*C*H-8,*C*H-12); 123.7 (*C*H-3,*C*H-5); 119.6 (*C*H-2/6/9/11); 114.0 (*C*H-2/6/9/11); 55.3 (*C*H_3-13); HPLC RT 11.2 min; >95% content; HRMS *m/z* (ASAP) [Found (M-N_2+H)$^+$ 277.0641; $C_{13}H_{13}N_2O_3SH$ requires (M-N_2+H)$^+$ 277.0647].

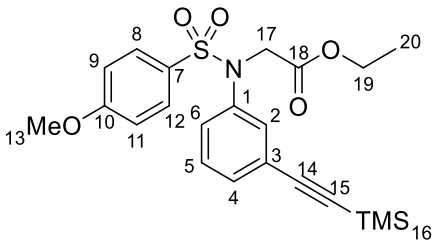


***N*-(3-Iodophenyl)-4-methoxybenzenesulphonamide (133).** Following general procedure A: 3-Iodoaniline **132** (2.0 g, 9.1 mmol), DMAP (0.11 g, 0.91 mmol), *p*-methoxybenzenesulphonyl chloride (2.3 g, 11 mmol), pyridine (1.5 mL, 18 mmol), DCM (40 mL, 0.23 M). Compound **133** was obtained as a pale yellow oil (2.9 g, 80%): R_f 0.19 [Hexane:EtOAc, 8:2]; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.76–7.75 (2H, m, 8-C*H*,12-C*H*); 7.44 (1H, dd, *J*= 1.9, 1.9 Hz, 2-C*H*); 7.43–7.41 (1H, m, 4-C*H*); 7.12 (1H, br s, 1-N*H*); 7.09 (1H, m, 6-C*H*); 6.97–6.92 (3H, m, 5-C*H*,9-C*H*,11-C*H*); 3.84 (3H, s, 11-CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 163.3 (*C*); 137.9 (*C*); 134.1 (*C*H-4); 131.0 (*C*H-5); 130.1 (*C*); 129.6 (*C*H-2); 129.4 (*C*H-8,*C*H-12); 120.1 (*C*H-6), 114.6 (*C*H-9,*C*H-11); 94.2 (*C*); 55.6 (CH_3-13); LCMS RT 4.30 min; 82% content; *m/z* 388.5 ([M-H]$^-$ 100%). The data are in good agreement with the literature values.[220]

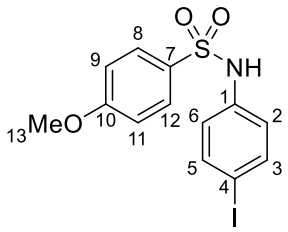


4-Methoxy-*N*-(3-((trimethylsilyl)ethynyl)phenyl)benzenesulphonamide (134). CuI (49 mg, 0.30 mmol) and Pd[$(PPh_3)_2Cl_2$] (90 mg, 0.10 mmol) were added to a stirred solution of sulphonamide **133** (1.0 g, 2.6 mmol) in dry THF (5.1 mL, 0.5 M). Et_3N (5.1 mL, 0.5 M) and TMS-acetylene (0.40 mL, 2.8 mmol) were added to the solution and the reaction was stirred overnight at room temperature. The resulting solution was filtered through a pad of Celite® and the solvent removed under vacuum. The crude residue was purified by column chromatography [Hexane:EtOAc, 19:1–8:2] to yield compound **134** as a red oil (0.76 g, 83%): R_f 0.23 [Hexane:EtOAc, 8:2]; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.73–7.70 (2H, m, 8-C*H*,12-C*H*); 7.21–7.15 (3H, m, 4-C*H*,5-C*H*,6-C*H*); 7.08–7.06 (2H, m, 2-C*H*); 6.92–6.90 (2H, m, 9-C*H*,11-C*H*); 6.78 (1H, br s, 1-N*H*); 3.84 (3H, s, 13-CH_3); 0.24 (9H, s, 16-CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 163.2 (*C*), 136.6 (*C*), 130.3 (*C*), 129.4 (*C*H-8,*C*H-12), 129.2 (*C*H-5), 128.7 (*C*H-4/6), 124.5 (*C*H-4/6), 124.3 (*C*), 121.6 (*C*H-2), 114.3 (*C*H-9,*C*H-11), 103.9 (*C*), 95.2 (*C*), 55.6 (CH_3-13), 0.14 (CH_3-16);

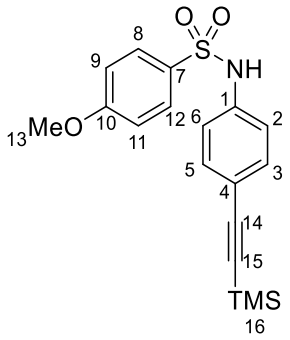
HPLC RT 14.6 min; >90% content; HRMS *m/z* (ESI$^+$) [Found $(M+NH_4)^+$ 377.1357; $C_{18}H_{25}N_2O_3SSiNH_4$ requires $(M+NH_4)^+$ 377.1350].



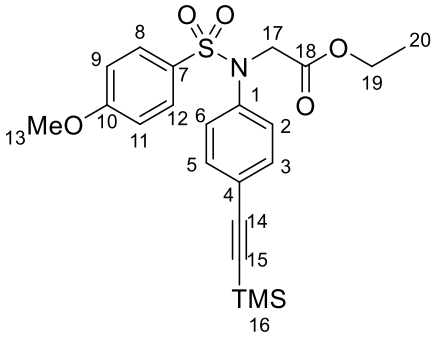
Ethyl-*N*-((4-methoxyphenyl)sulphonyl)-*N*-(3-((trimethylsilyl)ethynyl)phenyl)-glycinate (135). Following general procedure B: Ethyl bromoacetate (0.23 mL, 2.1 mmol), K_2CO_3 (0.29 g, 2.1 mmol), compound **134** (0.50 g, 1.4 mmol), DMF (7.0 mL, 0.2 M). Key proton signals indicated the crude residue contains compound **135** and was taken forward as crude: R_f 0.45 [Hexane:EtOAc, 7:3]; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.63–7.61 (2H, m, 8-C*H*,12-C*H*); 7.37 (1H, ddd, *J*= 7.6, 2.0, 1.7 Hz, 4-C*H*); 7.32 (1H, dd, *J*= 2.0, 1.7 Hz, 2-C*H*); 7.23–7.17 (2H, m, 5-C*H*, 6-C*H*); 6.93–6.92 (2H, m, 9-C*H*,11-C*H*); 4.37 (2H, s, 17-CH_2); 4.16 (2H, q, *J*= 7.2 Hz, 19-CH_2); 3.87 (3H, s, 13-CH_3); 1.24 (3H, t, *J*= 7.2 Hz, 20-CH_3); 0.24 (9H, s, 16-C*H*).



***N*-(4-Iodophenyl)-4-methoxybenzenesulphonamide (137).** Following general procedure A: 4-Iodoaniline **136** (1.0 g, 4.6 mmol), DMAP (56 mg, 0.50 mmol), *p*-methoxybenzenesulphonyl chloride (1.1 g, 5.5 mmol), pyridine (0.74 mL, 9.1 mmol), DCM (20 mL, 0.23 M). The crude residue was purified by column chromatography [Hexane:Acetone, 1:0–9:1] to yield compound **137** as a colourless solid (1.6 g, 91%): R_f 0.19 [Hexane:Acetone, 9:1]; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.72 (2H, d, *J*= 9.1 Hz, 8-C*H*,12-C*H*); 7.54 (2H, d, *J*= 8.5 Hz, 3-C*H*,5-C*H*); 6.96 (1H, br s, 1-N*H*); 6.91 (2H, d, *J*= 9.1 Hz, 9-C*H*,11-C*H*); 6.85 (2H, d, *J*= 8.5 Hz, 2-C*H*,6-C*H*); 3.84 (3H, s, 13-CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 163.3 (*C*); 138.3 (*C*H-3,*C*H-5); 136.5 (*C*); 130.1 (*C*); 129.4 (*C*H-8,*C*H-12); 123.1 (*C*H-2,*C*H-6); 114.3 (*C*H-9,*C*H-11); 89.1 (*C*); 55.6 (*C*H_3-13); LCMS RT 4.28 min; 88% content; *m/z*: 387.5 ([M-H]$^-$ 100%). The data are in good agreement with the literature values.[220]

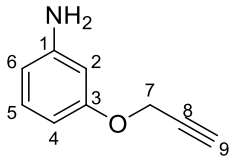


4-Methoxy-*N*-(4-((trimethylsilyl)ethynyl)phenyl)benzenesulphonamide (138). CuI (49 mg, 0.30 mmol) and Pd[$(PPh_3)_2Cl_2$] (90 mg, 0.10 mmol) were added to a stirred solution of sulphonamide **137** (1.0 g, 2.6 mmol) in dry THF (5.1 mL, 0.5 M). Et_3N (5.1 mL, 0.5 M) and TMS-acetylene (0.40 mL, 2.8 mmol) were added to the solution and the reaction was stirred overnight at room temperature. The resulting solution was filtered through a pad of Celite® and the solvent removed under vacuum. The crude residue was purified by column chromatography [Hexane:EtOAc, 19:1 - 8:2] to yield compound **138** as a red solid (0.61 g, 66%): R_f 0.32 [Hexane:EtOAc, 8:2]; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.72–7.69 (2H, m, 8-C*H*,12-C*H*); 7.35–7.33 (2H, m, 3-C*H*,5-C*H*); 7.03–7.01 (2H, m, 2-C*H*,6-C*H*); 6.90–6.88 (2H, m, 9-C*H*,11-C*H*); 6.87 (1H, br s, 1-N*H*); 3.81 (3H, s, 13-CH_3); 0.23 (9H, s, 16-CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 163.3 (*C*), 136.7 (*C*), 133.1 (*C*H-3,*C*H-5), 130.1 (*C*), 129.4 (*C*H-8,*C*H-12), 120.57 (*C*H-2,*C*H-6), 119.7 (*C*), 114.6 (*C*H-9,*C*H-11), 104.2 (*C*), 94.5 (*C*), 55.6 (*C*H_3-13), 0.10 (*C*H_3-16); HPLC >95% content; RT 14.9 min; HRMS *m/z* (ESI^+) [Found $(M+H)^+$ 360.1081; $C_{18}H_{22}NO_3SSiH$ requires $(M+H)^+$ 360.1084].

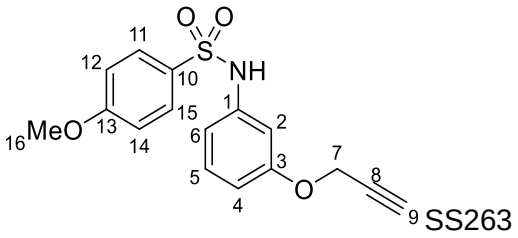


Ethyl-*N*-((4-methoxyphenyl)sulphonyl)-*N*-(4-((trimethylsilyl)ethynyl)phenyl)-glycinate (139). Following general procedure B: Ethyl bromoacetate (46 µL, 0.40 mmol), K_2CO_3 (58 mg, 0.40 mmol), compound **138** (100 mg, 0.28 mmol), DMF (1.4 mL, 0.2 M). The crude residue was purified by column chromatography [Hexane: EtOAc, 8:2 - 7:3] to yield compound **139** as a clear yellow gum (110 mg, 88%): R_f 0.37 [Hexane:Acetone, 8:2]; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.60–7.59 (2H, m, 8-C*H*,12-C*H*);

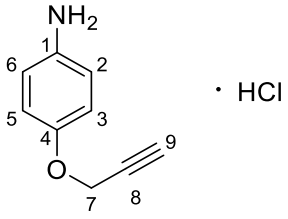
7.39–7.37 (2H, m, 3-C*H*,5-C*H*); 7.15–7.14 (2H, m, 2-C*H*,6-C*H*); 6.91–6.89 (2H, m, 9-C*H*,11-C*H*); 4.39 (2H, s, 17-CH_2); 4.16 (2H, q, *J*= 7.2 Hz, 19-CH_2); 3.86 (3H, s, 13-CH_3); 1.23 (3H, t, *J*= 7.2 Hz, 20-CH_3); 0.24 (9H, s, 16-CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 168.7 (*C*), 163.1 (*C*), 139.9 (*C*), 132.7 (*C*-Ar), 130.0 (*C*-Ar), 128.0 (*C*-Ar), 122.8 (*C*), 114.0 (*C*-Ar), 103.4 (*C*), 95.6 (*C*), 61.5 (*C*H_2-19), 55.6 (*C*H_3-13), 52.4 (*C*H_2-17), 14.1 (*C*H_3-20), 0.13 (*C*H_3-16); HPLC RT 16.1 min; >95% content; HRMS *m/z* (ASAP) [Found $(M+H)^+$ 446.1465; $C_{22}H_{28}NO_5SSiH$ requires $(M+H)^+$ 446.1458].



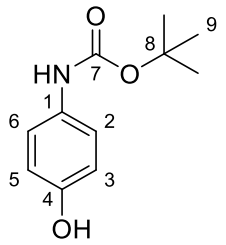
3-(Prop-2-yn-1-yloxy)aniline (141). 3-Hydroxyaniline **140** (1.0 g, 9.2 mmol) and NaOH (0.39 g, 9.8 mmol) were dissolved in dry MeOH (18 mL, 0.5 M) and placed under argon. The reaction mixture was heated to 40 °C for 1 hr, the solution was cooled and the organic solvent was removed under vacuum until a dry solid was obtained. The crude residue was placed under argon and suspended in dry MeCN (18 mL, 0.5 M). Propargyl bromide in toluene (80% *v/v* 1.6 mL, 11 mmol) was added in 3 portions over 1 hr. The reaction was stirred at room temperature overnight and monitored by TLC. Upon completion the reaction was diluted with diethyl ether and washed with 0.1 M aq. NaOH solution, H_2O, brine and the organic layer was dried over Na_2SO_4. The residue was dissolved in MeOH (~4 mL) and treated with 4 M HCl in dioxane (12 mL). The solvent was removed under vacuum to yield a peach solid which was washed with boiling ethyl acetate, dissolved in H_2O and adjusted to ~ pH 11 using 1 M aq. NaOH solution. The product was extracted with diethyl ether (2 x 100 mL) and washed with H_2O, brine and dried over Na_2SO_4. The solvent was removed under vacuum to yield compound **141** as a dark yellow oil (0.96 g, 74%): R_f 0.24 [Hexane:EtOAc, 8:2]; ^{1}H NMR (500 MHz, DMSO-d_6) δ: 6.90 (1H, dd, *J*= 8.0, 8.0 Hz, 5-C*H*); 6.19–6.16 (2H, m, 2-C*H*,4/6-C*H*); 6.13 (1H, dd, *J*= 8.0, 1.7 Hz, 4/6-C*H*); 5.07 (2H, br s, 1-NH_2); 4.65 (1H, d, *J*= 2.3 Hz, 7-C*H*); 3.52 (1H, t, *J*= 2.3 Hz, 9-C*H*); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 158.3 (*C*-1), 150.0 (*C*-3), 129.5 (*C*H-5), 107.5 (*C*H-4/6), 102.0 (*C*H-4/6), 100.3 (*C*H-2), 79.7 (*C*-8), 77.8 (*C*H-9), 54.9 (*C*H_2-7); HPLC RT 2.18 min; 84% content; *m/z* $(ESI)^+$ 189.05 ($[M+MeCN+H]^+$ 100%). The data are in good agreement with the literature values.[221]



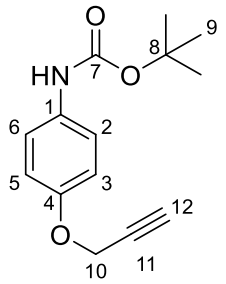
4-Methoxy-*N*-(3-(prop-2-yn-1-yloxy)phenyl)benzenesulphonamide (142). Following general procedure A: Compound **141** (0.8 g, 3.7 mmol), DMAP (45 mg, 0.37 mmol), *p*-methoxybenzenesulphonyl chloride (0.91 g, 4.4 mmol), pyridine (0.59 mL, 7.3 mmol), DCM (16 mL, 0.23 M). Compound **142** was isolated as a yellow oil (1.3 g, 91%): R_f 0.05 [Hexane:EtOAc, 8:2]; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.77–7.75 (2H, m, 11-C*H*,15-C*H*); 7.14 (1H, dd, *J*= 8.2, 8.2 Hz, 5-C*H*); 7.03 (1H, br s, 1-N*H*); 6.91–6.89 (2H, m, 12-C*H*,14-C*H*); 6.78 (2H, dd, *J*= 2.1, 2.2 Hz, 2-C*H*); 6.78–6.68 (2H, m, 4-C*H*,6-C*H*); 4.63 (2H, d, *J*= 2.4 Hz, 7-CH_2); 3.83 (3H, s, 16-CH_3); 2.50 (1H, t, *J*= 2.4 Hz, 9-C*H*); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 163.1 (*C*); 158.2 (*C*), 137.9 (*C*), 130.4 (*C*); 130.1 (*C*H-5), 129.4 (*C*H-11,*C*H-15), 114.2 (*C*H-12,*C*H-14), 114.1 (*C*H-4/6), 111.5 (*C*H-4/6), 107.7 (*C*H-2), 78.2 (C-8), 75.7 (*C*H-9), 55.8 (*C*H_2-7), 55.6 (*C*H_3-16); HPLC RT 10.6 min; >95% content; HRMS *m/z* (ESI$^+$) [Found $(M+H)^+$ 318.0795; $C_{16}H_{16}NO_4SH$ requires $(M+H)^+$ 318.0795].



4-(Prop-2-yn-1-yloxy)aniline (144). 4M HCl in dioxane (2 mL, 9.2 mmol) was added to a solution of compound **147** (0.46 g, 1.9 mmol) in MeOH (3 mL, 0.6 M). The solution was stirred at room temperature for 2 hour. The solvent was removed under vacuum to yield compound **144** as an orange solid (0.32 g, 95%): R_f 0.18 [Hexane:EtOAc, 7:3]; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 6.84 (2H, d, *J*= 8.8 Hz, 2-C*H*,6-C*H*); 6.66 (2H, d, *J*= 8.8 Hz, 3-C*H*,5-C*H*); 4.62 (2H, d, *J*= 2.5 Hz, 7-CH_2); 3.40 (2H, br s, 4-NH_2); 2.50 (1H, t, *J*= 2.5 Hz, 9-C*H*); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 150.7 (*C*), 140.9 (*C*), 116.3 (*C*H-2,*C*H-6), 116.2 (*C*H-3,*C*H-5), 79.1 (*C*-8), 75.1 (*C*H-9), 56.7 (*C*H_2-7); LCMS RT 0.76 min; 77% content; *m/z* (ESI)$^+$ 189.1 ([M+MeCN+H]$^+$ 100%). The data are in good agreement with the literature values.[222]

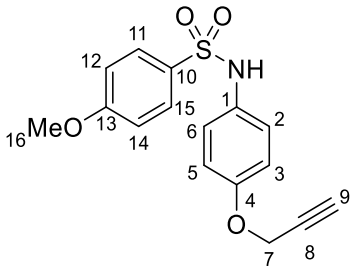


***tert-butyl*(4-Hydroxyphenyl)carbamate (146).** A solution of Boc_2O (2.0 g, 9.2 mmol) in THF (5 mL) was added dropwise over 10 mins to a solution of 4-hydroxyaniline **143** (1.0 g, 9.2 mmol) in THF (20 mL, 0.5 M) and stirred at room temperature for 18 hrs. The solvent was removed under vacuum to yield compound **146** as a pale brown solid (1.9 g, *quant.*): R_f 0.20 [Hexane:EtOAc, 7:3]; ^{1}H NMR (500 MHz, DMSO-d_6) δ: 9.02 (1H, br s, 4-N*H*); 8.97 (1H, br s, 4-O*H*); 7.20 (2H, d, *J*= 9.0 Hz, 2-C*H*,6-C*H*); 6.63 (2H, d, *J*= 9.0 Hz, 3-C*H*,5-C*H*); 1.44 (9H, s, 9-CH_3); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 153.0 (*C*), 152.5 (*C*), 131.0 (*C*), 120.0 (*C*H-2,*C*H-6), 115.0 (*C*H-3,*C*H-5), 78.4 (*C*-8), 28.2 (*C*H_3-9); HPLC RT 3.53 min; >95% content; *m/z* (ESI$^+$) 249.3 ([M+K]$^+$ 100%). The data are in good agreement with the literature values.[223]

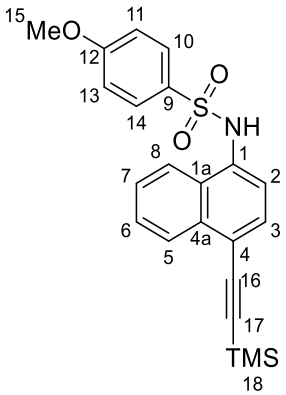


***tert-butyl*(4-(Prop-2-yn-1-yloxy)phenyl)carbamate (147).** Solid K_2CO_3 (0.36 g, 2.6 mmol) was added to a solution of compound **146** (0.50 g, 2.4 mmol) in DMF (8.0 mL, 0.3 M). The reaction was placed under an argon atmosphere and propargyl bromide in toluene (80% v/v, 0.39 mL, 2.6 mmol) was added, the reaction was stirred at room temperature overnight. Upon completion, the reaction was diluted with H_2O (~20 mL) and extracted with EtOAc (2 x 50 mL). The combined organic layers were washed with H_2O, brine and dried over $MgSO_4$. The crude residue was purified by column chromatography [Hexane:EtOAc, 9:1–8:2] to yield compound **147** as a pale yellow wax (0.47 g, 80%): R_f 0.69 [Hexane:EtOAc, 7:3]; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.29 (2H, d, *J*= 9.0 Hz, 2-C*H*,6-C*H*); 6.93 (2H, d, *J*= 9.0 Hz, 3-C*H*,5-C*H*); 6.37 (1H, br s, 4-N*H*); 4.66 (2H, d, *J*= 2.3 Hz, 10-CH_2); 2.51 (1H, t, *J*= 2.3 Hz, 12-C*H*); 1.52 (9H, s, 9-CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 153.5 (*C*), 153.0 (*C*), 132.3 (*C*), 120.3 (*C*H-2,*C*H-6), 115.5 (*C*H-3,*C*H-5), 80.3 (C-11), 78.6 (C-8), 75.4 (*C*H-12), 56.2 (*C*H_2-10), 28.3 (*C*H_3-9);

HPLC RT 11.3 min; >95% content; HRMS *m/z* (ESI$^+$) [Found (M+H)$^+$ 248.1281; $C_{14}H_{17}NO_3H$ requires (M+H)$^+$ 248.1281]. The data are in good agreement with the literature values.[224]

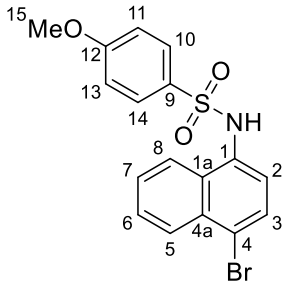


4-Methoxy-*N*-(4-(prop-2-yn-1-yloxy)phenyl)benzenesulphonamide (148). Following general procedure A: Compound **144** (0.30 g, 1.6 mmol), DMAP (20 mg, 0.16 mmol), *p*-methoxybenzenesulphonyl chloride (0.41 g, 2.0 mmol), pyridine (0.40 mL, 4.9 mmol), DCM (7.0 mL, 0.23 M). Compound **148** was obtained as a colourless solid (0.49 g, 95%): R_f 0.26 [Hexane:EtOAc, 7:3]; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.65–7.62 (2H, m, 11-C*H*,15-C*H*); 7.01–6.97 (2H, m, 2-C*H*,6-C*H*); 6.91–6.84 (4H, m, 3-C*H*,5-C*H*,12-C*H*,14-C*H*); 6.31 (1H, br s, 1-N*H*); 4.65 (2H, d, *J*= 2.4 Hz, 7-CH_2); 3.84 (3H, s, 16-CH_3); 3.53 (1H, t, *J*= 2.4 Hz, 9-C*H*); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 163.0 (*C*), 155.9 (*C*), 130.6 (*C*), 129.8 (*C*), 129.4 (*C*H-11,*C*H-15), 125.3 (*C*H-2,*C*H-6), 115.5 (*C*H-3,*C*H-5), 114.1 (*C*H-12,*C*H-14), 78.2 (*C*-8), 75.7 (*C*H-9), 55.6 (*C*H_2-7), 55.5 (*C*H_3-16); HPLC RT 10.4 min; >95% content; HRMS *m/z* (ESI$^+$) [Found (M+H)$^+$ 318.0795; $C_{16}H_{15}NO_4SH$ requires (M+H)$^+$ 318.0795].

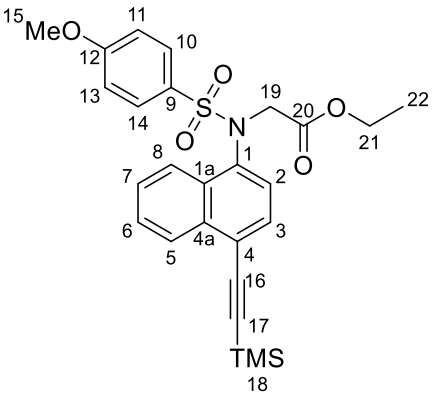


4-Methoxy-*N*-(4-((trimethylsilyl)ethynyl)naphthalen-1-yl)benzene-sulphonamide (149). CuI (0.15 g, 0.80 mmol) and Pd[(PPh_3)$_2$$Cl_2$] (0.54 g, 0.80 mmol) were added to a stirred solution of sulphonamide **150** (1.0 g, 2.6 mmol) in dry THF (5.1 mL, 0.5 M). Et_3N (5.1 mL, 0.5 M) and TMS-acetylene (1.1 mL, 7.6 mmol) were added to the solution and the reaction was heated using microwave radiation at 130 °C for 30 minutes. The

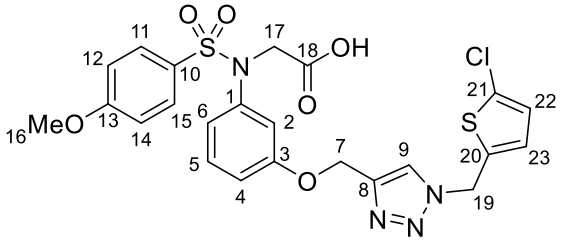
resulting solution was filtered through a pad of Celite® and the solvent removed under vacuum. The crude residue was purified by column chromatography [Hexane:Acetone, 9:1–8:2] to yield compound **149** as a dark orange solid (0.84 g, 80%): R_f 0.25 [Hexane:Acetone, 8:2]; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 8.32 (1H, d, *J*= 8.5 Hz, 5-C*H*); 7.79 (1H, d, *J*= 8.5 Hz, 8-C*H*); 7.68–7.76 (2H, m, 10-C*H*,14-C*H*); 7.61 (1H, d, *J*= 7.9 Hz, 3-C*H*); 7.56 (1H, ddd, *J*= 8.3, 6.8, 1.1 Hz, 6-C*H*); 7.48 (1H, ddd, *J*= 8.3, 6.8, 1.1 Hz, 7-C*H*); 7.39 (1H, d, *J*= 7.9 Hz, 2-C*H*); 6.90 (1H, br s, 1-N*H*); 6.83–6.80 (2H, m, 11-C*H*,13-C*H*); 3.79 (3H, s, 15-CH_3); 0.33 (9H, s, 18-CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 163.2 (*C*), 134.1 (*C*), 132.1 (*C*), 130.5 (*C*H-3), 130.4 (*C*), 129.4 (*C*H-10,*C*H-14), 127.1 (*C*H-6), 127.9 (*C*), 127.1 (*C*H-7), 127.0 (*C*H-5), 121.2 (*C*H-8), 121.0 (*C*H-2), 119.5 (*C*), 114.1 (*C*H-11,*C*H-13), 102.4 (*C*), 100.3 (*C*), 55.5 (*C*H_3-15), 0.0 (*C*H_3-18); HPLC RT 15.9 min; 80% content; HRMS *m/z* (ESI$^+$) [Found $(M+H)^+$ 410.1238; $C_{22}H_{23}NO_3SSiH$ requires $(M+H)^+$ 410.1241].



***N*-(4-Bromonaphthalen-1-yl)-4-methoxybenzenesulphonamide (150).** Following general procedure A: 4-Bromonaphthalen-1-amine **151** (2.0 g, 9.0 mmol), DMAP (110 mg, 0.90 mmol), *p*-methoxybenzenesulphonyl chloride (2.2 g, 11 mmol), pyridine (1.5 mL, 18 mmol), DCM (39 mL, 0.23 M). The crude product was triturated with ether to yield compound **150** as a pale purple solid (2.9 g, 83%): R_f 0.19 [Hexane:Acetone, 8:2]; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 8.23 (1H, d, *J*= 8.5 Hz, 5-C*H*); 7.87 (1H, d, *J*= 8.5 Hz, 8-C*H*); 7.69–7.66 (3H, m, 3-C*H*,10-C*H*,14-C*H*); 7.59 (1H, ddd, *J*= 8.5, 7.1, 1.2 Hz, 6-C*H*); 7.51 (1H, ddd, *J*= 8.5, 7.1, 1.2 Hz, 7-C*H*); 7.23 (1H, d, *J*= 8.2 Hz, 2-C*H*); 7.69–7.66 (3H, m, 1-N*H*, 11-C*H*,13-C*H*); 3.81 (3H, s, 15-CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 163.2 (*C*), 132.5 (*C*), 131.4 (*C*), 130.5 (*C*), 130.2 (*C*), 129.5 (*C*H-10,*C*H-14), 129.4 (*C*H-3), 127.8 (*C*H-5), 127.7 (*C*H-6), 127.4 (*C*H-7), 123.3 (*C*H-2), 122.1 (*C*H-8), 121.6 (*C*), 114.2 (*C*H-11,*C*H-13), 55.6 (*C*H_3-15); HPLC RT 13.1 min; >95% content; HRMS *m/z* (ESI$^+$) [Found $(M+NH_4)^+$ 409.0216; $C_{17}H_{18}BrN_2O_3SNH_4$ requires $(M+NH_4)^+$ 409.0216].

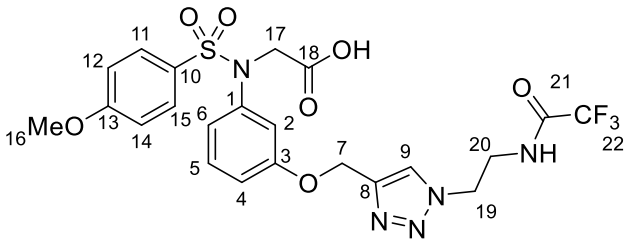


Ethyl-*N*-((4-methoxyphenyl)sulphonyl)-*N*-(4-((trimethylsilyl)ethynyl)-naphthalen-1-yl)glycinate (152). Following general procedure B: Ethyl bromoacetate (0.20 mL, 1.8 mmol), K_2CO_3 (0.25 g, 1.8 mmol), compound **149** (0.50 g, 1.2 mmol), DMF (6.1 mL, 0.2 M). The crude residue was purified by column chromatography [Hexane:EtOAc, 19:1–9:1] to yield compound **139** as a yellow powder (0.42 g, 70%): R_f 0.27 [Hexane:EtAOc, 8:2]; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 8.32 (1H, d, *J*= 8.3 Hz, 5-C*H*); 8.16 (1H, d, *J*= 8.1 Hz, 8-C*H*); 7.65–7.63 (2H, m, 10-C*H*,14-C*H*); 7.62–7.55 (3H, m, 3-C*H*,6-C*H*,7-C*H*); 7.16 (1H, d, *J*= 7.6 Hz, 2-C*H*); 6.93–6.90 (2H, m, 11-C*H*,13-C*H*); 4.64 (1H, d, *J*= 17.7 Hz, 19-CH_2); 4.33 (1H, d, *J*= 17.7 Hz, 19-CH_2); 4.16–4.05 (2H, m, 21-CH_2); 3.87 (3H, s, 15-CH_3); 1.17 (3H, t, *J*= 7.3 Hz, 22-CH_3); 0.34 (9H, s, 18-CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 168.8 (*C*), 163.2 (*C*), 136.7 (*C*), 134.6 (*C*), 131.7 (*C*), 130.3 (*C*H-10,*C*H-14), 130.2 (*C*), 130.0 (*C*H-3), 127.3 (*C*H-6,*C*H-7), 127.3 (*C*H-2), 126.6 (*C*H-5), 124.2 (*C*H-8), 122.2 (*C*), 113.9 (*C*H-11,*C*H-13), 102.3 (*C*), 101.1 (*C*), 61.4 (*C*H_2-21), 55.6 (*C*H_3-15), 53.3 (*C*H_2-19), 14.0 (*C*H_3-22), 0.01 (*C*H_3-18); HPLC RT 17.2 min; 82% content; HRMS *m/z* (ESI$^+$) [Found $(M+NH_4)^+$ 513.1861; $C_{26}H_{29}NO_5SSiNH_4$ requires $(M+NH_4)^+$ 513.1874].

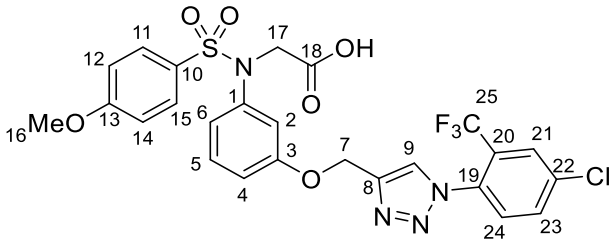


***N*-(3-((1-((5-Chlorothiophen-2-yl)methyl)-1*H*-1,2,3-triazol-4-yl)methoxy)-phenyl)-*N*-((4-methoxyphenyl)sulphonyl)glycine (153a).** Following general procedure D, method 1: Compound **119** (40 mg, 0.11 mmol), 2-(azidomethyl)-5-chlorothiophene (**AzD5-2**) (20 mg, 0.12 mmol), CuCl (5.3 mg, 55 µmol) and ^{t}BuOH: H_2O (1.1 mL, 0.1 M). Compound **153a** was isolated as a beige foam (32 mg, 55 %): R_f 0.64

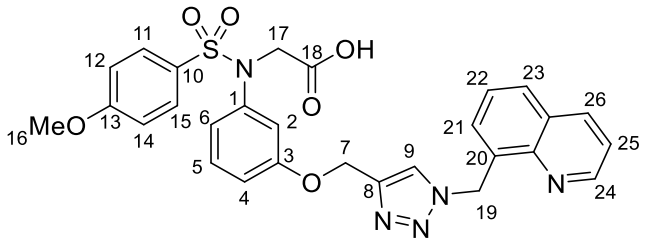
[EtOAc:MeOH:AcOH, 94:5:1]; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.71 (1H, s, 9-C*H*); 7.58 (2H, d, *J*= 8.8 Hz, 11-C*H*,15-C*H*); 7.16 (1H, dd, *J*= 8.0, 8.0 Hz, 5-C*H*); 6.98–6.97 (1H, br s, 2-C*H*); 6.93–6.88 (4H, m, 4/6-C*H*,12-C*H*,14-C*H*,22-C*H*); 6.80 (1H, d, *J*= 3.8 Hz, 23-C*H*); 6.69–6.67 (1H, m, 4/6-C*H*); 5.59 (2H, s, 19-CH_2); 5.13 (1H, s, 7-CH_2); 4.41 (2H, s, 17-CH_2); 3.86 (3H, s, 16-CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 163.2 (*C*), 158.4 (*C*), 140.8 (*C*), 141.1 (*C*), 134.4 (*C*), 131.9 (*C*H-5), 129.9 (*C*H-11,*C*H-15), 129.0 (*C*), 128.2 (*C*), 127.9 (*C*H-22), 126.3 (CH-23), 125.3 (C), 123.0 (*C*H-9); 120.5 (*C*H-4/6), 115.4 (*C*H-2), 115.3 (*C*H-4/6), 114.0 (*C*H-12,*C*H-14), 61.7 (*C*H_2-7), 55.6 (*C*H_3-16,*C*H_2-17), 48.8 (*C*H_2-19): HPLC RT 11.6 min; >90% content; HRMS *m/z* (ESI$^-$) [Found (M-H)$^-$ 547.0508; $C_{23}H_{20}ClN_4O_6S_2$ requires (M-H)$^-$ 547.0518].



***N*-((4-Methoxyphenyl)sulphonyl)-*N*-(3-((1-(2-(2,2,2-trifluoroacetamido)ethyl)-1*H*-1,2,3-triazol-4-yl)methoxy)phenyl)glycine (154a).** Following general procedure D, method 2: Compound **119** (0.10 g, 0.27 mmol *N*-(2-azidoethyl)-2,2,2-trifluoroacetamide (**AzB4-2**) (53 mg, 0.29 mmol), $Cu_2SO_4.5H_2O$ (33 mg, 0.13 mmol), sodium ascorbate (66 mg, 0.33 mmol) and ^{t}BuOH: H_2O (2.6 mL, 0.1 M). Compound **154a** was isolated as a beige solid (75 mg, 51%): R_f 0.38 [EtOAc:MeOH:AcOH, 94:5:1]; ^{1}H NMR (500 MHz, DMSO-d_6) δ: 12.87 (1H, br s, 18-O*H*); 9.62 (1H, t, *J*= 6.2 Hz, 20-N*H*); 8.22 (1H, s, 9-C*H*); 7.60 (2H, d, *J*= 8.7 Hz, 11-C*H*,15-C*H*); 7.23 (1H, dd, *J*= 7.9, 7.9 Hz, 5-C*H*); 7.07 (2H, d, *J*= 8.7 Hz, 12-C*H*,14-C*H*); 6.98–6.96 (1H, m, 4-C*H*); 6.80 (1H, br s, 2-C*H*); 6.76–6.75 (1H, m, 6-C*H*); 5.04 (2H, s, 7-CH_2); 4.54 (2H, t, *J*= 5.6 Hz, 19-CH_2); 4.37 (2H, br s, 17-CH_2); 3.82 (3H, s, 16-CH_3); 3.65 (2H, td, *J*= 5.6 Hz, *J*= 5.6 Hz, 20-CH_2); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 170.4 (*C*=O), 162.7 (*C*), 158.1 (*C*), 156.6 (q, *J*= 37 Hz, *C*=O), 142.3 (*C*), 141.0 (*C*), 130.1 (*C*), 129.6 (*C*H-5), 129.6 (*C*H-11,*C*H-15), 125.0 (*C*H-9), 120.0 (*C*H-4/6), 115.0 (q, *J*= 289 Hz, *C*F_3), 116.4 (*C*H-4/6), 115.8 (*C*H-2), 114.1 (*C*H-12,*C*H-14), 61.4 (*C*H_2-7), 55.6 (*C*H_3-16), 52.4 (*C*H_2-17), 48.6 (*C*H_2-19), 39.7 (*C*H_2-20); HPLC RT 9.53 min; >95% content; HRMS *m/z* (ESI$^-$) [Found (M-H)$^-$ 556.1108; $C_{22}H_{21}F_3N_5O_7S$ requires (M-H)$^-$ 556.1119].

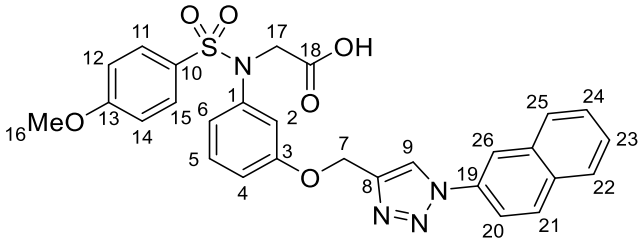


***N*-(3-((1-(4-Chloro-2-(trifluoromethyl)phenyl)-1*H*-1,2,3-triazol-4-yl)methoxy)-phenyl)-*N*-((4-methoxyphenyl)sulphonyl)glycine (155a).** Following general procedure D, method 1: Compound **119** (40 mg, 0.11 mmol), 1-azido-4-chloro-2-(trifluoromethyl)benzene (**AzD7-2**) (26 mg, 0.12 mmol), CuCl (5.3 mg, 55 µmol) and ^{t}BuOH: H_2O (1.1 mL, 0.1 M). Compound **155a** was isolated as a pale orange foam (54 mg, 84%): R_f 0.69 [EtOAc:MeOH:AcOH, 94:5:1]; ^{1}H NMR (500 MHz, DMSO-d_6) δ: 8.65 (1H, s, 9-C*H*); 8.17 (1H, d, *J*= 2.3 Hz, 21-C*H*); 8.04 (1H, dd, *J*= 8.3, 2.3 Hz, 23-C*H*); 7.83 (1H, d, *J*= 8.3 Hz, 24-C*H*); 7.61 (2H, d, *J*= 9.0 Hz, 11-C*H*,15-C*H*); 7.24 (1H, dd, *J*= 8.2, 8.2 Hz, 5-C*H*); 7.06 (2H, d, *J*= 9.0 Hz, 12-C*H*,14-C*H*); 7.02–7.00 (1H, m, 4/6-C*H*); 6.86 (1H, dd, *J*= 2.2, 2.2 Hz, 2-C*H*); 6.78–6.77 (1H, m, 4/6-C*H*); 5.16 (2H, s, 7-CH_2); 4.35 (2H, s, 17-CH_2); 3.82 (3H, s, 16-CH_3); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 162.6 (*C*), 158.0 (*C*), 142.7 (*C*), 141.1 (*C*), 135.8 (*C*), 134.0 (*C*H-23), 133.0 (*C*), 131.3 (*C*H-24), 130.2 (*C*), 129.6 (*C*H-5), 129.6 (*C*H-11,*C*H-15), 127.6 (q, *J*= 5.5 Hz, *C*H-21), 127.5 (*C*H-9), 126.5 (q, *J*= 32 Hz, *C*-20), 123.0 (*C*), 121.7 (q, *J*= 274.2 Hz, *C*F_3), 120.0 (*C*H-4/6), 114.4 (*C*H-2), 114.3 (*C*H-12,*C*H-14), 113.3 (*C*H-4/6), 60.9 (*C*H_2-7), 55.7 (*C*H_3-16), 52.2 (*C*H_2-17): HPLC RT 12.6 min; >90% content; HRMS *m/z* (ESI$^-$) [Found (M-H)$^-$ 595.0657; $C_{25}H_{19}ClF_3N_4O_6S$ requires (M-H)$^-$ 595.0671].

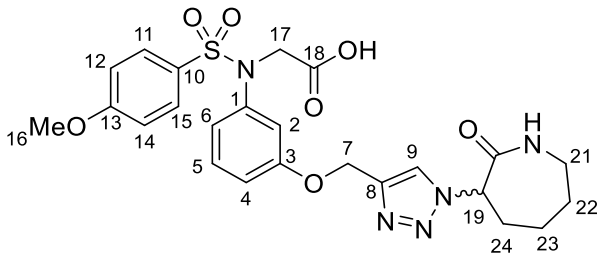


***N*-((4-Methoxyphenyl)sulphonyl)-*N*-(3-((1-(quinolin-8-ylmethyl)-1*H*-1,2,3-triazol-4-yl)methoxy)phenyl)glycine (156a).** Following general procedure D, method 1: compound **119** (40 mg, 0.11 mmol), 8-(azidomethyl)quinoline (**AzF3-2**) (22 mg, 0.12 mmol), CuCl (5.3 mg, 55 µmol) and ^{t}BuOH: H_2O (1.1 mL, 0.1 M). Compound **156a** was isolated as a colourless solid (14 mg, 24%): R_f 0.22 [EtOAc:MeOH:AcOH, 94:5:1]; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 9.07 (1H, dd, *J*= 4.4, 1.6 Hz, 24-C*H*); 8.31–8.29 (1H, m, 26-C*H*); 7.91–7.89 (1H, m, 23-C*H*); 7.84 (1H, s, 9-C*H*); 7.80 (1H, dd, *J*= 7.3, 0.6 Hz, 21-

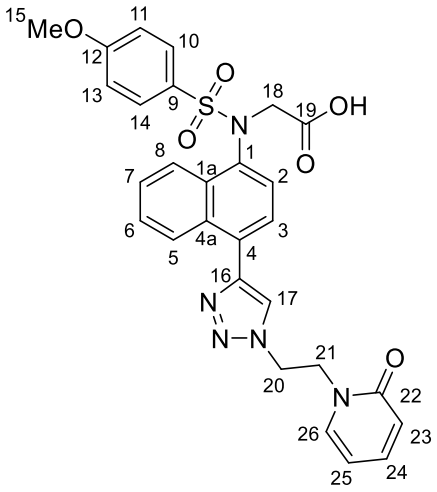
C*H*); 7.65 (2H, d, *J*= 8.8 Hz, 11-C*H*,15-C*H*); 7.60–7.56 (2H, m, 22-C*H*,25-C*H*); 7.14 (1H, dd, *J*= 8.4, 8.4 Hz, 5-C*H*); 6.90–6.86 (4H, m, 2-C*H*,4/6-C*H*,12-C*H*,14-C*H*); 6.71–6.60 (1H, m, 4/6-C*H*); 6.13 (2H, s, 19-CH_2); 5.09 (2H, s, 7-CH_2); 4.42 (2H, s, 17-CH_2); 3.85 (3H, s, 16-CH_3); ^{13}C NMR (125 MHz, $CDCl_3$) δ: 163.1 (*C*), 158.0 (*C*), 150.5 (*C*H-24), 145.0 (*C*), 143.1 (*C*), 140.7 (*C*), 137.9 (*C*H-26), 132.0 (*C*), 131.8 (*C*H-21), 130.4 (*C*), 130.0 (*C*H-11,*C*H-15), 130.0 (*C*H-5), 129.5 (*C*H-25), 128.8 (*C*), 128.2 (*C*), 126.9 (*C*H-22/25), 124.4 (*C*H-9), 121.9 (*C*H-22/25), 120.9 (*C*H-4/6), 116.8 (*C*H-2,*C*H-4/6), 115.0 (*C*H-21), 113.9 (*C*H-12,*C*H-14), 61.3 (*C*H_2-7), 55.6 (*C*H_3-16), 52.8 (*C*H_2-17), 49.6 (*C*H_2-19); HPLC RT 9.54 min; >95% content; HRMS *m/z* (ESI$^-$) [Found (M-H)$^-$ 558.1445; $C_{28}H_{24}N_5O_6S$ requires (M-H)$^-$ 558.1453].



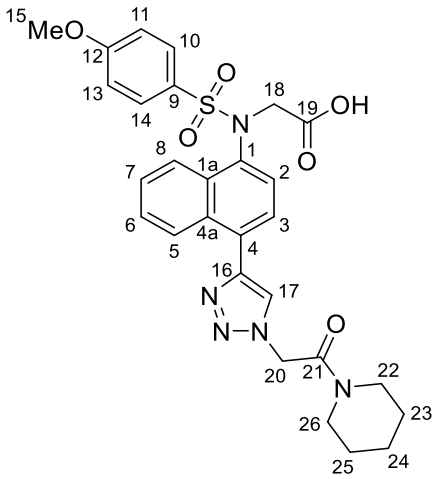
***N*-((4-Methoxyphenyl)sulphonyl)-*N*-(3-((1-(naphthalen-2-yl)-1*H*-1,2,3-triazol-4-yl)methoxy)phenyl)glycine (157a).** Following general procedure D, method 1: compound **119** (0.10 g, 0.27 mmol), 1-azidonaphthalene (**AzE2-2**) (50 mg, 0.29 mmol), CuCl (13 mg, 0.13 mmol) and ^{t}BuOH: H_2O (2.6 mL, 0.1 M). Compound **157a** was isolated as a colourless solid (55 mg, 38%): R_f 0.54 [EtOAc:MeOH:AcOH, 94:5:1]; ^{1}H NMR (500 MHz, CD_3OD) δ: 8.77 (1H, s, 9-C*H*); 8.38 (1H, m, 26-C*H*); 8.09 (1H, d, *J*= 8.8 Hz, 21-C*H*); 8.03 (1H, dd, *J*= 8.8, 1.9 Hz, 20-C*H*); 8.03–7.97 (2H, m, Ar-C*H*,Ar-C*H*); 7.62–7.58 (4H, m, Ar-C*H,*Ar-C*H*,11-C*H*,15-C*H*); 7.21 (1H, dd, *J*= 8.2, 8.2 Hz, 5-C*H*); 7.08 (1H, dd, *J*= 2.2, 2.2 Hz, 2-C*H*); 7.00–6.96 (3H, m, 4/6-C*H*,12-C*H*,14-C*H*); 6.75–6.73 (1H, m, 4/6-C*H*); 5.26 (2H, s, 7-CH_2); 4.41 (2H, s, 17-CH_2); 3.81 (3H, s, 16-CH_3); ^{13}C NMR (125 MHz, CD_3OD) δ: 164.9 (*C*), 159.8 (*C*), 145.6 (*C*), 142.7 (*C*), 138.9 (*C*), 135.7 (*C*), 134.8 (*C*), 134.5 (*C*), 131.4 (*C*), 131.2 (*C*H-21), 131.1 (*C*H-11,*C*H-15), 130.8 (*C*H-5), 129.5 (*C*H-Ar), 129.0 (*C*H-Ar), 128.6 (*C*H-Ar), 128.3 (*C*H-Ar), 124.0 (*C*H-9), 121.7 (*C*H-4/6), 119.9 (*C*H-20), 119.7 (*C*H-26), 116.5 (*C*H-2), 116.2 (*C*H-4/6), 115.1 (*C*H-12,*C*H-14), 62.4 (*C*H-7), 56.2 (*C*H_3-16), 49.9 (*C*H_2-17); HPLC RT 12.8 min; >95% content; HRMS *m/z* (ESI$^-$) [Found (M-H)$^-$ 543.1330; $C_{28}H_{23}N_4O_6S$ requires (M-H)$^-$ 543.1344].



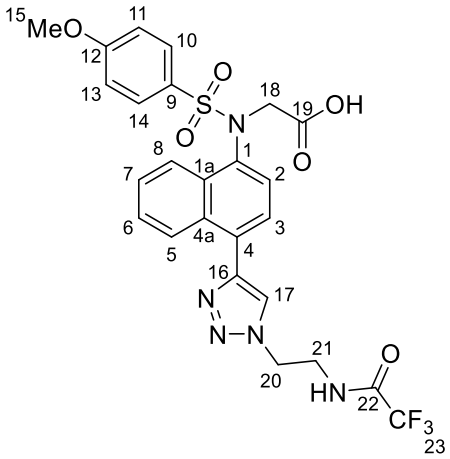
***N*-((4-Methoxyphenyl)sulphonyl)-*N*-(3-((1-(2-oxoazepan-3-yl)-1*H*-1,2,3-triazol-4-yl)methoxy)phenyl)glycine (158a).** Following general procedure D, method 2: compound **119** (0.10 g, 0.27 mmol), 3-azidoazepan-2-one (**AzA5-2**) (53 mg, 0.29 mmol), $Cu_2SO_4.5H_2O$ (33 mg, 0.13 mmol), sodium ascorbate (66 mg, 0.33 mmol) and ^{t}BuOH: H_2O (2.6 mL, 0.1 M). The crude residue was purified by reverse phase chromatography [H_2O + 0.1% FA:MeCN + 0.1% FA, 19:1 - 1] to yield compound **158a** as a beige solid (94 mg, 67%): R_f 0.13 [EtOAc:MeOH:AcOH, 94:5:1]; ^{1}H NMR (500 MHz, DMSO-d_6) δ: 12.85 (1H, s, 18-O*H*); 8.18 (1H, s, 9-C*H*); 8.03 (1H, dd, *J*= 7.3, 6.3 Hz, 21-N*H*); 7.60 (2H, d, *J*= 8.8 Hz, 11-C*H*,15-C*H*); 7.24 (1H, dd, *J*= 8.2, 8.2 Hz, 5-C*H*); 7.08 (2H, d, *J*= 8.8 Hz, 12-C*H*,14-C*H*); 7.01–6.99 (1H, m, 4/6-C*H*); 6.82 (1H, dd, *J*= 2.4, 2.4 Hz, 2-C*H*); 6.77–6.75 (1H, m, 4/6-C*H*); 5.74 (1H, d, *J*= 11.3 Hz, 19-C*H*); 5.04 (2H, s, 7-CH_2); 4.38 (2H, s, 17-CH_2); 3.83 (3H, s, 16-CH_3); 3.41–3.36 (1H, m, 21-C*H*); 3.17–3.09 (1H, m, 21-C*H*); 2.17–1.96 (2H, m, 24-CH_2); 1.84–1.71 (2H, m, 23-CH_2); 1.35–1.23 (2H, m, 22-CH_2); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 171.4 (*C*), 170.0 (*C*), 162.7 (*C*), 158.2 (*C*), 141.7 (*C*), 141.1 (*C*), 130.1 (*C*H-5), 129.6 (*C*H-11,*C*H-15), 125.0 (*C*H-9), 119.9 (*C*H-4/6), 114.3 (*C*H-12,*C*H-14), 114.3 (*C*H-2), 113.3 (*C*H-4/6), 62.0 (*C*H-19), 61.3 (*C*H_2-7), 55.7 (*C*$H_3$16), 52.0 (*C*H_2-17), 40.4 (*C*H_2-21), 30.3 (*C*H_2-24), 28.6 (*C*H_2-22), 27.2 (*C*H_2-23); HPLC RT 8.49 min; >95% content; HRMS *m/z* (ESI$^-$) [Found (M-H)$^-$ 528.1552; $C_{24}H_{26}N_5O_7S$ requires (M-H)$^-$ 528.1558].



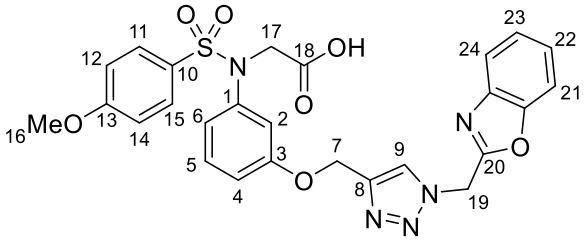
***N*-((4-Methoxyphenyl)sulphonyl)-*N*-(4-(1-(2-(2-oxopyridin-1(2*H*)-yl)ethyl)-1*H*-1,2,3-triazol-4-yl)naphthalen-1-yl)glycine (174a).** Following general procedure D, method 1: compound **125** (80 mg, 0.20 mmol), 1-(2-azidoethyl)pyridin-2(1*H*)-one (**AzD2-2**) (37 mg, 0.22 mmol), CuCl (10 mg, 0.10 mmol) and ^{t}BuOH: H_2O (2.0 mL, 0.1 M). The crude residue was purified by reverse phase chromatography [H_2O + 0.1% FA:MeCN + 0.1% FA, 19:1–9:1] to yield compound **174a** as a colourless solid (13 mg, 12%): R_f 0.62 [EtOAc:MeOH:AcOH, 94:5:1]; ^{1}H NMR (500 MHz, DMSO-d_6) δ: 12.84 (1H, br s, 19-O*H*); 8.65 (1H, s, 17-C*H*); 8.39 (1H, dd, *J*= 6.3, 3.2 Hz, 5-C*H*); 8.26 (1H, dd, *J*= 6.6, 3.2 Hz, 8-C*H*); 8.16 (1H, dd, *J*= 5.0, 1.8 Hz, 26-C*H*); 7.71 (1H, ddd, *J*= 8.5, 6.9, 1.8 Hz, 24-C*H*); 7.65–7.62 (3H, m, 3-C*H*,10-C*H*,14-C*H*); 7.60–7.56 (2H, m, 6-C*H*,7-C*H*); 7.22 (1H, d, *J*= 7.9 Hz, 2-C*H*); 7.12–7.07 (2H, d, *J*= 8.8 Hz, 11-C*H*,13-C*H*); 7.01–6.98 (1H, m, 25-C*H*); 6.84–6.83 (1H, m, 23-C*H*); 4.90 (1H, t, *J*= 5.1 Hz, 20-CH_2); 4.78 (1H, t, *J*= 5.1 Hz, 21-CH_2); 4.46 (1H, d, *J*= 18 Hz, 18-C*H*); 4.39 (1H, d, *J*= 18 Hz, 18-C*H*); 3.85 (3H, s, 15-C*H*); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 169.9 (*C*), 162.8 (*C*), 162.5 (*C*), 146.8 (*C*H-26), 144.9 (*C*), 139.5 (*C*H-23), 136.6 (*C*), 132.3 (*C*), 131.3 (*C*), 130.0 (*C*H-10,*C*H-14), 129.6 (*C*), 129.0 (*C*), 126.9 (*C*H-2), 126.8 (*C*H-6), 126.4 (*C*H-7), 126.2 (*C*H-3), 125.5 (*C*H-5), 124.9 (*C*H-17), 124.9 (*C*H-8), 117.6 (*C*H-25), 114.4 (*C*H-11,*C*H-13), 110.9 (*C*H-24), 63.7 (*C*H_2-21), 55.7 (*C*H_3-15), 53.2 (*C*H_2-18), 49.3 (*C*H_2-20); HPLC RT 9.75 min; >90% content; HRMS *m/z* (ESI$^-$) [Found $(M-H)^-$ 558.1446; $C_{28}H_{24}N_5O_6S$ requires $(M-H)^-$ 558.1453].



***N*-((4-Methoxyphenyl)sulphonyl)-*N*-(4-(1-(2-oxo-2-(piperidin-1-yl)ethyl)-1*H*-1,2,3-triazol-4-yl)naphthalen-1-yl)glycine (179a).** Following general procedure D, method 1: compound **125** (0.10 g, 0.25 mmol), 2-azido-1-(piperidin-1-yl)ethan-1-one (**AzE5-2**) (47 mg, 0.28 mmol), CuCl (13 mg, 0.13 mmol) and ^{t}BuOH:H_2O (2.5 mL, 0.1 M). The crude residue was purified by column chromatography [EtOAc:MeOH:AcOH, 99:0:1–94:5:1] to yield compound **179a** as a colourless solid (26 mg, 18%): R_f 0.41 [EtOAc:MeOH:AcOH, 94:5:1]; ^{1}H NMR (500 MHz, DMSO-d_6) δ: 12.85 (1H, br s, 19-O*H*); 8.49 (1H, s, 17-C*H*); 8.46–8.44 (1H, m, 5-C*H*); 8.29–8.27 (1H, dd, *J*= 6.6, 3.2 Hz, 8-C*H*); 7.67–7.57 (5H, m, 3-C*H*,6-C*H*,7-C*H*,10-C*H*,14-C*H*); 7.24 (1H, d, *J*= 7.9 Hz, 2-C*H*); 7.11–7.09 (2H, d, *J*= 8.8 Hz, 11-C*H*,13-C*H*); 5.57 (2H, s, 20-CH_2); 4.47 (1H, d, *J*= 18 Hz, 18-CH_2); 4.41 (1H, d, *J*= 18 Hz, 18-CH_2); 3.85 (3H, s, 15-CH_3); 3.52–3.47 (4H, m, 22-CH_2,26-CH_2); 1.63–1.60 (4H, m, 23-CH_2,25-CH_2); 1.51–1.47 (2H, m, 24-CH_2); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 170.0 (*C*), 163.4 (*C*), 162.9 (*C*), 144.6 (*C*), 136.6 (*C*), 132.3 (*C*), 131.3 (*C*), 130.0 (*C*H-10,*C*H-14); 129.6 (*C*), 129.0 (*C*), 126.9 (*C*H-2), 126.8 (*C*H-3/6/7), 126.4 (*C*H-3/6/7), 126.2 (*C*H-3/6/7), 126.2 (*C*H-17), 125.4 (*C*H-5), 124.9 (*C*H-8), 114.4 (*C*H-11,*C*H-13); 55.7 (*C*H_3-15), 53.2 (*C*H_2-18), 51.0 (*C*H_2-20), 45.2 (*C*H_2-22), 42.5 (*C*H_2-26), 25.8 (*C*H_2-25), 25.2 (*C*H_2-24), 23.8 (*C*H_2-23); HPLC RT 10.3 min; >95% content; HRMS *m/z* (ESI$^-$) [Found (M-H)$^-$ 562.1759; $C_{28}H_{28}N_5O_6S$ requires (M-H)$^-$ 562.1766].

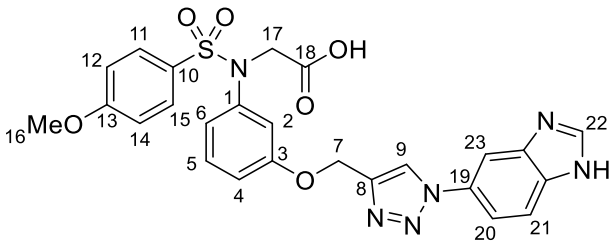


***N*-((4-Methoxyphenyl)sulphonyl)-*N*-(4-(1-(2-(2,2,2-trifluoroacetamido)ethyl)-1*H*-1,2,3-triazol-4-yl)naphthalen-1-yl)glycine (180a).** Following general procedure D, method 1: compound **125** (90 mg, 0.23 mmol), *N*-(2-azidoethyl)-2,2,2-trifluoroacetamide (**AzB4-2**) (46 mg, 0.25 mmol), CuCl (11 mg, 0.11 mmol) and ^{t}BuOH: H_2O (2.3 mL, 0.1 M). The crude residue was purified by column chromatography [EtOAc:MeOH:AcOH, 99:0:1–94:5:1] to yield compound **180a** as a pale yellow solid (49 mg, 37%): R_f 0.62 [EtOAc:MeOH:AcOH, 94:5:1]; ^{1}H NMR (500 MHz, DMSO-d_6) δ: 12.84 (1H, br s, 19-O*H*); 9.67 (1H, t, *J*= 5.4 Hz, 22-N*H*); 8.59 (1H, s, 17-C*H*); 8.42–8.40 (1H, m, 5-C*H*); 8.27–8.26 (1H, m, 8-C*H*); 7.64–7.61 (3H, m, 3-C*H*,10-C*H*,14-C*H*); 7.60–7.56 (2H, m, 6-C*H*,7-C*H*); 7.25–7.24 (1H, d, *J*= 7.9 Hz, 2-C*H*); 7.11–7.09 (2H, d, *J*= 8.8 Hz, 11-C*H*,13-C*H*); 4.65 (1H, t, *J*= 5.8 Hz, 20-CH_2); 4.47 (1H, d, *J*= 18 Hz, 18-CH_2); 4.40 (1H, d, *J*= 18 Hz, 18-CH_2); 3.85 (3H, s, 15-CH_3); 3.78–3.74 (2H, m, 21-CH_2); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 169.9 (*C*), 162.9 (*C*), 156.6 (q, J= 37.6 Hz, *C*=O), 144.8 (*C*), 136.7 (*C*), 132.3 (*C*), 131.4 (*C*), 130.0 (*C*H-10,*C*H-14), 129.6 (*C*), 129.1 (*C*H-3), 128.2 (C), 126.9 (*C*H-6/7), 126.8 (*C*H-2), 126.4 (C-6/7), 126.2 (*C*H-5), 125.5 (*C*H-8), 124.9 (*C*H-17), 115.5 (q, *J*= 286 Hz, *C*F_3), 114.4 (*C*H-11,*C*H-13), 55.7 (*C*H_3-15), 53.2 (*C*H_2-18), 48.2 (*C*H_2-20), 39.6 (*C*H_2-21); HPLC RT 10.1 min; >95% content; HRMS *m/z* (ESI$^-$) [Found (M-H)$^-$ 576.1162; $C_{25}H_{21}F_3N_5O_6S$ requires (M-H)$^-$ 576.1170].

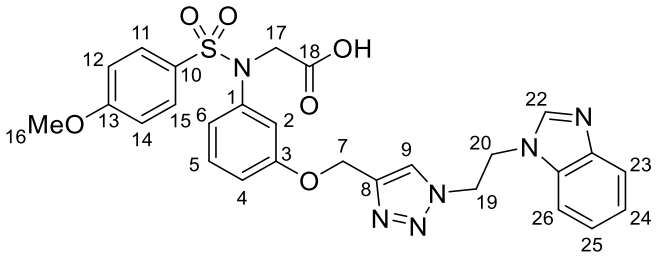


***N*-(3-((1-(Benzoxazol-2-ylmethyl)-1*H*-1,2,3-triazol-4-yl)methoxy)phenyl)-*N*-((4-methoxyphenyl)sulphonyl)glycine (192a).** Following general procedure D, method 1:

compound **119** (90 mg, 0.24 mmol), 2-(azidomethyl)-benzoxazole (**AzB6-2**) (46 mg, 0.26 mmol), CuCl (12 mg, 0.21 mmol) and ^{t}BuOH: H_2O (2.4 mL, 0.1 M). The crude residue was purified by column chromatography [EtOAc:MeOH:AcOH, 99:0:1–94:5:1] to yield compound **192a** as an orange solid (42 mg, 32%): R_f 0.44 [EtOAc:MeOH:AcOH, 94:5:1]; ^{1}H NMR (500 MHz, CD_3OD) δ: 8.15 (1H, s, 9-C*H*); 7.80–7.79 (1H, m, Ar-C*H*); 7.60–7.57 (2H, m, 11-C*H*,15-C*H*); 7.20 (1H, dd, *J*= 8.0, 8.0 Hz, 5-C*H*); 7.01–6.94 (5H, m, Ar-C*H*,2-C*H*,4/6-C*H*,12-C*H*,14-C*H*); 6.86–6.85 (1H, m, Ar-C*H*); 6.81–6.75 (2H, m, Ar-C*H*,4/6-C*H*); 5.42 (2H, s, 19-CH_2); 5.14 (2H, s, 7-CH_2); 4.40 (2H, s, 17-CH_2); 3.84 (3H, s, 16-CH_3); ^{13}C NMR (125 MHz, CD_3OD) δ: 172.5 (*C*), 166.2 (*C*), 165.0 (*C*), 160.1 (*C*), 149.7 (*C*), 144.8 (*C*), 142.8 (*C*), 131.6 (*C*), 131.3 (*C*H-11,*C*H-15), 130.9 (*C*H-5), 127.5 (*C*H-9), 126.9 (*C*H-Ar), 126.6 (*C*), 123.8 (*C*H-Ar), 122.1 (*C*H-4/6), 120.7 (*C*H-Ar), 116.6 (*C*H-2), 116.5 (*C*H-4/6), 116.1 (*C*H-Ar), 115.3 (*C*H-12,CH-14), 62.6 (*C*H_2-7), 58.4 (*C*H_3-16), 53.7 (*C*H_2-19); 53.6 (*C*H_2-17); HPLC RT 9.40 min; >90% content; HRMS *m/z* (ESI^-) [Found $(M-H)^-$ 548.1247; $C_{26}H_{22}N_5O_7S$ requires $(M-H)^-$ 548.1245].



***N*-(3-((1-(1*H*-Benzoimidazol-5-yl)-1*H*-1,2,3-triazol-4-yl)methoxy)phenyl)-*N*-((4-methoxyphenyl)sulphonyl)glycine (196a).** Following general procedure D, method 1: compound **119** (40 mg, 0.11 mmol), compound **AzE4-2** (19 mg, 0.12 mmol), CuCl (5.3 mg, 55 µmol) and ^{t}BuOH: H_2O (1.1 mL, 0.1 M). Compound **196a** was purified by reverse phase column chromatography [H_2O + 0.1% FA:MeCN + 0.1% FA, 19:1–1] and isolated as colourless solid (15 mg, 26%): ^{1}H NMR (500 MHz, DMSO-d_6) δ: 12.83 (1H, br s, 18-O*H*); 8.93 (1H, s, 9-C*H*); 8.39 (1H, br s, 22-C*H*); 8.11 (1H, br s, Ar-C*H*); 7.80–7.73 (2H, m, Ar-C*H*,Ar-C*H*); 7.61–7.60 (2H, m, 11-C*H*,15-C*H*); 7.26 (1H, dd, *J*= 8.2, 8.2 Hz, 5-C*H*); 7.07–7.02 (3H, m, 4/6-C*H*,12-C*H*,14-C*H*); 6.89 (1H, br s, 2-C*H*); 6.78–6.76 (1H, m, 4/6-C*H*); 5.16 (2H, s, 7-CH_2); 4.40 (2H, s, 17-CH_2); 3.81 (3H, s, 16-CH_3); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 170.0 (*C*), 162.7 (*C*), 158.1 (*C*), 143.3 (*C*), 141.1 (*C*), 130.1 (*C*), 129.7 (*C*H-5), 129.6 (*C*H-11,*C*H-15), 123.4 (*C*H-9), 119.9 (*C*H-4/6), 114.5 (*C*H-2), 114.3 *(C*H-12,*C*H-14), 113.5 (*C*H-4/6), 61.2 (*C*H-7), 55.7 (*C*H_3-16), 52.0 (*C*H_2-17); HPLC RT 7.47 min; >95% content; HRMS *m/z* (ESI^-) [Found $(M-H)^-$ 553.1246; $C_{25}H_{21}N_6O_6S$ requires $(M-H)^-$ 553.1249].



***N*-(3-((1-(2-(1*H*-Benzimidazol-1-yl)ethyl)-1*H*-1,2,3-triazol-4-yl)methoxy)-phenyl)-*N*-((4-methoxyphenyl)sulphonyl)glycine (198a).** Following general procedure D, method 2: compound **119** (0.10 g, 0.27 mmol 1-(2-azidoethyl)-benzoimidazole (**AzC6-2**) (55 mg, 0.29 mmol), $Cu_2SO_4.5H_2O$ (33 mg, 0.13 mmol), sodium ascorbate (66 mg, 0.33 mmol) and ^{t}BuOH: H_2O (2.6 mL, 0.1 M). The crude residue was purified by reverse phase chromatography [H_2O + 0.1% FA:MeCN + 0.1% FA, 19:1–9:1] to yield compound **198a** as a pale brown solid (15 mg, 10%): R_f 0.05 [EtOAc:MeOH:AcOH, 94:5:1]; ^{1}H NMR (500 MHz, DMSO-d_6) δ: 8.13 (1H, s, 9-C*H*); 7.94 (1H, br s, Ar-C*H*); 7.61–7.59 (3H, m, Ar-C*H*,11-C*H*,15-C*H*); 7.47 (1H, br s, Ar-C*H*); 7.23 (1H, dd, *J*= 8.9, 8.3 Hz, 5-C*H*); 7.20–7.18 (2H, m, Ar-C*H*,Ar-C*H*); 7.08 (2H, d, *J*= 8.8 Hz, 12-C*H*,14-C*H*); 6.95–6.93 (1H, m, 4/6-C*H*); 6.80 (1H, dd, *J*= 2.2, 2.2 Hz, 2-C*H*); 6.77–6.75 (1H, m, 4/6-C*H*); 4.98 (2H, s, 7-CH_2); 4.85–4.77 (4H, m, 19-CH_2,20-CH_2); 4.37 (2H, br s, 7-CH_2); 4.40 (2H, s, 17-CH_2); 3.62 (3H, s, 16-CH_3); ^{13}C NMR (125 MHz, DMSO-d_6) δ: 162.7 (*C*), 158.0 (*C*), 142.5 (*C*), 141.0 (*C*), 130.1 (*C*), 129.6 (*C*H5), 129.6 (*C*H-11,*C*H-15), 124.9 (*C*H-9), 122.5 (*C*H-Ar), 122.5 (*C*H-Ar), 121.7 (*C*H-Ar), 121.6 (*C*H-Ar), 121.5 (*C*H-Ar), 119.9 (*C*H-4/6), 114.5 (*C*H-2), 114.3 (*C*H-12,*C*H-14), 113.1 (*C*H-4/6), 61.1 (*C*H_2-7), 55.7 (*C*H_3-16), 52.0 (*C*H_2-17), 49.1(*C*H_2-20); 44.1 (*C*H_2-19); HPLC RT 7.46 min; >95% content; HRMS *m/z* (ESI$^-$) [Found $(M-H)^-$ 561.1557; $C_{27}H_{25}N_6O_6S$ requires $(M-H)^-$ 561.1562].

References

1. Baylin, S. B.; Schuebel, K. E. Genomic biology - The epigenomic era opens. *Nature* **2007,** 448, 548-549.

2. Gelato, K. A.; Fischle, W. Role of histone modifications in defining chromatin structure and function. *Biological Chemistry* **2008,** 389, 353-363.

3. Starkman, B. G.; Sakharkar, A. J.; Pandey, S. C. Epigenetics - Beyond the Genome in Alcoholism. *Alcohol Research-Current Reviews* **2012,** 34, 293-305.

4. Waddington, C. H. The genetic control of wing development in Drosophila. *Journal of Genetics* **1940,** 41, 75-139.

5. Bird, A. Perceptions of epigenetics. *Nature* **2007,** 447, 396-398.

6. Schulz, L. C. The Dutch Hunger Winter and the developmental origins of health and disease. *Proceedings of the National Academy of Sciences of the United States of America* **2010,** 107, 16757-16758.

7. Fraga, M. F.; Ballestar, E.; Paz, M. F.; Ropero, S.; Setien, F.; Ballestart, M. L.; Heine-Suner, D.; Cigudosa, J. C.; Urioste, M.; Benitez, J.; Boix-Chornet, M.; Sanchez-Aguilera, A.; Ling, C.; Carlsson, E.; Poulsen, P.; Vaag, A.; Stephan, Z.; Spector, T. D.; Wu, Y. Z.; Plass, C.; Esteller, M. Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** 102, 10604-10609.

8. Martin, G. M. Epigenetic drift in aging identical twins. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** 102, 10413-10414.

9. Golbabapour, S.; Abdulla, M. A.; Hajrezaei, M. A Concise Review on Epigenetic Regulation: Insight into Molecular Mechanisms. *International Journal of Molecular Sciences* **2011,** 12, 8661-8694.

10. Arrowsmith, C. H.; Bountra, C.; Fish, P. V.; Lee, K.; Schapira, M. Epigenetic protein families: a new frontier for drug discovery. *Nature Reviews Drug Discovery* **2012,** 11, 384-400.

11. Turner, B. M. Defining an epigenetic code. *Nature Cell Biology* **2007,** 9, 2-6.

12. Collings, C. K.; Waddell, P. J.; Anderson, J. N. Effects of DNA methylation on nucleosome stability. *Nucleic Acids Research* **2013,** 41, 2918-2931.

13. Rivera, C.; Gurard-Levin, Z. A.; Almouzni, G.; Loyola, A. Histone lysine methylation and chromatin replication. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **2014,** 1839, 1433-1439.

14. Winter, S.; Fischle, W. Epigenetic markers and their cross-talk. *Essays in Biochemistry: Epigenetics, Disease and Behaviour* **2010,** 48, 45-61.

15. McGinty, R. K.; Koehn, M.; Chatterjee, C.; Chiang, K. P.; Pratt, M. R.; Muir, T. W. Structure-Activity Analysis of Semisynthetic Nucleosomes: Mechanistic

Insights into the Stimulation of Dot1L by Ubipuitylated Histone H2B. *ACS Chemical Biology* **2009,** 4, 958-968.

16. Struck, A.-W.; Thompson, M. L.; Wong, L. S.; Micklefield, J. S-Adenosyl-Methionine-Dependent Methyltransferases: Highly Versatile Enzymes in Biocatalysis, Biosynthesis and Other Biotechnological Applications. *ChemBioChem* **2012,** 13, 2642-2655.

17. Copeland, R. A.; Solomon, M. E.; Richon, V. M. Protein methyltransferases as a target class for drug discovery. *Nature Reviews Drug Discovery* **2009,** 8, 724-732.

18. Allis, C. D.; Berger, S. L.; Cote, J.; Dent, S.; Jenuwien, T.; Kouzarides, T.; Pillus, L.; Reinberg, D.; Shi, Y.; Shiekhattar, R.; Shilatifard, A.; Workman, J.; Zhang, Y. New nomenclature for chromatin-modifying enzymes. *Cell* **2007,** 131, 633-636.

19. Chan-Penebre, E.; Kuplast, K. G.; Majer, C. R.; Boriack-Sjodin, P. A.; Wigle, T. J.; Johnston, L. D.; Rioux, N.; Munchhof, M. J.; Jin, L.; Jacques, S. L.; West, K. A.; Lingaraj, T.; Stickland, K.; Ribich, S. A.; Raimondi, A.; Scott, M. P.; Waters, N. J.; Pollock, R. M.; Smith, J. J.; Barbash, O.; Pappalardi, M.; Ho, T. F.; Nurse, K.; Oza, K. P.; Gallagher, K. T.; Kruger, R.; Moyer, M. P.; Copeland, R. A.; Chesworth, R.; Duncan, K. W. A selective inhibitor of PRMT5 with in vivo and in vitro potency in MCL models. *Nature Chemical Biology* **2015,** 11, 432-437.

20. Zhang, X.; Zhou, L.; Cheng, X. D. Crystal structure of the conserved core of protein arginine methyltransferase PRMT3. *Embo Journal* **2000,** 19, 3509-3519.

21. Barski, A.; Cuddapah, S.; Cui, K.; Roh, T.-Y.; Schones, D. E.; Wang, Z.; Wei, G.; Chepelev, I.; Zhao, K. High-resolution profiling of histone methylations in the human genome. *Cell* **2007,** 129, 823-837.

22. Snowden, A. W.; Gregory, P. D.; Case, C. C.; Pabo, C. O. Gene-specific targeting of H3K9 methylation is sufficient for initiating repression in vivo. *Current Biology* **2002,** 12, 2159-2166.

23. Kaniskan, H. Ü.; Martini, M. L.; Jin, J. Inhibitors of Protein Methyltransferases and Demethylases. *Chemical Reviews* **2017**, Article ASAP. DOI: 10.1021/acs.chemrev.6b00801.

24. Bracken, A. P.; Pasini, D.; Capra, M.; Prosperini, E.; Colli, E.; Helin, K. EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *Embo Journal* **2003,** 22, 5323-5335.

25. ClinicalTrials.gov Identifier: NCT02082977. A Study to Investigate the Safety, Pharmacokinetics, Pharmacodynamics and Clinical Activity of GSK2816126 in Subjects With Relapsed/Refractory Diffuse Large B Cell and Transformed Follicular Lymphoma, Transformed Follicular Lymphoma, Other Non-Hodgkin's Lymphomas, Solid Tumors and Multiple Myeloma (accessed Aug 2017).

26. Knutson, S. K.; Kawano, S.; Minoshima, Y.; Warholic, N. M.; Huang, K.-C.; Xiao, Y.; Kadowaki, T.; Uesugi, M.; Kuznetsov, G.; Kumar, N.; Wigle, T. J.; Klaus, C. R.; Allain, C. J.; Raimondi, A.; Waters, N. J.; Smith, J. J.; Porter-Scott, M.; Chesworth, R.; Moyer, M. P.; Copeland, R. A.; Richon, V. M.; Uenaka, T.; Pollock, R. M.; Kuntz, K. W.; Yokoi, A.; Keilhack, H. Selective Inhibition of EZH2

by EPZ-6438 Leads to Potent Antitumor Activity in EZH2-Mutant Non-Hodgkin Lymphoma. *Molecular Cancer Therapeutics* **2014,** 13, 842-854.

27. Morera, L.; Lubbert, M.; Jung, M. Targeting histone methyltransferases and demethylases in clinical trials for cancer therapy. *Clinical Epigenetics* **2016,** 8.

28. Bradley, W. D.; Arora, S.; Busby, J.; Balasubramanian, S.; Gehling, V. S.; Nasveschuk, C. G.; Vaswani, R. G.; Yuan, C. C.; Hatton, C.; Zhao, F.; Williamson, K. E.; Iyer, P.; Mendez, J.; Campbell, R.; Cantone, N.; Garapaty-Rao, S.; Audia, J. E.; Cook, A. S.; Dakin, L. A.; Albrecht, B. K.; Harmange, J. C.; Daniels, D. L.; Cummings, R. T.; Bryant, B. M.; Normant, E.; Trojer, P. EZH2 Inhibitor Efficacy in Non-Hodgkin's Lymphoma Does Not Require Suppression of H3K27 Monomethylation. *Chemistry & Biology* **2014,** 21, 1463-1475.

29. ClinicalTrials.gov Identifier: NCT02395601. A Study Evaluating CPI-1205 in Patients With B-Cell Lymphoma (accessed Aug 2017).

30. ClinicalTrials.gov Identifier: NCT02141828. A Phase 1 Dose Escalation and Expanded Cohort Study of EPZ-5676 in the Treatment of Pediatric Patients With Relapsed/Refractory Leukemias Bearing a Rearrangement of the MLL Gene (accessed Aug 2017).

31. ClinicalTrials.gov Identifier: NCT02783300. Dose Escalation Study of GSK3326595 in Subjects with Solid Tumors and Non-Hodgkin's Lymphoma (accessed Aug 2017).

32. Kramer, J. M. Regulation of cell differentiation and function by the euchromatin histone methyltranserfases G9a and GLP. *Biochemistry and Cell Biology* **2016,** 94, 26-32.

33. Xiong, Y.; Li, F. L.; Babaultlt, N.; Dong, A. P.; Zeng, H.; Wu, H.; Chen, X.; Arrowsmith, C. H.; Brown, P. J.; Liu, J.; Vedadi, M.; Jin, J. Discovery of Potent and Selective Inhibitors for G9a-Like Protein (GLP) Lysine Methyltransferase. *Journal of Medicinal Chemistry* **2017,** 60, 1876-1891.

34. Siarheyeva, A.; Senisterra, G.; Allali-Hassani, A.; Dong, A. P.; Dobrovetsky, E.; Wasney, G. A.; Chau, I.; Marcellus, R.; Hajian, T.; Liu, F.; Korboukh, I.; Smil, D.; Bolshan, Y.; Min, J. R.; Wu, H.; Zeng, H.; Loppnau, P.; Poda, G.; Griffin, C.; Aman, A.; Brown, P. J.; Jin, J.; Al-awar, R.; Arrowsmith, C. H.; Schapira, M.; Vedadi, M. An Allosteric Inhibitor of Protein Arginine Methyltransferase 3. *Structure* **2012,** 20, 1425-1435.

35. Kaniskan, H. U.; Szewczyk, M. M.; Yu, Z. T.; Eram, M. S.; Yang, X. B.; Schmidt, K.; Luo, X.; Dai, M.; He, F.; Zang, I.; Lin, Y.; Kennedy, S.; Li, F. L.; Dobrovetsky, E.; Dong, A. P.; Smil, D.; Min, S. J.; Landon, M.; Lin-Jones, J.; Huang, X. P.; Roth, B. L.; Schapira, M.; Atadja, P.; Barsyte-Lovejoy, D.; Arrowsmith, C. H.; Brown, P. J.; Zhao, K. H.; Jin, J.; Vedadi, M. A Potent, Selective and Cell-Active Allosteric Inhibitor of Protein Arginine Methyltransferase 3 (PRMT3). *Angewandte Chemie-International Edition* **2015,** 54, 5166-5170.

36. Butler, K. V.; Ma, A. Q.; Yu, W. Y.; Li, F. L.; Tempel, W.; Babault, N.; Pittella-Silva, F.; Shao, J.; Wang, J. Y.; Luo, M. K.; Vedadi, M.; Brown, P. J.; Arrowsmith, C. H.; Jin, J. Structure-Based Design of a Covalent Inhibitor of the SET Domain-Containing Protein 8 (SETD8) Lysine Methyltransferase. *Journal of Medicinal Chemistry* **2016,** 59, 9881-9889.

37. He, S. H.; Senter, T. J.; Pollock, J.; Han, C. H.; Upadhyay, S. K.; Purohit, T.; Gogliotti, R. D.; Lindsley, C. W.; Cierpicki, T.; Stauffer, S. R.; Grembecka, J. High-Affinity Small-Molecule Inhibitors of the Menin-Mixed Lineage Leukemia (MLL) Interaction Closely Mimic a Natural Protein-Protein Interaction. *Journal of Medicinal Chemistry* **2014,** 57, 1543-1556.

38. Feng, Q.; Wang, H. B.; Ng, H. H.; Erdjument-Bromage, H.; Tempst, P.; Struhl, K.; Zhang, Y. Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Current Biology* **2002,** 12, 1052-1058.

39. Woo Park, J.; Kim, K. B.; Kim, J. Y.; Chae, Y. C.; Jeong, O. S.; Seo, S. B. RE-IIBP Methylates H3K79 and Induces MEIS1-mediated Apoptosis via H2BK120 Ubiquitination by RNF20. *Scientific Reports* **2015,** 5, 1-12.

40. Wong, M.; Polly, P.; Liu, T. The histone methyltransferase DOT1L: regulatory functions and a cancer therapy target. *American Journal of Cancer Research* **2015,** 5, 2823-2837.

41. Zhou, L. J.; Holt, M. T.; Ohashi, N.; Zhao, A. S.; Muller, M. M.; Wang, B. Y.; Muir, T. W. Evidence that ubiquitylated H2B corrals hDot1L on the nucleosomal surface to induce H3K79 methylation. *Nature Communications* **2016,** 7, 10589.

42. Nguyen, A. T.; Xiao, B.; Neppl, R. L.; Kallin, E. M.; Li, J. A.; Chen, T. P.; Wang, D. Z.; Xiao, X. A.; Zhang, Y. DOT1L regulates dystrophin expression and is critical for cardiac function. *Genes & Development* **2011,** 25, 263-274.

43. Nguyen, A. T.; Zhang, Y. The diverse functions of Dot1 and H3K79 methylation. *Genes & Development* **2011,** 25, 1345-1358.

44. Nguyen, A. T.; He, J.; Taranova, O.; Zhang, Y. Essential role of DOT1L in maintaining normal adult hematopoiesis. *Cell Research* **2011,** 21, 1370-1373.

45. Min, J. R.; Feng, Q.; Li, Z. Z.; Zhang, Y.; Xu, R. M. Structure of the catalytic domain of human DOT1L, a Non-SET domain nucleosomal histone methyltransferase. *Cell* **2003,** 112, 711-723.

46. Schapira, M. Structural Chemistry of Human SET Domain Protein Methyltransferases. *Current Chemical Genomics* **2011,** 5, 85-94.

47. Cho, M.-H.; Park, J.-H.; Choi, H.-J.; Park, M.-K.; Won, H.-Y.; Park, Y.-J.; Lee, C. H.; Oh, S.-H.; Song, Y.-S.; Kim, H. S.; Oh, Y.-H.; Lee, J.-Y.; Kong, G. DOT1L cooperates with the c-Myc-p300 complex to epigenetically derepress CDH1 transcription factors in breast cancer progression. *Nature Communications* **2015,** 6, 7821-7835.

48. Betancourt, M. C. C.; Cailotto, F.; Kerkhof, H. J.; Cornelis, F. M. F.; Doherty, S. A.; Hart, D. J.; Hofman, A.; Luyten, F. P.; Maciewicz, R. A.; Mangino, M.; Metrustry, S.; Muir, K.; Peters, M. J.; Rivadeneira, F.; Wheeler, M.; Zhang, W. Y.; Arden, N.; Spector, T. D.; Uitterlinden, A. G.; Doherty, M.; Lories, R. J. U.; Valdes, A. M.; van Meurs, J. B. J. Genome-wide association and functional studies identify the DOT1L gene to be involved in cartilage thickness and hip osteoarthritis. *Proceedings of the National Academy of Sciences of the United States of America* **2012,** 109, 8218-8223.

49. Shen, C.; Jo, S. Y.; Liao, C.; Hess, J. L.; Nikolovska-Coleska, Z. Targeting Recruitment of Disruptor of Telomeric Silencing 1-like (DOT1L) characterizing the interactions between DOT1L and mixed lineage leukemia (MLL) fusion proteins. *Journal of Biological Chemistry* **2013,** 288, 30585-30596.

50. Campagna-Slater, V.; Mok, M. W.; Nguyen, K. T.; Feher, M.; Najmanovich, R.; Schapira, M. Structural Chemistry of the Histone Methyltransferases Cofactor Binding Site. *Journal of Chemical Information and Modeling* **2011,** 51, 612-623.

51. Daigle, S. R.; Olhava, E. J.; Therkelsen, C. A.; Majer, C. R.; Sneeringer, C. J.; Song, J.; Johnston, L. D.; Scott, M. P.; Smith, J. J.; Xiao, Y. H.; Jin, L.; Kuntz, K. W.; Chesworth, R.; Moyer, M. P.; Bernt, K. M.; Tseng, J. C.; Kung, A. L.; Armstrong, S. A.; Copeland, R. A.; Richon, V. M.; Pollock, R. M. Selective Killing of Mixed Lineage Leukemia Cells by a Potent Small-Molecule DOT1L Inhibitor. *Cancer Cell* **2011,** 20, 53-65.

52. Travers, J.; Blagg, J.; Workman, P. Epigenetics: Targeting leukemia on the DOT. *Nature Chemical Biology* **2011,** 7, 663-5.

53. Daigle, S. R.; Olhava, E. J.; Therkelsen, C. A.; Basavapathruni, A.; Jin, L.; Boriack-Sjodin, P. A.; Allain, C. J.; Klaus, C. R.; Raimondi, A.; Scott, M. P.; Waters, N. J.; Chesworth, R.; Moyer, M. P.; Copeland, R. A.; Richon, V. M.; Pollock, R. M. Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. *Blood* **2013,** 122, 1017-1025.

54. Waters, N. J.; Smith, S. A.; Olhava, E. J.; Duncan, K. W.; Burton, R. D.; O'Neill, J.; Rodrigue, M.-E.; Pollock, R. M.; Moyer, M. P.; Chesworth, R. Metabolism and disposition of the DOT1L inhibitor, pinometostat (EPZ-5676), in rat, dog and human. *Cancer Chemotherapy and Pharmacology* **2016,** 77, 43-62.

55. Yu, W.; Smil, D.; Li, F.; Tempel, W.; Fedorov, O.; Nguyen, K.; Bolshan, Y.; Al-Awar, R.; Knapp, S.; Arrowsmith, C.; Vedadi, M.; Brown, P.; Schapira, M. Bromo-deaza-SAH: A potent and selective DOT1L inhibitor. *Bioorganic & Medicinal Chemistry* **2013,** 21, 1787-1794.

56. Yu, W.; Chory, E. J.; Wernimont, A. K.; Tempel, W.; Scopton, A.; Federation, A.; Marineau, J. J.; Qi, J.; Barsyte-Lovejoy, D.; Yi, J.; Marcellus, R.; Iacob, R. E.; Engen, J. R.; Griffin, C.; Aman, A.; Wienholds, E.; Li, F.; Pineda, J.; Estiu, G.; Shatseva, T.; Hajian, T.; Al-Awar, R.; Dick, J. E.; Vedadi, M.; Brown, P. J.; Arrowsmith, C. H.; Bradner, J. E.; Schapira, M. Catalytic site remodelling of the DOT1L methyltransferase by selective inhibitors. *Nature communications* **2012,** 3, 1288-1288.

57. Anglin, J. L.; Deng, L. S.; Yao, Y.; Cai, G. B.; Liu, Z.; Jiang, H.; Cheng, G.; Chen, P. H.; Dong, S.; Song, Y. C. Synthesis and Structure-Activity Relationship Investigation of Adenosine-Containing Inhibitors of Histone Methyltransferase DOT1L. *Journal of Medicinal Chemistry* **2012,** 55, 8066-8074.

58. Yao, Y.; Chen, P.; Diao, J.; Cheng, G.; Deng, L.; Anglin, J. L.; Prasad, B. V. V.; Song, Y. Selective Inhibitors of Histone Methyltransferase DOT1L: Design, Synthesis, and Crystallographic Studies. *Journal of the American Chemical Society* **2011,** 133, 16746-16749.

59. Kim, Y. A.; Sharon, A.; Chu, C. K.; Rais, R. H.; Al Safarjalani, O. N.; Naguib, F. N. M.; el Kouni, M. H. Structure-activity relationships of 7-deaza-6-benzylthioinosine analogues as ligands of Toxoplasma gondii adenosine kinase. *Journal of Medicinal Chemistry* **2008,** 51, 3934-3945.

60. Anderson, J. D.; Bontems, R. J.; Geary, S.; Cottam, H. B.; Larson, S. B.; Matsumoto, S. S.; Smee, D. F.; Robins, R. K. Synthesis of tubercidin, 6-chlorotubercidin and related nucleosides. *Nucleosides & Nucleotides* **1989,** 8, 1201-1216.

61. Deng, L.; Zhang, L.; Yao, Y.; Wang, C.; Redell, M. S.; Dong, S.; Song, Y. Synthesis, activity and metabolic stability of non-ribose containing inhibitors of histone methyltransferase DOT1L. *Medicinal Chemical Communications* **2013,** 4, 822-826.

62. Lee, J. M.; Choi, M. J. Compounds and compositions for modulating histone methyltransferase activity. Patent Application WO2014035140 A2. **2014**.

63. Yi, J. S.; Federation, A. J.; Qi, J.; Dhe-Paganon, S.; Hadler, M.; Xu, X.; St Pierre, R.; Varca, A. C.; Wu, L.; Marineau, J. J.; Smith, W. B.; Souza, A.; Chory, E. J.; Armstrong, S. A.; Bradner, J. E. Structure-Guided DOT1L Probe Optimization by Label-Free Ligand Displacement. *ACS Chemical Biology* **2015,** 10, 667-674.

64. Cottam, H. B.; Wasson, D. B.; Shih, H. C.; Raychaudhuri, A.; Dipasquale, G.; Carson, D. A. New adenosine kinase inhibitors with oral antiinflammatory activity - synthesis and biological evaluation. *Journal of Medicinal Chemistry* **1993,** 36, 3424-3430.

65. Meanwell, N. A. Synopsis of Some Recent Tactical Application of Bioisosteres in Drug Design. *Journal of Medicinal Chemistry* **2011,** 54, 2529-2591.

66. Jones, L.; Summerhill, N.; Swain, N.; Mills, J. Aromatic chloride to nitrile transformation: medicinal and synthetic chemistry. *Medicinal Chemistry Communications* **2010,** 1, 309-318.

67. Matthews, D. P.; McCarthy, J. R.; Whitten, J. P.; Kastner, P. R.; Barney, C. L.; Marshall, F. N.; Ertel, M. A.; Burkhard, T.; Shea, P. J.; Kariya, T. Synthesis and Cardiotonic Activity of Novel Biimidazoles. *Journal of Medicinal Chemistry* **1990,** 33, 317-327.

68. Bell, A. S.; Campbell, S. F.; Morris, D. S.; Roberts, D. A.; Stefaniak, M. H. 2(1H)-Quinolinones with cardiac stimulant activity .3. Synthesis and biological properties of 6-imidazol-1-yl derivatives. *Journal of Medicinal Chemistry* **1989,** 32, 1552-1558.

69. Cody, V.; Wojtczak, A.; Davis, F. B.; Davis, P. J.; Blas, S. D. Structure-Activity-Relationships of Milrinone Analogs Determined In-Vitro in a Rabbit Heart Membrane Ca(2+) Atpase Model. *Journal of Medicinal Chemistry* **1995,** 38, 1990-1997.

70. Browne, L. J.; Gude, C.; Rodriguez, H.; Steele, R. E.; Bhatnager, A. Fadrozole Hydrochloride - A Potent, Selective, Nonsteroidal Inhibitor of Aromatase for the Treatment of Estrogen-Dependent Disease. *Journal of Medicinal Chemistry* **1991,** 34, 725-736.

71. Bromidge, S. M.; Brown, F.; Cassidy, F.; Clark, M. S. G.; Dabbs, S.; Hadley, M. S.; Hawkins, J.; Loudon, J. M.; Naylor, C. B.; Orlek, B. S.; Riley, G. J. Design of R-(Z) -(+)-alpha-(methoxyimino)-1-azabicyclo 2.2.2 octane-3-acetonitrile (SB 202026), a functionally selective azabicyclic muscarinic M1 agonist incorporating the N-methoxy imidoyl nitrile group as a novel ester bioisostere. *Journal of Medicinal Chemistry* **1997,** 40, 4265-4280.

72. Arnott, J. A.; Planey, S. L. The influence of lipophilicity in drug discovery and design. *Expert Opinion on Drug Discovery* **2012,** 7, 863-875.

73. Leeson, P. D.; Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nature Reviews Drug Discovery* **2007,** 6, 881-90.

74. Abraham, M. H.; Chadha, H. S.; Whiting, G. S.; Mitchell, R. C. Hydrogen Bonding .32. An Analysis of Water-Octanol and Water-Alkane Partitioning and the Δlog*P* Parameter of Seiler. *Journal of Pharmaceutical Sciences* **1994,** 83, 1085-1100.

75. Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics* **2012,** 4, 1-17.

76. Spurr, S. S.; Bayle, E. D.; Yu, W. Y.; Li, F. L.; Tempel, W.; Vedadi, M.; Schapira, M.; Fish, P. V. New small molecule inhibitors of histone methyl transferase DOT1L with a nitrile as a non-traditional replacement for heavy halogen atoms. *Bioorganic & Medicinal Chemistry Letters* **2016,** 26, 4518-4522.

77. Ushijima, S.; Togo, H. Metal-Free One-Pot Conversion of Electron-Rich Aromatics into Aromatic Nitriles. *Synlett* **2010**, 1067-1070.

78. Zhang, N.; Zhang, X.; Zhu, J.; Turpoff, A.; Chen, G.; Morrill, C.; Huang, S.; Lennox, W.; Kakarla, R.; Liu, R.; Li, C.; Ren, H.; Almstead, N.; Venkatraman, S.; Njoroge, F.; Gu, Z.; Clausen, V.; Graci, J.; Jung, S.; Zheng, Y.; Colacino, J.; Lahser, F.; Sheedy, J.; Mollin, A.; Weetall, M.; Nomeir, A.; Karp, G. Structure-Activity Relationship (SAR) Optimization of 6-(Indol-2-yl)pyridine-3-sulfonamides: Identification of Potent, Selective, and Orally Bioavailable Small Molecules Targeting Hepatitis C (HCV) NS4B. *Journal of Medicinal Chemistry* **2014,** 57, 2121-2135.

79. Corbett, M.; Freeman-Cook, K.; Kauffman, G.; Lippa, B.; Luzzio, M.; Morris, J. Amine derivatives useful as anticancer agents. U.S. Patent Application 11/780,654. **2007**.

80. Rokade, B. V.; Prabhu, K. R. Chemoselective Schmidt Reaction Mediated by Triflic Acid: Selective Synthesis of Nitriles from Aldehydes. *Journal of Organic Chemistry* **2012,** 77, 5364-5370.

81. Ren, P.; Liu, Y.; Li, L.; Chan, K.; Wilson, T. E.; Castro, A. C.; Evans, C. A.; Snyder, D. A. Heterocyclic compounds and uses thereof. U.S. Patent 8,901,133. **2012**.

82. Scott, R. W.; Fox, D. E.; Williams, D. K. β-selective nucleoside analog synthesis from chlorofuranoses. *Tetrahedron Letters* **2000,** 41, 8207-8210.

83. Wang, Z. Vorbrüggen Glycosylation. In *Comprehensive Organic Name Reactions and Reagents*, Wang, Z., Ed. Wiley & Sons: 2010; Vol. 652, pp 2915-2919.

84. Porcari, A.; Townsend, L. Total synthesis of the naturally occurring antibiotic toyocamycin using new and improved synthetic procedures. *Nucleosides & Nucleotides* **1999,** 18, 153-159.

85. Xiao, C.; Sun, C.; Han, W.; Pan, F.; Dan, Z.; Li, Y.; Song, Z.; Jin, Y. Synthesis of 6-(het) ary Xylocydine analogues and evaluating their inhibitory activities of CDK1 and CDK2 in vitro. *Bioorganic & Medicinal Chemistry* **2011,** 19, 7100-7110.

86. Leonard, P.; Ingale, S.; Ding, P.; Ming, X.; Seela, F. Studies on the glycosylation of pyrrolo[2,3-*d*]pyrimidines with 1-*O*-acetyl-2,3,5-tri-*O*-benzoyl-β-D-ribofuranose: The formation of regioisomers during toyocamycin and 7-deazainosine syntheses. *Nucleosides Nucleotides & Nucleic Acids* **2009,** 28, 678-694.

87. Kim, Y.; Kwon, S.; Bae, I.; Kim, B. Selectivity between N-1 and N-7 nucleosides: regioselective synthesis of BMK-Y101, a potent CDK7 and 9 inhibitor. *Tetrahedron Letters* **2013,** 54, 5484-5488.

88. Bradner, J. E.; Qi, J.; Federation, A. DOT1L Inhibitors, US Patent Application 2016/0060269 Al. **2016**.

89. Stein, E. M.; Garcia-Manero, G.; Rizzieri, D. A.; Tibes, R.; Berdeja, J. G.; Jongen-Lavrencic, M.; Altman, J. K.; Dohner, H.; Thomson, B.; Blakemore, S. J.; Daigle, S.; Fine, G.; Waters, N. J.; Krivstov, A. V.; Koche, R.; Armstrong, S. A.; Ho, P. T.; Lowenberg, B.; Tallman, M. S. A Phase 1 Study of the DOT1L Inhibitor, Pinometostat (EPZ-5676), in Adults with Relapsed or Refractory Leukemia: Safety, Clinical Activity, Exposure and Target Inhibition. *Blood* **2015,** 126, 2547.

90. Chen, C.; Zhu, H.; Stauffer, F.; Caravatti, G.; Vollmer, S.; Machauer, R.; Holzer, P.; Mobitz, H.; Scheufler, C.; Klumpp, M.; Tiedt, R.; Beyer, K. S.; Calkins, K.; Guthy, D.; Kiffe, M.; Zhang, J.; Gaul, C. Discovery of Novel Dot1L Inhibitors through a Structure-Based Fragmentation Approach. *ACS Medicinal Chemistry Letters* **2016,** 7, 735-740.

91. Scheufler, C.; Mobitz, H.; Gaul, C.; Ragot, C.; Be, C.; Fernandez, C.; Beyer, K. S.; Tiedt, R.; Stauffer, F. Optimization of a Fragment-Based Screening Hit toward Potent DOT1L Inhibitors Interacting in an Induced Binding Pocket. *ACS Medicinal Chemistry Letters* **2016,** 7, 730-734.

92. Chen, S. J.; Li, L. J.; Chen, Y. T.; Hu, J. C.; Liu, J. Q.; Liu, Y. C.; Liu, R. F.; Zhang, Y. Y.; Meng, F. W.; Zhu, K. K.; Lu, J. Y.; Zheng, M. Y.; Chen, K. X.; Zhang, J.; Jiang, H. L.; Yao, Z. Y.; Luo, C. Identification of Novel Disruptor of Telomeric Silencing 1-like (DOT1L) Inhibitors through Structure-Based Virtual Screening and Biological Assays. *Journal of Chemical Information and Modeling* **2016,** 56, 527-534.

93. Tolman, R. L.; Robins, R. K.; Townsend, L. B. Pyrrolopyrimidine nucleosides .3. Total synthesis of toyocamycin sangivamycin tubercidin and related derivatives. *Journal of the American Chemical Society* **1969,** 91, 2102-2108.

94. Earl, R. A.; Townsend, L. B. Pyrazolopyrimidine nucleosides .7. Synthesis of certain pyrazolo 3,4-*d* pyrimidine nucleosides related to nucleoside antibiotics toyocamycin and sangivamycinPyrazolopyrimidine nucleosides .7. Synthesis of certain pyrazolo 3,4-d pyrimidine nucleosides related to nucleoside antibiotics toyocamycin and sangivamycin. *Journal of Heterocyclic Chemistry* **1974,** 11, 1033-1039.

95. Lin, W. Q.; Li, H.; Ming, X.; Seela, F. 1,N-6-Etheno-7-deaza-2,8-diazaadenosine: syntheses, properties and conversion to 7-deaza-2,8-diazaadenosine. *Organic & Biomolecular Chemistry* **2005,** 3, 1714-1718.

96. Song, Y.; Ding, H.; Dou, Y.; Yang, R.; Sun, Q.; Xiao, Q.; Ju, Y. Efficient and Practical Synthesis of 5'-Deoxytubercidin and Its Analogues via Vorbruggen Glycosylation. *Synthesis* **2011,** 9, 1442-1446.

97. Castro, A. C. E., C.A. Janardanannair, S.; Lescarbeau, A.; Liu, T.; Snyder, D. A.; Tremblay, M. R.; Ren, P.; Liu, Y.; Li, L. Heterocyclic compounds and uses thereof. Patent Application WO2013012915. **2013**.

98. Shen, W.; Kim, J.-S.; Hilfinger, J. Expedient total synthesis of triciribine and its prodrugs. *Synthetic Communications* **2012,** 42, 358-374.

99. Taylor, E. C.; Hendess, R. W. Synthesis of pyrrolo 2,3-d pyrimidines. The aglycone of toyocamycin. *Journal of the American Chemical Society* **1965,** 87, 1995-2003.

100. Porcari, A.; Townsend, L. An improved synthesis of the versatile heterocycle, 4-amino-6-bromo-5-cyanopyrrolo[2,3-d]pyrimidine. *Synthetic Communications* **1998,** 28, 3835-3843.

101. Porcari, A. R.; Ptak, R. G.; Borysko, K. Z.; Breitenbach, J. M.; Vittori, S.; Wotring, L. L.; Drach, J. C.; Townsend, L. B. Deoxy sugar analogues of triciribine: Correlation of antiviral and antiproliferative activity with intracellular phosphorylation. *Journal of Medicinal Chemistry* **2000,** 43, 2438-2448.

102. Bourdier, T.; Fookes, C. J. R.; Pham, T. Q.; Greguric, I.; Katsifis, A. Synthesis and stability of S-(2- F-18 fluoroethyl)-L-homocysteine for potential tumour imaging. *Journal of Labelled Compounds & Radiopharmaceuticals* **2008,** 51, 369-373.

103. Imlay, J. A. The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. *Nature Reviews Microbiology* **2013,** 11, 443-454.

104. Carocho, M.; Ferreira, I. A review on antioxidants, prooxidants and related controversy: Natural and synthetic compounds,. screening and analysis methodologies and future perspectives. *Food and Chemical Toxicology* **2013,** 51, 15-25.

105. Itoh, K.; Chiba, T.; Takahashi, S.; Ishii, T.; Igarashi, K.; Katoh, Y.; Oyake, T.; Hayashi, N.; Satoh, K.; Hatayama, I.; Yamamoto, M.; Nabeshima, Y. An Nrf2 small Maf heterodimer mediates the induction of phase II detoxifying enzyme genes through antioxidant response elements. *Biochemical and Biophysical Research Communications* **1997,** 236, 313-322.

106. Itoh, K.; Mimura, J.; Yamamoto, M. Discovery of the Negative Regulator of Nrf2, Keap1: A Historical Overview. *Antioxidants & Redox Signaling* **2010,** 13, 1665-1678.

107. Abed, D. A.; Goldstein, M.; Albanyan, H.; Jin, H. J.; Hu, L. Q. Discovery of direct inhibitors of Keap1-Nrf2 protein-protein interaction as potential therapeutic and preventive agents. *Acta Pharmaceutica Sinica B* **2015,** 5, 285-299.

108. McMahon, M.; Thomas, N.; Itoh, K.; Yamamoto, M.; Hayes, J. D. Redox-regulated turnover of Nrf2 is determined by at least two separate protein domains, the redox-sensitive Neh2 degron and the redox-insensitive Neh6 degron. *Journal of Biological Chemistry* **2004,** 279, 31556-31567.

109. Itoh, K.; Tong, K. I.; Yamamoto, M. Molecular mechanism activating Nrf2-Keap1 pathway in regulation of adaptive response to electrophiles. *Free Radical Biology and Medicine* **2004,** 36, 1208-1213.

110. Kobayashi, A.; Kang, M. I.; Okawa, H.; Ohtsuji, M.; Zenke, Y.; Chiba, T.; Igarashi, K.; Yamamoto, M. Oxidative stress sensor Keap1 functions as an adaptor for Cul3-based E3 ligase to regulate for proteasomal degradation of Nrf2. *Molecular and Cellular Biology* **2004,** 24, 7130-7139.

111. Kobayashi, M.; Li, L.; Iwamoto, N.; Nakajima-Takagi, Y.; Kaneko, H.; Nakayama, Y.; Eguchi, M.; Wada, Y.; Kumagai, Y.; Yamamoto, M. The Antioxidant Defense System Keap1-Nrf2 Comprises a Multiple Sensing Mechanism for Responding to a Wide Range of Chemical Compounds. *Molecular and Cellular Biology* **2009,** 29, 493-502.

112. Zhang, D. D.; Hannink, M. Distinct cysteine residues in Keap1 are required for Keap1-dependent ubiquitination of Nrf2 and for stabilization of Nrf2 by chemopreventive agents and oxidative stress. *Molecular and Cellular Biology* **2003,** 23, 8137-8151.

113. Rachakonda, G.; Xiong, Y.; Sekhar, K. R.; Starner, S. L.; Liebler, D. C.; Freeman, M. L. Covalent modification at Cys151 dissociates the electrophile sensor Keap1 from the ubiquitin ligase CUL3. *Chemical Research in Toxicology* **2008,** 21, 705-710.

114. Fujii, S.; Sawa, T.; Ihara, H.; Tong, K. I.; Ida, T.; Okamoto, T.; Ahtesham, A. K.; Ishima, Y.; Motohashi, H.; Yamamoto, M.; Akaike, T. The Critical Role of Nitric Oxide Signaling, via Protein S-Guanylation and Nitrated Cyclic GMP, in the Antioxidant Adaptive Response. *Journal of Biological Chemistry* **2010,** 285, 23970-23984.

115. Zipper, L. M.; Mulcahy, R. T. The Keap1 BTB/POZ dimerization function is required to sequester Nrf2 in cytoplasm. *Journal of Biological Chemistry* **2002,** 277, 36544-36552.

116. Li, X. C.; Zhang, D.; Hannink, M.; Beamer, L. J. Crystal structure of the Kelch domain of human Keap1. *Journal of Biological Chemistry* **2004,** 279, 54750-54758.

117. McMahon, M.; Thomas, N.; Itoh, K.; Yamamoto, M.; Hayes, J. D. Dimerization of substrate adaptors can facilitate cullin-mediated ubiquitylation of proteins by

a "Tethering" mechanism - A two-site interaction model for the Nrf2-Keap1 complex. *Journal of Biological Chemistry* **2006,** 281, 24756-24768.

118. Tong, K. I.; Padmanabhan, B.; Kobayashi, A.; Shang, C.; Hirotsu, Y.; Yokoyama, S.; Yamamoto, M. Different electrostatic Potentials define ETGE and DLG motifs as hinge and latch in oxidative stress response. *Molecular and Cellular Biology* **2007,** 27, 7511-7521.

119. Canning, P.; Sorrell, F. J.; Bullock, A. N. Structural basis of Keap1 interactions with Nrf2. *Free Radical Biology and Medicine* **2015,** 88, 101-107.

120. Kansanen, E.; Kuosmanen, S. M.; Leinonen, H.; Levonen, A. L. The Keap1-Nrf2 pathway: Mechanisms of activation and dysregulation in cancer. *Redox Biology* **2013,** 1, 45-49.

121. de Vries, H. E.; Witte, M.; Hondius, D.; Rozermuller, A. J. M.; Drukarch, B.; Hoozemans, J.; van Horssen, J. Nrf2-induced antioxidant Protection: A Promising target to counteract ROS-mediated damage in neurodegenerative disease? *Free Radical Biology and Medicine* **2008,** 45, 1375-1383.

122. Linker, R. A.; Lee, D. H.; Ryan, S.; van Dam, A. M.; Conrad, R.; Bista, P.; Zeng, W.; Hronowsky, X.; Buko, A.; Chollate, S.; Ellrichmann, G.; Bruck, W.; Dawson, K.; Goelz, S.; Wiese, S.; Scannevin, R. H.; Lukashev, M.; Gold, R. Fumaric acid esters exert neuroprotective effects in neuroinflammation via activation of the Nrf2 antioxidant pathway. *Brain* **2011,** 134, 678-692.

123. Cho, H. Y.; Kleeberger, S. R. Nrf2 protects against airway disorders. *Toxicology and Applied Pharmacology* **2010,** 244, 43-56.

124. Li, J. Q.; Ichikawa, T.; Janicki, J. S.; Cui, T. X. Targeting the Nrf2 pathway against cardiovascular disease. *Expert Opinion on Therapeutic Targets* **2009,** 13, 785-794.

125. Jimenez-Osorio, A. S.; Gonzalez-Reyes, S.; Pedraza-Chaverri, J. Natural Nrf2 activators in diabetes. *Clinica Chimica Acta* **2015,** 448, 182-192.

126. Zhang, H. Q.; Davies, K. J. A.; Forman, H. J. Oxidative stress response and Nrf2 signaling in aging. *Free Radical Biology and Medicine* **2015,** 88, 314-336.

127. Wells, G. Peptide and small molecule inhibitors of the Keap1-Nrf2 protein-protein interaction. *Biochemical Society Transactions* **2015,** 43, 674-679.

128. Kensler, T. W.; Egner, P. A.; Agyeman, A. S.; Visvanathan, K.; Groopman, J. D.; Chen, J. G.; Chen, T. Y.; Fahey, J. W.; Talalay, P. Keap1-Nrf2 Signaling: A Target for Cancer Prevention by Sulforaphane. *Natural Products in Cancer Prevention and Therapy* **2013,** 329, 163-177.

129. Bomprezzi, R. Dimethyl fumarate in the treatment of relapsing-remitting multiple sclerosis: an overview. *Therapeutic Advances in Neurological Disorders* **2015,** 8, 20-30.

130. Campbell, D.; Weir, M. R. Defining, Treating, and Understanding Chronic Kidney Disease - A Complex Disorder. *Journal of Clinical Hypertension* **2015,** 17, 514-527.

131. ClinicalTrials.gov Identifier: NCT01351675. Bardoxolone Methyl Evaluation in Patients With Chronic Kidney Disease and Type 2 Diabetes (BEACON) (accessed Aug 2017).

132. Wang, Y. Y.; Yang, Y. X.; Zhe, H.; He, Z. X.; Zhou, S. F. Bardoxolone methyl (CDDO-Me) as a therapeutic agent: an update on its pharmacokinetic and pharmacodynamic properties. *Drug Design Development and Therapy* **2014,** 8, 2075-2088.

133. Gupta, S. C.; Patchva, S.; Aggarwal, B. B. Therapeutic Roles of Curcumin: Lessons Learned from Clinical Trials. *Journal of the American Association of Pharmaceutical Scientists* **2013,** 15, 195-218.

134. Scott, D. E.; Bayly, A. R.; Abell, C.; Skidmore, J. Small molecules, big targets: drug discovery faces the protein-protein interaction challenge. *Nature Reviews Drug Discovery* **2016,** 15, 533-550.

135. Lo, S. C.; Li, X. C.; Henzl, M. T.; Beamer, L. J.; Hannink, M. Structure of the Keap1 : Nrf2 interface provides mechanistic insight into Nrf2 signaling. *Embo Journal* **2006,** 25, 3605-3617.

136. Hancock, R.; Bertrand, H. C.; Tsujita, T.; Naz, S.; El-Bakry, A.; Laoruchupong, J.; Hayes, J. D.; Wells, G. Peptide inhibitors of the Keap1-Nrf2 protein-protein interaction. *Free Radical Biology and Medicine* **2012,** 52, 444-451.

137. Lu, M. C.; Chen, Z. Y.; Wang, Y. L.; Jiang, Y. L.; Yuan, Z. W.; You, Q. D.; Jiang, Z. Y. Binding thermodynamics and kinetics guided optimization of potent Keap1-Nrf2 peptide inhibitors. *RSC Advances* **2015,** 5, 85983-85987.

138. Steel, R.; Cowan, J.; Payerne, E.; O'Connell, M. A.; Searcey, M. Anti-inflammatory Effect of a Cell-Penetrating Peptide Targeting the Nrf2/Keap1 Interaction. *ACS Medicinal Chemistry Letters* **2012,** 3, 407-410.

139. Hancock, R.; Schaap, M.; Pfister, H.; Wells, G. Peptide inhibitors of the Keap1-Nrf2 protein-protein interaction with improved binding and cellular activity. *Organic & Biomolecular Chemistry* **2013,** 11, 3553-3557.

140. Hu, L. Q.; Magesh, S.; Chen, L.; Wang, L. L.; Lewis, T. A.; Chen, Y.; Khodier, C.; Inoyama, D.; Beamer, L. J.; Emge, T. J.; Shen, J.; Kerrigan, J. E.; Kong, A. N. T.; Dandapani, S.; Palmer, M.; Schreiber, S. L.; Munoz, B. Discovery of a small-molecule inhibitor and cellular probe of Keap1-Nrf2 protein-protein interaction. *Bioorganic & Medicinal Chemistry Letters* **2013,** 23, 3039-3043.

141. Wen, X.; Thorne, G.; Hu, L. Q.; Joy, M. S.; Aleksunes, L. M. Activation of NRF2 Signaling in HEK293 Cells by a First-in-Class Direct KEAP1-NRF2 Inhibitor. *Journal of Biochemical and Molecular Toxicology* **2015,** 29, 261-266.

142. Jnoff, E.; Albrecht, C.; Barker, J. J.; Barker, O.; Beaumont, E.; Bromidge, S.; Brookfield, F.; Brooks, M.; Bubert, C.; Ceska, T.; Corden, V.; Dawson, G.; Duclos, S.; Fryatt, T.; Genicot, C.; Jigorel, E.; Kwong, J.; Maghames, R.; Mushi, I.; Pike, R.; Sands, Z. A.; Smith, M. A.; Stimson, C. C.; Courade, J. P. Binding Mode and Structure-Activity Relationships around Direct Inhibitors of the Nrf2-Keap1 Complex. *ChemMedChem* **2014,** 9, 699-705.

143. Marcotte, D.; Zeng, W. K.; Hus, J. C.; McKenzie, A.; Hession, C.; Jin, P.; Bergeron, C.; Lugovskoy, A.; Enyedy, I.; Cuervo, H.; Wang, D. P.; Atmanene, C.; Roecklin, D.; Vecchi, M.; Vivat, V.; Kraemer, J.; Winkler, D.; Hong, V.; Chao, J. H.; Lukashev, M.; Silvian, L. Small molecules inhibit the interaction of Nrf2 and the Keap1 Kelch domain through a non-covalent mechanism. *Bioorganic & Medicinal Chemistry* **2013,** 21, 4011-4019.

144. Jiang, Z. Y.; Lu, M. C.; Xu, L. L.; Yang, T. T.; Xi, M. Y.; Xu, X. L.; Guo, X. K.; Zhang, X. J.; You, Q. D.; Sun, H. P. Discovery of Potent Keap1-Nrf2 Protein-Protein Interaction Inhibitor Based on Molecular Binding Determinants Analysis. *Journal of Medicinal Chemistry* **2014,** 57, 4406-4406.

145. Jiang, Z. Y.; Xu, L. L.; Lu, M. C.; Chen, Z. Y.; Yuan, Z. W.; Xu, X. L.; Guo, X. K.; Zhang, X. J.; Sun, H. P.; You, Q. D. Structure-Activity and Structure-Property Relationship and Exploratory in Vivo Evaluation of the Nanomolar Keap1-Nrf2 Protein-Protein Interaction Inhibitor. *Journal of Medicinal Chemistry* **2015,** 58, 6410-6421.

146. Jain, A. D.; Potteti, H.; Richardson, B. G.; Kingsley, L.; Luciano, J. P.; Ryuzoji, A. F.; Lee, H.; Krunic, A.; Mesecar, A. D.; Reddy, S. P.; Moore, T. W. Probing the structural requirements of non-electrophilic naphthalene-based Nrf2 activators. *European Journal of Medicinal Chemistry* **2015,** 103, 252-268.

147. Winkel, A. F.; Engel, C. K.; Margerie, D.; Kannt, A.; Szillat, H.; Glombik, H.; Kallus, C.; Ruf, S.; Gussregen, S.; Riedel, J.; Herling, A. W.; von Knethen, A.; Weigert, A.; Brune, B.; Schmoll, D. Characterization of RA839, a Noncovalent Small Molecule Binder to Keapl and Selective Activator of Nrf2 Signaling. *Journal of Biological Chemistry* **2015,** 290, 28446-28455.

148. Sun, H. P.; Jiang, Z. Y.; Zhang, M. Y.; Lu, M. C.; Yang, T. T.; Pan, Y.; Huang, H. Z.; Zhang, X. J.; You, Q. D. Novel protein-protein interaction inhibitor of Nrf2-Keap1 discovered by structure-based virtual screening. *MedChemComm* **2014,** 5, 93-98.

149. Satoh, M.; Saburi, H.; Tanaka, T.; Matsuura, Y.; Naitow, H.; Shimozono, R.; Yamamoto, N.; Inoue, H.; Nakamura, N.; Yoshizawa, Y.; Aoki, T.; Tanimura, R.; Kunishima, N. Multiple binding modes of a small molecule to human Keap1 revealed by X-ray crystallography and molecular dynamics simulation. *Febs Open Bio* **2015,** 5, 557-570.

150. Shimozono, R.; Asaoka, Y.; Yoshizawa, Y.; Aoki, T.; Noda, H.; Yamada, M.; Kaino, M.; Mochizuki, H. Nrf2 Activators Attenuate the Progression of Nonalcoholic Steatohepatitis-Related Fibrosis in a Dietary Rat Model. *Molecular Pharmacology* **2013,** 84, 62-70.

151. Bertrand, H. C.; Schaap, M.; Baird, L.; Georgakopoulos, N. D.; Fowkes, A.; Thiollier, C.; Kachi, H.; Dinkova-Kostova, A. T.; Wells, G. Design, Synthesis, and Evaluation of Triazole Derivatives That Induce Nrf2 Dependent Gene Products and Inhibit the Keap1-Nrf2 Protein-Protein Interaction. *Journal of Medicinal Chemistry* **2015,** 58, 7186-7194.

152. Davies, T. G.; Wixted, W. E.; Coyle, J. E.; Griffiths-Jones, C.; Hearn, K.; McMenamin, R.; Norton, D.; Rich, S. J.; Richardson, C.; Saxty, G.; Willems, H. M. G.; Woolford, A. J. A.; Cottom, J. E.; Kou, J. P.; Yonchuk, J. G.; Feldser, H. G.; Sanchez, Y.; Foley, J. P.; Bolognese, B. J.; Logan, G.; Podolin, P. L.; Yan,

H. X.; Callahan, J. F.; Heightman, T. D.; Kerns, J. K. Monoacidic Inhibitors of the Kelch-like ECH-Associated Protein 1: Nuclear Factor Erythroid 2-Related Factor 2 (KEAP1:NRF2) Protein Protein Interaction with High Cell Potency Identified by Fragment-Based Discovery. *Journal of Medicinal Chemistry* **2016,** 59, 3991-4006.

153. Valkov, E.; Sharpe, T.; Marsh, M.; Greive, S.; Hyvoenen, M. Targeting Protein-Protein Interactions and Fragment-Based Drug Discovery. *Fragment-Based Drug Discovery and X-Ray Crystallography* **2012,** 317, 145-179.

154. Murray, C. W.; Rees, D. C. The rise of fragment-based drug discovery. *Nature Chemistry* **2009,** 1, 187-192.

155. Macarron, R.; Banks, M. N.; Bojanic, D.; Burns, D. J.; Cirovic, D. A.; Garyantes, T.; Green, D. V. S.; Hertzberg, R. P.; Janzen, W. P.; Paslay, J. W.; Schopfer, U.; Sittampalam, G. S. Impact of high-throughput screening in biomedical research. *Nature Reviews Drug Discovery* **2011,** 10, 188-195.

156. Veber, D. F.; Johnson, S. R.; Cheng, H. Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular properties that influence the oral bioavailability of drug candidates. *Journal of Medicinal Chemistry* **2002,** 45, 2615-2623.

157. Clark, D. E.; Pickett, S. D. Computational methods for the prediction of 'drug-likeness'. *Drug Discovery Today* **2000,** 5, 49-58.

158. Rees, D.; Congreve, M.; Murray, C.; Carr, R. Fragment-based lead discovery. *Nature Reviews Drug Discovery* **2004,** 3, 660-672.

159. Irwin, J. J.; Duan, D.; Torosyan, H.; Doak, A. K.; Ziebart, K. T.; Sterling, T.; Tumanian, G.; Shoichet, B. K. An Aggregation Advisor for Ligand Discovery. *Journal of Medicinal Chemistry* **2015,** 58, 7076-7087.

160. Cramer, J.; Schiebel, J.; Wulsdorf, T.; Grohe, K.; Najbauer, E. E.; Ehrmann, F. R.; Radeva, N.; Zitzer, N.; Linne, U.; Linser, R.; Heine, A.; Klebe, G. A False-Positive Screening Hit in Fragment-Based Lead Discovery: Watch out for the Red Herring. *Angewandte Chemie-International Edition* **2017,** 56, 1908-1913.

161. Oueis, E.; Sabot, C.; Renard, P. Y. New insights into the kinetic target-guided synthesis of protein ligands. *Chemical Communications* **2015,** 51, 12158-12169.

162. Erlanson, D. A.; Fesik, S. W.; Hubbard, R. E.; Jahnke, W.; Jhoti, H. Twenty years on: the impact of fragments on drug discovery. *Nature Reviews Drug Discovery* **2016,** 15, 605-619.

163. Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A rule of three for fragment-based lead discovery? *Drug Discovery Today* **2003,** 8, 876-877.

164. Koster, H.; Craan, T.; Brass, S.; Herhaus, C.; Zentgraf, M.; Neumann, L.; Heine, A.; Klebe, G. A Small Nonrule of 3 Compatible Fragment Library Provides High Hit Rate of Endothiapepsin Crystal Structures with Various Fragment Chemotypes. *Journal of Medicinal Chemistry* **2011,** 54, 7784-7796.

165. Hann, M. M.; Leach, A. R.; Harper, G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *Journal of Chemical Information and Computer Sciences* **2001,** 41, 856-864.

166. Cang, S. D.; Iragavarapu, C.; Savooji, J.; Song, Y. P.; Liu, D. L. ABT-199 (venetoclax) and BCL-2 inhibitors in clinical development. *Journal of Hematology & Oncology* **2015,** 8.

167. Bollag, G.; Tsai, J.; Zhang, J. Z.; Zhang, C.; Ibrahim, P.; Nolop, K.; Hirth, P. Vemurafenib: the first drug approved for BRAF-mutant cancer. *Nature Reviews Drug Discovery* **2012,** 11, 873-886.

168. Crick, F. Central Dogma of Molecular Biology. *Nature* **1970,** 227, 561-&.

169. Sharpless, K. B.; Manetsch, R. In situ click chemistry: a powerful means for lead discovery. *Expert Opinion on Drug Discovery* **2006,** 1, 525-538.

170. Krasinski, A.; Radic, Z.; Manetsch, R.; Raushel, J.; Taylor, P.; Sharpless, K. B.; Kolb, H. C. In situ selection of lead compounds by click chemistry: Target-guided optimization of acetylcholinesterase inhibitors. *Journal of the American Chemical Society* **2005,** 127, 6686-6692.

171. Rideout, D. Self-assembling Cytotoxins. *Science* **1986,** 233, 561-563.

172. Lehn, J. M. Dynamic combinatorial chemistry and virtual combinatorial libraries. *Chemistry-a European Journal* **1999,** 5, 2455-2463.

173. Corbett, P. T.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J. L.; Sanders, J. K. M.; Otto, S. Dynamic combinatorial chemistry. *Chemical Reviews* **2006,** 106, 3652-3711.

174. Mondal, M.; Hirsch, A. K. H. Dynamic combinatorial chemistry: a tool to facilitate the identification of inhibitors for protein targets. *Chemical Society Reviews* **2015,** 44, 2455-2488.

175. Hu, X. D.; Manetsch, R. Kinetic target-guided synthesis. *Chemical Society Reviews* **2010,** 39, 1316-1324.

176. Bosc, D.; Jakhlal, J.; Deprez, B.; Deprez-Poulain, R. Kinetic target-guided synthesis in drug discovery and chemical biology: a comprehensive facts and figures survey. *Future Medicinal Chemistry* **2016,** 8, 381-404.

177. Nguyen, R.; Huc, I. Using an enzyme's active site to template inhibitors. *Angewandte Chemie-International Edition* **2001,** 40, 1774-1776.

178. Gelin, M.; Poncet-Montange, G.; Assairi, L.; Morellato, L.; Huteau, V.; Dugue, L.; Dussurget, O.; Pochet, S.; Labesse, G. Screening and In Situ Synthesis Using Crystals of a NAD Kinase Lead to a Potent Antistaphylococcal Compound. *Structure* **2012,** 20, 1107-1117.

179. Lewis, W. G.; Green, L. G.; Grynszpan, F.; Radic, Z.; Carlier, P. R.; Taylor, P.; Finn, M. G.; Sharpless, K. B. Click chemistry in situ: Acetylcholinesterase as a reaction vessel for the selective assembly of a femtomolar inhibitor from an array of building blocks. *Angewandte Chemie* **2002,** 114, 1095-1101.

180. Mock, W. L.; Irra, T. A.; Wepsiec, J. P.; Manimaran, T. L. Cycloaddition Induced By Cucurbituril - A Case Of Pauling Principle Catalysis. *Journal of Organic Chemistry* **1983,** 48, 3619-3620.

181. Hu, X. D.; Sun, J. Z.; Wang, H. G.; Manetsch, R. Bcl-X-L-templated assembly of its own protein-protein interaction modulator from fragments decorated with thio acids and sulfonyl azides. *Journal of the American Chemical Society* **2008,** 130, 13820-13821.

182. Kulkarni, S. S.; Hu, X. D.; Doi, K.; Wang, H. G.; Manetsch, R. Screening of Protein-Protein Interaction Modulators via Sulfo-Click Kinetic Target-Guided Synthesis. *ACS Chemical Biology* **2011,** 6, 724-732.

183. Willand, N.; Desroses, M.; Toto, P.; Dirie, B.; Lens, Z.; Villeret, V.; Rucktooa, P.; Locht, C.; Baulard, A.; Deprez, B. Exploring Drug Target Flexibility Using in Situ Click Chemistry: Application to a Mycobacterial Transcriptional Regulator. *ACS Chemical Biology* **2010,** 5, 1007-1013.

184. Manetsch, R.; Krasinski, A.; Radic, Z.; Raushel, J.; Taylor, P.; Sharpless, K. B.; Kolb, H. C. In situ click chemistry: Enzyme inhibitors made to their own specifications. *Journal of the American Chemical Society* **2004,** 126, 12809-12818.

185. Tieu, W.; da Costa, T. P. S.; Yap, M. Y.; Keeling, K. L.; Wilce, M. C. J.; Wallace, J. C.; Booker, G. W.; Polyak, S. W.; Abell, A. D. Optimising in situ click chemistry: the screening and identification of biotin protein ligase inhibitors. *Chemical Science* **2013,** 4, 3533-3537.

186. Wischeler, J. S.; Sun, D.; Sandner, N. U.; Linne, U.; Heine, A.; Koert, U.; Klebe, G. Stereo- and Regioselective Azide/Alkyne Cycloadditions in Carbonic Anhydrase II via Tethering, Monitored by Crystallography and Mass Spectrometry. *Chemistry-a European Journal* **2011,** 17, 5842-5851.

187. Whiting, M.; Muldoon, J.; Lin, Y. C.; Silverman, S. M.; Lindstrom, W.; Olson, A. J.; Kolb, H. C.; Finn, M. G.; Sharpless, K. B.; Elder, J. H.; Fokin, V. V. Inhibitors of HIV-1 protease by using in situ click chemistry. *Angewandte Chemie-International Edition* **2006,** 45, 1435-1439.

188. Dimitrios-Georgakopoulos, N. The development of direct inhibitors of the Keap1-Nrf2 protein-protein interaction as modulators of mitochondrial function and quality control. PhD thesis. UCL School of Pharmacy, London, 2017.

189. Kosynkin, D. V.; Tour, J. M. Phenylene ethynylene diazonium salts as potential self-assembling molecular devices. *Organic Letters* **2001,** 3, 993-995.

190. Zhang, W. J.; Li, Z.; Zhou, M.; Wu, F.; Hou, X. Y.; Luo, H.; Liu, H.; Han, X.; Yan, G. Y.; Ding, Z. Y.; Li, R. Synthesis and biological evaluation of 4-(1,2,3-triazol-1-yl)coumarin derivatives as potential antitumor agents. *Bioorganic & Medicinal Chemistry Letters* **2014,** 24, 799-807.

191. Patrick, G. L. Protein Structure and Function. In *An Introduction to Medicinal Chemistry*, Fifth ed.; Patrick, G. L., Ed. OUP Higher Education Divison: 2013; pp 17-29.

192. Scott, D. E.; Spry, C.; Abell, C. Differential Scanning Fluorimetry as Part of a Biophysical Screening Cascade. In *Fragment-based Drug Discovery Lessons and Outlook*, Erlanson, D. A.; Jahnke, W., Eds. Wiley-VCH Verlag GmbH & Co. KGaA: Germany, 2016; pp 139-172.

193. Niesen, F. H.; Berglund, H.; Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols* **2007,** 2, 2212-2221.

194. Kranz, J. K.; Schalk-Hihi, C. Protein Thermal Shifts to Identify Low Molecular Weight Fragments. *Fragment-Based Drug Design: Tools, Practical Approaches, and Examples* **2011,** 493, 277-298.

195. Jameson, D. M.; Ross, J. A. Fluorescence Polarization/Anisotropy in Diagnostics and Imaging. *Chemical Reviews* **2010,** 110, 2685-2708.

196. Grimster, N. P.; Stump, B.; Fotsing, J. R.; Weide, T.; Talley, T. T.; Yamauchi, J. G.; Nemecz, A.; Kim, C.; Ho, K. Y.; Sharpless, K. B.; Taylor, P.; Fokin, V. V. Generation of Candidate Ligands for Nicotinic Acetylcholine Receptors via in situ Click Chemistry with a Soluble Acetylcholine Binding Protein Template. *Journal of the American Chemical Society* **2012,** 134, 6732-6740.

197. Tan, C. J.; Chua, M. G.; Ker, K. H.; Tong, Y. W. Preparation of bovine serum albumin surface-imprinted submicrometer particles with magnetic susceptibility through core-shell miniemulsion polymerization. *Analytical Chemistry* **2008,** 80, 683-692.

198. Deprez-Poulain, R.; Hennuyer, N.; Bosc, D.; Liang, W. G. G.; Enee, E.; Marechal, X.; Charton, J.; Totobenazara, J.; Berte, G.; Jahklal, J.; Verdelet, T.; Dumont, J.; Dassonneville, S.; Woitrain, E.; Gauriot, M.; Paquet, C.; Duplan, I.; Hermant, P.; Cantrelle, F. X.; Sevin, E.; Culot, M.; Landry, V.; Herledan, A.; Piveteau, C.; Lippens, G.; Leroux, F.; Tang, W. J.; van Endert, P.; Staels, B.; Deprez, B. Catalytic site inhibition of insulin-degrading enzyme by a small molecule induces glucose intolerance in mice. *Nature Communications* **2015,** 6, 8250-8263.

199. Barcellos, G. B.; Pauli, I.; Caceres, R. A.; Timmers, L.; Dias, R.; de Azevedo, W. F. Molecular Modeling as a Tool for Drug Discovery. *Current Drug Targets* **2008,** 9, 1084-1091.

200. Durrant, J. D.; McCammon, J. A. AutoClickChem: Click Chemistry in Silico. *Plos Computational Biology* **2012,** 8, 1-7.

201. Trott, O.; Olson, A. J. Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *Journal of Computational Chemistry* **2010,** 31, 455-461.

202. Rajagopalan, S. Protein Kinase Inhibitors. Patent Application WO2014/162039 A1. 2014.

203. Huisgen, R.; Seidl, H.; Bruning, I. 1,3-Dipolar Cycloadditions .49. Kinetics and Mechanism of Nitrone Additions to Unsaturated Compounds. *Chemische Berichte-Recueil* **1969,** 102, 1102-1116.

204. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. A stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective ligation of azides and terminal alkynes. *Angewandte Chemie-International Edition* **2002,** 41, 2596-2599.

205. Tornoe, C. W.; Christensen, C.; Meldal, M. Peptidotriazoles on solid phase: 1,2,3 -triazoles by regiospecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *Journal of Organic Chemistry* **2002,** 67, 3057-3064.

206. Worrell, B. T.; Malik, J. A.; Fokin, V. V. Direct Evidence of a Dinuclear Copper Intermediate in Cu(I)-Catalyzed Azide-Alkyne Cycloadditions. *Science* **2013,** 340, 457-460.

207. Boren, B. C.; Narayan, S.; Rasmussen, L. K.; Zhang, L.; Zhao, H. T.; Lin, Z. Y.; Jia, G. C.; Fokin, V. V. Ruthenium-catalyzed azide-alkyne cycloaddition: Scope and mechanism. *Journal of the American Chemical Society* **2008,** 130, 8923-8930.

208. Zhang, L.; Chen, X. G.; Xue, P.; Sun, H. H. Y.; Williams, I. D.; Sharpless, K. B.; Fokin, V. V.; Jia, G. C. Ruthenium-catalyzed cycloaddition of alkynes and organic azides. *Journal of the American Chemical Society* **2005,** 127, 15998-15999.

209. Smith, C. D.; Greaney, M. F. Zinc Mediated Azide-Alkyne Ligation to 1,5-and 1,4,5-Substituted 1,2,3-Triazoles. *Organic Letters* **2013,** 15, 4826-4829.

210. Kwok, S. W.; Fotsing, J. R.; Fraser, R. J.; Rodionov, V. O.; Fokin, V. V. Transition-Metal-Free Catalytic Synthesis of 1,5-Diaryl-1,2,3-triazoles. *Organic Letters* **2010,** 12, 4217-4219.

211. Krasinski, A.; Fokin, V. V.; Sharpless, K. B. Direct synthesis of 1,5-disubstituted-4-magnesio-1,2,3-triazoles, revisited. *Organic Letters* **2004,** 6, 1237-1240.

212. Hancock, R.; Bertrand, H.; Naz, S.; Tsujita, T.; Hayes, J.; Wells, G. Design, synthesis and biological evaluation of peptide inhibitors of the Keap1-Nrf2 protein-protein interaction. *Cancer Research* **2011,** 71.

213. Molyneaux, C. A. Developing novel antibacterials targeting the glyoxalase system using kinetic-target guided synthesis. PhD thesis. UCL School of Pharmacy, London, 2017.

214. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *Journal of Computational Chemistry* **2009,** 30, 2785-2791.

215. Vainio, M. J.; Johnson, M. S. Generating conformer ensembles using a multiobjective genetic algorithm. *Journal of Chemical Information and Modeling* **2007,** 47, 2462-2474.

216. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *Journal of Cheminformatics* **2011,** 3, 1-16.

217. MOPAC2012, James J. P. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA, HTTP://OpenMOPAC.net (2012).

218. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **2004,** 25, 1605-1612.

219. Chauhan, A.; Paladhi, S.; Debnath, M.; Dash, J. Selective recognition of c-MYC G-quadruplex DNA using prolinamide derivatives. *Organic & Biomolecular Chemistry* **2016,** 14, 5761-5767.

220. Bouchain, G.; Leit, S.; Frechette, S.; Abou Khalil, E.; Lavoie, R.; Moradei, O.; Woo, S. H.; Fournel, M.; Yan, P. T.; Kalita, A.; Trachy-Bourget, M. C.; Beaulieu, C.; Li, Z. M.; Robert, M. F.; MacLeod, A. R.; Besterman, J. M.; Delorme, D. Development of potential antitumor agents. Synthesis and biological evaluation of a new set of sulfonamide derivatives as histone deacetylase inhibitors. *Journal of Medicinal Chemistry* **2003,** 46, 820-830.

221. Tiew, K. C.; Dou, D. F.; Teramoto, T.; Lai, H. G.; Alliston, K. R.; Lushington, G. H.; Padmanabhan, R.; Groutas, W. C. Inhibition of Dengue virus and West Nile virus proteases by click chemistry-derived benz d isothiazol-3(2H)-one derivatives. *Bioorganic & Medicinal Chemistry* **2012,** 20, 1213-1221.

222. Magilavy, D. Treatment for Vitilogo. U.S. Patent US2014073659 A1. 2014.

223. Tahtouh, T.; Elkins, J. M.; Filippakopoulos, P.; Soundararajan, M.; Burgy, G.; Durieu, E.; Cochet, C.; Schmid, R. S.; Lo, D. C.; Dehommel, F.; Oberhozer, A. E.; Pearl, L. H.; Carreaux, F.; Bazureau, J. P.; Knapp, S.; Meijer, L. Selectivity, Cocrystal Structures, and Neuroprotective Properties of Leucettines, a Family of Protein Kinase Inhibitors Derived from the Marine Sponge Alkaloid Leucettamine B. *Journal of Medicinal Chemistry* **2012,** 55, 9312-9330.

224. Muller, A.; Lindhorst, T. K. Synthesis of Hetero-bifunctional Azobenzene Glycoconjugates for Bioorthogonal Cross-Linking of Proteins. *European Journal of Organic Chemistry* **2016**, 1669-1672.

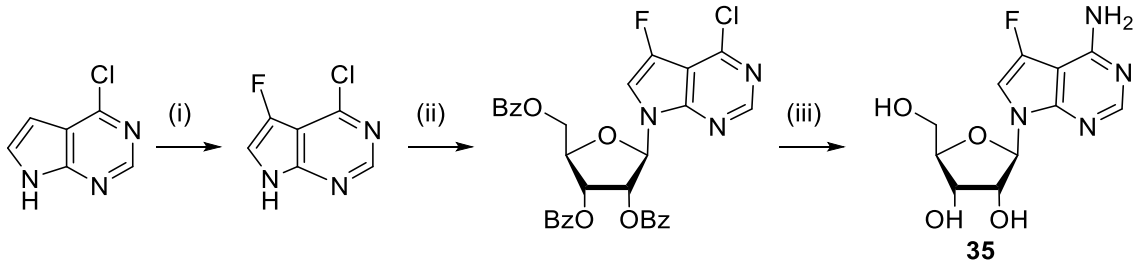
225. Buchowiecki, P. K. Epigenetic Drug Targets: Design and Synthesis of New Inhibitors of the Histone Methyltransferase DOT1L. Masters thesis. UCL School of Pharmacy, London, 2015.

226. Zhang, Y.; Ye, W.; Wang, H. S.; Schneller, S. W. Carbocyclic 4-Deazaformycins. *Synthesis* **2012,** 44, 723-730.

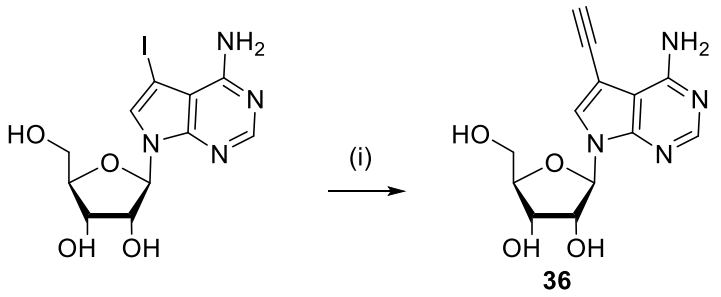
Appendices

Appendix 1: Summary of synthetic routes to compounds 35-40

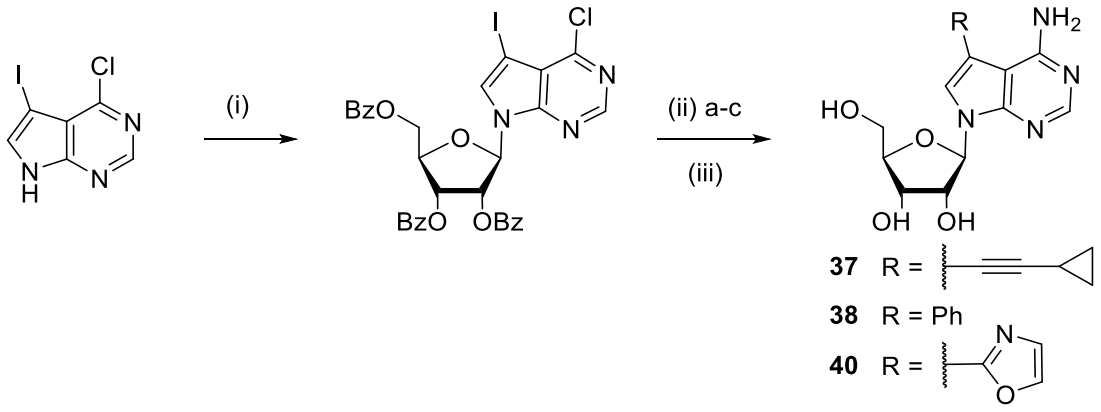
All synthetic procedures were carried out by Dr Elliott Bayle.



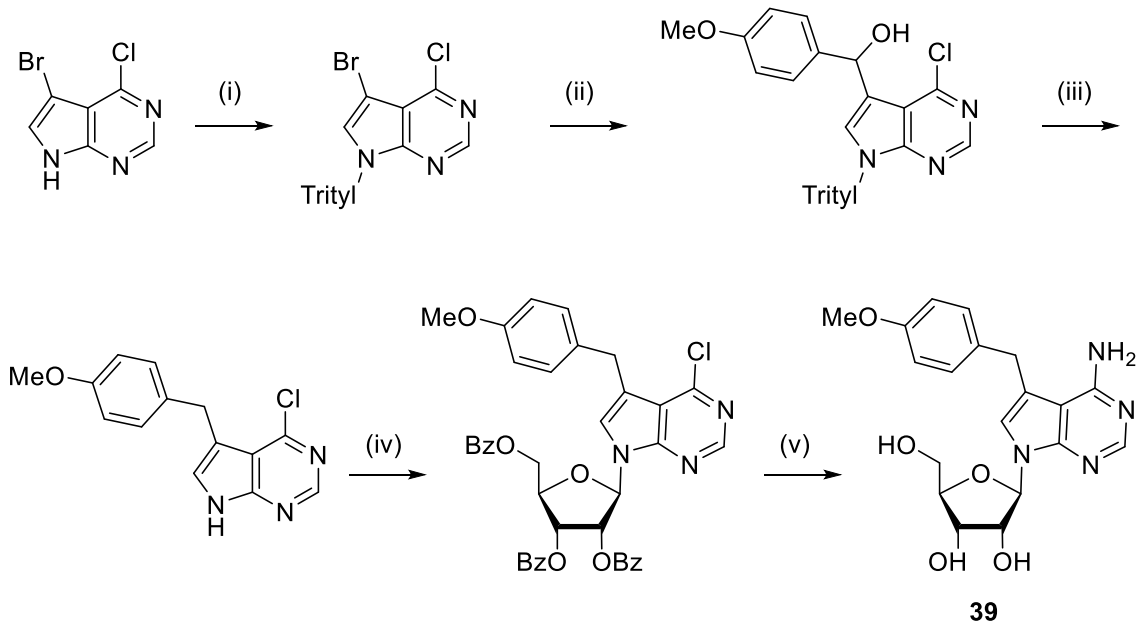
Scheme 10.1 Synthetic route to compound **35**. *Reagents and conditions*: (i) 1-(Chloromethyl)-4-fluoro-1,4-diazoniabicyclo-[2.2.2]octane bis(tetrafluoro-borate), MeCN-AcOH (5:1), 70 °C, 14 hr (14 %); (ii) 1-O-Acetyl-2,3,5-tri-O-benzoyl-β-D-ribofuranose, *N*,*O*-bis(trimethylsilyl)acetamide, TMSOTf, MeCN, rt to 80 °C, 3 hr, (63 %); (iii) 7 M NH_3 in MeOH, 120 °C, 2 h, sealed vessel (59 %).



Scheme 10.2 Synthetic route to compound **36**. *Reagents and conditions*: (i) Trimethylsilylacetylene, CuI, $Pd(PPh_3)_2Cl_2$, Et_3N, DMF, rt, 14 hr *then* K_2CO_3, MeOH, rt, 4 hr, (68 %).



Scheme 10.3 Synthetic route to compound **37**, **38** and **40**. *Reagents and conditions*: (i) 1-O-Acetyl-2,3,5-tri-O-benzoyl-β-D-ribofuranose, *N,O*-bis(trimethylsilyl)-acetamide TMSOTf, MeCN, rt to 80 °C (52%); (ii.a) **37**: Cyclopropylacetylene, CuI, $Pd(PPh_3)_2Cl_2$, DMF-Et_3N (3:1), rt, 14 hr (74%); (ii.b) **38**: $PhSnMe_3$, CuI, $Pd(PPh_3)_4$, DMF, 90 °C (6%). (ii.c) **40**: 2-(Tri-*n*-butyl)stannyloxazole, $Pd(PPh_3)_4$, DMF, 90 °C (25%); (iii) 7 M NH_3 in MeOH, 120 °C, 2 h, sealed vessel, (**37**: 37%, **38**: 70%, **40**: 81%).



Scheme 10.4 Synthetic route to compound **39**. *Reagents and conditions*: (i) Triphenylmethyl chloride, NaH (60 % w/w), THF, 0 °C to rt, 16 hr (43%); (ii) *n*-BuLI, *p*-anisaldehyde, THF, 0 °C to rt, (51%); (iii) Chlorodiphenylsilane, $InCl_3$, CH_2ClCH_2Cl, rt, 6 hr *then* trifluoroacetic acetic, 0 °C to rt, 14 hr, (87%); (iv) 1-O-Acetyl-2,3,5-tri-O-benzoyl-β-D-ribofuranose, *N*,*O*-bis(trimethylsilyl)acetamide, TMSOTf, MeCN, rt to 80 °C, 3 hr, (63%); (v) 7 M NH_3 in MeOH, 120 °C, sealed vessel, 2 h, (59%).

Appendix 2: Synthetic approach to CN-SGC946

To isolate CN-SGC946 (**80**), three different strategies were employed: a modular approach where each fragment is separately assembled and subsequently coupled; a step-wise approach where the target compound is built step by step from the nucleoside toyocamycin (**29**); and synthesis of SGC-946 (**16**) followed by Pd-catalysed insertion of 5-CN. For each synthetic route the isolated products were characterised by proton NMR before progression to the next step, however, as indicated previously, full experimental data is not presented for these structures. The reactions were performed on a small scale to explore the viability of each synthetic route with the intention that, after identification of a successful route, complete characterisation would be acquired upon subsequent scale-up and re-synthesis.

Strategy 1 (a): A modular approach was adopted initially as it was viewed as a more attractive method than the alternative step-wise approach because it would limit the quantity of nucleoside fragment required, which is complicated and costly to synthesise. It was envisaged toyocamycin, activated at 5' (Y), could be coupled to the urea fragment **216** (Figure 10.2).

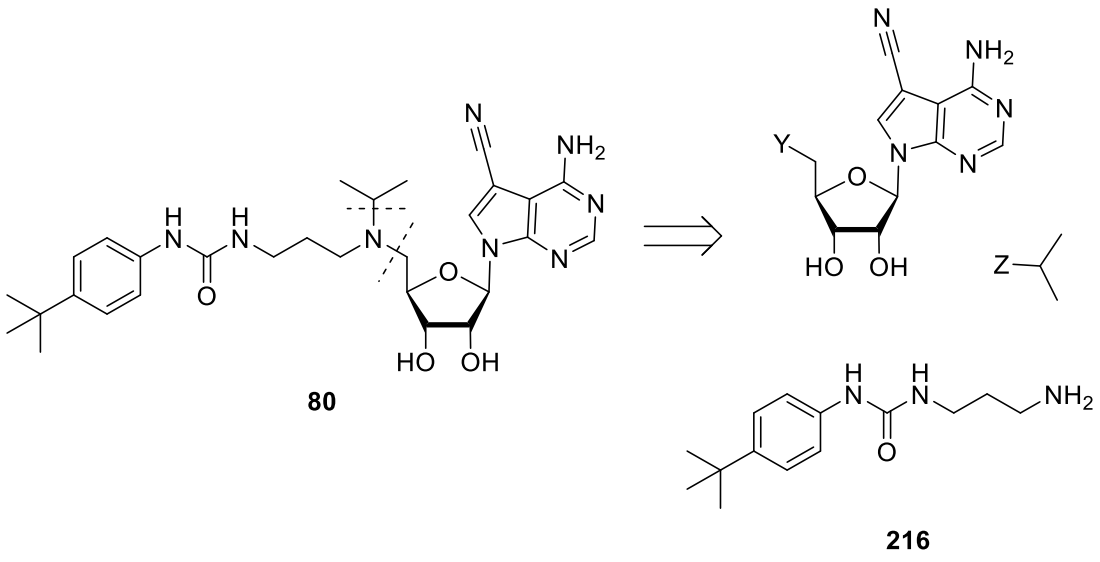
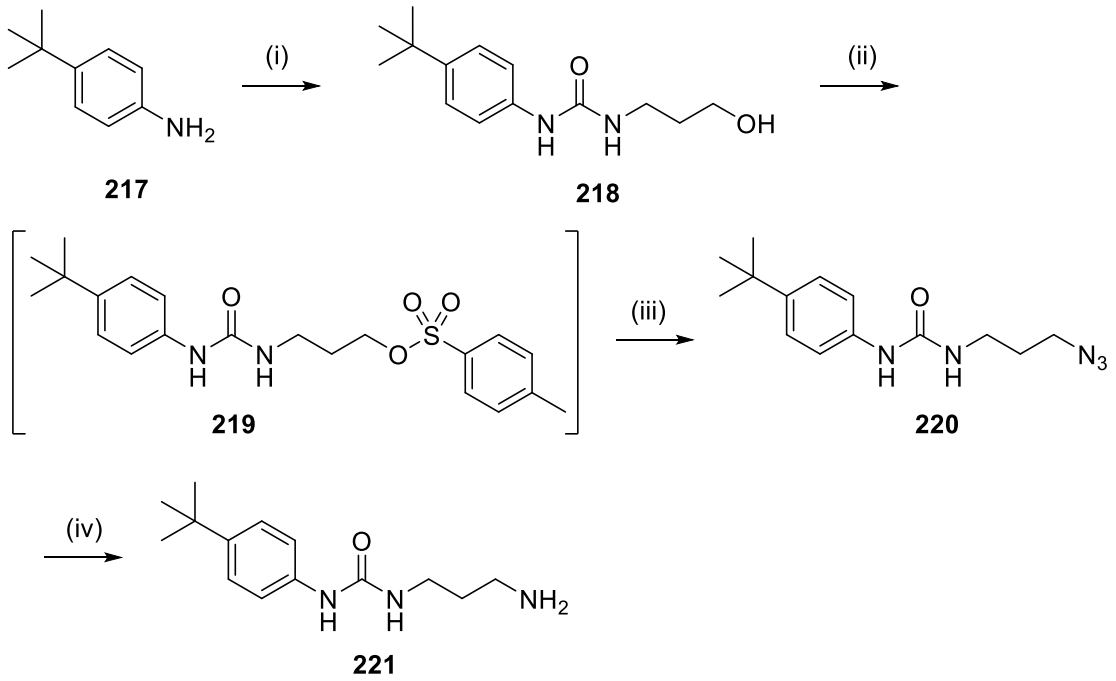


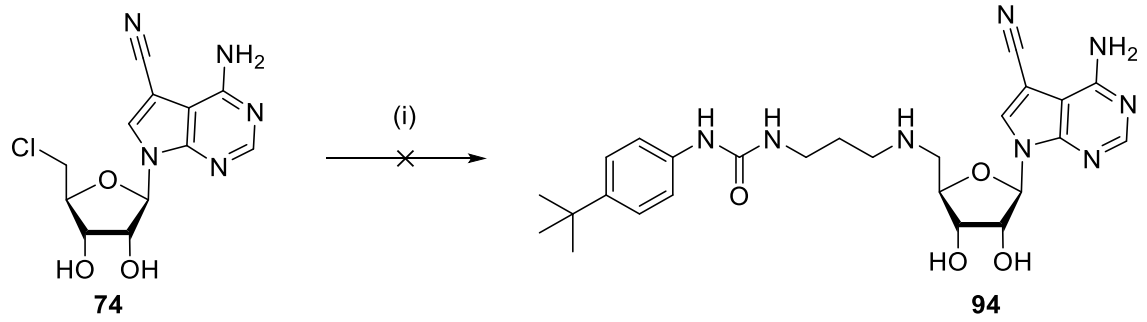
Figure 10.2 Retrosynthesis of the modular approach proposed to obtain **80**. Where Y and Z are appropriate leaving groups, for example -Cl, -I, -OTs etc.

First, isolation of the urea fragment **216** was achieved as set out in Scheme 10.5; intermediate **218** was isolated in a moderate yield from the commercially available aniline **217** following a literature procedure (Scheme 10.5, step (i)).[56] The primary amine was installed by introduction of a tosyl group followed by substitution to an azide and subsequent reduction (Scheme 10.5, step (ii)-(iv)) resulting in isolation of urea **221.**

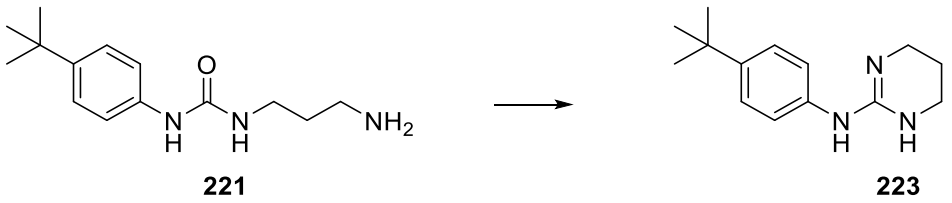
With an activated 5'-chlororibose **74** analogue of toyocamycin already in hand from the synthesis of CNSAH (**48**), this approach was attempted by coupling **74** to urea fragment **221** to yield intermediate **222,** however, this reaction was unsuccessful (Scheme 10.6). It was hypothesised the primary amine in urea fragment **221** was cyclising to form **223**, consequently hindering the reaction (Figure 10.3).



Scheme 10.5 Synthesis of urea **221**. *Reagents and conditions*: (i) 1,1'-carbonyldiimazole (CDI), 3-aminopropanol, THF, 0 °C to rt, 2 hrs (90%); (ii) Tosyl chloride, pyridine, DCM, 0 °C to rt, 6 hrs. Taken forward as crude (iii) NaN_3, DMF, rt, 5 hrs (73%); (iv) Pd/C, H_2 (balloon), EtOH, rt, 16 hours (55%).



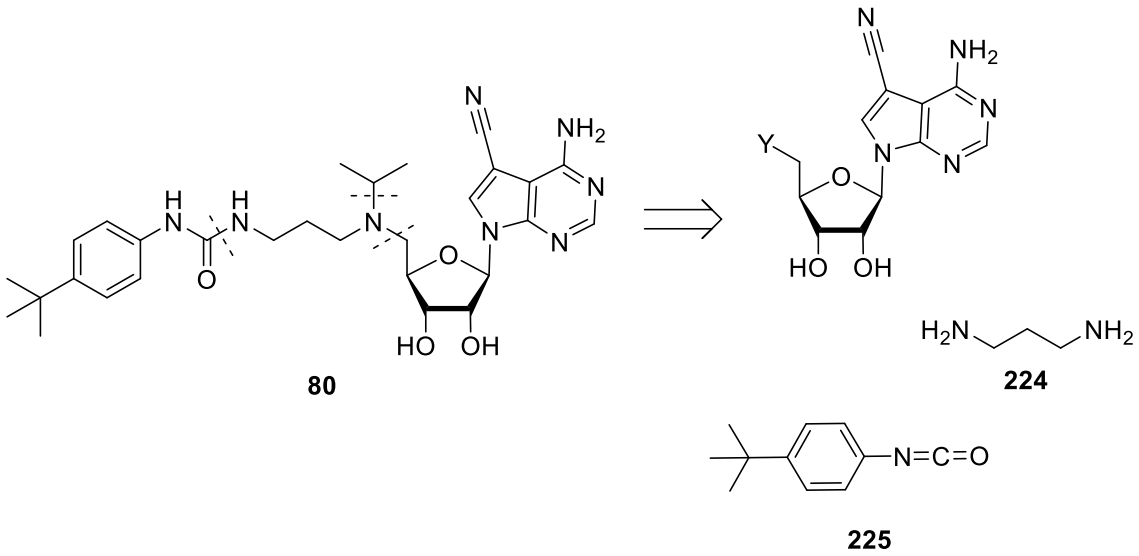
Scheme 10.6 A modular approach to isolate intermediate **222** towards the synthesis of **80**. *Reagents and conditions:* (i) Compound **221**, KI, K_2CO_3, DMF, 80 °C, 5 hrs. No reaction observed.



221 **223**

Figure 10.3 Hypothesised cyclisation of **223** forming a guanidine moiety.

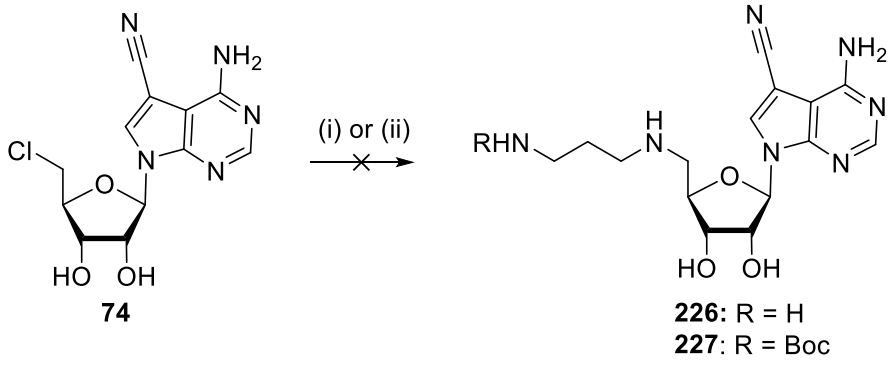
Strategy 2: To circumvent potential cyclisation of urea fragment **221**, a stepwise approach was employed. This encompassed addition of the urea portion of the inhibitor onto the activated toyocamycin fragment in two stages, using two commercially available fragments, **224** and **225** (Figure 10.4).



80 **224** **225**

Figure 10.4 Retrosynthesis of a step wise approach to obtain **80.** Both **224** [CAS: 109-76-2] and **225** [CAS: 1943-67-5] were purchased from Sigma Aldrich. Where Y is an appropriate leaving group, for example -Cl, -I, -OTs etc.

The first step involved addition of the di-amine **224**, this was attempted using 1,3-diaminopropane and employing conditions published by Yu *et al.* (which had been employed to isolate an analogous substrate with a 5-H substitution) (Scheme 10.7).[56] However, multiple products were observed and the desired intermediate, **226**, was not isolated. It was unclear why the reaction did not proceed as the published analogue. Furthermore, difficulties arose with removal of 1,3-diaminopropane which was used as both reagent and solvent. It was theorised using an *N*-Boc protected analogue of 1,3-diaminopropane would overcome both of these issues, however no reaction was observed.



Scheme 10.7 Step-wise approach towards isolation of **80**. *Reagents and conditions*: (i) 1,3-diaminopropane (anhydrous), 65 °C, 16 hrs (ii) *tert*-butyl-(3-aminopropyl)carbamate, THF, rt, 16 hrs. No reaction observed

Strategy 1 (b): In parallel to strategy 2, an alternative modular approach was employed, also described by Yu *et al*., which utilises a masked aldehyde **228** to install the urea part using reductive amination conditions.[56] This would circumvent potential cyclisation of urea fragment **221** as the amine would be installed onto the nucleoside fragment (5-NH_2), thus requiring synthesis of 5'-NH_2 **229** and as well as the masked aldehyde **228** (Figure 10.5).

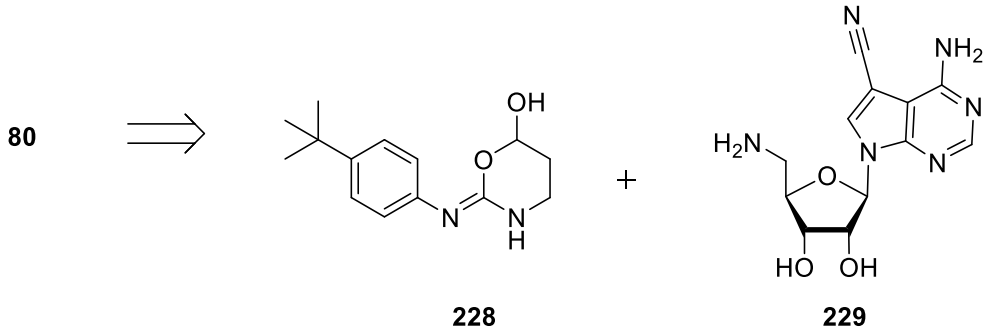
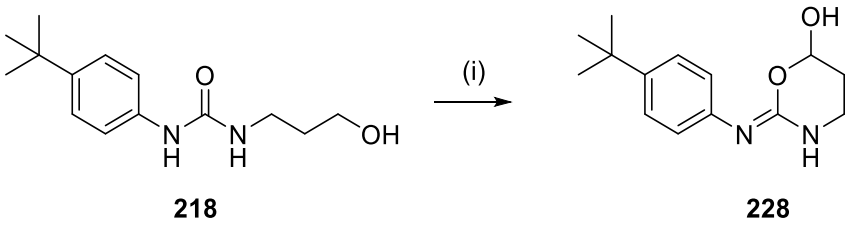


Figure 10.5 Retrosynthesis of an alternative modular approach to isolate **80** using reductive amination conditions published by Yu *et al*.[56]

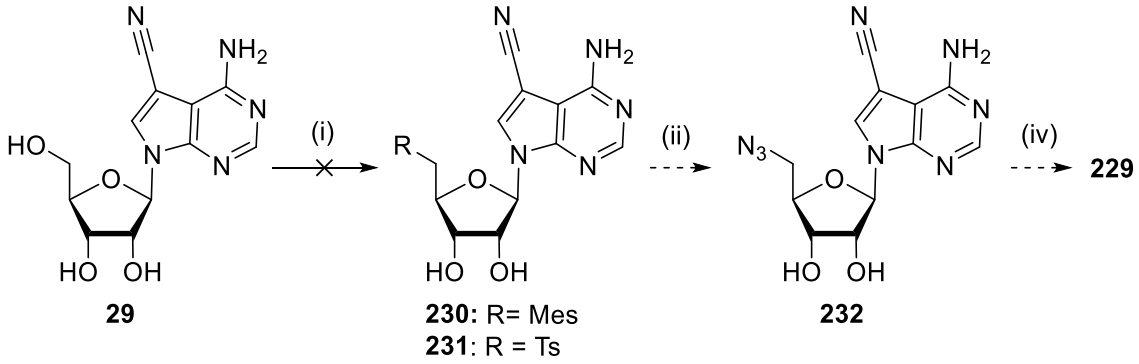
Compound **228** was isolated following published conditions using the intermediate compound previously synthesised **218** (Scheme 10.5, step (i)) in a moderate yield (Scheme 10.8), attention was then turned to isolation of 5-NH_2, **229**.



218 **228**

Scheme 10.8 Synthesis of compound **100**. *Reagents and conditions*: (i) Oxalyl chloride ($(COCl)_2$), DMSO, Et_3N, anhydrous DCM, -78 °C to rt (55%).

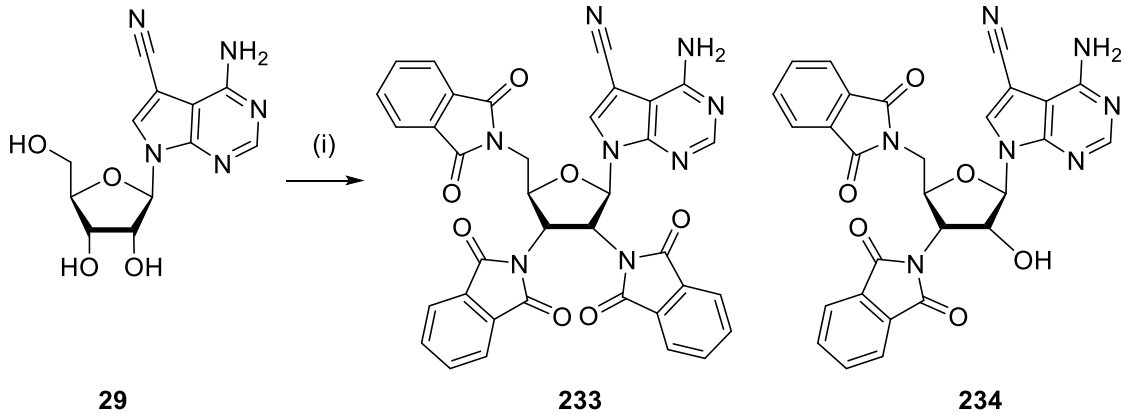
It was envisaged 5'-NH_2 **229** could be obtained through activation of the 5'-OH in toyocamycin (**29**), followed by instalment of an azide via substitution using NaN_3 and finally reduction to the amine (Scheme 10.9). Activation of the 5'-OH was initially attempted via sulphonation using mesyl chloride to form 5'-OMes **230**, however multiple products were observed to form (Scheme 10.9, step (i)). The same route was attempted using tosyl chloride to form the more stable analogue 5'-OTos **231**, however again, multiple products were observed to form. It was hypothesised the reactions were not selective and multiple hydroxyl groups on the ribofuranse were reacting.



29 **230:** R= Mes **231**: R = Ts **232**

Scheme 10.9 Attempted synthetic route to isolate 5'-NH_2 (**229**) via instalment of 5-OMes **230** or 5-OTs **231**. *Reagents and conditions*: (i) MesCl or TosCl, 0 °C, 0.5–1 hr. No product isolated. (ii) Suggested conditions: NaN_3, DMF (iv) Pd/C, H_2, MeOH, rt.

To overcome this, an alternative route employing the Gabriel synthesis was attempted as it was envisaged this would selectively react with the primary alcohol (Scheme 10.10). However, it was apparent by NMR and LRMS that the reaction had not occurred exclusively on the 5'-OH but had resulted in the addition of multiple phthalimide groups.



Scheme 10.10 Approach to install 5'-NH2 via Gabriel synthesis. *Reagents and conditions*: (i) PPh_3, phthalimide, THF, -15 °C *then* DIAD, 0 °C to rt, 2.5 hrs. Two products with multiple phthalimide substitutions were observed by NMR, **233** and **234**.

Thus, to direct selective addition, the 2'-OH and 3'-OH on the ribofuranose were protected with an acetonide group using camphor sulphonic acid (CSA) (Scheme 10.11, step (i)). This allowed isolation of 5'-phthalimide **235** (Scheme 10.11, step (ii)) and subsequent deprotection using hydrazine resulted in isolation of 5'-NH_2 **236** (Scheme 10.11, step (iii))**.** The *N*-iPr moiety was installed using reductive amination with acetone to yield **237** (Scheme 10.11, step (vi)). Coupling of amine **236** to the masked aldehyde **228** using reductive amination conditions as published by Yu *et al* (Scheme 10.10, step (v)), did not result in isolation of the desired product **228**; although new nucleobase peaks were observed to form by NMR indicating formation of a novel product, the compound was lost on purification.

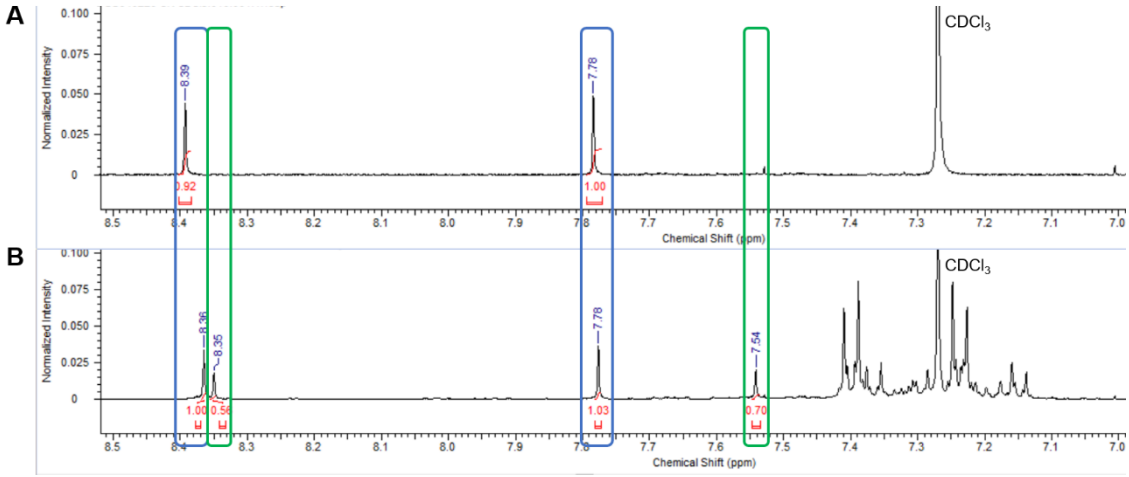
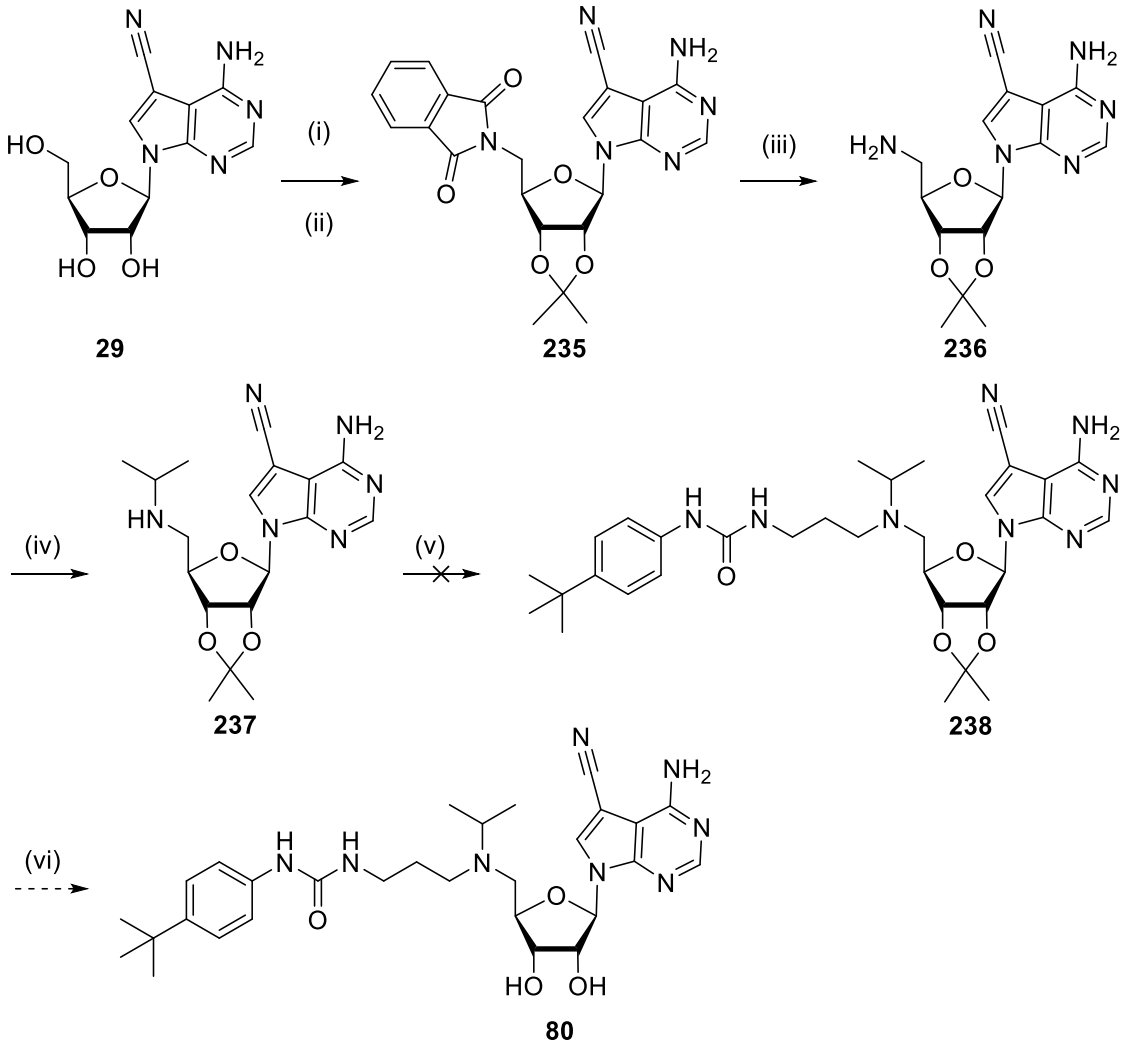
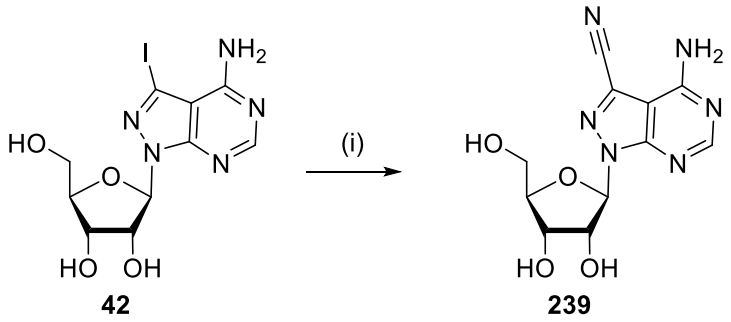


Figure 10.6 Comparison of the proton NMR of the crude reaction mixture of **110** (B) versus the starting material **237** (A). The formation of novel nucleobase peaks (highlighted in green) indicates the potential formation of the desired product **228**.



Scheme 10.11 Modular approach where the amine is installed on the nucleoside rather than the urea fragment. *Reagents and conditions*: (i) camphor sulphonic acid, 2,2-dimethoxyporpane, rt, 16 hrs (66%) (ii) PPh_3, phthalimide, THF, -15 °C *then* DIAD, 0 °C to rt, 2.5 hrs (46%); (iii) $H_2NNH_2.2H_2O$, MeOH, 12 hrs (18%); (iv) $Na(OAc)_3BH$, Acetone, AcOH, DCM. Taken through as crude; (v) Aldehyde **228**, $Na(OAc)_3BH$, AcOH, Intermediate **238** not isolated; (vi) Proposed conditions: 4 M HCl in dioxane, 0 °C, 7 hrs.

Strategy 3: In parallel to the above synthetic routes, an alternative approach was attempted which assessed the direct instalment of the 5-CN onto SGC-946, or a 5-iodo analogue by Pd-mediated cyanation. The approach was investigated using the pyrazolo[3,4-*d*]pyrimidine scaffold as a model system, **42**. Various metals, ligands and sources of cyanide were screened, resulting in the identification of preferred conditions as set out in Scheme 10.12.[225] It was hoped these conditions could be translated to the pyrrolo[2,3-*d*]pyrimidine scaffold and applied directly to the synthesis of **80**. However, this work was not performed because due to the publication of **81**. This work was carried out by Mr Peter K Buchowiecki, a masters student in the PVF group, while under supervision.

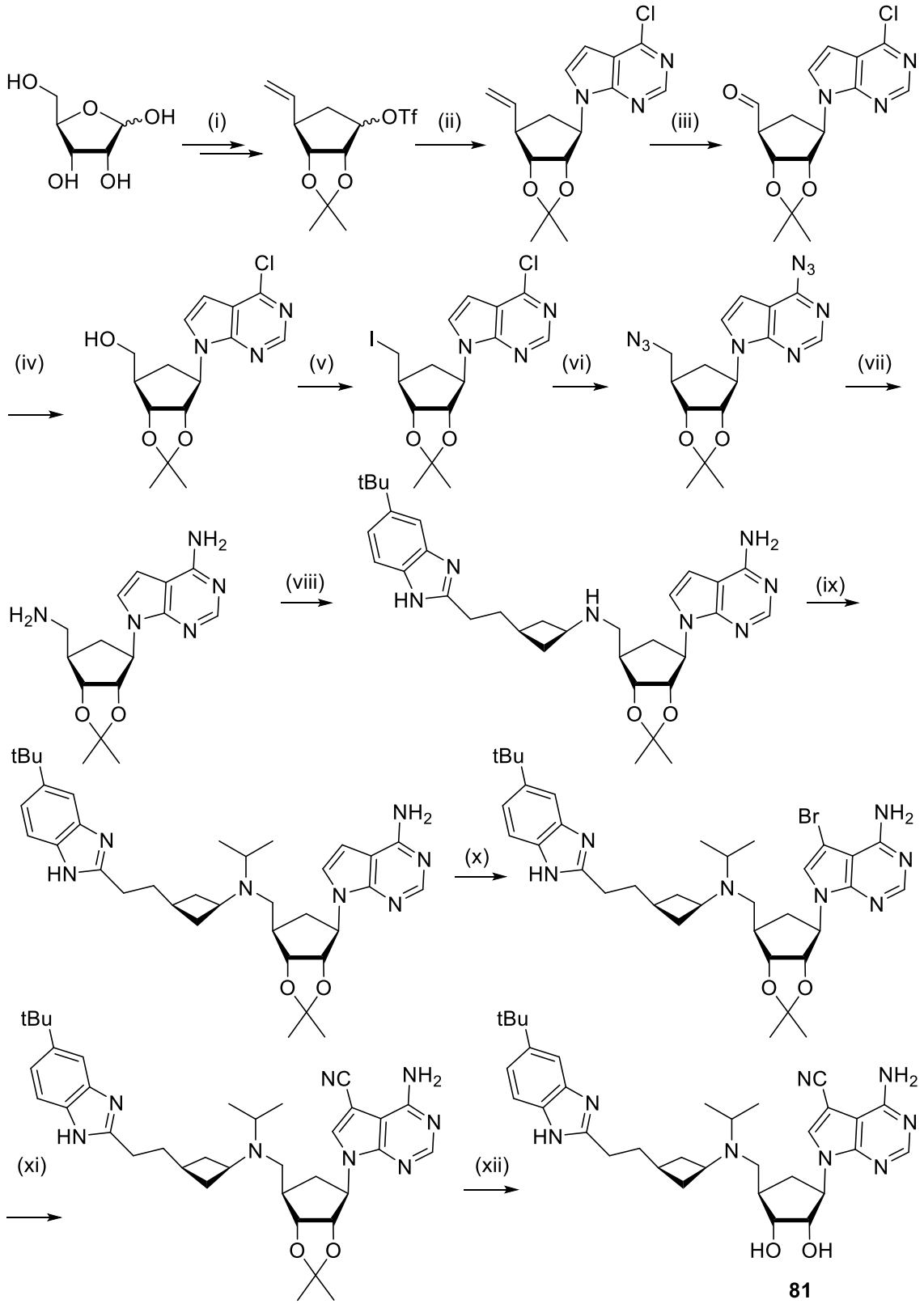


42 **239**

Scheme 10.12 Model system used to assess introduction of the nitrile moiety via palladium catalysed insertion. *Reagents and conditions:* (i) $Pd_2(dba)_3$, $Pd(dppf)Cl_2$·DCM, $Zn(CN)_2$, anhydrous DMF, 120 °C, rt. Synthesised, under supervision, by Mr Peter K. Buchowiecki a masters student in the PVF group.

Appendix 3: Reported synthetic route to compound 81

Synthetic route to **81**, published by Bradner *et al*., is presented in Scheme 10.13. [88]



Scheme 10.13 Synthetic route published by Bradner *et al*. to obtain CN-analogue **76**. *Reagents and conditions*: (i) 9-steps as previously published[226] (ii) 4-chloro-7*H*-pyrrolo[2,3-d]pyrimidine, NaH, DMF, rt, 2 hrs (iii) $NaIO_4$, OsO_4, MeOH/H_2O, 0 °C to rt, 3 hrs (iv) $NaBH_4$, MeOH, 0 °C, 4 hrs (v) I_2, PPh_3, imidazole, THF, rt, 5 hrs (vi) NaN_3, DMF, 85 °C, 3 hrs (vii) H_2, Pd/C EtOH, rt, 16 hrs (viii) 3-(2-(5-(*tert*-butyl)-1*H*-benzo[d]imidazol-2-yl)ethyl)cyclobutan-1-one, $NaBH_3CN$, AcOH, MeOH, rt, 5 hrs (ix) iPrI, K_2CO_3, MeCN, 65 °C, 2 hrs (x) *N*-bromosuccinimde, DMF, 5 hrs, rt (xi) $Zn(CN)_2$, $Pd(PPh_3)_4$, NMP, microwave, 1 hr (xii) TFA/H_2O, 0 °C to rt, 1 hr.

Appendix 4: Peer reviewed letter in *Bioorg. Med. Chem. Lett*.

NB: Removed due to copyright issues (DOI: 10.1016/j.bmcl.2016.07.041)