University of Sussex (logo: US)

A University of Sussex PhD thesis

Available online via Sussex Research Online:

http://sro.sussex.ac.uk/

Synthetic Strategies Towards Challenging PPI Drug Targets

UNIVERSITY OF SUSSEX

Thomas O. Moore

Supervisor: Prof. Simon E. Ward

Submitted to the University of Sussex in part fulfilment of the requirements of the degree of Doctor of Philosophy, May 2016

Declaration

I hereby declare that the work presented in this thesis was carried out at the University of Sussex under the supervision of Professor Simon Ward between the dates of September 2012 and March 2016. The work presented in this thesis is my own, unless otherwise stated, and has not been submitted in whole or in part form for award of another degree.

Thomas O. Moore

Acknowledgements

Abstract

Protein-protein interactions (PPIs) present a challenging target for the development of inhibitors, due to the large size and lipophilicity of protein binding sites. Because of this, work that is able to increase our understanding of how PPI inhibitors can be better developed is of great value.

In unusual cases, approved drug compounds are able to exhibit excellent pharmacological (PK) profiles despite having unfavourable physical properties. New methodology was developed whereby the solubilising side chains found in one such compound, daclatasvir, could be incorporated onto aryl-bromides using palladium-catalysed C-H activation. The scope of the reaction was tested, and in 18 different examples the C-H activation was found to be successful 11 times.[1] This methodology has application in the potential synthesis of cell-permeable PPI inhibitors.

Inhibiting the PALB2/BRCA2 PPI has the potential to produce synthetic lethality in cancerous cells with damaged DNA repair mechanisms. Computational techniques were used to design small-molecule and miniprotein mimetics of BRCA2, which could potentially competitively inhibit PALB2/BRCA2. This resulted in the purchase of 10 miniprotein sequences ready for testing, one of which has been proven to have significant helical secondary structure. The synthesis of small molecule α-helical mimetics of BRCA2 has also been carried out, with one compound ready for biological testing and the synthesis of two others close to completion.

Publications:

1   T. O. Moore, M. Paradowski and S. E. Ward, *Org. Biomol. Chem.*, 2016, **14**, 3307–3313.

List of Abbreviations

Ac	Acetyl, CH_3CO
ACPC	Aminocyclopentylcarboxylate, $C_6H_{11}NO_2$
AIDS	Acquired immune deficiency syndrome
AMBER	Assisted model building with energy refinement
ANP	atrial natriuretic peptides
aPP	Avian pancreatic peptide
AR	Androgen receptor
Ar	Aryl
AUC	Area-under-the-curve
Bcl	B-cell lymphoma
BER	Base excision repair
[BMIm][PF$_6$]	1-Butyl-3-methylimidazolium hexafluorophosphate, $C_8H_{15}F_6N_2P$
BMS	Bristol Myers-Squibb
Boc	tert-Butoxycarbonyl, $C_4H_9O_2$
BRCA1	Breast cancer 1
BRCA2	Breast cancer 2
brm	Broad medium adsorption (IR spectroscopy)
brs	Broad singlet (^{1}H NMR spectroscopy)
brw	Broad weak adsorption (IR spectroscopy)
CaM	Calmodulin
Cbz	Carboxybenzyl, $C_8H_7O_2$
CD	Circular dichroism
ChTx	Charybdotoxin
CMD	Concerted metalation-deprotonation
DFT	Density functional theory
DIPEA	Diisopropylethylamine, $C_2H_5N(C_3H_7)_2$
DMA	Dimethyl acetamide, C_3H_7NO
DMF	Dimethyl formamide, C_4H_9NO
DMSO	Dimethyl sulfoxide, C_2H_6OS
DNA	Deoxyribose nucleic acid
DSB	Double Strand Break
DSF	Differential scanning fluorimetry

EDAC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide, $C_8H_{17}N_3$
ER	Oestrogen receptor
Et	Ethyl, C_2H_5
F	Bioavailability
Fmoc	Fluorenylmethyloxycarbonyl, $C_{15}H_{11}O_2$
FT	Fourier Transform
GPCR	G-protein coupled receptor
HBTU	*N,N,N′,N′*-Tetramethyl-*O*-(1H-benzotriazol-1-yl)uronium hexafluorophosphate, $C_{11}H_{16}F_6N_5OP$
hCD4	Human cluster of differentiation 4
HCV	Hepatitis C
Hdm2	Human double minute 2 homolog
HIF	Hypoxia-inducible factor
HIV	Human immunodeficiency virus
HOBt	Hydroxybenzotriazole, $C_6H_5N_3O$
HPLC	High-pressure liquid chromatography
HSA	Human serum albumin
HWE	Horner-Wadsworth-Emmons
IBX	2-Iodoxybenzoic acid, $C_7H_5IO_4$
IgG$_1$	Immunoglobin G1
IL-2	Interleukin-2
*i*Pr	*iso*-Propyl, $CH(CH_3)_2$
KHMDS	Potassium hexamethyldisilazide, $KSi_2C_6NH_{18}$
KNIME	Konstanz Information Miner
LCMS	Liquid-chromatography/mass-spectroscopy
MCL-1	Induced myeloid leukemia cell differentiation protein
Mdm2	Mouse double minute 2 homolog
Me	Methyl, CH_3
MIDA	Methyliminodiacetic acid, $CH_3N(CH_2CO_2H)_2$
MMFF94	Merck molecular force field
MOE	Molecular Operating Environment
MWT	Molecular weight
NER	Nucleotide excision repair

NHC	*N*-Heterocyclic carbene
NHEJ	Non-homologous end joining
NMDA	*N*-methyl-D-aspartate, $C_5H_9NO_4$
NMI	*N*-methyl imidazole, $C_3H_3N_2CH_3$
NMP	*N*-methyl-2-pyrrolidone, C_5H_9NO
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NR	Nuclear receptor
NS5A	Non-structural protein 5A
PALB2	Partner and localiser of BRCA2
PARP	Poly ADP ribose polymerase
PEG-IFN	Pegylated interferon α
Ph	Phenyl, C_5H_6
PPI	Protein-protein interaction
PPII	Polyproline-II
RNA	Ribose nucleic acid
SAR	Structure activity relationship
SEM	2-(Trimethylsilyl)ethoxymethyl, $CH_2OC_2H_4Si(CH_3)_3$
SMAC	Second mitochondria-derived activator of caspases
smMLCK	Smooth muscle myosin light-chain kinase
SSB	Single Strand Break
TBAC	Tetrabutylammonium chloride, $(C_4H_9)_4NCl$
TBAF	Tetrabutylammonium fluoride, $(C_4H_9)_4NCl$
tBu	*tert*-Butyl, $C(CH_3)_3$
TCEP	Tris(2-carboxyethyl)phosphine, $C_9H_{15}O_6P$
TEMPO	(2,2,6,6-Tetramethylpiperidin-1-yl)oxyl, $C_9H_{18}NO$
Tf	Trifluoromethanesulfonic (triflic), CF_3O_2F
TFA	Trifluoroacetic acid, CF_3COOH
TFIIH	Transcription factor II H
THF	Tetrahydrofuran, $(CH_2)_4O$
TPSA	Topological polar surface area
XIAP	X-linked inhibitor of apoptosis protein
Zif286	Zinc finger 286

Contents

1 Introduction

1.1 The Search for a New Generation of Drug Targets

It is no surprise that the easiest to target and best-understood proteins have often been the focus of drug discovery programmes. In their seminal paper published in 2002, Hopkins and Groom demonstrated the skew that this bias had afforded on the targets of available drugs,[1] with over 75 % of drugs on the market designed to target either an enzyme or a G-protein coupled receptor (GPCR) (Figure 1.1).
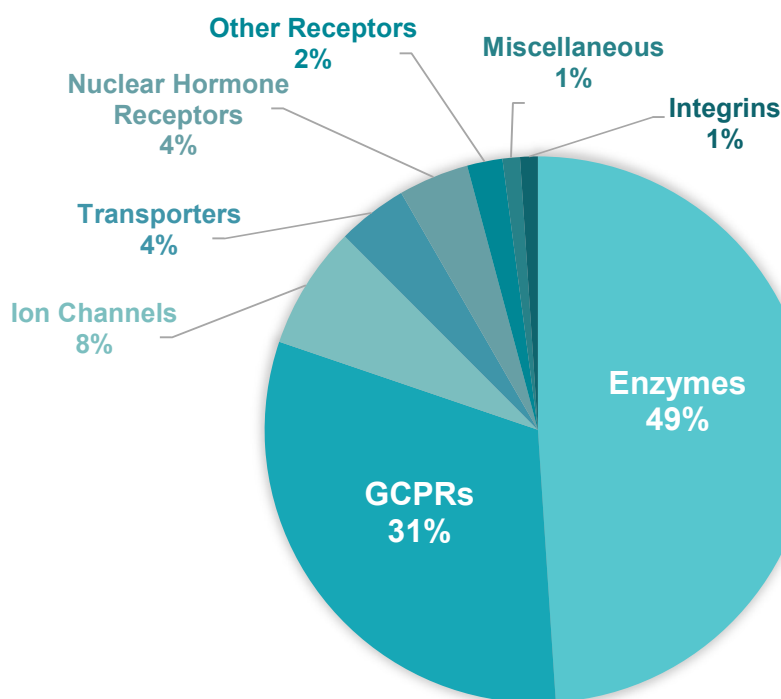
Figure 1.1: The target protein class of marketed drugs in 2002, adapted from a publication by Hopkins and Groom.[1]

Enzymes and GPCRs make ideal drug targets.[2] Proteins from these classes bind to small molecules to drive their function within the body. For example, enzymes bind to a substrate and catalyse a chemical transformation at the active site, while GPCRs bind to a chemical messenger and trigger a response within the cell. The existence of these endogenous small molecule ligands for the proteins provides two key benefits for the medicinal chemist. Firstly, it means that there must already be a pocket on the protein surface of the correct size and with the potential for the correct intermolecular interactions to bind to a drug molecule (enzyme pockets are typically 300-1000 $Å^2$, ideal for binding to a drug).[3] Secondly, the natural ligand can be used as a lead for drug design. Indeed,

before developments in biochemistry which allowed us to better understand the three-dimensional structure of proteins, this was a commonly used practice for drug design. For example, the psychoactive tryptamine and amphetamine families of compounds were designed by using serotonin and dopamine as chemical leads respectively, both of which are ligands for GPCRs within the brain.[4]

Given the relative ease with which the medicinal chemist can design drug compounds for enzymes and GPCRs, it is no real surprise that they comprise such a large portion of marketed drug classes in 2002.[1] However, as time has moved on, the number of readily targeted proteins without marketed drugs has decreased. As a result, medicinal chemists are left to look towards less conventional drug targets, which provide fewer straightforward opportunities to develop inhibitors. In the words of Wells and McClendon, we are left 'reaching for high-hanging fruit in drug discovery'.[5] One such class of underdeveloped targets for drugs that has emerged in recent years is protein-protein interactions (PPIs).

1.1.1   Protein-Protein Interactions (PPIs) Make Difficult Drug Targets

The interface between two proteins is not a traditional target for small molecules. Unlike the intrinsic pockets on enzymes and GPCRs, a PPI is much larger (typically 1500-4500 $Å^2$ compared to 300-1000 $Å^2$).[6,7] They are also mostly governed by lipophilic interactions and London dispersion forces, and usually lack the high numbers of hydrogen bond donors and acceptors required to achieve selectivity with small molecule ligands.[8] Finally, they are often quite flat in shape, and do not always provide an environment which allows a drug to be enveloped by the protein. Each of these factors cause PPIs to be less well suited to inhibition by small molecules than enzymes or GCPRs.[9]

Although they may be difficult to target, PPIs could have great therapeutic potential. There are predicted to be around 40,000 unique PPIs in humans,[10] each of which contribute to support cell function in some way. When considering that currently only around 3000 proteins exist which belong to families for which small molecule inhibitors have been developed, there is great potential to be gained from inhibiting just a small fraction of available PPIs. Unlike enzymes and GPCRs, they are involved in a large variety of different roles within the cell, and therefore offer a much wider range of

potential therapeutic effects when targeted by small molecules. For example, very few enzymes have been discovered to have therapeutic potential for the treatment of cancer, although there are notable exceptions such as receptor tyrosine kinases.[11] Moreover, by inhibiting PPIs, there is a large amount of choice of potential drug targets for cancer, and with it the potential for cancer specific small-molecule drugs.[12]

Given the huge number of PPIs from which a drug target can be selected, it stands to reason that if even a small proportion of them are found to be druggable, a large number of potential new 'high-hanging fruit' targets may be accessed. This brings two major challenges for the medicinal chemist in the search for PPI inhibitors. Firstly, a protein target must be selected with care, to ensure that it is both druggable and has therapeutic potential. Secondly, the design of potential inhibitors must be approached with creativity, as PPI inhibitors often differ in structure to conventional drugs. This chapter will briefly discuss the important considerations to take when selecting a PPI, followed by an in depth review into the different methods developed to design PPI inhibitors.

1.2 Assessing PPIs for Druggability

Historically, PPIs were considered to be too large and flat[13] and with electrostatic and hydrogen bonding groups too evenly spread across the interface[6,14] to allow selective binding of small molecules. However, using techniques such as alanine-scanning mutagenesis, it became increasingly clear that the affinity of many PPIs was governed by a small selection of key residues, known as binding 'hot-spots'.[15,16] If a small molecule was able to selectively bind at the site of the hot spot, then it could be possible to inhibit the PPI without needing a ligand with an intolerably high molecular weight. However, not all hot-spots are amenable to binding by a small-molecule ligand, and thus assessing the druggability of the PPI is very important.

Assessing the druggability of PPIs in order to select a target for a drug discovery programme is difficult. The hot-spots are typically flat and the implicated residues lipophilic, meaning that conventional libraries may not show any binding to the PPI, and physical and computational tests may not be sufficiently sensitive to identify positive hits. Because of this, there has been a great deal of work into developing methods for screening

druggability that are more specific to PPIs.[17–19] These can generally be categorised into either biophysical or computational screening methods.

1.2.1 Using Computational Methods to Test PPI Druggability

In the last two decades, advances in computing have meant that tasks that would have previously been performed on a supercomputer can now be carried out on the average desktop. This has meant that computational chemistry and bioinformatics have become increasingly valuable as the resource costs become negligible. Furthermore, they are also more scaleable than wet lab experiments, with the possibility of analysing larger fragment libraries quickly against many protein targets.

However, there is much greater room for error in predictions made by computational experiments. Computational models rely on physical approximations and each of these increases error margins in the results obtained. Because of this, it is common practice to only use results obtained computationally in support, often helping to reduce the number of expensive biophysical experiments. There are several different approaches to predicting the druggability of PPIs computationally.

1.2.1.1 Statistical Analysis of Existing Non-Crystallographic Data

One approach that has traditionally proven useful for assessing the druggability of proteins is the statistical analysis of non-crystallographic data, normally protein homology analysis. This is the most basic form of computational druggability analysis, as it does not require three-dimensional information about the protein structure. For example, the protein sequence may be analysed to find domains that are shared with other families of proteins. If there is high homology with protein domains that have known inhibitors, then the target protein might also be of interest for inhibitor development. Online resources such as BLAST[20,21] enable the user to conduct this analysis quickly and simply on any protein.[22] Other attempts have been made to determine hot-spot residues through this approach. By analysing which residues are conserved between protein domains similar to those involved in a PPI, the Consurf algorithm attempts to determine which are the functional regions, and thus the likelihood of individual residues to be hot-spots.[23]

Statistical analysis such as this has relatively low demand on computational processing power, and so large amounts of analysis can be carried out very quickly. However, it also relies on having access to a large amount of existing data, for example on known ligands, and pre-established druggability of protein domain families. As the field of PPI inhibition is not yet well established, there is little information of this kind to support their analysis, and thus purely statistical approaches to assessing druggability are less useful.

1.2.1.2   Predicting Druggability Based on Protein Surface Topology

A very common method for assessing the druggability of a protein is to use algorithms that analyse the topology of a protein surface in order assess the likelihood that a drug-like molecule could interact at any site. However, the standard algorithms for detecting such pockets on a protein, such as Fpocket,[24] may not be able to detect the pockets and hot spots at a PPI as they are too shallow or lipophilic by the normal metrics used.[25] Because of this, esoteric algorithms geared towards the detection of pockets on PPIs have been developed.

Unfortunately by reducing the stringency of pocket definition to enable their detection at PPI surfaces, the error margins are increased drastically, resulting in increased false positives. One approach that has been taken to mitigate this is to use topological pocket finding algorithms in combination with statistical analysis such as that discussed above. Sugaya and co-workers developed a systematic method for narrowing down protein domains using sequential and interactome data to find PPIs which were likely to be druggable. They then used pocket-finding algorithms to assess the 3D structures of these proteins in order to further predict the druggability.[26,27] This resulted in the database of potentially druggable PPIs 'Dr. PIAS'.[28] However, their method was unable to correctly predict many PPIs proven to be druggable, such as BAK/B-cell lymphoma (BCL)2.

Another method for assessing pockets at a PPI interface was developed by Koes and Camacho. PocketQuery uses the residues embedded in the pocket by the complementary protein at the PPI to predict druggability.[29] They developed a web-based software that would select clusters of three amino acid residues that were embedded into a pocket on the protein partner at the PPI. Rather than then using the topology of the pocket itself to score the PPI for druggability, the software uses the embedded residues, assessing the likelihood of being able to design a small molecule mimetic. This is an excellent method

for assessing the druggability of a PPI, however, it relies on the existence of a co-crystal structure of both partners of the PPI, which is usually only available in cases where a lot of work has already been done on the protein towards developing inhibitors. It also requires a lot of human input, which could provide a source of bias.

1.2.1.3 Alanine Scanning Mutagenesis

Alanine scanning mutagenesis is a computational technique that is particularly tailored towards assessing the druggability of PPIs due to the importance of amino acid hotspots.[30] This is the computational equivalent of biochemical experiments which use protein mutagenesis to discover hotspots in PPIs, by making point mutations of amino acid residues to alanine, and testing how these affect the strength of binding at the PPI.[15] DrugScore[PPI] is a software developed to carry out this experiment computationally.[31] Using the protein crystal structure, amino acids from one protein partner in the PPI are sequentially mutated to an alanine and the effect that this has on $\Delta G_{binding}$ of the two protein partners is predicted. This is then plotted and any outliers with a high $\Delta\Delta G_{binding}$ can be predicted as potential hot-spots. This is an excellent tool for predicting the presence of hotspots and thus the potential druggability of a PPI crystal structure,[8] however, as with PocketQuery, a co-crystal structure of both partners in the PPI must be available.

1.2.1.4 Docking 'Solvent' Molecules over the Protein Surface

One docking technique that has emerged recently in order to evaluate the druggability of proteins at PPIs involves the docking of small solvent-like molecules over the whole surface of a protein crystal structure.[32–34] Regions where many binding modes of a solvent molecule (a 'cluster') are found in one confined area of the protein over a certain threshold are designated as hot spots. For example, FTMap[34] is a web base application which can be used to analyse structures from the protein data-bank (PDB) and search for clusters of 5 solvent molecules or more, which are then output as a 3D structure (Figure 1.2). This has the great advantage over using surface topology to search for pockets, in that it is able to detect flatter areas of the protein that may still have the correct electrostatics to bind a small molecule. Although not originally intended for the purpose of specifically testing for hot-spots on PPIs, this method has been shown to be particularly amenable towards them.[35]
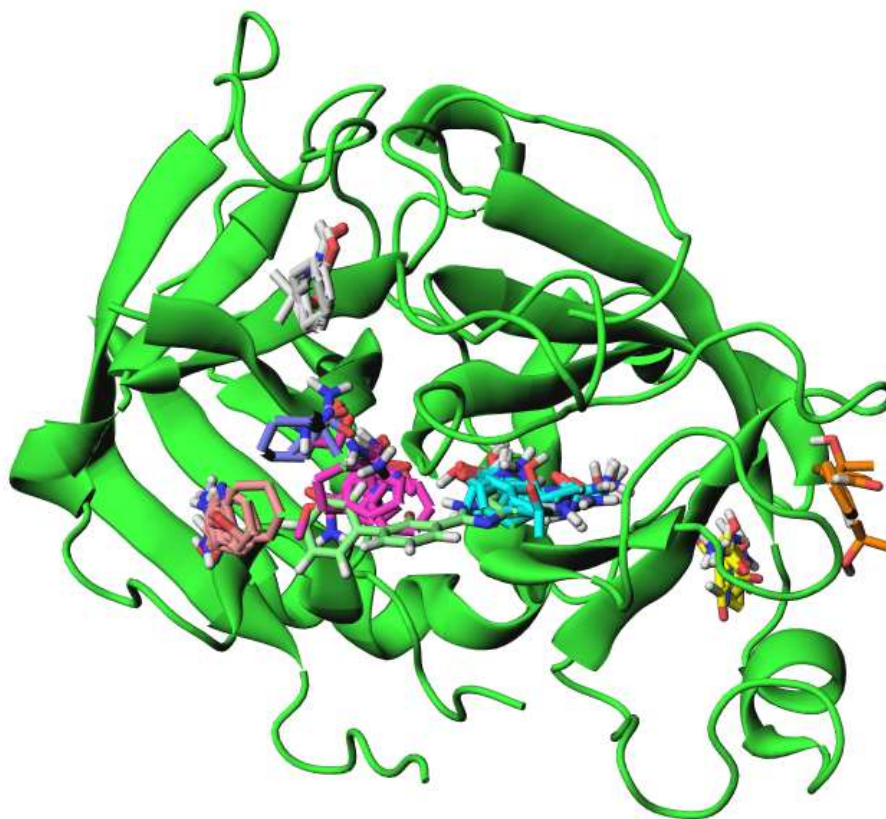
Figure 1.2: The hotspots of Urokinase Type Plasminogen Activator (PDB Code: 2O8T), using FTMap. Each colour of molecule represents a different solvent cluster, indicating a hot-spot. Clusters of 5 or more are shown.

Recently, this method was taken one step further in an attempt to predict druggability. Kozakov and co-workers made an attempt to mimic the pioneering work discussed previously by Hadjuk and coworkers,[36] where indices were proposed in order to translate the direct result of NMR fragment binding studies into a metric of druggability. Kozokov analysed a series of FTMap[34] outputs of proteins with known small-molecule binders, and created a series of criteria which must be met by the FTMap clusters to designate the protein as druggable.[37] For PPIs they devised a separate set of criteria for proteins that are druggable by non-canonically large molecules. Using these, they were able to correctly assign 12 out of 15 PPIs with known inhibitors as druggable. Overall, this seems to be the most promising computational method to date with regard to predicting druggability of unconventional protein targets.

1.3 Designing Protein Mimetics to Inhibit PPIs

As mentioned previously, one reason PPIs make difficult targets for drug discovery programmes is that there is no native small-molecule ligand to the protein, such as in the

case of enzymes or GPCRs. Instead the native ligand for the design of PPI inhibitors is primarily a set of hot-spot residues on a protein, and mimicking its activity with a small molecule requires much more ingenuity.

In order to create mimetics of one half of a PPI, it is necessary to understand the protein structures involved. The most common protein secondary structure at a PPI interface is the α-helix, which features in 62% of the PPI structures submitted in the PDB.[8] Because of this, α-helical mimetics are hugely valuable in the development of therapeutics for PPIs. Fortunately, α-helices have very defined secondary structures. They consist of a coiled peptide backbone, with 3.6 amino acid residues per turn of the helix.[38] This configuration allows for a hydrogen bond between the carbonyl oxygen of one amino acid, and the amide N-H of the amino acid four residues later, denoted as $i \rightarrow i+4$. As a result, the structure is highly stable.

Although the structure of the α-helix is incredibly well defined, the associated short chain polypeptides do not typically spontaneously exist as stable helices. The correct folding relies on the conformational influence of the rest of the protein. However, delivering full-length proteins therapeutically is not an option for intracellular targets, due to lack of membrane permeability, and metabolic instability, combined with the high cost of expression of sufficiently large quantities. This has led to significant scientific development in identifying mimetics for α-helices to use as tool or therapeutic leads for the development of PPI inhibitors.

1.3.1 Using Foldamers to Inhibit PPIs

A foldamer is an oligopeptide containing unnatural amino acid residues which folds in such a way as to mimic the confirmation of biological polymers, in particular protein secondary structures.[39] This is achieved most simply by modifying the structural elements of an amino acid in a small enough way that it retains the ability to fold into the same secondary structure when part of an oligomer. These modifications, although subtle, should be adequate to prevent the foldamer from being recognised by the enzymes involved in biological processes such as proteolysis.

The most basic form of amino acid backbone modification is the insertion of a methylene group, to transform an α-amino acid into the respective β-amino acid (Figure 1.3). Early

pioneering work in this area was done by D. Seebach,[40] who developed the synthesis of β-amino acid residues using the Wolff rearrangement. In the same work, Seebach also found that hexamers of β-alanine were highly resistant to peptin mediated proteolysis. This led to an increase of interest towards these compounds as stable *in vivo* therapeutic polypeptides.

α–residue β3-residue β2-residue

ACPC residue ACHC residue
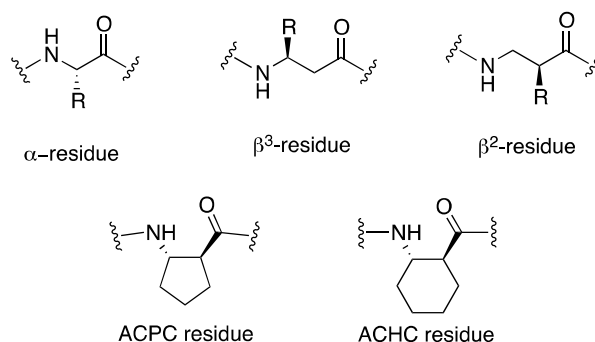
Figure 1.3: The structure of an α-amino acid residue, with some examples of β-residues

Initially it was hypothesized that as β-polypeptides are less conformationally restricted they would be less likely to form secondary structures than their α-polypeptide counterparts. Surprisingly however, this was found not to be the case, and furthermore it was observed that sometimes β-amino acid foldamers had secondary structures with increased stability. Over the last 20 years Gellman has carried out research into how to tune the backbone of a synthetic foldamer to achieve the desired secondary structure.[41,42] He has found that close mimetics to the α-helical secondary structure can be achieved by utilising alternating α/β subunits. He discovered that by using β-amino acids containing hydrocarbon rings to restrict conformational rotation, he could reliably form foldamers with the desired 'α-helical-like' secondary structure. To do this, he used subunits such as aminocyclopentylcarboxylate (ACPC) (Figure 1.3), which could also be seen as a β- form of proline. The α/β-polypeptides were also found to be proteolytically stable, akin to β-polypeptides.[43]

This reproducible method to design foldamers that will mimic α-helices has led to the discovery of several biologically active foldamers that target α-helical PPIs. Gellman used cyclic β-amino acids in the design of a foldamer to target human cytomegalovirus.[44] He was able to develop a foldamer with an IC$_{50}$ of 30 μM, an increase in potency over the wild type α-helix which has an IC$_{50}$ of >100 μM. The technique has also been used to target other more conventional PPI interfaces. For example, Schepartz developed an α/β peptide mimicking p53 that has only slightly lower affinity than the wild type p53/mouse

double minute 2 homolog (Mdm2) interaction (K_d = 368 ± 76 nM and 233 ± 32 nM respectively).[45,46]

Despite the success of these examples, there are still factors that prevent foldamers from making effective therapeutic compounds. Although their stability against proteolysis in the cell has increased, their large size prevents them from having good membrane transport properties. They are also incredibly expensive to produce, as homologation reactions used to synthesize β-amino acids use hazardous reagents not scalable to an industrial level.

1.3.2 Using Designed Miniproteins to Inhibit PPIs

With the development of solid phase peptide synthesis[47,48] and high-pressure liquid chromatography (HPLC) purification[49] it became possible to synthesise polypeptide oligomers. This meant that for the first time small protein domains or hormones could be used as synthetically accessible leads. However, the technique was limited by polypeptide chain length, as each synthetic step used instilled a greater proportion of side-products, and made the desired oligomer harder to purify. Due to the limited amount of published literature reviewing the use of miniproteins to target PPIs, an in depth review has been carried out here.

1.3.2.1 Early Use of Designed Miniproteins to Target ANP receptors

Several groups have attempted to mimic the action of atrial natriuretic peptides (ANPs) using synthetic peptide oligomers.[50–54] ANP's are a family of hormones that are released as part of the body's response to raised venal blood pressure, and trigger a series of response measures such as vasorelaxation.[55] A synthetic compound that mimics the action of ANPs should therefore have the potential to act as a treatment for symptoms such as raised blood pressure. As ANP has a chain length of 28 amino acid residues it is within the upper boundary of what is accessible by synthetic polypeptide chemistry, and therefore amenable to a drug discovery programme.
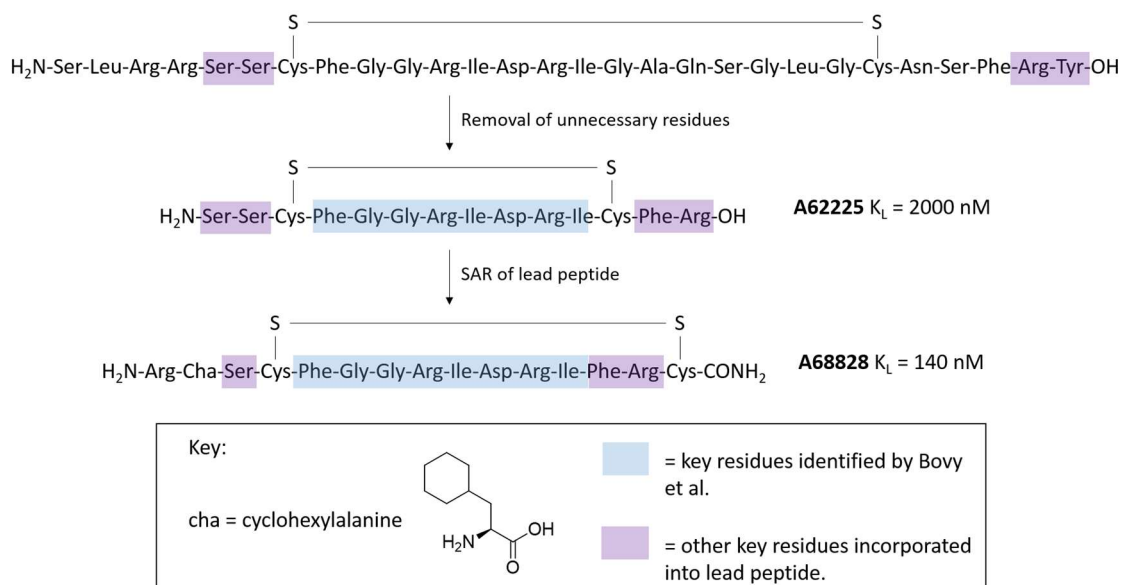
Figure 1.4: Development of lead compound A62225 and candidate A68828 by Geldern and coworkers,[50] using key residues identified by Bovy and coworkers.[53] K_L is the measured binding of the lowest affinity ANP receptor, presumed to be ANP-R1.

In a successful study, Geldern and co-workers at Abbott used solid phase peptide synthesis to investigate the structure-activity relationship (SAR) of ANP. They took a sequence of amino acids, 8-15, previously identified by Bovy and co-workers as the ANP hotspot residues,[53] combined it with other key parts of ANP, and reduced the disulfide ring size, to produce lead peptide A62225 (Figure 1.4). In doing so, they managed to half the size of the peptide from 28 to 14 residues, whilst maintaining some activity (K_L = 2000 nM). From here, they carried out SAR studies where they substituted and added both natural and unnatural amino acids and changed the size of disulfide ring in attempt to improve the efficacy. By enlarging the disulfide ring size, adding an arginine to the *N*-terminus, substituting a serine for a cyclohexylalanine and capping the *C*-terminus with an –NH₂ group they discovered A68828. This 15 residue peptide is much shorter than the 28 residue ANP, however, with a K_L of 140 nM it is only 65-fold less potent. Geldern and co-workers had shown that it was possible to synthetically create a miniprotein mimetic of a hormone using solid-phase synthetic techniques, allowing them to carry out SAR studies on oligopeptides.

The solid phase synthesis techniques used to design A68828 meant that there were limitations to the number of compounds that could be produced given the time and funding available. In a seminal study by Li and co-workers, they were able to vastly increase the number of peptides tested using much fewer resources by employing a phage-

display peptide screen.[56] This technique involves inserting the DNA encoding a protein of interest into phage DNA such that the protein is displayed on the surface of the virus. This means that the phage will now exhibit binding affinity to any partner protein of that displayed.[57,58]

This technique can be used to simultaneously test the binding affinity of many different proteins to a given binding partner. For example, consider that a 'library' of phages is synthesised in which point mutations have been made at three different places on the surface displayed protein, and each of these mutations randomly involves any one of the 20 amino acids. This 'library' culture therefore consists of 3^{20} unique phages. If this culture is added to a dish with the partner protein of interest fixed to the wall, any phages with a surface protein with high affinity will also become fixed to the dish. Low affinity phages can be washed away. Allowing the remaining phages to infect bacterial cells and propagate will generate a culture with a higher quantity of those phages displaying a high affinity surface protein.

If this process is repeated 3 or 4 times, only the very best binding phages will remain in high concentrations. The DNA of these phages can then be sequenced in order to determine the amino acid sequence of the highest affinity proteins.
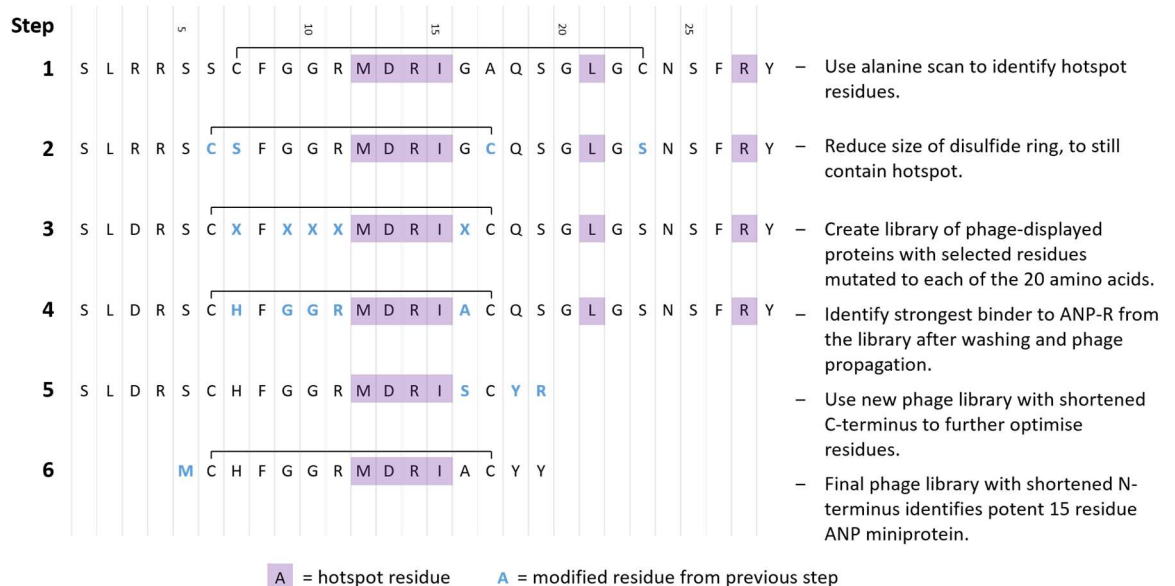
Figure 1.5: Process of optimising ANP using phage display to identify a high affinity 15-residue miniprotein.[56]

Li and co-workers used this technique to identify a very high affinity ANP miniprotein (Figure 1.5).[56] As in previous studies hotspot residues were identified using alanine-

scanning techniques. The disulfide ring size of ANP was reduced, and a set of residues randomly selected for mutation to build a phage display library. The phage display technique was then used to identify the amino acids best suited at each of these positions, to achieve the highest binding affinity. Following this, they were able to decrease the chain length of the enhanced peptide and point mutate at different positions in repeated rounds of phage display scanning, to finally achieve a 15 amino acid ANP miniprotein with an EC_{50} of 0.48 nM (compared to 140 nM for A68828). This is seven times lower than wild-type ANP. Due to the large number of proteins that can be tested in each phage display experiment it is highly efficient, and it was possible to discover an optimised miniprotein with much higher affinity than that from the conventional medicinal chemistry approach carried out by Geldern and co-workers, using far fewer experiments.[50]

1.3.2.2 Designing a Protein-A Miniprotein with Defined Secondary Structure

Although the work done by Li and co-workers was seminal towards developing methodology to optimise the amino acid sequence, their miniproteins did not contain specific secondary structure. As an estimated 62 % of all PPIs involve at least one α-helix,[8] the mimicry of secondary structure using miniproteins is important. In an important development, Braisted and Wells were able to design and optimise a miniprotein mimetic of a domain of *S. aureus* protein A, which contains an α-helical secondary structure.[59]

In this work, the authors aimed to develop peptidic inhibitors to target the interface between the F_C portion of immunoglobulin G1 (IgG$_1$) and the Z-domain of Staphylococcus aureus 'protein A'. From crystallographic data of the bound proteins it was known that the binding domain of protein A comprised of a bundle of 3 α-helices, with two helices bound to IgG$_1$ and the third seemingly not involved at the protein interface (Figure 1.6).[60]
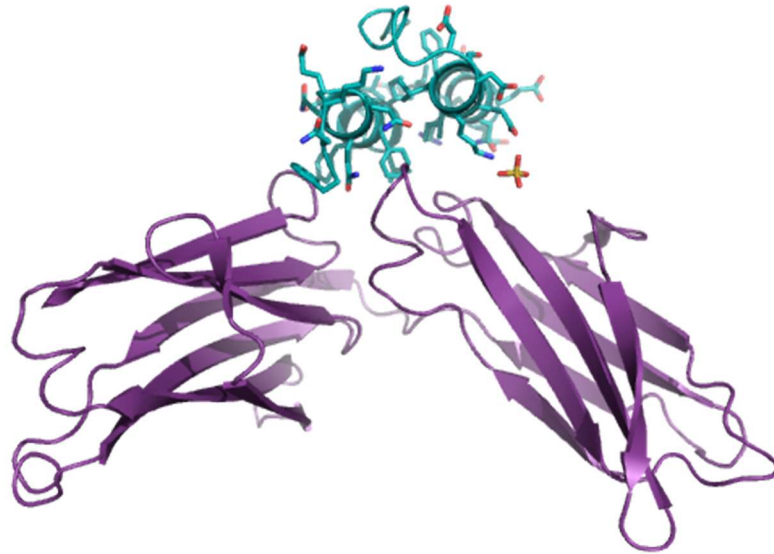
Figure 1.6: Crystal structure of IgG$_1$ (purple) bound to protein A (cyan), PDB code 1FC2.[60]

Due to the apparent lack of involvement of the third α-helix at the binding interface, Braisted and Wells hypothesised that deleting this helix would only have minimal effect on the protein binding energy. However, surprisingly they found that this deletion reduced the binding energy by 10^5 fold (from a K_d of 10 nM to a K_d of >1 mM). In order to remedy this they optimised the sequence of the truncated protein A to re-establish efficacy for IgG$_1$.

It had been previously shown that protein A can be displayed on the surface of M13 phage particles, rendering phage-display optimisation well suited for this protein.[61,62] Braisted and Wells used site-directed mutagenesis to build phage display libraries with random mutations at specific positions on the truncated protein A. They first optimised residues that would have faced the deleted third α-helix (the exoface), and unsurprisingly found it beneficial that some lipophilic residues (Leu-20 and Phe-31) were replaced with charged residues Asp and Lys, which are much more commonly found on the solvent-exposed face. After this, they built a new phage-display library, mutating residues between the two remaining α-helices (the intraface) in order to stabilise the core of the truncated protein A. They found three mutations (A13R, I17A and L35I), all positioned at the open end of the two helices, which seem to stabilise the protein folding. Finally, once these changes had been made to stabilise the secondary structure of the truncated protein A, several phage libraries were built to optimise all 19 residues at the interface with IgG$_1$. Although these residues were largely best conserved, there were a few mutations to

charged residues, such as N12R, for which presumably the resulting improved solubility outweighed any potentially negative effect on the binding to IgG_1.

Ultimately, the optimised truncated form of the protein A Z-domain contained 12 point mutations from the wild type protein, which increased the K_d more than 10^4 fold, from >1 mM to 42 nM. Perhaps more importantly, the authors were able to show that by optimising the structure they had increased the α-helicity in solution from 11% for the original truncated peptide, to 56% for the fully optimised mutant (a maximum helical content of 63% was calculated from the 1FC2 crystal structure). This demonstrates that it is possible to engineer secondary structure into artificial miniproteins.

1.3.2.3 Mimicking A Zinc Finger Secondary Structure with A Designed Miniprotein

At the same time as the Braisted and Wells work on protein A, a similar study was underway regarding artificially recreating secondary structure in a synthetic miniprotein. Here the aim was to create a zinc finger domain which could form the desired secondary structure without reliance on zinc presence.[59] The work, carried out by Imperiali and co-workers, began with optimisation of the naturally occurring zinc finger 286 (Zif286).[63] Rather than using a systematic phage-display library screen in order to optimise the structure, they employed rational design facilitated by nuclear Overhauser effect (NOE)-derived NMR structures.[64] Initially the Zif286 was modified and shortened, taking into consideration residues known to be key for zinc finger domain formation, and enforcing the β-hairpin formation using a D-serine residue (Figure 1.7, step 2). Eight point mutations were then made to stabilise the secondary structure further, with the aim of decreasing the reliance of the protein on zinc, but requiring that it retain the correct conformation. By using a D-proline residue instead of a D-serine residue, a type-II' β-hairpin was encouraged over a type-II turn, resulting in the protein secondary structure being sufficiently stable to form independently of zinc ions (Figure 1.7, step 4). Finally, mutations at the C-terminus of the protein were used to prevent the formation of an unwanted *trans* isomer of the D-proline amide bond, visible by NMR spectroscopy. The final result was a 23-amino acid polypeptide with well-defined secondary structure, similar to that of a zinc finger, although metal independent.
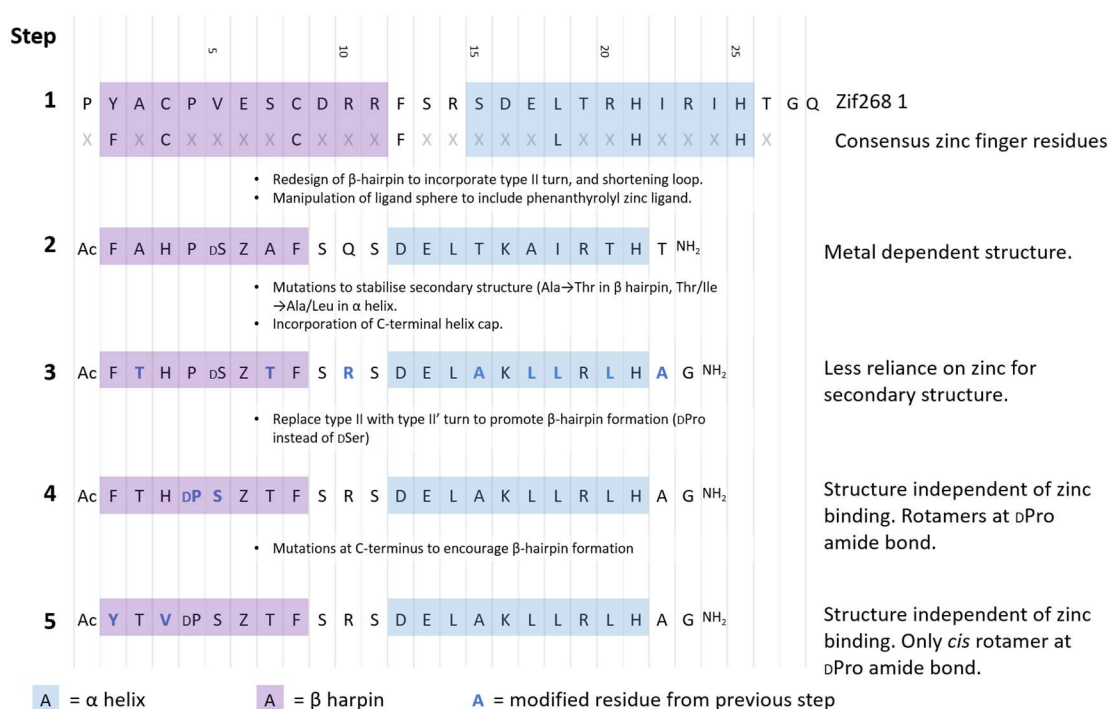
Figure 1.7: The modification of Zif268 by rational design in order to discover a zinc-independent miniprotein exhibiting zinc finger-like secondary structure.

1.3.2.4 Transferring the Human cluster of differentiation 4 (hCD4) Active Site onto a Naturally Occurring Miniprotein Scaffold

Despite good progress in the design of artificial miniproteins, all examples had thus far involved the optimisation of an existing protein or domain with some inherent stability. Unfortunately however, this method is not applicable to most PPIs, as typically the hot-spot residues are part of a domain that is too large to optimise into a stable miniprotein. In breakthrough work Vita and co-workers were able to transfer the hot-spot residues from a protein to a naturally occurring stable miniprotein and maintain some binding at the PPI interface.[65]

When analysing a set of 13 scorpion toxin structures it was observed that despite conservation of just six cysteine residues, each of the miniproteins formed a very consistent secondary structure, where an α-helix was joined to a β-sheet by three disulfide bridges. It was therefore hypothesised that it should be possible to alter the protein sequences significantly while maintaining a very defined secondary structure, as long as the six key cysteine residues were conserved.[66]
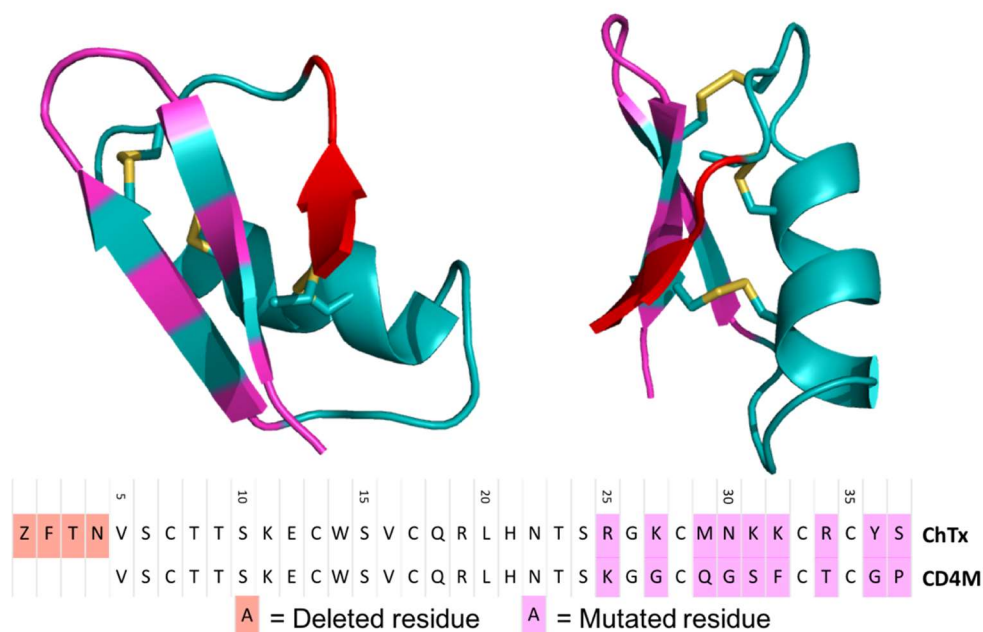
Figure 1.8: The secondary structure of charybdotoxin (ChTx, PDB source: 2CRD), showing the truncated residues (red) and the mutated residues (magenta) forming the human CD4 mimetic CD4M.[65]

Human cluster of differentiation 4 (hCD4) interacts with human immunodeficiency virus-1 (HIV-1), allowing the virus to bind to the cell surface and subsequently infect the cell.[67] Blocking this interaction should therefore hinder the spread of HIV-1 around the body. The interaction of hCD4 with HIV-1 gp20 involves a β-hairpin at residues 35-46. Vita and co-workers transposed these residues onto the β-hairpin portion of scorpion venom charybdotoxin (ChTx, Figure 1.8), and truncated the protein by removing four structurally insignificant residues from the C-terminus. In doing so they created CD4M, a 33 residue miniprotein mimetic of the hCD4 binding domain, which could be synthesised using solid phase peptide chemistry.

Circular dichroism (CD) of CD4M was used to assess the mix of α-helices and β-sheets and indicated that the desired secondary structure had formed. The IC_{50} of CD4M binding to gp20 was measured as 20 μM, and when the CD4M sequence was incorporated into a CD4M-human serum albumin (HSA) chimera, the IC_{50} improved to 0.8 nM, likely due to solubilisation. As such it was possible to design a miniprotein with high affinity for its binding partner by making significant changes to the ChTx protein sequence while retaining the secondary structure. This process required very little optimisation, and demonstrates how robust the secondary structure of scorpion venom can be as a result of disulfide bond formation.

In a further study, Vita and co-workers showed that this technique could be adapted to other scorpion venom toxins.[68] Whereas in the first study they used the 37-amino acid ChTx as the surrogate structure for their miniprotein design, here they showed that the 31-amino acid scyllatoxin could also be used. Scyllatoxin also contains the 3 disulfide bridges, and the resulting α-helix and β-hairpin that are conserved in scorpion venom proteins.[69]
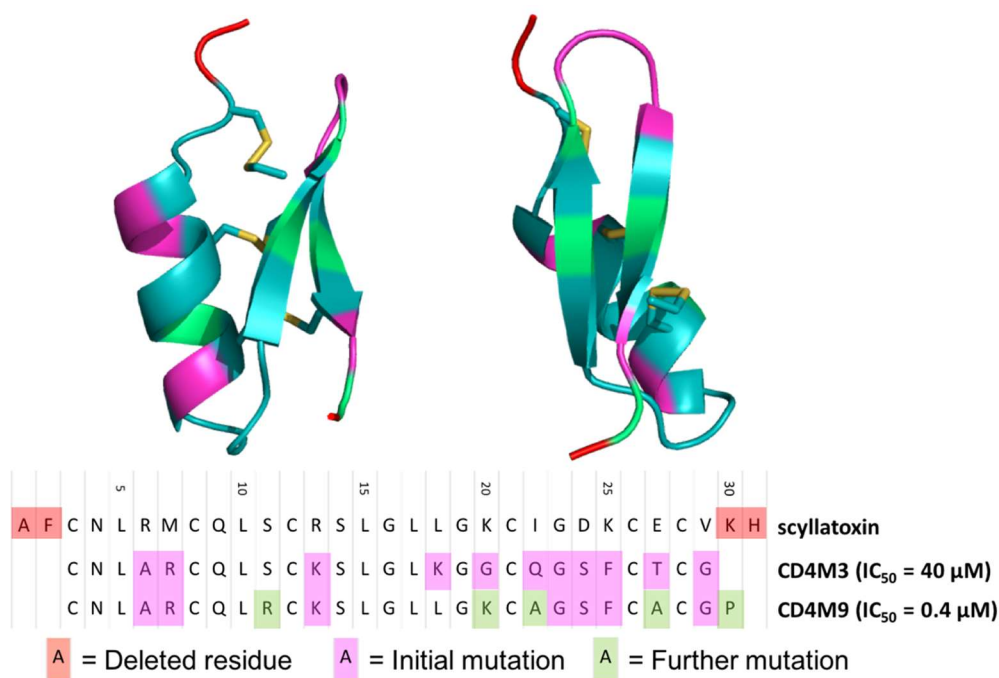
Figure 1.9: The secondary structure of scyllatoxin (PDB source: 1SCY), showing the truncated residues (red) and the initial mutated residues (magenta) forming the human CD4 mimetic CD4M3 (IC$_{50}$ = 40 μM). The further mutations (green) made to optimise this sequence to CD4M9 (IC$_{50}$ = 0.4 μM) are also shown.[65]

Similar to previous examples they made a series of mutations to incorporate the β-hairpin from the hCD4 binding site into scyllatoxin (Figure 1.9). This resulted in compound CD4M3, which had an IC$_{50}$ of 40 μM. This is comparable to the previously discovered ChTx derivative CD4M (IC$_{50}$ = 20 μM). However, CD4M3 is six amino acids shorter, with an overall length of 27 residues. This demonstrates how robust the technique is, as using an entirely new scorpion toxin 'surrogate' produced a miniprotein with incredibly similar efficacy. Vita and co-workers went on to optimise CD4M3. They tested a series of derivatives, employing structural and binding information to predict advantageous mutations, and finally arrived at CD4M9 (IC$_{50}$ = 0.4 μM). This sequence has a 100-fold improvement in activity compared to CD4M3, a result of five different amino acid

mutations. Overall, a miniprotein had been designed that optimised hCD4 to roughly a quarter of the size, whilst improving its activity from being undetectable to 0.4 μM.

1.3.2.5 *Designing a DNA-Binding Miniprotein Based on Avian Pancreatic Peptide (aPP)*

Schepartz and Zondlo used a similar technique to Vita and co-workers to design a DNA-binding miniprotein.[70] They based their miniproteins on the structure of avian pancreatic peptide (aPP), a 36-amino acid protein which has a defined α-helical secondary structure without any disulfide bridges. This is possible due to the presence of a highly stable three turn polyproline II (PPII) helix,[71] which binds to one face of the aPP α-helix through a series of lipophilic interactions. This encourages the secondary structure to form in solution.
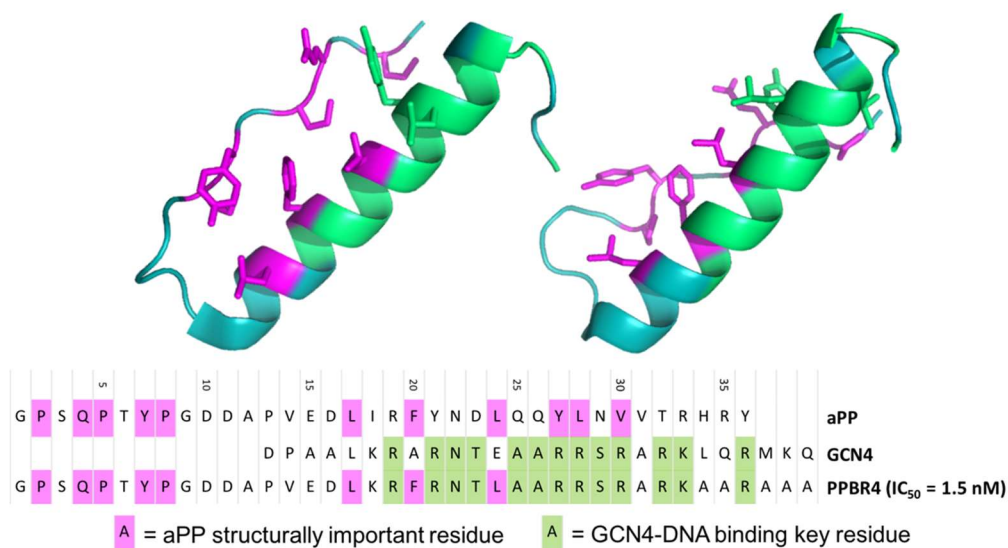
Figure 1.10: The structure of aPP (PDB source: 2BF9), showing the structurally important residues as sticks, the structurally important residues conserved in PPBR4 (magenta), and the key DNA binding residues in PPBR4 (green).[70]

GCN4 is a protein which binds to DNA through a long α-helix, with 13 key residues interacting with the DNA.[72,73] Schepartz and Zondlo set out to transfer different combinations of these key residues to the aPP α-helix, in an attempt to discover a DNA binding miniprotein. They began by identifying the residues that were key for the formation of the aPP secondary structure (Figure 1.10). They then aligned the GCN4 and aPP α-helical secondary structure and designed new sequences incorporating GCN4 residues, avoiding mutation of structurally important aPP residues. Unfortunately, they found they could not achieve DNA binding activity by doing this, and were forced to make mutations which disrupted some structurally important residues on the α-helix. By

incorporating mutations Y27R, L28R, and V30R, they discovered PPBR4, a miniprotein able to bind to DNA with an IC_{50} of 1.5 nM. However, because of the loss of three key structural residues, it appeared that the α-helix did not form unaided in solution by CD spectroscopy. Fortunately though the influence of the PPII helix was sufficient to encourage the formation of the required α-helix in the presence of DNA, resulting in a miniprotein with very high efficacy.

This work by Schepartz and Zondlo demonstrated that disulfide bond formation is not a prerequisite in proteins which make suitable starting points for miniprotein design. However, it also highlights the drastic effect of losing structurally important residues on the ability of a protein to form a defined secondary structure in solution.

In order to restore the secondary structure of PPBR4 in solution, Schepartz carried out a further study, using phage display library techniques, to optimise the structure of the PPII helix.[74,75] By stabilising the PPII helix, they aimed to re-establish the α-helical stability observed in aPP, thereby increasing the binding affinity of their miniproteins to DNA. Two phage display libraries were built. In library A, three residues were randomly mutated in the PPII helix region and the proline residues conserved, whereas in library B, four mutations were made and two of the proline residues disrupted to allow increased flexibility in this region of the protein (Figure 1.11). Library A produced no proteins with improved affinity to the PPBR4 parent protein, however, surprisingly library B produced several very high affinity proteins.

Figure 1.11: Phage display libraries designed to optimise the PPII helix on PPBR4. Randomly mutated residues within the phage libraries are shown in green, and the optimised mutated residues in p007 in magenta.[74]

The most active of the Library B proteins was p007, with a 100-fold increased activity compared to PPBR4. Despite having lost two PPII prolines, and thus most of the PPII secondary structure, p007 was found to adopt an α-helical conformation at room

temperature. It also afforded good specificity to the required DNA sequences, unlike PPBR4 which was less selective.

1.3.2.6 Targeting Bcl-2/Bcl-X$_L$ With An aPP Derived Bak Mimetic

Given their success in using aPP as a surrogate structure for miniprotein design of a GCN4 mimetic, Schepartz and co-workers set about designing a mimetic for Bak, in order to target Bcl-2 and Bcl-X$_L$.[76] Bcl-2 and Bcl-X$_L$ are cell death antagonists, so inhibiting them may encourage fragile cells to enter apoptosis. This has potential in treatment of cancers, as cancer cells are naturally more fragile and therefore vulnerable to the apoptosis-inducing effects of Bcl-2/Bcl/X$_L$ inhibition,[77] and moreover Bcl-2 levels are elevated in 80 % of B-cell lymphomas, as well as other cancers.[78] As Bak interacts with Bcl-2/Bcl-X$_L$ via a single face of an α-helix, it was identified as a suitable candidate for hot-spot transposition onto aPP.
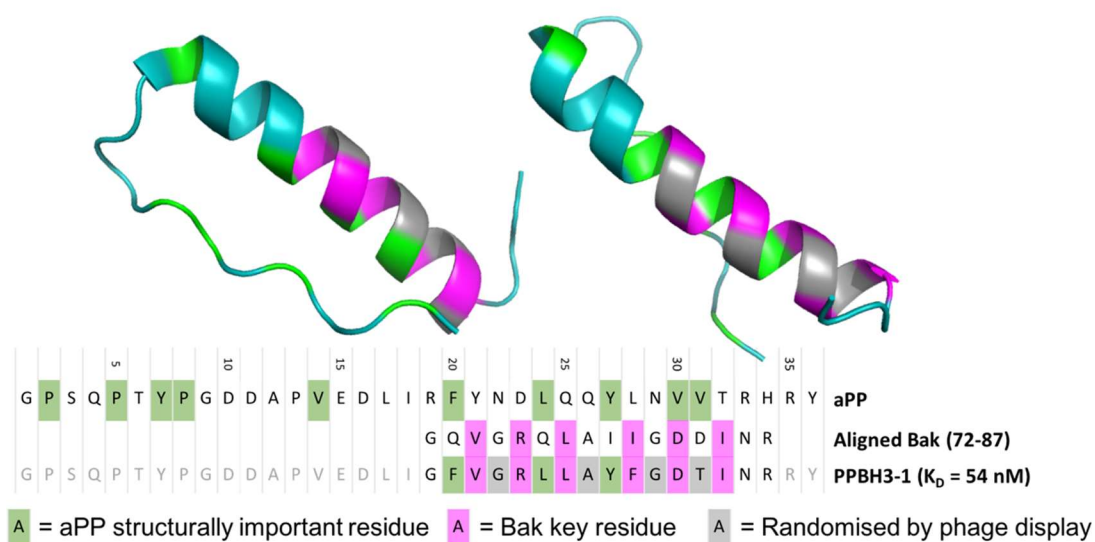
Figure 1.12: The aPP sequence aligned with Bak residues 72-87, and the most active designed miniprotein (PPBH3-1) discovered by phage display analysis.[76] The structurally important residues are shown in green, the key Bak residues are shown in magenta, and the residues randomised in phage display libraries in grey.

Firstly INSIGHT II was used to determine the best alignment of the aPP and Bak helices, by evaluating any steric clashing that could occur between each resulting miniprotein and Bcl-2 when they interact. A phage display library was synthesised, where six key residues from the Bak helix were incorporated into aPP, and then the sequence randomised in four other positions (Figure 1.12). This library contained three sequences with measurable binding to Bcl-2; the best of which being PPBH3-1, with a K$_d$ of 52 ± 5 nM. Furthermore, CD spectroscopy of PPBH3-1 at 4 ºC confirmed the formation of an α-helix in solution.

Schepartz and co-workers had again succeeded in designing an active miniprotein utilising the aPP sequence to induce spontaneous secondary structure.

1.3.2.7 *Designing an EVH1 Binding Miniprotein to Interact Through the aPP PPII helix*

Although the α-helix of aPP shows the most versatility for integration of hotspot residues from different proteins, in another study Schepartz also showed that the PPII helix could be modified to afford binding affinity to a target.[79] EVH1 domains can be found in a wide range of signalling proteins involved with the actin cytoskeleton.[80] They recognise polyproline rich peptide sequences that are folded into PPII helices,[81] and could potentially therefore have affinity for a modified aPP PPII helix.
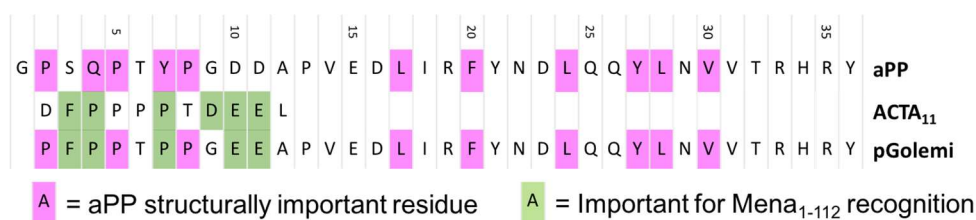
Figure 1.13: Modifications made to the PPII helix of aPP to produce the EVH1 domain binding miniprotein pGolemi. The structurally important residues are shown in magenta, and the residues important for EVH1 affinity in green.[79]

To test this hypothesis, the authors identified ActA, a short peptide sequence which had been shown to bind EVH1 domains, in particular that of of Mena.[82] They incorporated the ActA residues onto the aPP PPII helix (Figure 1.13) to produce pGolemi. This sequence had good affinity for the Mena EVH1 domain, with a K_d of 700 ± 30 nM, 10 fold higher than the ActA sequence itself. The CD spectra of pGolemi showed strong PPII and α-helical secondary structure. This clearly indicates that the structural stability offered by miniproteins can improve the efficacy of other stable structures, such as the PPII helix, by restricting the conformation of the hot-spot residues, and therefore lowering the entropic loss on binding.

1.3.2.8 *Targeting p53/Hdm2 Miniprotein P53 Mimetics*

One PPI of particular interest as a potential drug target is the p53/human double minute 2 homolog (Hdm2) interaction. p53 has many roles in DNA damage repair and can ultimately trigger apoptosis of a damaged cell.[83] The activity of p53 is regulated by the antagonist Hdm2. Disrupting this interaction could therefore increase the activity of p53, which could lead to apoptosis of cancerous cells in conjunction with other treatments.[12]
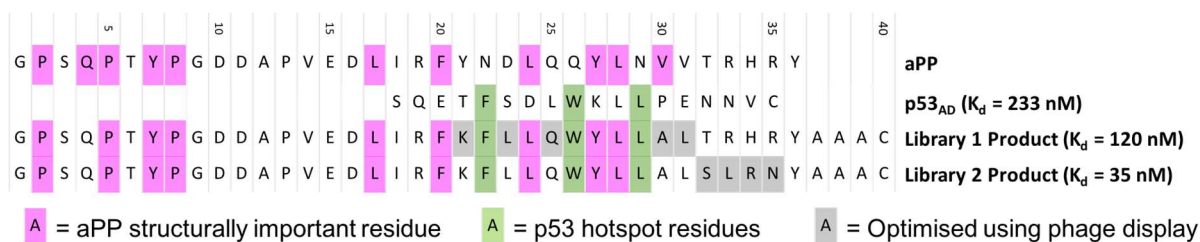
Figure 1.14: Optimisation of aPP-p53 chimera miniprotein by phage display to improve affinity. Structurally important residues are shown in magenta, p53 hotspot residues in green, and the proteins randomised in each phage display library in grey.[84]

As p53 interacts with Hdm2 through three hotspot residues on an α-helix, Schepartz and co-workers hypothesised that it could be possible to use aPP to produce a miniprotein inhibitor of p53/ Hdm2. Phage display techniques were utilised to design an appropriate peptide sequence.[84] A phage display library was constructed based on a hybrid p53/aPP sequence, randomised in five positions on the α-helix near the hotspot residues (Figure 1.14). Unfortunately, the best affinity sequence in this library only offered a two-fold increase in activity from the unmodified p53 α-helix. Because of this a second phage display library was designed, incorporating the optimised residues and randomising at a further four positions on the α-helix. The highest affinity miniprotein from this library offered a further four-fold increase in activity to 35 nM. This is likely due to a drastic increase in α-helicity, from 28% for the product of library one, to 54% for the library two product, as measured by CD spectroscopy. Both libraries contained AAAC four-peptide sequence adjoined to the N-terminus, to be reacted with 5-iodoacetamidofluoracein for fluorescence polarisation assay analysis.

Li and co-workers also developed a miniprotein inhibitor of the p53/ Hdm2 interaction, using the scorpion venom toxin BmBKTx1 as the surrogate structure for their sequences.[85] The short chain K^+ channel toxin BmBKTx1 consists of an N-terminal α-helix connected to a C-terminal antiparallel β-sheet by three disulfide bonds.[86] The hotspot residues from the α-helical p53 active site can therefore be transposed onto the BmBKTx1 α-helix to create a p53 miniprotein mimetic.
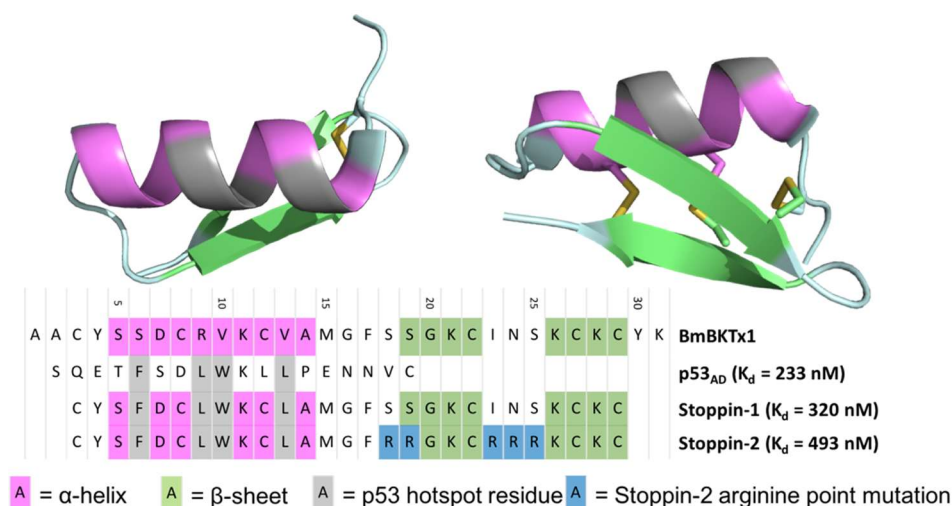
Figure 1.15: Design of Stoppin-1 and Stoppin-2 based on the BmBKTx1 sequence. The α-helical sections are shown in magenta, β-sheets in green, key p53 hotspot residues in grey and arginine mutations in stoppin-2 in blue.[85]

Li and co-workers utilised this to design stoppin-1, by transposing four hotspot residues onto the BmBKTx1 sequence and truncating the peptide by two residues at the C- and N-termini (Figure 1.15).[85] This peptide was synthesised by solid phase techniques and bound to Hdm2 with a K_d of 320 nM, comparable to that of the unmodified p53 sequence, 233 nM. However, when the HTC116 cancer cell line was treated with stoppin-1, there was no drop in cell viability as would be expected of a p53/Hdm2 inhibitor. Li and co-workers reasoned that this may be due to poor cell permeability of the stoppin-1, so modified the sequence with addition of five arginine residues to facilitate cross-membrane transport, producing stoppin-2. Although this resulted in a slight decrease in efficacy, K_d 493 nM, the improved cell permeability resulted in a significant decline in viability of HTC116 cells exposed to stoppin-2.

1.3.2.9 *Designing Miniprotein Inhibitors of the Androgen Receptor (AR) Co-activator Interaction*

Co-activator proteins for nuclear receptors (NRs) make good targets for the design of mimetics. They regularly follow the same pattern of hotspot residues, binding through an LXXLL motif, and regulate the interaction between the NR and its respective hormone. Androgen receptor (AR) co-activators bind through a FXXFL motif and so also make good targets for miniprotein design. Regulating the binding of AR co-activators is thought to have potential therapeutic benefit for some cancers, such as prostate cancer.[87]

Brunsveld and co-workers took a different approach to the design of AR co-activator mimetics,[88] where they designed inhibitors based on several different surrogate protein sequences. The proteins they chose to base their designs on were apamin (bee venom), κ-heuftoxin1, CD4M3 (from work previously described by Vita and co-workers), and Om-toxin3. From these four proteins they designed eight different sequences, each with the FXXLF motif installed as part of an α-helix (Figure 1.16). Occasionally an additional charged residue was mutated to arginine to avoid unfavourable charge-charge interactions.
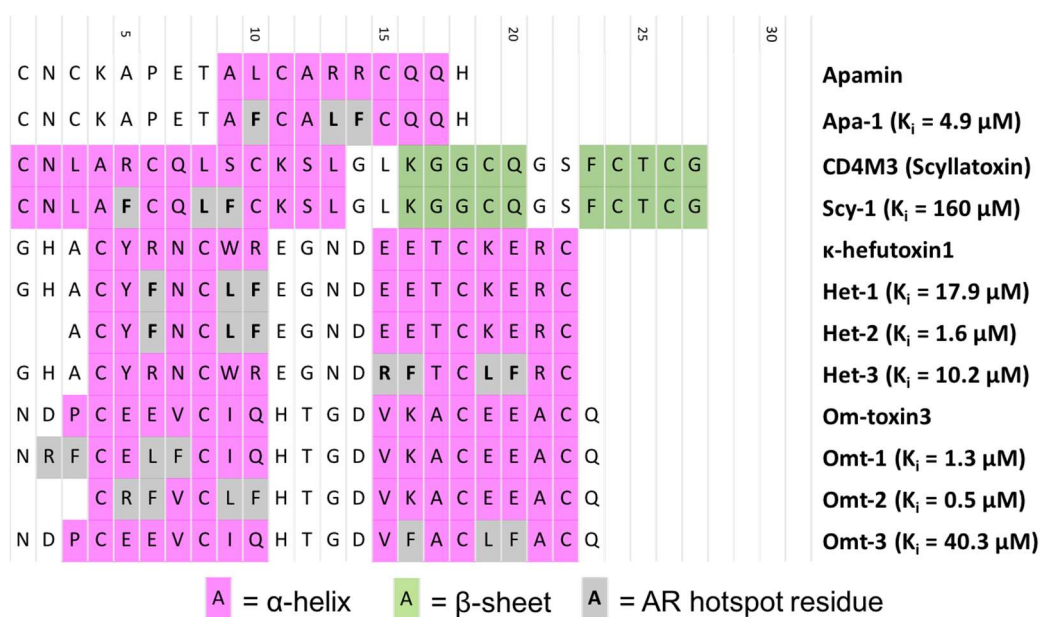
Figure 1.16: Design of AR co-activator FXXLF mimetics from the structure of four different miniproteins. α-helical (magenta) and β-sheet (green) secondary structure is shown, along with mutations from the original sequence (grey).[88]

Of the eight analogues designed, only one, Scy-1, showed low binding affinity. All seven others showed moderate activity, with the best sequence, Omt-2, having a K_i of 0.5 µM. The high hit rate is likely due to the short length of the FXXLF binding motif, as well as the promiscuity of the androgen receptor, which interacts with many natural co-activators. The authors also found that truncating the protein to shorten the α-helix, and thus limit the amount of unused α-helical residues on the miniprotein, impacted the activity significantly.

Brunsveld and co-workers had limited success in improving the activity of the best candidates from this series of miniproteins.[89] They used rational design to introduce mutations with the aim to both stabilise the secondary structure of the miniprotein, and

improve the affinity for the hotspot residues, to develop an understanding of the SAR of each residue with respect to AR-coactivator affinity. Despite producing a series of 31 sequences based on the structures of Het-2 and Omt-1, only a select few sequences showed improved affinity (Figure 1.17). It was found for both Het-2 and Omt-1 that replacing the second PXXLF phenylalanine residue with the unnatural amino acid *p*-chlorophenylalanine improved activity 2-3 fold. The N → E mutation in Het-2e improved activity ten-fold to 0.2 μM. However, all other attempts resulted in a small to moderate loss in activity, suggesting that the secondary structure of the surrogate miniproteins is already optimised for interaction with the AR. Notably, the phage display techniques regularly used in miniprotein activity studies were not employed here, and this may have afforded greater success.

Figure 1.17: SAR studies on the AR-coactivator series developed by Brunsveld, showing the sequences with improved affinity. A-helical secondary structure (magenta), FXXLF hot-spot residues (grey) and point mutations (blue) are labelled.[89]

1.3.2.10 Designing Miniprotein Inhibitors of the Oestrogen Receptor (ER) Co-activator Interaction

Another nuclear receptor which attracts a great deal of attention in developing miniprotein inhibitors is the oestrogen receptor (ER). ER co-activators interact through a five-peptide α-helical LXXLL motif. Design of inhibitors for the ER-co-activator interaction has helped to improve our understanding of the receptor function.

Wittiliff and co-workers designed oligopeptide ER co-activator mimetics containing the LXXLL motif.[90] In order to induce helicity amide or disulfide bridges were incorporated between residues spaced apart by one helical turn. They synthesised a series of oligopeptides which incorporated the LXXLL leucine residues, as well as an isoleucine and lysine that had been identified as important when reviewing known natural LXXLL

containing proteins. Of these peptides, which all had activity in the nanomolar range, the best was PERM-1 with a K_i to ERα of 25 nM (Figure 1.18).

Figure 1.18: Crystal structure of PERM-1 bound to ERα (PDB Code: 1PGC), showing the key binding leucine residues in magenta, and other important residues in blue.[90]

Interestingly, CD spectroscopy of PERM-1 indicates a lack of helical secondary structure in solution. However, crystallographic evidence clearly shows defined helical structure when it is bound to ERα (Figure 1.18, PDB code: 1PGC). The increase in activity between the five-amino acid LGGLL sequence and PERM-1 suggests that the disulfide bridge contributes significantly to the formation of the α-helical conformation required for binding. By using bridging amino acid residues spaced at one helical turn from each other, Wittilif and co-workers were able to produce a short, nine-amino acid, ER co-activator mimetic with remarkable affinity for its target.

Brunsveld and co-workers also developed a miniprotein mimetic of the LXXLL motif based on the peptide sequence of apamin, a protein in bee venom.[91] Apamin consists of an 18-amino acid sequence containing an α-helix, which is held by two disulfide bridges to a loop on the opposite face of the protein. Brunsveld and co-workers designed two phage display libraries in attempt to optimise the sequence to mimic ER co-activators (Figure 1.19). Interestingly, they opted to randomise most of the residues on the α-helix sequence, including the LXXLL hotspot residues, although they limited the randomisation to either charged/hydrophobic residues (Figure 1.19, blue) or small residues (Figure 1.19, pink).

28

Figure 1.19: The design of ER-Apamin-1 and ER-Apamin-2 from the apamin protein sequence, using two phage display libraries. The α-helical portion of apamin is shown in magenta, and the residues randomised by phage display in blue and grey. LXXLL hotspot residues are highlighted in bold.[91]

Enrichment of the phage display libraries established two high affinity miniprotein ER co-activator mimetics. The first library yielded ER-Apamin-1, with an IC_{50} of 2.1 μM towards ERα. The second yielded ER-Apamin-2, with an IC_{50} of 1.3 mM. In both libraries the LXXLL configuration of the hotspot residues was found to be retained. There was a notable preference for hydrophobic amino acids, with leucine particularly favoured in library 2. Computational overlays with a natural LXXLL crystal structure suggested that ER-Apamin-2 may be able to bind via either of the overlapping LXXLL motifs.

1.3.3 Using Stapled Helices to Inhibit PPIs

There is an increasing interest in using stapled peptides to generate stable α-helices that can be utilised in PPI inhibition.[92–95] A peptide staple refers to a covalent link between two residues on a polypeptide which fixes them a set distance apart. Use of a well-designed staple of the correct size can encourage the formation of an α-helix, or stabilise a pre-existing one (Figure 1.20).
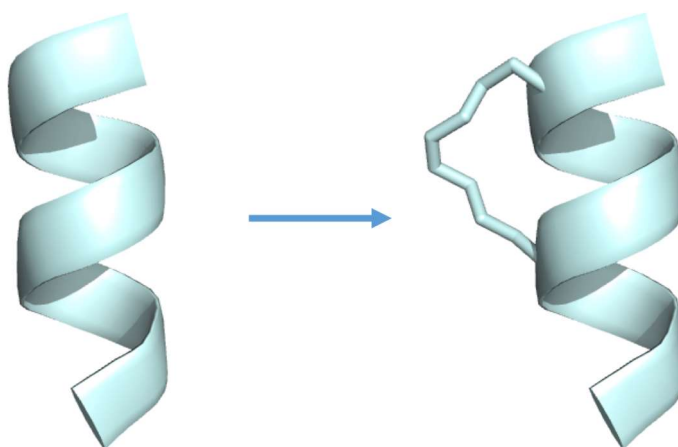
Figure 1.20: Attaching a staple across one turn of an α-helix.

An early example was published by Mierke and co-workers in 1997.[96] The authors postulated that a disulfide bridge between two cysteine residues in the *i* and *i* + 3 positions could bestow stabilisation similar to the *i*, *i* + 3 H-bonding observed in an α-helix. This technique offers the advantage that it utilises the naturally occurring amino acid, cysteine. As a result the desired peptides are compatible with unmodified peptide synthesis machines, and potentially even bacterial production, keeping the cost low. On the other hand, the disulfide bond would be quickly metabolised in the body, resulting in short half lives.

A second example of peptide stapling was developed by Rosenblatt. This involves the formation of a lactam bridge between two amino acid residues.[97] Again, naturally occurring amino acids can be utilised, as a lactam bridge can be formed between lysine and aspartic or glutamic acid, which facilitates the synthesis. However, as with the disulfide bridge, this results in susceptibility to degradation by biological processes, in this case peptins.
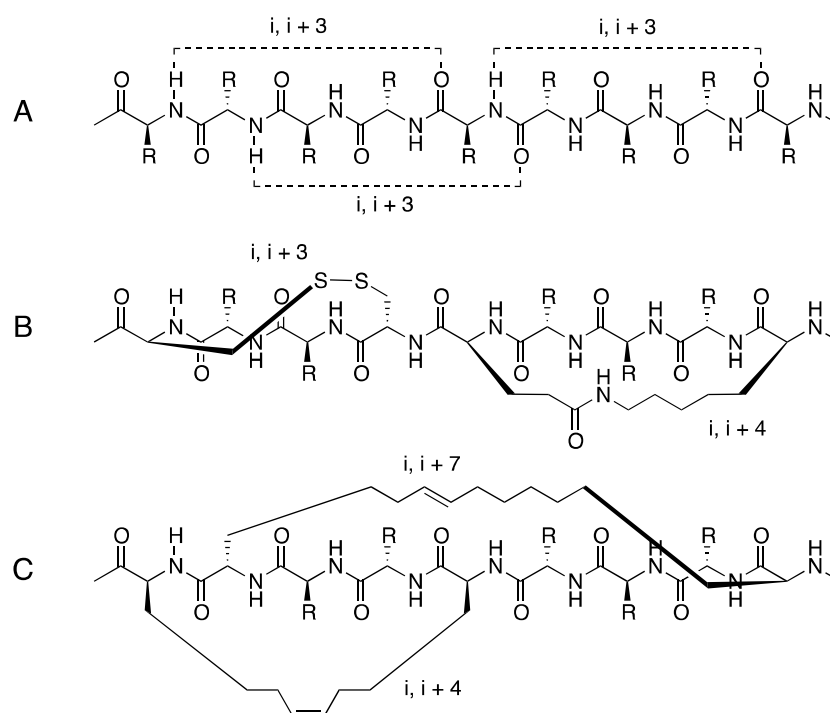
Figure 1.21: A) A polypeptide demonstrating the *i*, *i* + 3 H-bonding observed in α-helices. B) A polypeptide demonstrating an i, i + 3 disulfide bridge,[96] and an i, i + 4 lactam bridge.[97] C) A polypeptide demonstrating an i, i + 4 and an i, i + 7 hydrocarbon bridge.[98,99]

Probably the most popular form of peptide stapling to date is the hydrocarbon stapling method developed by Grubbs.[98] Unlike the previous examples, this method does not use naturally occurring amino acids. Instead amino acids containing a hydrocarbon side chain

with a terminal olefin are required. Once these modified residues are incorporated into the polypeptide, treatment with Grubbs catalyst[100–102] induces a cross metathesis reaction to construct the staple. The product is an all-hydrocarbon staple containing an alkene. These staples are much more inert to biological processes than the previous examples and as terminal alkenes do not exist in naturally occurring amino acids, the reaction also offers improved selectivity, although at the expense of requiring an increased number of synthetic steps.

1.3.3.1   Using Stapled BH3 Helices to Inhibit Bcl-2/X_L

The BH3 domain α-helix has inspired the development of many stapled derivatives.[103–112] It consists of an α-helix which binds to many different apoptosis inducing domains through a single lipophilic face. In general, BH3 helices bind *via* a series of lipophilic residues spaced one helical turn apart.[113,114] For example, the Bim-BH3 domain contains binding hotspots at Val74, Leu78 and Ile81 (Figure 1.22).[115] It has been hypothesised that BH3 mimetics may have the ability to selectively induce apoptosis and thus be utilised therapeutically in reducing the viability of cancer cells.
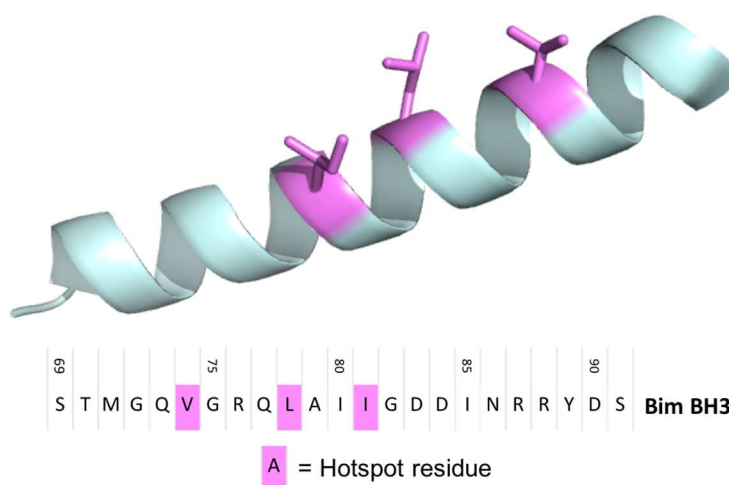
Figure 1.22: The Bim-BH3 domain showing Val74, Leu78 and Ile81 hotspot residues in magenta (PDB code: 2XPX).[115]

In a seminal study by Walensky and coworkers,[103] it was shown that by applying a 'staple' to the Bid-BH3 α-helix, it was possible to improve the helicity and thus the efficacy towards Bcl-2. They used the hydrocarbon stapling method, developed by Grubbs,[98] and three different olefin-containing amino acids with varying chain length and chirality to make up their staples (Figure 1.23). This allowed them to design staples which bridged one helical turn, from the *i* to *i + 4* positions, or two helical turns, from the *i* to *i + 7*

positions. Of the sequences that were synthesised, the most helical was SABH$_A$, with a helicity of 87.5% in solution, compared to the unstapled Bid-BH3 sequence which has a helicity of 15.7%, as measured by CD spectroscopy. The longer *i, i + 7* staple proved to be less effective, resulting in only a small increase in helicity to 35.6%. The increased helicity in SABH$_A$ lead to a six-fold increase in affinity for the Bcl-2 binding site, to give a K$_d$ of 38.8 nM compared to 269 nM for the unstapled Bid-BH3 peptide.

	1				5					10					15					20					% helicity	K$_d$
Bid-BH3	E	D	I	I	R	N	I	A	R	H	L	A	Q	V	G	D	S	N$_L$	D	R	S	I	W	15.7 %	269 nM
SAHB$_A$	E	D	I	I	R	N	I	A	R	H	L	A	S5	V	G	D	S5	N$_L$	D	R	S	I	W	87.5 %	38.8 nM
SAHB$_{A(G>E)}$	E	D	I	I	R	N	I	A	R	H	L	A	S5	V	E	D	S5	N$_L$	D	R	S	I	W	77.8 %	-
SAHB$_B$	E	D	I	I	R	N	I	S5	R	H	L	S5	Q	V	G	D	S	N$_L$	D	R	S	I	W	85.5 %	-
SAHB$_C$	E	D	I	I	R	N	I	A	S5	H	L	A	S5	V	G	D	S	N$_L$	D	R	S	I	W	59.7 %	-
SAHB$_D$	E	D	I	I	R	N	I	A	R	R5	L	A	Q	V	G	D	S8	N$_L$	D	R	S	I	W	35.6 %	-

N$_L$ = Norleucine

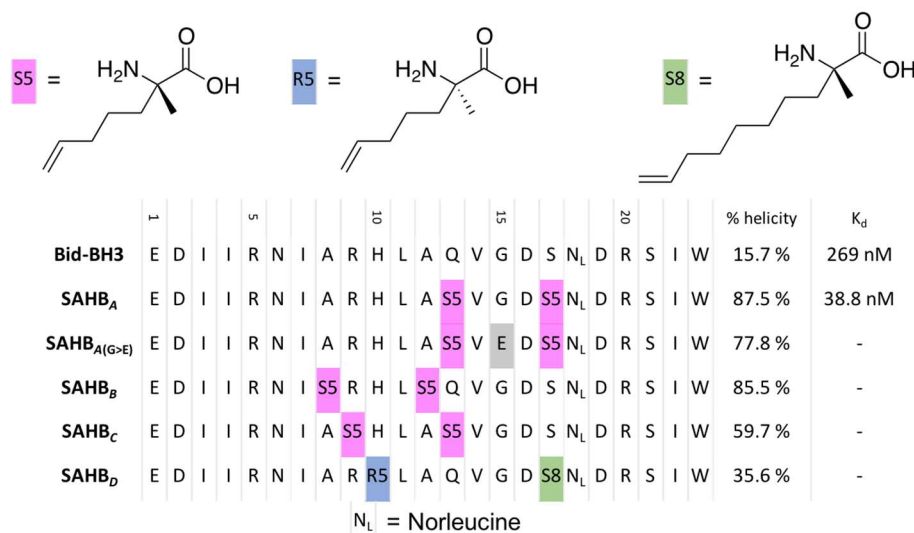**Figure 1.23:** Series of stapled helices synthesised by Walensky and coworkers,[103] demonstrating the effect of staple size and position on the α-helicity measured by CD spectroscopy.

The SAHB$_A$ was applied to Jurkat T-cell lymphoma cells in order to test its ability to activate apoptosis. After 20 hours, cells treated with SAHB$_A$ demonstrated staining by annexin V, an apoptotic marker, whereas the control Bid-BH3 with SAHB$_{A(G>E)}$ peptides did not. *In vivo* studies were conducted in mice transplanted with RS4:11 leukemia cells. Intravenous treatment of SAHB$_A$ dosed at 10 mg.kg^{-1} lead to a median survival rate of 11 days, compared to 5 days in the control. Walensky and co-workers had demonstrated that stapled α-helices could be used as PPI inhibitors with an observable medical effect when administered intravenously in mice.

Alleman and co-workers were able to modify the BH3 helix staple such that the secondary structure became sensitive to light exposure.[105] They designed a staple containing an azo double bond (Figure 1.24). By exposing this azo bond to light it is possible to switch between the *trans* and *cis* conformations, thereby changing the length of the staple. With the staple placed at particular locations along the polypeptide sequence, this could affect

whether the helicity was turned 'on' or 'off'. The trans form of the staple encouraged helicity in an *i, i*+11 conformation, and the cis form in *i, i*+7 or *i, i* + 4 conformations.
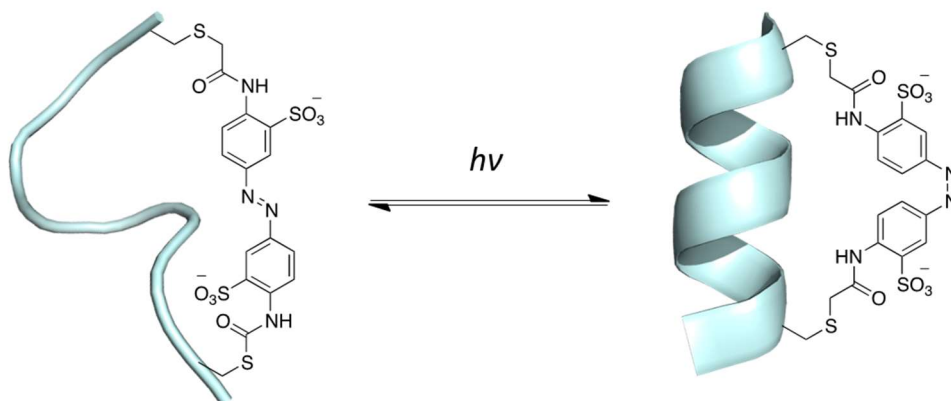

Figure 1.24: Using light-sensitive staples to change the helicity of a polypeptide.[105]

Alleman found that by inserting the staple into the Bak and Bid BH3 helices, the affinity towards Bcl-X_L became dependent on the extent of light exposure. For the *i, i*+4 and *i, i*+7 variants, where the *cis* conformation post irradiation promotes helix formation, a particularly dramatic increase in affinity was observed. Both cases exhibited at least a 20-fold increase in affinity towards Bcl-X_L after irradiation (Table 1.1). The *i, i*+11 variant, where the non-irradiated trans conformation tended towards helix formation, did not demonstrate as much variation in activity. The non-irradiated conformer was only two-fold more potent than the irradiated form. Every peptide tested showed at least a two-fold increase in potency from their unstapled parent sequence. This phenomenon is useful in mechanistic studies on Bcl family proteins *in vitro*, where the activity can be switched on and off using a light source. However, it has limited potential *in vivo* as a therapeutic treatment, as the required wavelength of light cannot penetrate as deep as would be necessary into tissue.

Table 1.1: The effect of light on the affinity of the different light-sensitive stapled BH3 peptides designed by Alleman and co-workers.[105]

	K_d (nM)		
Peptide form	Bak *i*+7	Bak *i*+11	Bid *i*+4
parent	134±16	328±19	117±48
dark-adapted	825±157	21±1	1275±139
irradiated	42±9	48±10	55±4

The development of staple stabilised BH3 helices has aided our understanding of the BH3 interactome within the cell,[104,107,109] and the structural biology of BH3 binding sites,

including how structural changes can affect the protein function.[106,108,110,111] However, the stapled BH3 peptides have shown limited success in therapeutic applications due to poor cell membrane permeability and metabolic instability. A recent study even demonstrated a negligible difference in the effects of stapled peptides and their non-stapled parents.[112]

1.3.3.2   Using a Stapled p53 Helix to Inhibit Mdm2

Another target which attracts attention in developing stapled α-helices is the Hdm2 binding p53 helix.[116–125] p53 is a transcription factor responsible for triggering apoptosis in response to genetic damage, and is thus highly important in preventing cancers from developing.[126] Hdm2 modulates p53 through a binding interaction that neutralises p53 activity by encouraging ubiquitylation and eventual degradation.[127] The interaction of p53 with Hdm2 occurs via an α-helical segment of 16 amino acid residues and relies on the hotspot residues Phe19, Trp23 and Leu26, which are in an *i*, *i*+4, *i*+7 conformation.[128] It is hoped that by stapling this short α-helix, to encourage secondary structure in solution, a stable tool compound or even a drug can be identified for use in targeting the p53/Hdm2 PPI.
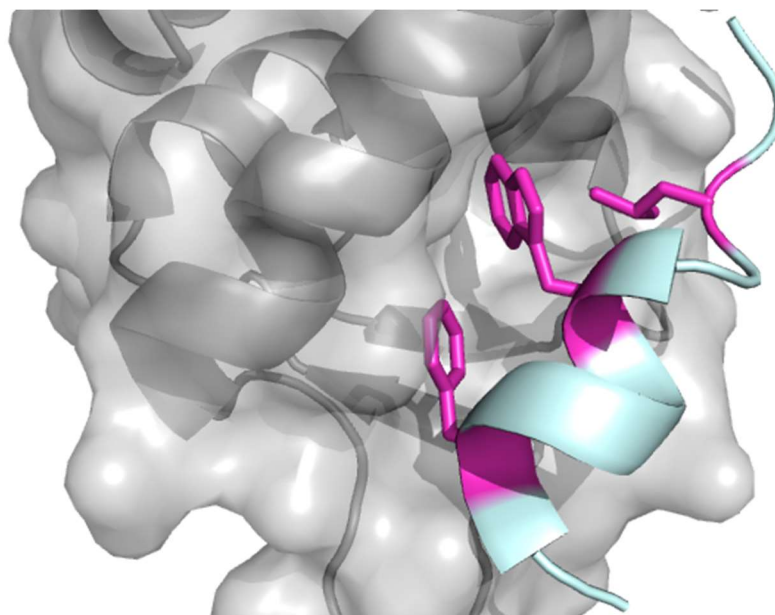
Figure 1.25: The p53 transactivation domain (cyan) binding to Mdmx (grey). The hotspot residues are shown in magenta. (PDB code: 3DAC)[128]

Bernal and co-workers used the 16-amino acid transactivation domain to design stapled α-helical p53 mimetics.[116] They began by designing four different analogues of p53, each with an *i, i*+7 hydrocarbon staple in a different position (Figure 1.26). Of these, only SAH-p53-4 demonstrated significant α-helicity (59%) by CD spectroscopy, and this resulted in remarkable efficacy, with the K_d measured as 92 pm. However, despite this,

when Jurkat T-cell lymphoma cells were treated with SAH-p53-4, no cell death was observed. It was hypothesised that this was due to a lack of cellular uptake of the compound, so point mutations were made to give the peptide a positive charge at physiological conditions (pH 7.4) and facilitate cellular uptake.[129] After making five point mutations, the authors discovered SAH-p53-8. Although the K_d was lower, at 55 nM, this stapled helix successfully promoted cell death in Jurkat T-cells due to improved cellular uptake.

	15					20					25					30		% helicity	K_d (nM)	Charge (pH 7.4)	Cell permeable	Cell death
WT p53	L	S	Q	E	T	F	S	D	L	W	K	L	L	P	E	N	NH₂	11 %	410	-2		
SAH-p53-1	L	S	Q	E	T	F	S	D	R8	W	K	L	L	P	E	S5	NH₂	25 %	100	-2		
SAH-p53-2	L	S	Q	E	R8	F	S	D	L	W	K	S5	L	P	E	N	NH₂	10 %	400	-2		
SAH-p53-3	L	S	Q	R8	T	F	S	D	L	W	S5	L	L	P	E	N	NH₂	12 %	1200	-2		
SAH-p53-4	L	S	Q	E	T	F	R8	D	L	W	K	L	L	S5	E	N	NH₂	59 %	0.92	-2		
SAH-p53-5	L	S	Q	E	T	F	R8	N	L	W	K	L	L	S5	Q	N	NH₂	20 %	0.80	0	✓	
SAH-p53-6	L	S	Q	Q	T	F	R8	N	L	W	R	L	L	S5	Q	N	NH₂	14 %	56	+1	✓	
SAH-p53-7	Q	S	Q	Q	T	F	R8	N	L	W	K	L	L	S5	Q	N	NH₂	36 %	50	+1	✓	
SAH-p53-8	Q	S	Q	Q	T	F	R8	N	L	W	R	L	L	S5	Q	N	NH₂	85 %	55	+1	✓	✓
SAH-p53-8$_{F19A}$	Q	S	Q	Q	T	A	R8	N	L	W	R	L	L	S5	Q	N	NH₂	39 %	>4000	+1	✓	
UAH-p53-8	Q	S	Q	Q	T	F	R8	N	L	W	R	K	K	S5	Q	N	NH₂	36 %	100	+1	✓	

A = Basic residue change
A = Key binding residue

Figure 1.26: Discovery of SAH-p53-8 by modifying the Hdm2 binding helix from p53. The stapled residues are shown in blue and magenta, and point mutations made to basic residues for the point of cell permeability in green. Key binding residues are in orange.[116]

In later work it was shown that SAH-p53-8 acted through specific binding to Hdm2 and Hdmx, thereby restoring p53 activity in cancer cells that overexpressed these proteins.[117] The crystal structure of SAH-p53-8 bound to Mdm2 shows that not only does it bind at the p53 site, but additional interactions are picked up between the staple and the protein, further increasing the affinity (Figure 1.27).[122]
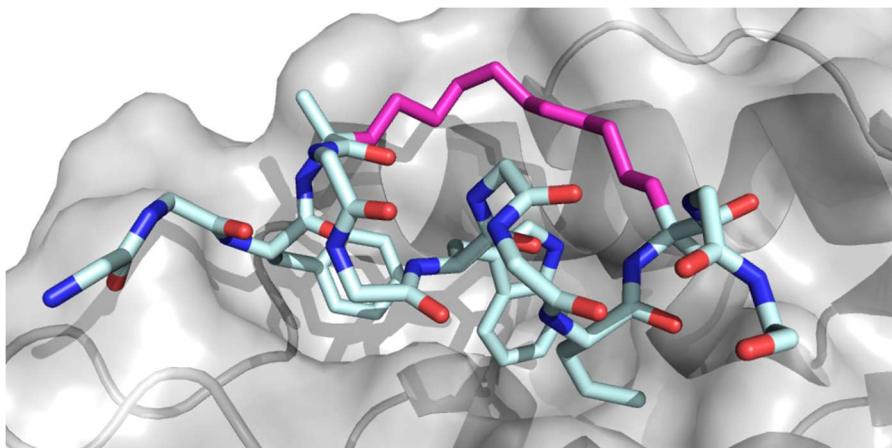
Figure 1.27: Crystal structure of SAH-p53-8 bound to Mdm2, showing the interaction of the staple with the protein pocket.[122] The staple is shown in magenta, and non-hotspot residues have been hidden for clarity. (PDB code: 3V3B).

Further work has been carried out to develop stapled peptides to target Hdm2, based on the structure of pDI (Figure 1.28).[124,130,131] PDI is an unstapled peptide sequence discovered by Pazgier and co-workers using phage display optimisation methods.[130] The number of comprising amino acids is reduced compared to the wild type p53 helix, with 12 instead of 16, while the affinity is increased from a K_d of 410 nm to 19.7 nm. Two different research groups simultaneously developed stapled peptides based on the structure of pDI. Chang and co-workers developed ATSP-7041 using an *i, i*+7 staple.[131] By introducing the staple and making several modifications to help boost solubility (a H21E mutation and two alanines incorporated at the C-terminus) and binding affinity (a cyclobutyl hotspot residue) they achieved an affinity of 0.9 nM. The stapled helix was cell-permeable and reactivated p53 function in cancer cells at low μM concentrations. In the second, very similar investigation, Brown and co-workers developed sMTide02A, an 11-amino acid stapled peptide with a Kd of 6.76 nM.[124] They also substituted one of the hotspot residues to improve binding affinity; this time Trp23 was replaced with the 6-chloro substituted isomer, and the same H21E mutation was made to improve solubility. The resulting peptide could cross cell membranes and encouraged apoptosis in cancer cells with reduced p53 function.
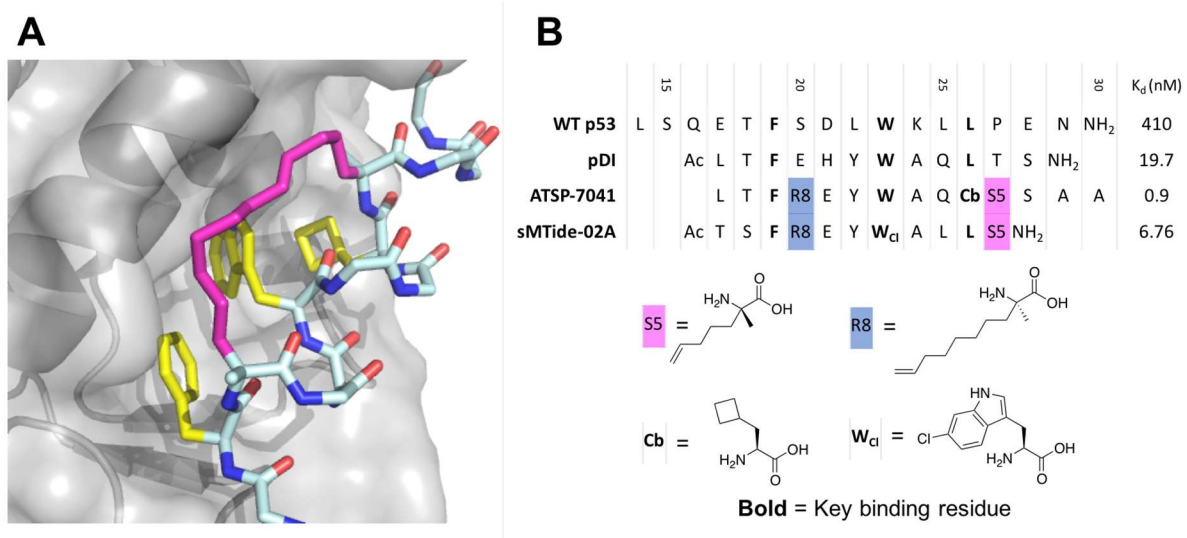
A

B

					15				20					25				30		K_d (nM)
WT p53	L	S	Q	E	T	**F**	S	D	L	**W**	K	L	**L**	P	E	N	NH₂			410
pDI		Ac	L	T	**F**	E	H	Y	**W**	A	Q	**L**	T	S	NH₂					19.7
ATSP-7041		L	T	**F**	R8	E	Y	**W**	A	Q	**Cb**	S5	S	A	A					0.9
sMTide-02A		Ac	T	S	**F**	R8	E	Y	**W$_{cl}$**	A	L	**L**	S5	NH₂						6.76

S5 =

R8 =

Cb =

W$_{Cl}$ =

Bold = Key binding residue

Figure 1.28: A) Crystal structure of ATSP-7041 bound to Mdmx (PDB code: 4N5T).[131] The hydrocarbon staple is shown in magenta, and the key residues in yellow. B) Peptide sequences of ATSP-7041[131] and sMTide-02A[124], both of which are derived from pDI.[130]

The use of stapled peptides has proven to be a very effective strategy for developing cell-permeable inhibitors of the p53/Hdm2 PPI. Optimisation of the stapled helix sequences has yielded several helices with nanomolar activity, which are able to activate p53 and promote cell death in cancer cell lines. The stapled peptides have also been shown to maintain high efficacy towards Hdm2 mutants that have previously proven resistant to small molecule p53/Mdm2 inhibitors, such as nutlin-3a.[125]

1.3.3.3 Further Potential of Stapled Helices as PPI Inhibitors.

The stapling of α-helices has been highlighted as a versatile method for developing inhibitors of PPIs. As well as the two examples described they have also been used to target other oncogenic PPIs such as induced myeloid leukemia cell differentiation protein (MCL-1),[132] notch,[133] β-catenin,[134–136] EED[137] and eIF4G;[138] targets for infectious diseases such as human immunodeficiency virus (HIV)-1[139–142] and hepatitis-C;[143] endocrine modulators such as oestrogen receptor α;[144] and neurological targets such as the *N*-methyl-D-aspartate (NMDA) receptor.[145] Despite their adaptability, poor cell-membrane permeability and high rate of metabolism by peptins has meant there has been little success so far in using them as therapeutic compounds.

1.3.4 Using small molecule α-helical mimetics to inhibit PPIs

Another approach that has been explored in the search for inhibitors of PPIs is the development of small molecule α-helical mimetics.[92,146,147] In principle a core organic framework is employed that can be functionalised in order to express moieties in the same

positions as if they were on an α-helix (Figure 1.29). A small molecule mimetic is designed by selecting an appropriate scaffold that is able to display functionality in the same positions as hot-spot residues on the PPI α-helix.
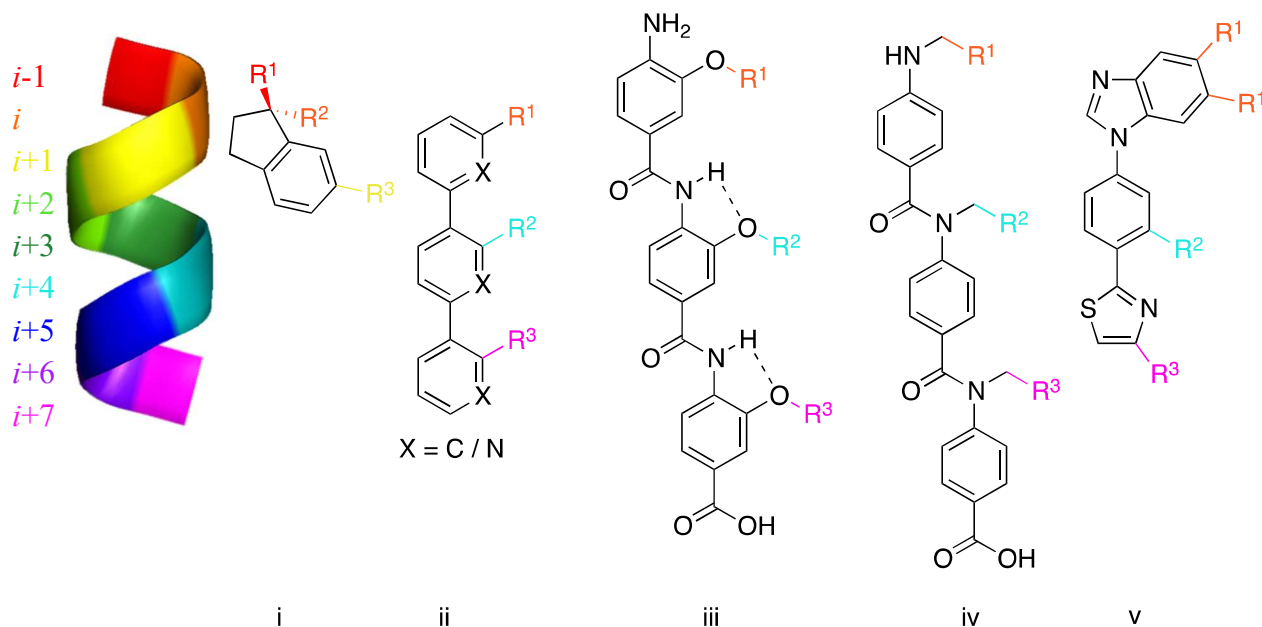

i	ii	iii	iv	v

Figure 1.29: Examples of α-helical mimetics: i) 1,1,6-substituted indane,[148] ii) terphenyl,[149] iii) oligobenzamide,[150] iv) *N*-alkylated oligobenzamide,[151] and v) 5-6-5 imidazole-phenyl-thiazole.[152]

1.3.4.1 *Using 1,6-Substituted Indanes to Mimic i, i+1 Residues*

The first example of an α-helical mimetic was developed by Horwell and co-workers.[148,153] They were looking to design a mimetic with affinity for tachykinin neuroreceptors NK_1 and NK_3. These receptors bind to an α-helix *via* either Phe-Phe or Trp-Phe residues in an *i*, *i+1* configuration. Molecular modelling studies were used to select a 1,6-indane structure which overlaid well with the helix and displayed the indane functionalities at the same vector as the *i*, *i+1* amino acid residues (Figure 1.30, A). A small library of indane based *i*, *i+1* mimetics was synthesised, and several compounds identified that had μM affinity for the NK_1 and NK_3 receptors. The most potent of these (**1**, Figure 1.30) had a measured IC_{50} of 2.4 μM for NH_1 and 3.0 μM for NK_3, demonstrating the potential of using α-helical scaffold mimetics towards development of small molecule PPI inhibitors.
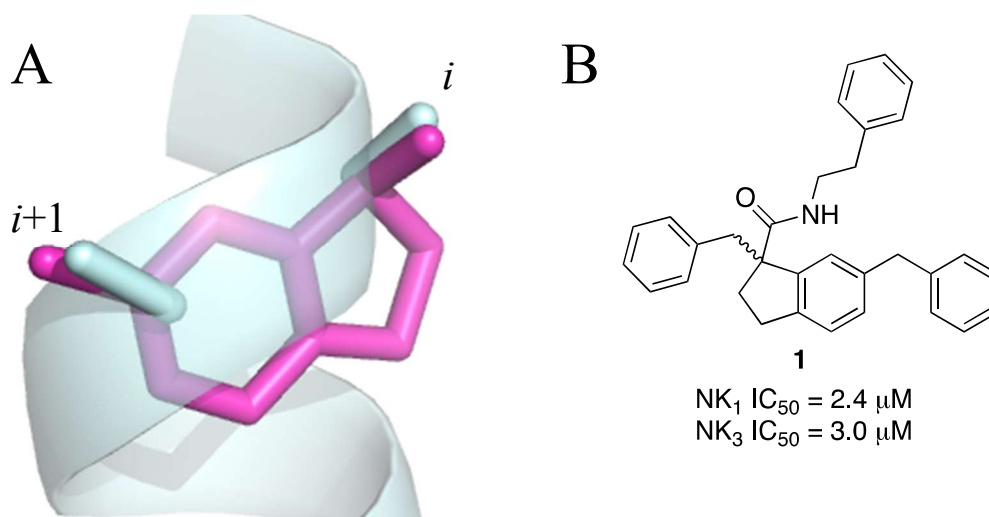
Figure 1.30: A) Alignment of Horwell's indane scaffold with the *i*, *i* + 1 positions on an α-helix. B) Structure of the most efficacious indane mimetic discovered.[148,153]

1.3.4.2 Development of Terphenyl i, i+3/4, i+7 Mimetics

Although it had been shown that α-helical structure could be successfully mimicked using a small-molecule scaffold, this was limited to a small subset of PPIs where the hotspot residues are in an *i*, *i*+1 conformation. Hamilton realised that if a terphenyl scaffold was functionalised at each phenyl ring to display an amino acid side chain, these would align with the *i*, *i*+3/4 and *i*+7 positions on an α-helix.[149] This is due to *ortho*-positioning of the functional groups on the phenyl rings, which cause the terphenyl scaffold to lose planarity and adopt a staggered conformation (Figure 1.31). This is applicable to a much larger set of PPI targets as it is a much more common hot-spot configuration.
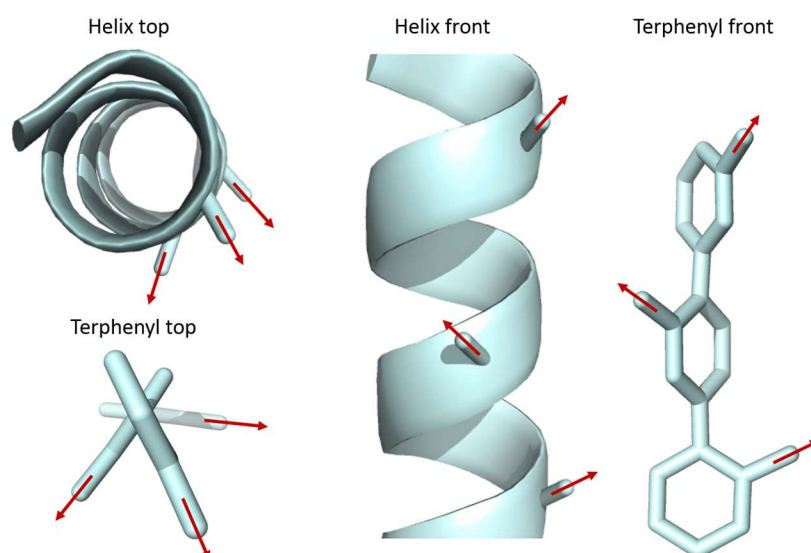

Figure 1.31: The terphenyl functionality aligned with the *i*, *i* + 4 and *i* + 7 positions on an α-helix.

The first target selected was calmodulin (CaM), which binds to an α-helical section of smooth muscle myosin light-chain kinase (smMLCK). Alanine scanning mutagenesis had

shown that smMLCK binds *via* three hot-spot residues (Trp800, Thr803 and Val807) in an *i*, *i*+3, *i*+7 configuration. Hamilton designed a direct mimetic of these three residues on the terphenyl scaffold and found that it inhibited CaM with an IC_{50} of 800 nM (the IC_{50} of the 20-aa smMLCK peptide is 80 nM). By changing the tryptophan mimetic to a 1-napthyl group, they discovered compound **2** (Figure 1.32), which was able to better mimic the binding of smMLCK to CaM to result in an IC_{50} of 9 nM. This demonstrated the potential of using small molecule mimetics of *i*, *i*+3/4, *i*+7 α-helical residues to inhibit PPIs.
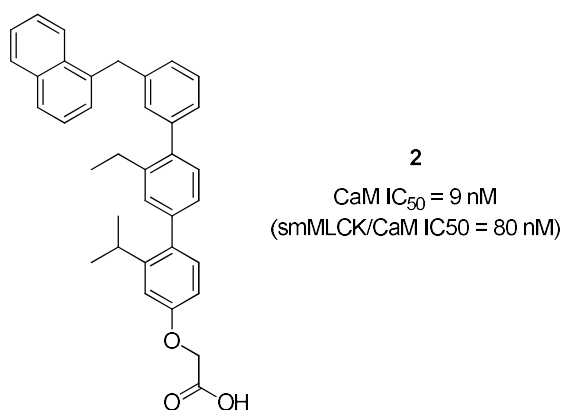

2
CaM IC_{50} = 9 nM
(smMLCK/CaM IC50 = 80 nM)

Figure 1.32: Structure of smMLCK mimetic discovered by Hamilton which binds to CaM with an IC_{50} of 9 nM.[149]

1.3.4.3 Using Oligamides to Ease Synthesis and Enforce Conformational Structure of i, i+3/4, i+7 Mimetics

Although the terphenyl structures could be used successfully in mimicking an α-helix, they were difficult to synthesise and their high lipophilicity resulted in poor aqueous solubility, meaning that they were difficult to use in cell based assays, reducing their value as chemical tools. In order to counteract this, in several examples benzamides have been used instead to mimic the α-helix *i*, *i*+3/4, *i*+7 positions (Figure 1.33). The amide bonds introduced facilitate synthesis, making the compounds amenable to solid-phase techniques, and also lower the pKa significantly, increasing solubility and 'drug-likeness'. The first example of an oligobenzamide mimetic was synthesised by Hamilton.[154] It was calculated that the hydrogen-bonding present in an oligopicolinamide based molecular scaffold (Figure 1.33, i) would hold the *O*-functionalised side chains on the same side of the molecule, and induce an amount of inwards backbone bending. Indeed this resulted in a conformation that closely mimicked the terphenyl *i*, *i*+3/4, *i*+7 mimetics. Using this scaffold they were able to design moderate μM inhibitors of the Bcl-

X_L/BAK interaction. However, the functionalised picolinic acid monomers proved synthetically challenging, limiting the diversity of oligomers achievable.
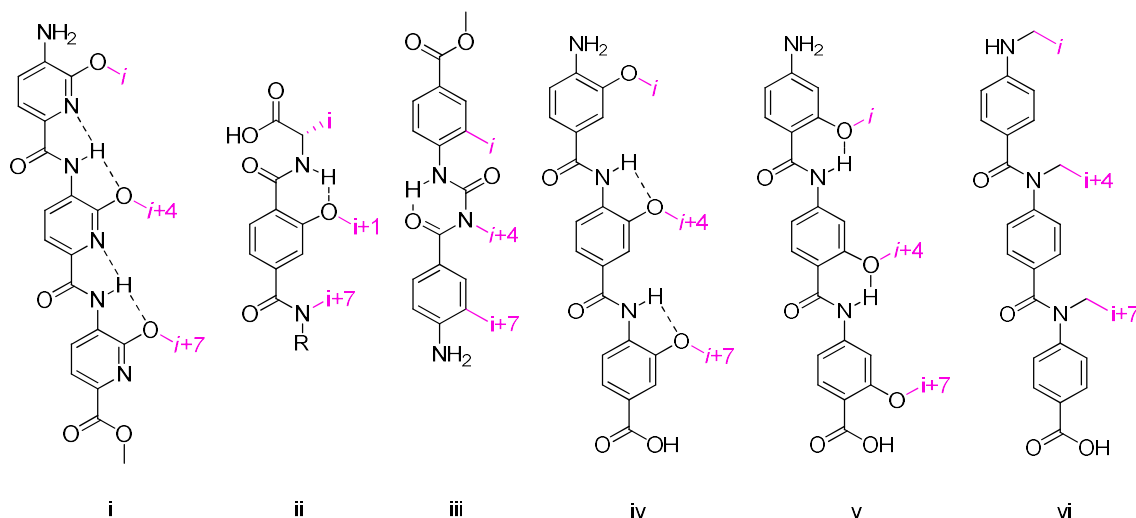

Figure 1.33: Different structures of oligobenzamide α-helical mimetics, including i) oligopicalinamide,[154] ii) terephthalamide,[155] iii) benzoylurea,[156–158] iv) 3-substituted oligamide,[150,159] v) 2-substituted oligobenzamide,[160] and vi) N-substituted oligobenzamide.[151]

Wilson investigated the mimetic effects of simpler oligobenzamides, with monomers that are more synthetically accessible. He developed the synthetic methodology required to make 3-substituted oligamides (Figure 1.33, iv)[150] and 2-substituted oligamides (Figure 1.33, v)[160] from amino benzoic acid monomers, which could be made in far fewer synthetic steps than the picolinic acid anologues. It was initially found that by masking the aniline as a nitro group he could carry out iterative peptide coupling/nitro reduction steps to reach the desired oligomeric product. However, although the trimer was easily synthesised using this methodology, the yields of the tetramer and pentamer oligomides were drastically lower due to the decreased solubility of the intermediates. Solid phase synthesis techniques using fluorenylmethyloxycarbonyl (Fmoc) protected starting monomers gave improved yields, but required a more lengthy synthesis of starting materials.[161]

Both oligamide regioisomers afforded inhibitors for the p53/Mdm2 PPI when functionalised with residues to mimic the p53 i, $i+4$ and $i+7$. The 3-substituted oligobenzamides produced the most efficacious p53 mimetics and the most potent example had an IC$_{50}$ of 1.0±0.11 μM, comparable to the free p53$_{15-31}$ oligopeptide, with a measured IC$_{50}$ of 1.2±0.04 μM.[159] The 2-substituted isomers were generally less potent, with the most efficacious of these having an IC$_{50}$ of 4.15±0.20 μM.[160] N-substituted oligobenzamides were also developed which were amenable to solid phase synthesis.[151]

Although these contained no conformationally restricting H-bonds, they showed moderate activity against p53/Mdm2 of up to 2.8±0.8 μM. Wilson hypothesised that the differing efficacy could be due to the changing degree of backbone bending as a result of weaker H-bonds, meaning the side chains do not align as well to the i, $i+3/4$, $i+7$ conformation. Of the three oligamide examples, the 2-substituted and *N*-substituted oligobenzamides exhibited the least backbone bending. Hamilton's H-bond rich oligopicalinamides had more backbone bending than either of these.

Hamilton *et al.* produced two further scaffolds, similar to the oligobenzamide α-helical mimetics, which relied on H-bonding to adopt the necessary conformation; the terephthalamide (Figure 1.33, ii)[155] and benzoylurea (Figure 1.33, iii)[156–158] scaffolds. Both of these serve to display functional groups so as to mimic the residues in the i, $i+3/4$ and $i+7$ positions to produce a Bcl-X$_L$/Bak inhibitor. A functionalised terephthalamide was able to inhibit Bcl-X$_L$/Bak with an IC$_{50}$ of 0.78±0.07 μM,[155] and a benzoylurea with a K$_i$ of 2.4 nM.[157]

More recently, Hamilton *et al.* showed that monomers from the various oligamide mimetics could be combined to create hybrid compounds, to customise the degree of backbone bending caused by intramolecular H-bonds.[162] However, these molecules tended to have lower efficacy for their target, and they were only able to achieve an IC$_{50}$ of 9.2±0.4 μM against p53/Hdm2.

1.3.4.4 *Other Examples of α-Helical Mimetic Small Molecule Scaffolds.*

Terphenyl and oligamide inspired α-helical mimetics are not the only scaffolds that have been developed, and some of the other examples have also afforded compounds with activity against PPIs (Figure 1.34).
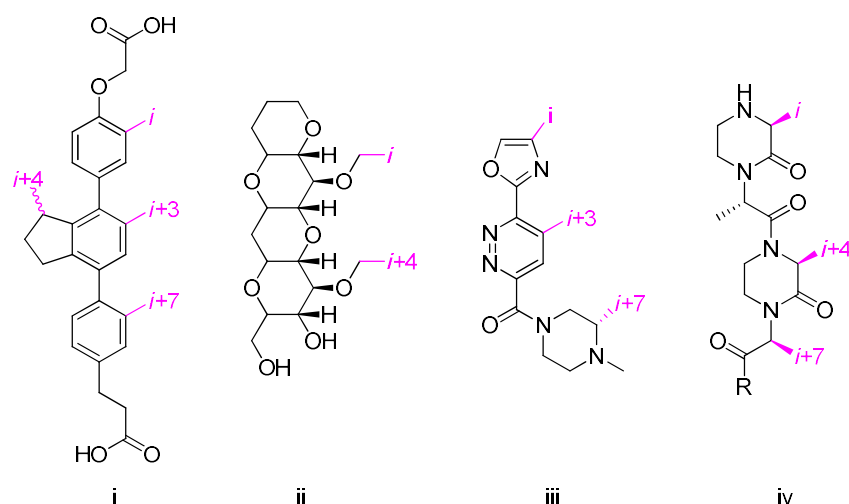
Figure 1.34: Other examples of α-helical mimetic scaffolds including: i) Diphenylindane,[163] ii) polycyclic ether,[164] iii) pyridazine-based polyheterocycle,[165,166] and oligooxopiperazine.[167]

Hamilton *et al.* combined his previous work on terphenyls[149] with Horwell's indane based *i*, *i*+1 mimetics[153] to design and synthesise a hybrid *i*, *i*+3, *i*+4, *i*+7 scaffold (Figure 1.34, i).[163] Although the molecule was designed to mimic these positions on the p53 helix to interact with Hdm2, no data on the efficacy of the molecule towards its intended target was published. Oguri and co-workers synthesised another scaffold which mimicked the *i*, *i*+4 positions (Figure 1.34, ii).[164] This design was based on a tetracyclic ether scaffold, which was selected to bestow aqueous solubility. The scaffold was functionalised with guanidinium groups to enable binding with the carboxylate groups on a polypeptide helix which contained aspartic acid residues in the *i* and *i*+4 positions. Reasonable selectivity was achieved over polypeptides with aspartic acid residues in different positions, and CD spectroscopy indicated that addition of their mimetic stabilised the helicity of the complementary peptide when bound.

Another scaffold designed with aqueous solubility in mind was a series of pyridazine based scaffolds from Rebek and co-workers (Figure 1.34, iii).[165,166] They attempted to instil a 'wet' edge on the face of the scaffold opposite to the lipophilic amino acid side chains, which contained heteroatoms to lower the clogP. As this part of the molecule is in theory solvent exposed when bound to the protein, this should have improved solubility without lowering lipophilicity. However, although some binding was detected for Bcl-X_L by fluorescence polarisation, it was not high enough to calculate an IC_{50} value.

Arora had more luck in his development of oxopiperazine scaffolds as α-helical mimetics (Figure 1.34, iv).[167] These heteroatom rich molecules were found to be soluble in aqueous

media at μM levels. By functionalising the scaffold with larger lipophilic residues to mimic p53 they successfully developed compounds with affinity for Mdm2, with their best having a K_d of 0.3±0.01 μM. They were also able to develop potent HIF mimetics, the best of which bound with a K_d of 0.16±0.06 μM, making these some of the most potent small molecule PPI inhibitors discovered at that time.

1.3.5 Targeting PPIs using the Conventional Drug Discovery Process

Although there have been successes in targeting PPIs using carefully designed α-helical mimetics, very few candidates for therapeutic application have been generated using the methods described previously. In the meantime the pharmaceutical industry has also had some small successes towards a therapeutic compound inhibiting a PPI.[168] As discussed in part 1.1, this is often a difficult and expensive process, due to the less druggable nature of the PPI. However, by combining traditional fragment screening to identify leads, with SAR to develop a candidate compound, progress has been made in some cases. Here, a few of the more successful case studies will be discussed.

1.3.5.1 Case Study 1: p53/Mdm2 and the Nutlins

The nutlins are named after Nutley, New Jersey, where they were discovered by Vassilev and co-workers at Hoffman-La Roche.[169,170] A high-throughput screen using surface plasmon resonance identified a series of *cis*-imidazolines with activity against Mdm2. These were the racemic nutlin 1 and nutlin 2, and the separated enantiomers nutlin 3a and Nutlin 3b (Figure 1.35), with IC_{50} values of 0.26, 0.14, 0.09, and 13.60 μM, respectively. Crystal-structure analysis of nutlin 2 (PDB code: 1RV1) indicated that the nutlins bound to the same pocket on Mdm2 as the p53 α-helix. To some extent, the nutlins mimic the interaction of p53 protein residues with Mdm2, with a bromophenyl in both the Trp and Leu pockets, and an ether side chain filling the Phe pocket. The most active of the nutlins (3a) was taken into in vivo testing, and was well tolerated on oral administration in mice to give plasma levels above the in vitro IC_{90}. This gave a 90% inhibition of tumour growth in mice compared to the control group. It is worth noting that nutlin 3a, breaks Lipinski's rules on account of its high molecular weight (581.5 Daltons) and logP (5.19).
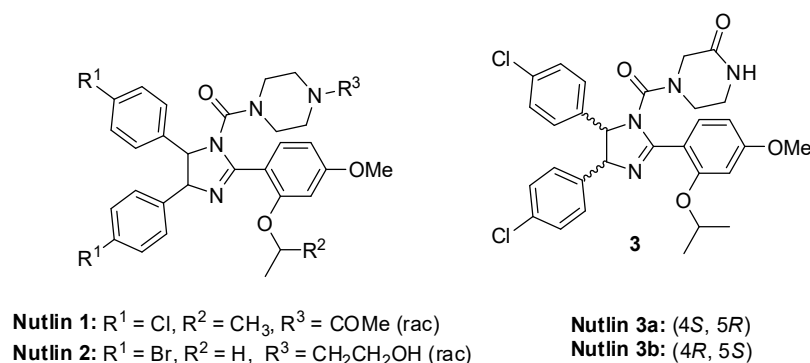
Nutlin 1: R^1 = Cl, R^2 = CH$_3$, R^3 = COMe (rac)
Nutlin 2: R^1 = Br, R^2 = H, R^3 = CH$_2$CH$_2$OH (rac)

Nutlin 3a: (4*S*, 5*R*)
Nutlin 3b: (4*R*, 5*S*)

Figure 1.35: Structure of nutlins 1, 2, 3a and 3b.[169,170]

1.3.5.2 Case Study 2: Targeting Bcl-X$_L$ with ABT-737

Researchers at Abbott used nuclear magnetic resonance fragment screening, and subsequently structural biology, to discover compounds with high binding affinity for the hydrophobic groove on Bcl-X$_L$, to act as Bcl2 inhibitors.[171] Initial lead compounds identified from this process were highly active (K$_i$ = 36 ± 1.6 nM), however the inhibitory effect was effectively nullified in the presence of HSA. After identifying the portion of the molecule responsible for binding to domain III of HSA, the subsequent structural modifications and optimisation regarding cytotoxicity effects led to the discovery of ABT-737 (Figure 1.36). This compound had affinity for both Bcl-2 (IC$_{50}$ < 1 nM) and Bcl-X$_L$ (IC$_{50}$ < 0.5 nM), and was maintained in 10% human albumin serum, but unfortunately was not orally bioavailable. However, modifications to this structure gave ABT-263, which retains a lot of the affinity (ABT-263 K$_i$ = 0.4 nM, ABT-737 K$_i$ = 0.08 nM) and exhibits oral bioavailability in animal models.[172]
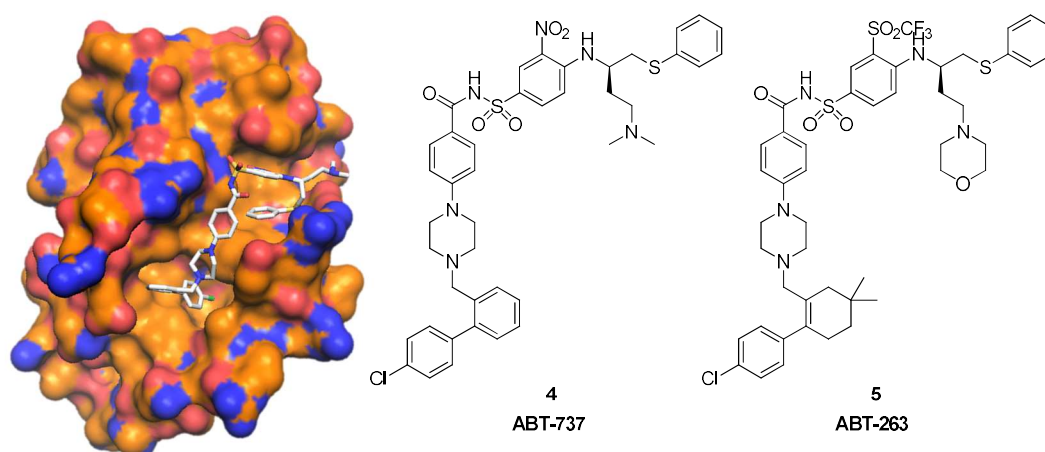

Figure 1.36: Left: Crystal structure of ABT-737 bound to Bcl-XL (PDB: 2YXJ). Right: Structures of ABT-737 and ABT 263.[172]

1.3.5.3   Case Study 3: The Road to Bivalent SMAC Mimetics to Target XIAP

The protein 'Second mitochondria-derived activator of caspases' (SMAC) is a regulatory protein which binds to X-linked inhibitor of apoptosis protein (XIAP), blocking its interaction with caspase 3, 7 or 9, and thereby inhibiting the apoptotic cycle. SMAC binds to XIAP through four residues on the *N*-terminus (Figure 1.37; SMAC AVPI). Early studies toward creating a therapeutic to target this interaction focused on developing SAR of the SMAC AVPI peptide. Vucic and co-workers developed compound CS3 (Figure 1.37), which is based on the SMAC *N*-terminus and is highly potent, with a K_i of 16 nM.[173] The structure of SMAC mimetics dramatically changed direction when the first dimeric mimetics were discovered accidentally in 2005, as the result of the unintended homodimerisation of an alkyne group by a Glaser coupling (Figure 1.37).[174] The compound resulted from an accidental dimerization of a terminal alkyne to form a diyne bridge and had higher potency even than the SMAC protein, with an estimated K_d of 0.3 nM. From this point on new SMAC mimetics have typically incorporated this bivalent frame. The most successful of these is birinapant, an intravenously administered molecule which has been taken through to phase II testing by Tetralogic.[175]
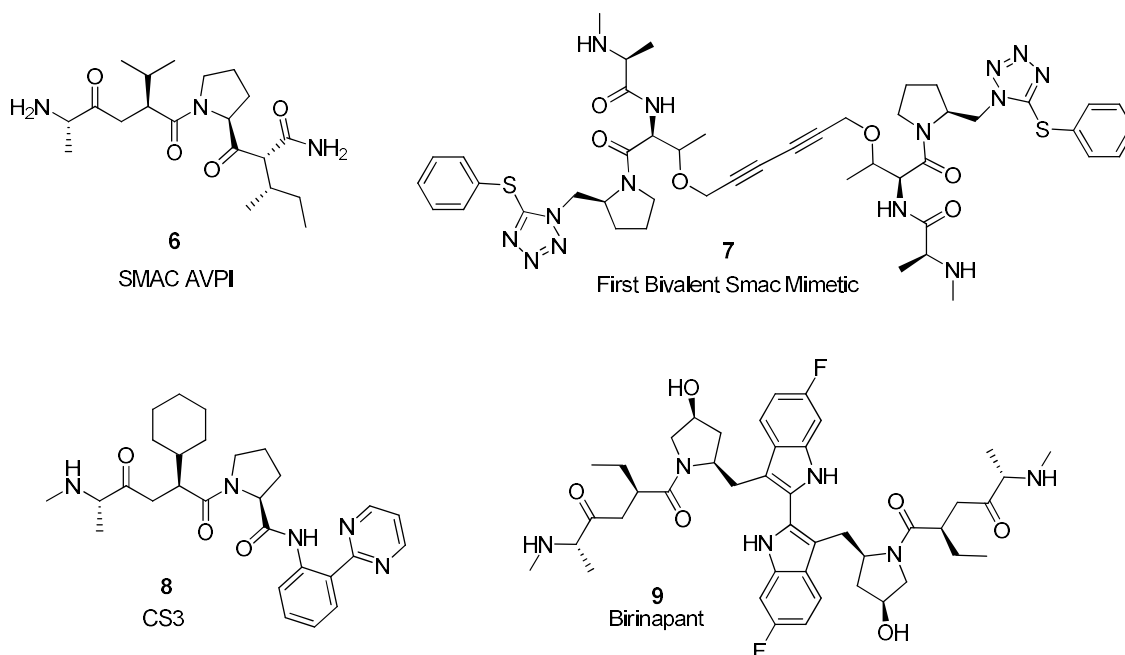

Figure 1.37: The structure of the *N*-terminus of SMAC, as well as three designed SMAC mimetics.

1.4   Conclusion

In conclusion, PPIs have proved a challenging target by traditional drug discovery methods. This is due to the fact that the protein binding site is typically large in size, flat, and very lipophilic in nature. As a result, typical fragment and small molecule libraries

designed for lead identification by high throughput screening have very low hit rates against PPI targets.

As a result, more esoteric methods have been developed to facilitate the *ab initio* design of PPI inhibitors, primarily through the mimicry of one of the proteins at the binding site. As α-helices are involved at the binding site of approximately 62% of protein-protein interactions, a great deal of attention has been given to the development of α-helical mimetics. Different approaches to accomplish this include synthesising peptide foldamers using unnatural amino acid residues, designing structurally stable miniproteins which are forced into a restricted conformation in solution, covalently stapling one face of the α-helix in order to encourage secondary structure, and designing small molecule scaffolds which can be functionalised as to hold functional groups in positions that mimic the α-helical amino acid side-chains.

The medicinal chemist hoping to discover a novel chemical lead for the inhibition of a PPI now has the choice of many different techniques, facilitating the development of tool compounds and drug molecules.

2 Instilling 'daclatasvir-like' functional groups using palladium catalysed C-H activation.

2.1 Chapter Summary

This chapter outlines the development of methodology to functionalise aryl bromides with 'daclatasvir-like' functional groups using C-H activation. Daclatasvir is a hepatitis C anti-viral drug that has exceptionally high oral bioavailability, despite having a high molecular weight. It was hypothesised that this functionalization of structurally related cores with 'daclatasvir-like' termini could be used as a strategy to improve cell permeation and drug-like properties of PPI inhibitors.

The synthesis of the reagents required for the C-H activation chemistry is described, in particular optimising the conditions to reduce the amount of homo-coupling within the reaction. The scope of the C-H activation was then tested on 18 different aryl dibromides and was successful in 11 cases. The chemistry developed has applications in the late stage core-functionalization of HCV non-structural protein 5A (NS5A) inhibitors, as well as the potential to introduce improved PK to polyaryl α-helical inhibitors.

2.2 The Unusual Pharmacokinetics of Daclatasvir, and potential application to PPIs.

10
daclatasvir

Molecular Weight: 738.9
*clog*P: 4.18
EC$_{50}$ (NS5A genotype 1a): 50 ± 13 pM
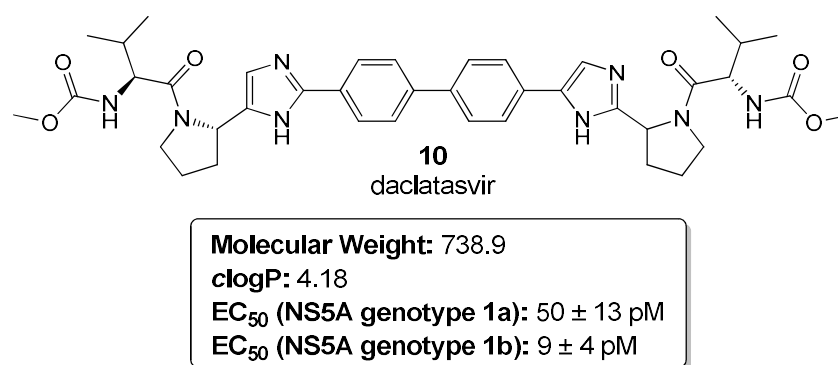EC$_{50}$ (NS5A genotype 1b): 9 ± 4 pM

Figure 2.1: Structure of the HCV NS5A inhibitor daclatasvir.

Although we are starting to see the development of more and more PPI inhibitors, these compounds consistently produce poor pharmacokinetic (PK) profiles. Due to the size and lipophilicity inherent in PPI inhibitors, the compounds will regularly have poor (if any) oral bioavailability, poor transport properties into cells, significant metabolism, significant excretion, and poor aqueous solubility.[168,176] These factors are often governed by whether or not a drug meets the 'Lipinski rule-of-five' criteria.[177,178] These state that

for a drug to have good oral PK properties, it must normally have MWT≤500, logP≤5, number of H-bond donors ≤ 5, and number of H-bond acceptors ≤ 10. PPI inhibitors regularly break the rules regarding molecular weight and lipophilicity.[177]

Examples of existing drug molecules that break the rule-of-five criteria, yet still exhibit high oral-bioavailability and solubility, were analysed. In doing so it was hoped that insight would be developed into how to incorporate these physical properties into the design of novel PPI inhibitors. One such molecule is the HCV non-structural protein 5A (NS5A) inhibitor, daclatasvir.[179]

Not only does daclatasvir exhibit exceptional potency (as low as 9 pM against NS5A genotype 1b), but it has also been found to have a surprisingly good pharmacokinetic profile.[180] Despite having a molecular weight of 739 Daltons, it exhibits excellent bioavailability (F) of 38-108% in animal tests (Table 2.1).[181]

Table 2.1: Bioavailability of Daclatasvir in animal tests.[181]

	Rat	Dog	Monkey	
AUC	4.8	11	1.93	μM.h
F	50	108	38	%

There has been little research into what structural properties give rise to the excellent PK profile exhibited by daclatasvir. Based on research into other compounds of similar structure, it is reasonable to hypothesise that internal hydrogen bonding may be the cause of the high membrane permeability of the compound.[182] Low energy conformations of daclatasvir modelled without the presence of water exhibit internal hydrogen bonding which causes the polar side chains to wrap inwards (Figure 2.2). This could potentially mask polar sites on the molecule when in a lipophilic environment, allowing inter-membrane diffusion, without decreasing solubility in aqueous media.
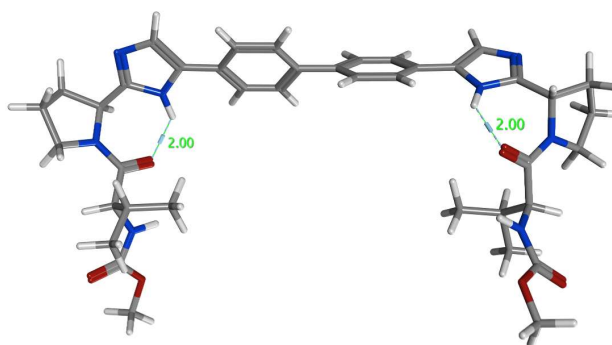
Figure 2.2: The 3D structure of daclatasvir, conformationally minimised in MOE using LowModeMD calculations.

Due to the polar nature of the side chains, they contribute only a small net gain to the clogP and clogD of the molecule (Table 2.2). The overall clogP over daclatasvir is only 0.56 higher (3.62 to 4.16) than the unfunctionalised biphenyl core of the molecule, despite a significant contribution to molecular weight. Interestingly, due to the addition of imidazole rings, the side chains contribute a significant gain to the solubility forecast index (SFI)[183,184] of daclatasvir, which is predicted to be 8.16 at physiological pH. Statistical analysis of known compounds indicates that only 5% of compounds with equivalent SFI values exhibit good solubility when tested.[183] This is an indication of how much of an outlier the PK properties of daclatasvir are considering the large molecular weight.

Table 2.2: Calculated measures of lipophilicity and solubility for daclatasvir and biphenyl.

Compound	$c$logP	$c$logD$_{pH7.4}$	SFI[a]
	4.18	4.16	8.16
	3.62	3.62	5.62

[a]Solubility forecast index (SFI) calculated as clogD$_{pH7.4}$ plus number of aromatic rings.[183]

It was hypothesised that the polar side chains were responsible for giving daclatasvir its excellent PK profile. Considering this, it was felt there could be value to instilling these side chains onto known PPI mimetics. For example, a terphenyl α-helical mimetic such as those developed by Hamilton could be functionalised in this way, potentially leading to greater permeability, aqueous solubility and bioavailability.[149] This chapter explores the chemistry required to achieve this.

2.3 Retrosynthetic Analysis of Daclatasvir Side-chains
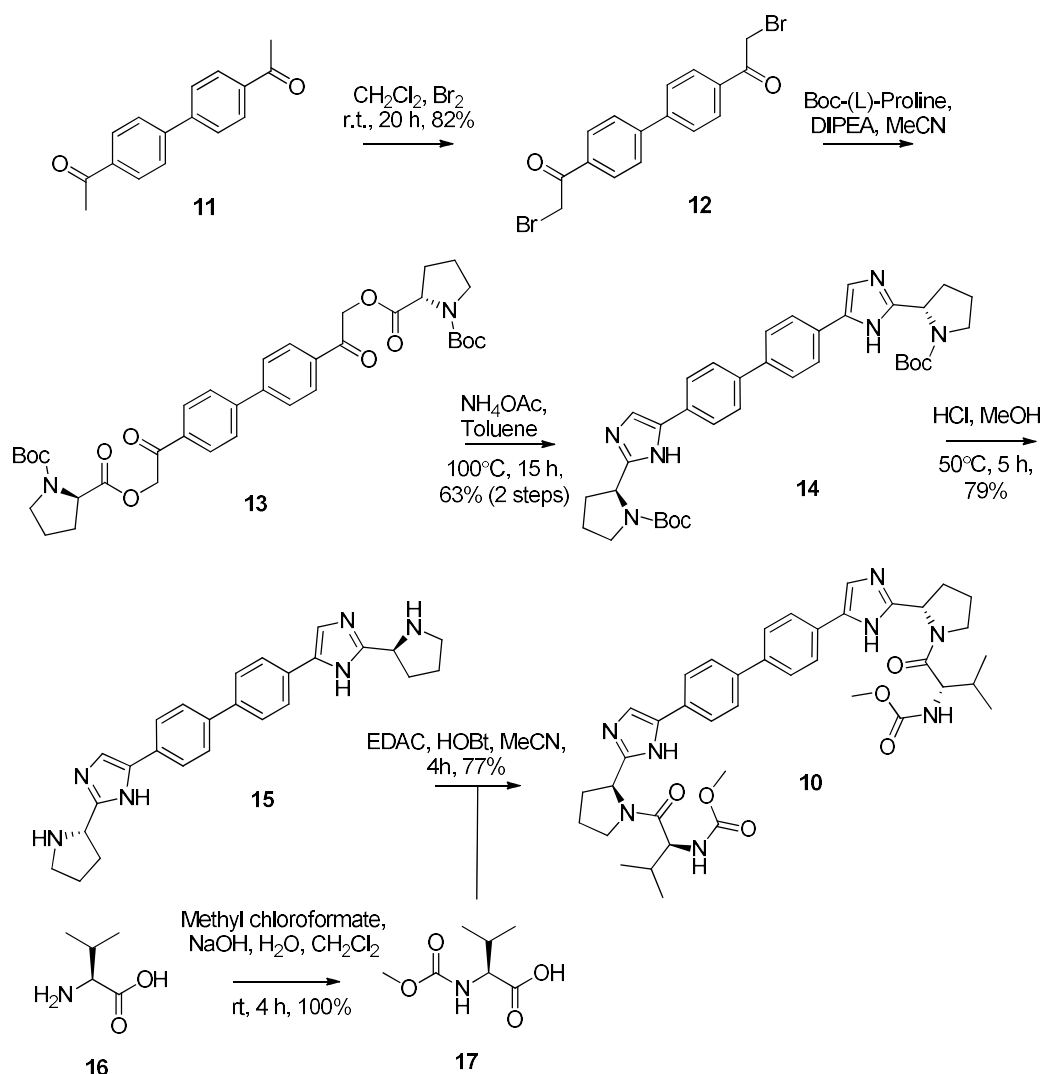

Scheme 2.1: Synthesis of daclatasvir developed by BMS.[185]

The synthetic route to daclatasvir developed by Bristol-Myers Squibb is an efficient, linear synthesis, reaching the product in a 31% yield over six steps (Scheme 2.1).[186] 1,1'-Diacetylbisphenyl **11** is dibrominated, reacted with Boc-L-proline, then transformed to the imidazole **14** upon reaction with ammonium acetate *via* an interesting rearrangement step. The butyloxycarbonyl (Boc) groups are then removed with acid, and the resulting amines subjected to amide coupling to install the valine sidechain present in the product, daclatasvir (**10**).

Although this is an efficient route, and has proven to be scalable for large-scale production of the drug, the linearity means that the core bisphenyl functionality within the molecule has been incorporated within the first step. To establish methodology to explore α-helical mimetics with the daclatasvir side chains, a more convergent synthetic

route in which the aryl-imidazole bond would be formed in the final stages was required. Retrosynthetic analysis was therefore carried out on the structure of daclatasvir (Scheme 2.2). There have been examples in the literature where this key bond in NS5A inhibitors has been formed using Suzuki cross-coupling chemistry between an imidazole-boronic acid, and an aryl bromide.[187,188] However, the instability of the imidazole-boronic acid was found to be an issue. This lead us to consider a C-H activation strategy, with C-H activation directed to the imidazole 2-position (Scheme 2.2).[189]
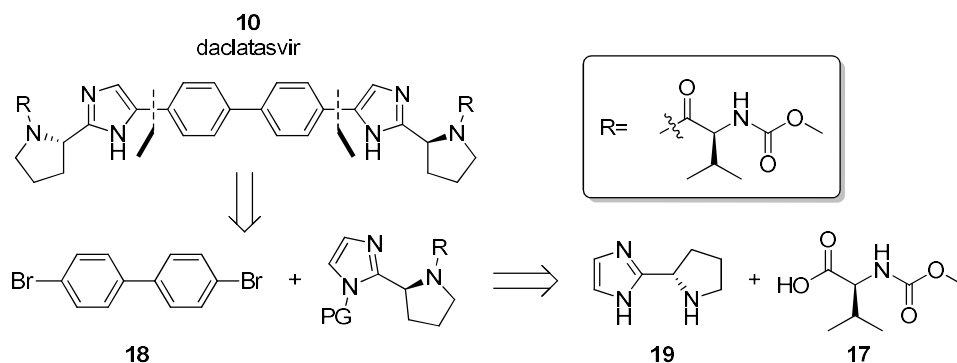
Scheme 2.2: Retrosynthetic analysis of daclatasvir using C-H activation.

### 2.3.1   C-H Activation of Imidazoles in the 2-Position.

Metal-catalysed C-H activation chemistry has gained a great deal of attention in recent years.[189–191] It exists as an attractive alternative to conventional metal-catalysed coupling reactions, such as Suzuki cross-coupling. When carrying out conventional cross-coupling reactions, both reagents must be functionalised with reaction specific functional groups, such as a boronic acid and an aryl halide for a Suzuki coupling. C-H activation provides an atom-efficient alternative, by making use of existing C-H bonds within molecules. In doing so, it reduces the need for additional reaction steps to functionalise reagents with the necessary functionalities, potentially making the synthetic routes quicker and greener.

However, the ubiquitous nature of C-H bonds brings about the problem of selectivity. In order to be synthetically useful, the reaction must proceed so that only the desired position is functionalised. Because of this, a great deal of research has been carried out into how to direct C-H activation to specific positions, in particular on 5 and 6-membered heteroaromatic rings. A few recent examples of significant C-H activations are described in the following paragraphs.

2.3.1.1 Controlling C-H Activation Using Directing Groups

The pyridyl nitrogen lone pair has been shown to direct palladium catalysed C-H activation reactions on arenes.[191] One of the most studied C-H activation reactions is palladium(II) acetate-catalysed acetoxylation using diacetoiodobenzene, pioneered by Sandford.[192] The catalytic cycle follows the sequence of C-H activation, followed by oxidation by diacetoiodobenzene, and finally a reductive elimination. The cycle involves transitions between Pd(II) and Pd(IV) oxidation states (Scheme 2.3).[193]
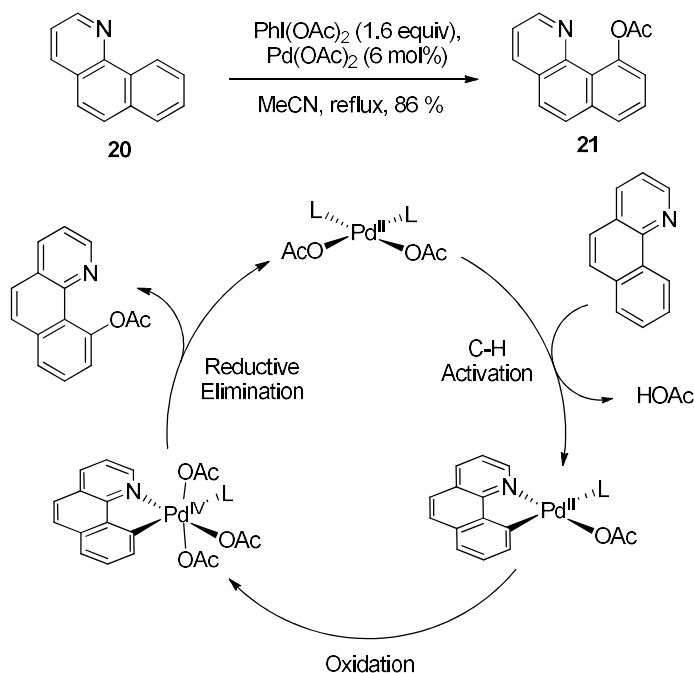
Scheme 2.3: The N-directed acetoxylation of benzoquinoline by diacetoiodobenzene catalysed by palladium(II) acetate, showing the catalytic cycle.[193]

With modified conditions, this C-H activation can also be used for the functionalization of various different groups. By carrying out the reaction in alcohols as solvents, it is possible to functionalise with aryl ethers, such as OMe, OEt, O-*i*Pr and OCH_2CF_3 in good yields.[193] It is also possible to functionalise with a range of aryl groups using hypervalent *bis*-aryl iodonium salts as the oxidant. As well as the benzoquinoline directing group, Sanford has also shown that a wide range of other functional groups are able to *ortho*-direct the C-H activation through ligation with the palladium metal centre (Figure 2.3).[191]
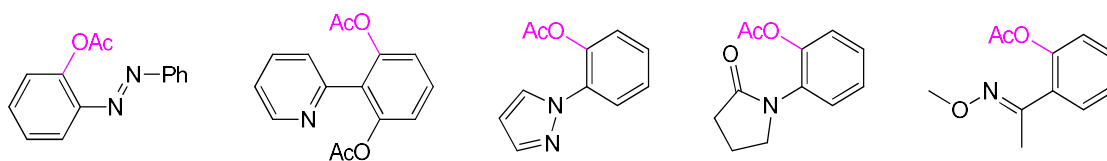

Figure 2.3: Examples of ortho directing groups for C-H activated acetoxylation.[191]

2.3.1.2 C-H Activation on 5-Membered Heterocycles.

Unlike for arenes, where a directing group is usually required for regioselective C-H activation, 5 membered heterocycles have an inherent electronic bias within the molecule which is often sufficient to achieve regioselective C-H activation.[190] The C-H activation will normally occur at the most nucleophilic position on the ring. The reaction was traditionally predicted to proceed through electrophilic aromatic substitution, or Heck-type mechanism (Scheme 2.4).[190] However, more recently, density functional theory (DFT) calculations carried out by Fagnou has shown that the mechanism is more likely to proceed through the concerted metalation-deprotonation (CMD) mechanism, particularly when in the presence of a carboxylate ligand (Scheme 2.4).[194] It is possible to C-H activate the position with selectivity to the position with the lowest pKa, by metalation using copper iodide, followed by a normal cross-coupling (Scheme 2.4).
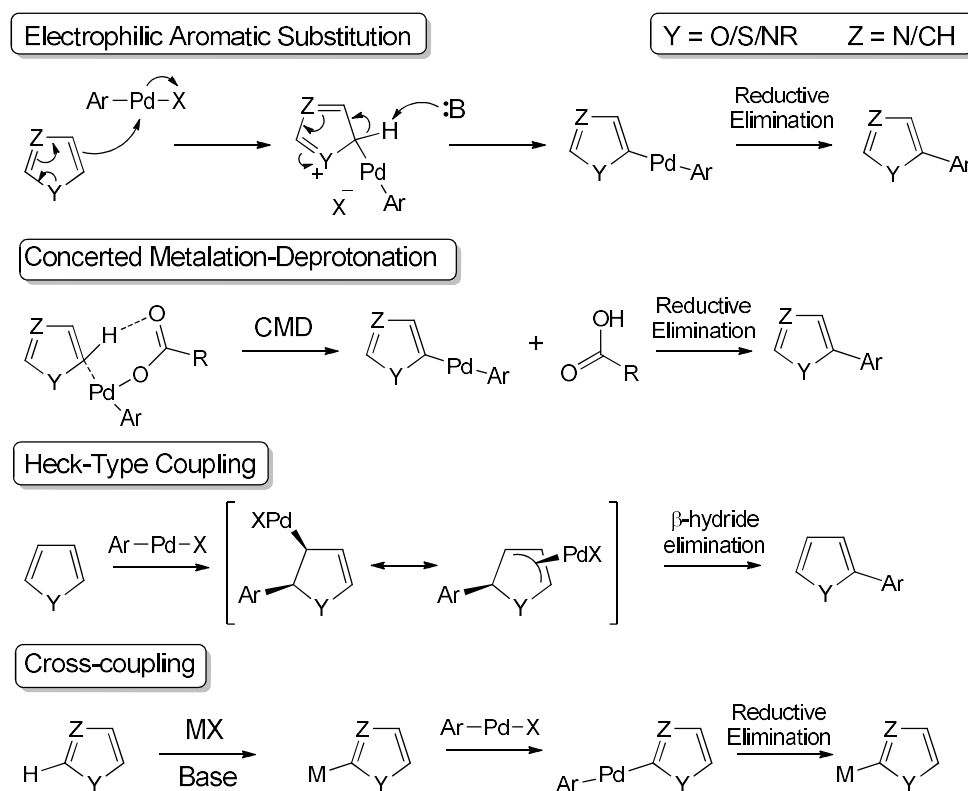

Scheme 2.4: Different proposed mechanisms of C-H activation on 5-membered heterocycles.[190,194]

As early as 1985, the C-3 arylation of *N*-protected indoles by C-H activation had been developed by Ohta and co-workers.[195] By refluxing a solution of the tosylated indole with a 2-chloropyrazine and potassium acetate in *N,N'*-dimethyl acetamide (DMA) for 12 hours in the presence of Pd(PPh$_3$)$_4$, they were able to achieve C-H activation in yields of up to 98% (Scheme 2.5, **A**).[196] The reaction could be explained by general ability of the

indole 3-position to react with electrophiles, combined with the ability for 2-choloropyridines to undergo S_NAr.

Suzuki reported another early example of C-H activation on the C-3 position of an imidazole by intramolecular reaction with an aryl bromide (Scheme 2.5, **B**).[197] Again the conditions were forcing, 150 °C in DMA. Interestingly, during the optimisation process, it was found that the addition of tetrabutylammonium chloride (TBAC) increased the yield by around 10 %. This could be due to the fact that TBAC has been shown to stabilise the formation of palladium nanoparticles, which act as an effective catalytic species.[198]
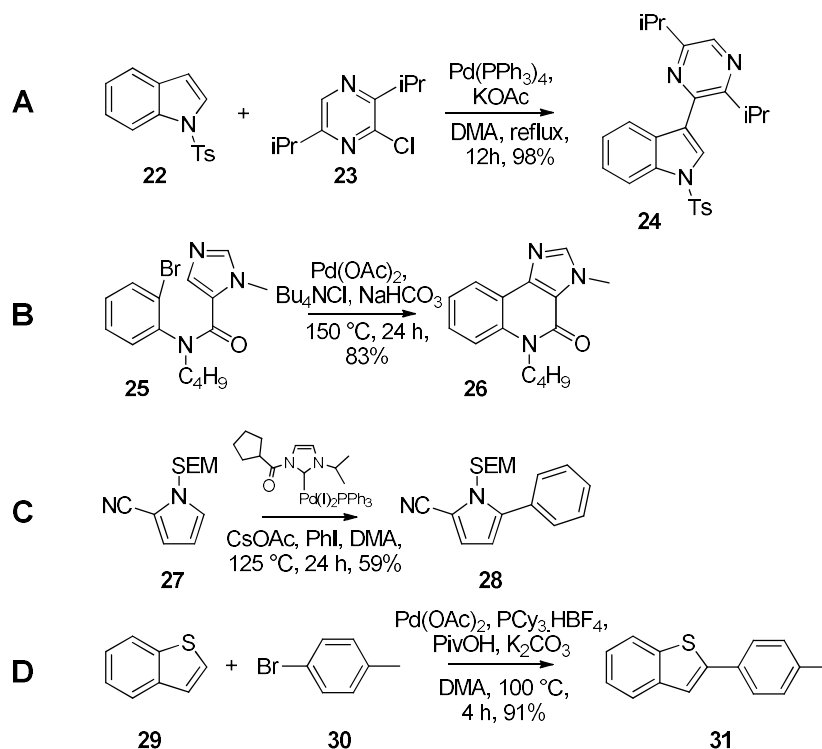

Scheme 2.5: Selected C-H activation reactions, including **A)** Early C-3 arylation of protected indole by Ohta,[196] **B)** Intramolecular C-3 arylation of imidazole by Suzuki,[197] **C)** Arylation of SEM protected pyrrole by Sames,[199] and **D)** Use of PivOH for broadly applicable C-H activation conditions by Fagnou.[200]

Sames and co-workers breakthrough work pioneered C-H activation on *N*-protected pyrroles (normally a difficult heteroaromatic for this type of reaction). The use of a trimethylsilylehtoxymethyl (SEM) protecting group enabled C-H activation in the C-5 position with a yield of 59% (Scheme 2.5, **C**).[199] However the esoteric N-heterocyclic carbene palladium complex utilised as catalyst limited the applicability of the reaction. Nevertheless, the discovery of the SEM protecting group as facilitating the C-H activation of 5-membered heterocyclic rings was an important one.

A major breakthrough towards developing C-H activation conditions that would have broad applications across many different heterocycles was the introduction of pivalic acid as an additive (Scheme 2.5, D).[200] Pivalic acid has been shown to facilitate the CMD transition state that occurs during the C-H activation step of the catalytic cycle (Scheme 2.4).[194] This allowed C-H activation to occur using commercially available catalyst. Although yields were sometimes moderate, this reaction conditions had incredibly wide scope.

Sames developed C-H activation conditions to access each position on an imidazole ring (Scheme 2.6).[201] He shows that it is possible to use C-H activation as an alternative to cross coupling on each position on the imidazole ring, by choosing appropriate conditions and using SEM protective group strategy to drive regioselectivity.
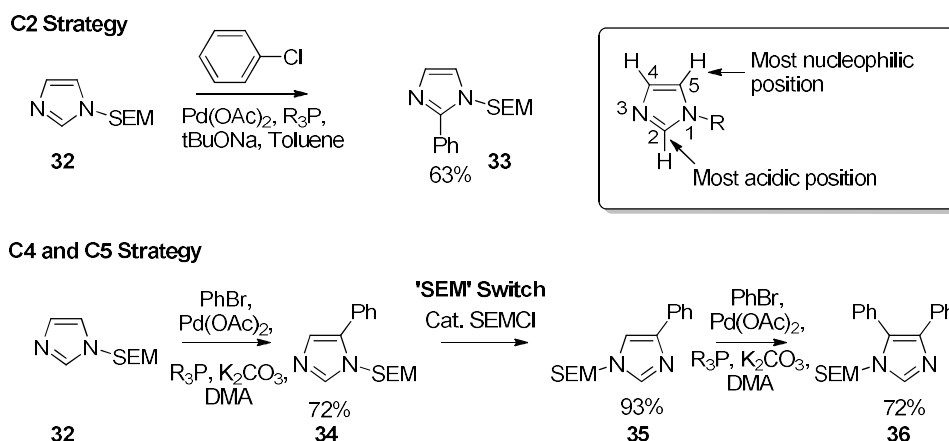
Scheme 2.6: Strategy to access all positions on an imidazole ring developed by Sames.[201]

These recent developments in imidazole C-H activation strategy were considered a viable option for the functionalisation of aryl groups with daclatasvir-like side chains, and thus synthesis of the imidazole intermediates identified in the above retrosynthesis was required to test the conditions, (Scheme 2.2).

2.4 Synthesis of the Imidazole Reagent for C-H Activation
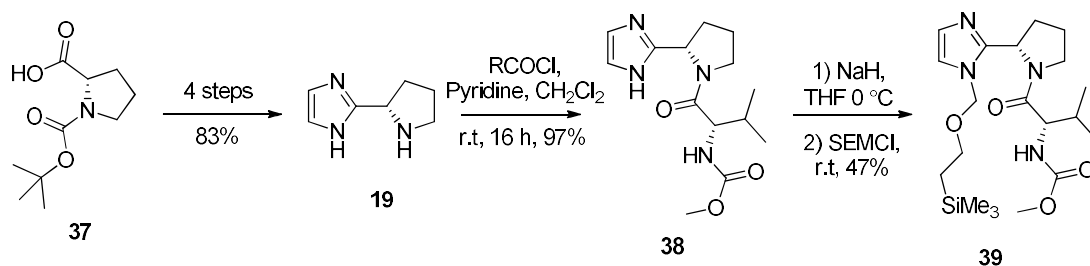
Scheme 2.7: A summary of the synthesis of C-H activation reagent **39.**

Scheme 2.7 shows the synthesis of CH-activation reagent **45** that will be discussed in detail in this section. Firstly, the synthesis of imidazole **25** following procedures similar to those reported in the literature will be described, followed by the amide coupling to introduce the valyl carbamate side chain, and finally the SEM protection to give the desired imidazole reagent **45**.

2.4.1   Synthesis of Imidazole 19
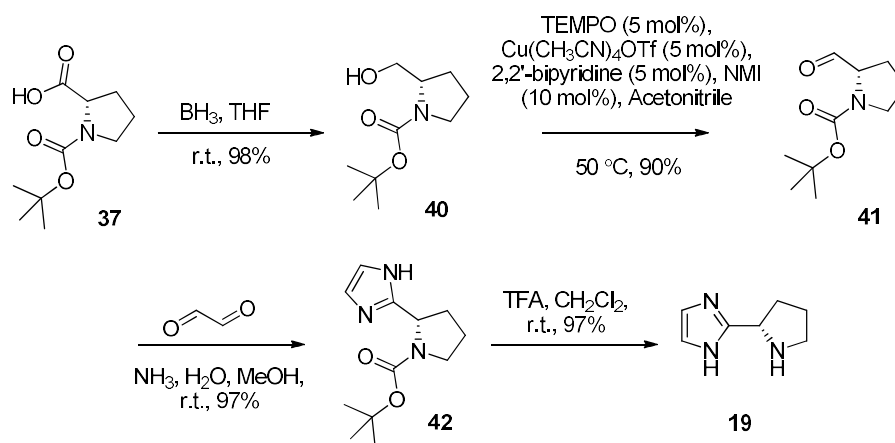
Scheme 2.8: Synthesis of imidazole **19**.

The synthesis of imidazole **19** was carried out following a similar route to that described previously in the literature (Scheme 2.8).[202,203] Firstly Boc-L-proline **37** was reduced to the corresponding alcohol **40** using borane in tetrahydrofuran (THF). This method was scaled up to 20 g of starting material giving the product **40** in a 98% yield. The oxidation of the Boc-prolinol to the necessary aldehyde **41** was carried out with particular care due to the potential racemisation of the aldehyde α-carbon. The final conditions chosen involved a (2,2,6,6-tetramethylpiperidin-1-yl)oxyl (TEMPO)/Cu(I) catalytic system due to the mild nature of the reaction and the ease of scale up, which gave the product **41** in 90% yield. The different conditions investigated are discussed in more detail in section 2.4.1.2. The aldehyde was then cyclised to the corresponding imidazole **42** without purification, to reduce the chance of racemisation. This was achieved by reaction with glyoxal and ammonia, and gave the imidazole in a 97% yield. The final step to give compound **19** was the cleavage of the Boc group using triflouroacetic acid.[204] The crude triflouroacetate salt was purified by use of an SCX-2 ion exchange column, to give the free-base pyrrolidine **19** in a 97% yield. Overall, this synthetic route allowed for the synthesis of compound **19** in an 83% yield over 4 steps, and was successfully performed on a 20 g scale.

2.4.1.1 Rotamers around the Boc-pyrrolidine bond

The NMR characterisation of compounds containing the Boc carbamate group was challenging due to the presence of rotamers at room temperature. Carbamate groups are known to exist in either the *syn* or *anti* conformations (Figure 2.4),[205] with the *syn* usually being the thermodynamically favoured.
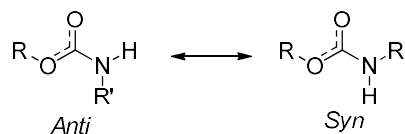

Figure 2.4: *Syn* and *anti* carbamate rotamers

However, because the carbamate in the compounds synthesized in this project are formed off a pyrrolidine ring, neither the *syn* or *anti* forms are significantly more favoured. The interconversion between either forms by rotation around the carbamate C-N bond is slow compared to the NMR timescale, and splitting of NMR peaks between rotamers is consistently observed. For example, in compound **41**, the two carbamate rotamers (Figure 2.5) led to a complex NMR spectrum, with splitting at both the aldehyde and *tert*-butyl peaks.
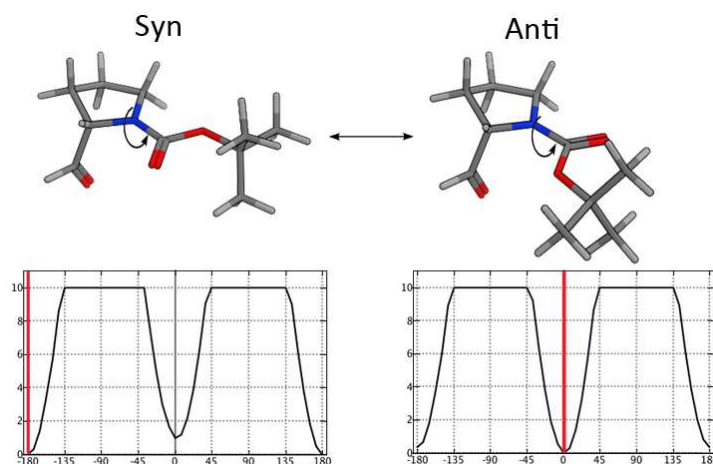

Figure 2.5: Rotation around the carbamate C-N bond, and the energy profile of ration, demonstrating the two energetically favoured rotamers

*2.4.1.2 Oxidation of (S)-tert-butyl 2-(hydroxymethyl)pyrrolidine-1-carboxylate **40** to (S)-tert-butyl 2-formylpyrrolidine-1-carboxylate **41***

The published synthetic route to imidazole **19** described the use of catalytic TEMPO, with NaOCl for the oxidation of prolinol **40** to prolinal **41**. Although near-quantitative yields were given for this reaction in the literature, it was not possible to push the reaction past 60% completion as measured by NMR, with the crude material containing a mixture of

starting material and product. As there was reluctance to purify the aldehyde using chromatographic techniques due to the risk of α-carbon racemisation, the crude material was reacted directly to form imidazole **42**, but the isolated yield was low. Because of this, it was important to further optimise the oxidation step.

Table 2.3: Conditions tested for the oxidation of Boc-prolinol **40** to the respective aldehyde.

	Conditions	Yield
1	TEMPO, NaBr, NaOCl, NaHCO$_3$	60%
2	(COCl)$_2$, DMSO, Et$_3$N, CH$_2$Cl$_2$, -78 °C → r.t	98%
3	TEMPO (5 mol%), Cu(CH3CN)4OTf (5 mol%), 2,2'-bipyridine (5 mol%), NMI (10 mol%), acetonitrile	90%

Initially, Swern oxidation conditions were used as an alternative to the low yielding TEMPO oxidation (Table 2.3). This reaction was very high yielding (99%), however, the presence of triethylamine gave rise to concerns over the racemisation of the aldehyde α-carbon centre (Scheme 2.9). The calculated pKa of the carbonyl α position is 16.8 (ChemAxon MarvinSketch 6.2.0). Triethylamine used in the reaction is basic enough to racemise this position by catalysing the formation of the enolate form of the aldehyde (Scheme 2.9). Therefore, the use of milder conditions was investigated, in which the risk of racemisation would be mitigated.
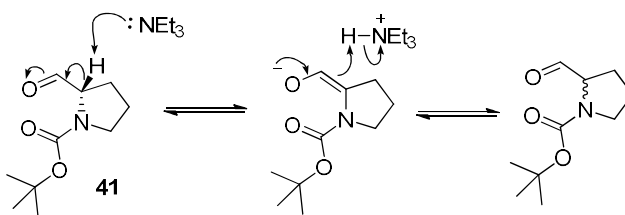
Scheme 2.9: Mechanism for triethylamine-mediated racemisation under Swern oxidation conditions.

Shannon Stahl recently released conditions whereby a primary alcohol could be oxidised to the respective aldehyde using atmospheric oxygen.[206,207] This is done using a co-catalytic system of a bipyridyl-copper-*N*-methylimidazole (NMI) complex, as well as the stable radical species, TEMPO. Under these conditions, oxygen will oxidise the copper(I) to a copper(II) peroxide species. TEMPO then catalytically mediates the ability for this species to oxidise alcohols to aldehydes. It does this by allowing the formation of a copper hydroxyl species, which undergoes substitution with the alcohol, and finally provides the respective aldehyde by reductive elimination (Scheme 2.10).
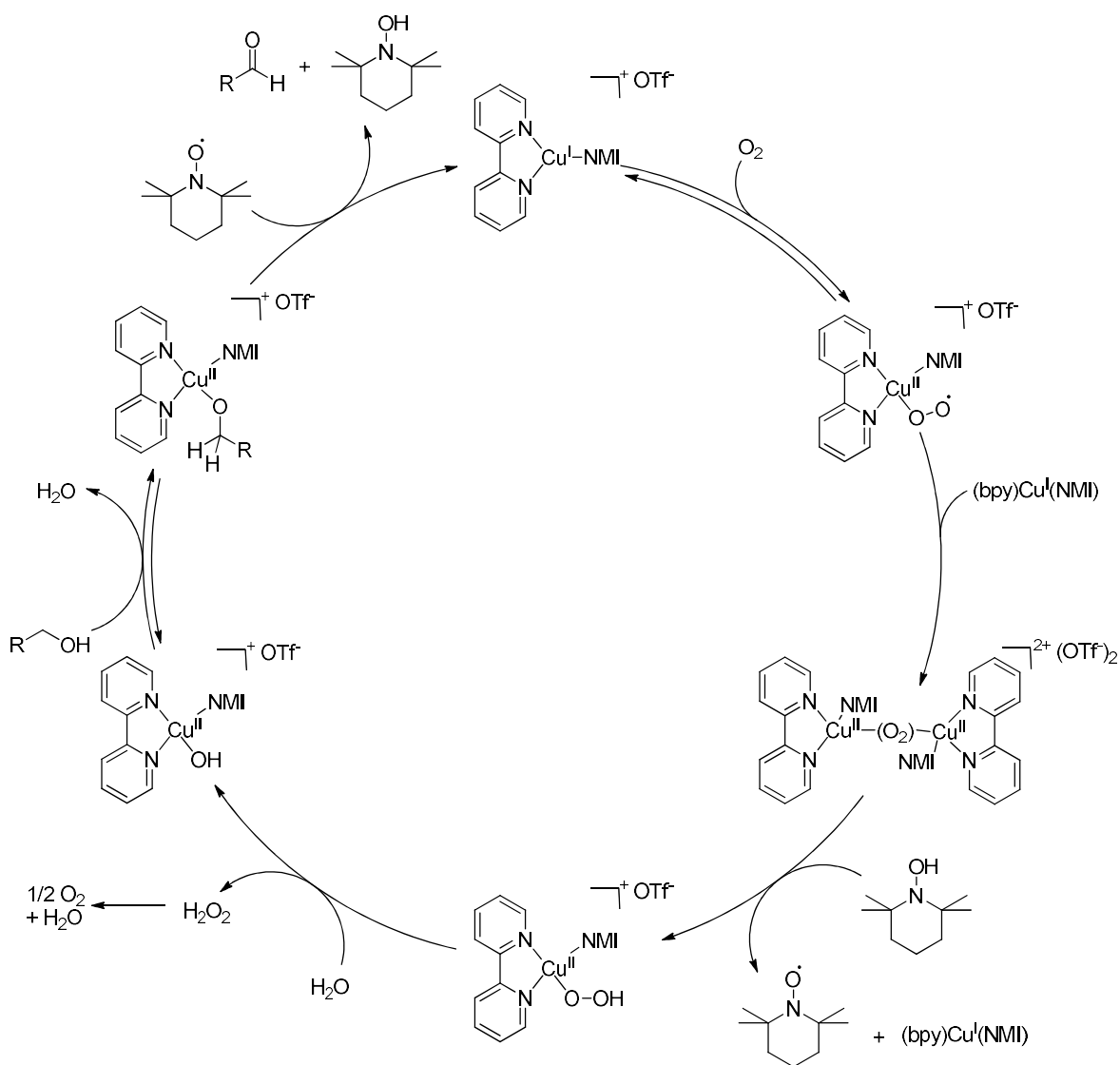
Scheme 2.10: Mechanism of oxidation of alcohols using a Cu(I)/TEMOPO co-catalytic system, as proposed by Stahl.[208]

As the conditions are very mild compared to the Swern oxidation, with NMI only having a pKa of 7.4, there will be a significantly lower rate of racemisation of the aldehyde product. Most importantly, these conditions were also found to be much easier to carry out on a large scale reactions.

This reaction was carried out on a 10 g scale, with 90% yield. In order to confirm that racemisation had not occurred, the optical rotation of the starting material and product were measured. The product was found to have an $[\alpha]_D$ of -94.5 deg·mL·g^{-1}·dm^{-1}, which corresponds to the literature value of -96.1 deg·mL·g^{-1}·dm^{-1}.[209] Despite the fact that the yield was slightly lower than Swern conditions, this reaction was still considered favourable, and adopted into the synthetic route.

2.4.2 Peptide coupling to form amide 38

Once compound **19** had been synthesized, the next step was to form the amide bond between the pyrrolidine with the valyl-carbamate side chain. The valyl carbamate **17** was synthesized using the Schotten-Baumann conditions described in the original Bristol Myers-Squibb publications.[181,186] L-Valine is reacted with methylchloroformate in a biphasic system water/dichloroethane in the presence of sodium hydroxide. The resulting water-soluble acid was then dried by lyophilisation to give the product in an 84% yield.
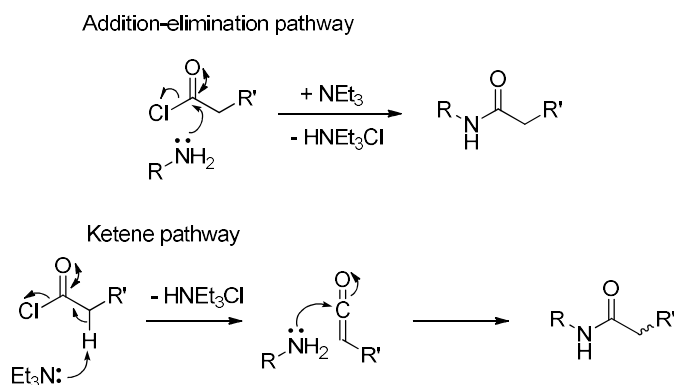

Scheme 2.11: Ketene formation from an acyl chloride during amide coupling.

To form peptide **38**, many different methods of peptide coupling were analysed.[210] Firstly, reagents that could transform the carboxylic acid group into a reactive acyl chloride were considered. These include thionyl chloride, phosphoryl chloride, and more commonly oxalyl chloride with dimethylformamide (through the Vilsmeier-Haack intermediate). The acyl chloride is highly reactive and the reaction is very clean and atom efficient, and is therefore a good option.

The use of acid chlorides, however, was not favoured in this synthesis due to the risk of racemisation due to intermediate ketene formation (Scheme 2.11). Instead, milder coupling reagents were initially selected, with which the reactive intermediate is generated *in situ* before reacting directly with the amine to produce a peptide bond, thus preventing racemisation from occurring. The most commonly used modern example of these are the guanidinium (once uronium)[211] salts, such as *N,N,N',N'*-tetramethyl-*O*-(1*H*-benzotriazol-1-yl)uronium hexafluorophosphate (HBTU). These react with the carboxylic acid to produce an acyl-triazolate intermediate, which can undergo further reaction with an amine to produce a peptide bond (Scheme 2.12). It has been extensively shown that these reagents are less prone to cause racemization, despite the fact that

triethylamine is still present.[210] It is because of this that use of HBTU was chosen as the first method to attempt synthesis of peptide **38**.

The reaction between pyrrolidine **19** and carboxylic acid **17** mediated by HBTU was carried out in THF at room temperature. The reaction was found to have proceeded to completion by LCMS, however, problems were encountered in purifying the product. Under normal workup conditions, the benzotriazole side product is removed by basic aqueous wash, due to the high water solubility of the corresponding salts. However, it was found that peptide **38** also exhibited high aqueous solubility, which preclude this work-up option. Instead, purification of the compound was attempted by column chromatography. Here, two problems were encountered. Firstly, the benzotriazole side product was found to co-elute with amide **38**. As well as this, amide **38** was not found to be partially unstable under silica chromatography conditions. The initial 31 g of crude material, which was shown to contain roughly 30% product by ^{1}H NMR, yielded only 1.49 g of product after the three consecutive purifications required. These difficulties rendered this route unviable.
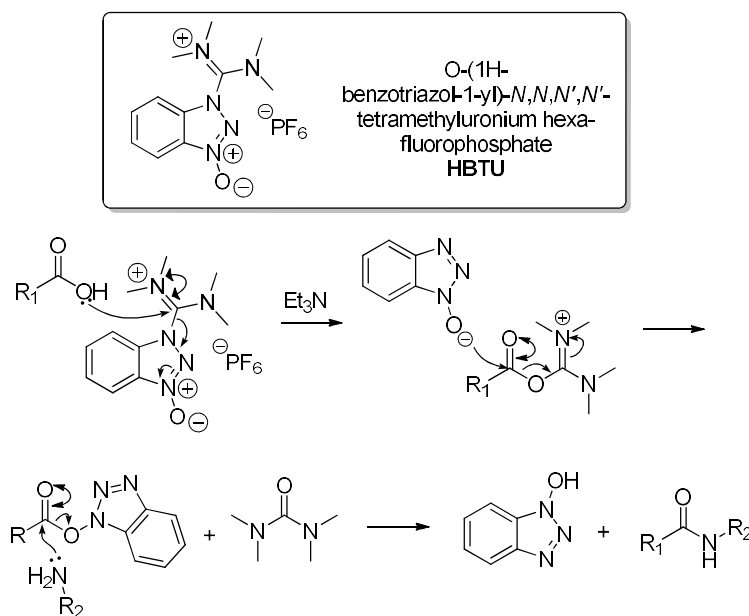
Scheme 2.12: Mechanism of guanidinium mediated peptide coupling.

Because of the difficulties encountered in removing the guanidinium salt by-products in the HBTU reaction, the acid chloride method was reconsidered in order to form amide **38**. Using this method, there were two factors which required careful consideration. The first was the aforementioned risk of racemization of the valyl chiral centre. The second was the methyl carbamate group on the acid, which is labile under acidic conditions,

therefore also a risk when carrying out this reaction. Because of this, pyridine was chosen as a base over triethylamine. With a pKa of 5.3, as opposed to 10.6, pyridine is not basic enough to deprotonate next to the acid chloride to form the ketene.[212] It instead facilitates the reaction by forming a reactive pyridinium intermediate with the acid chloride.

Initial attempts to form the acid chloride using oxalyl chloride were unsuccessful, and product appeared to decompose upon evaporation of the solvent (CH$_2$Cl$_2$) under reduced pressure. Carrying out a literature survey it was found that previous use of this acid chloride had required the addition to the amine *in situ* without isolation of the acyl chloride.[213] By making this modification to the reaction procedure, and adding the acid chloride dropwise to pyrrolidine **19** in the presence of pyridine in dichloromethane, the desired product was formed in excellent yield by LCMS. The reaction was then quenched with ammonia in isopropanol, and the resultant ammonium salt removed by filtration. The crude filtrate could then be lyophilized from water to remove any residual pyridine, and give the product in acceptable purity without the need for purification by silica chromatography.
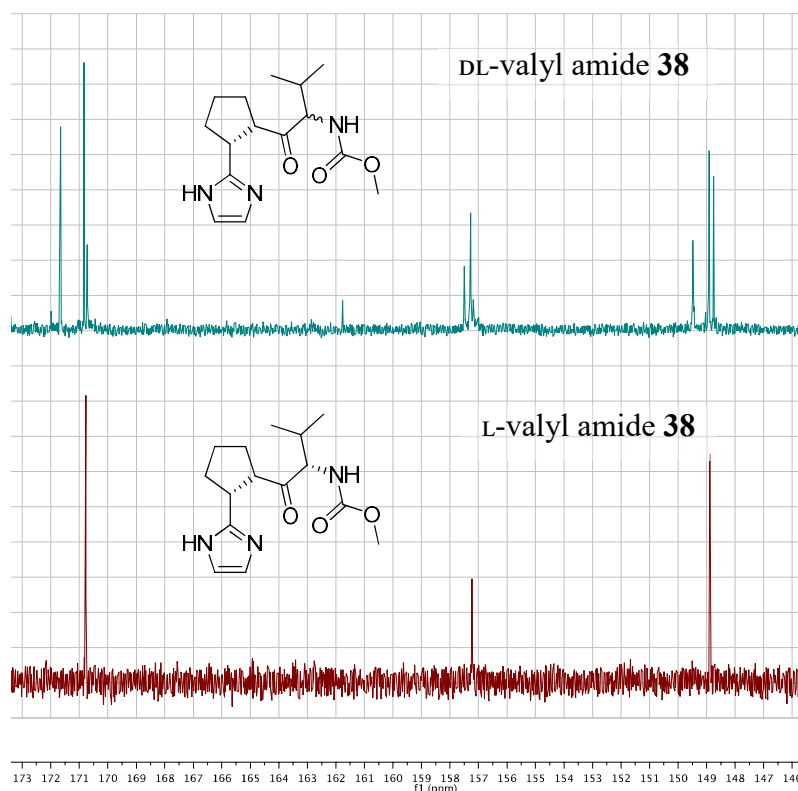

Figure 2.6: ^{13}C NMR Comparison between racemic amide **38** and acid chloride coupling product, showing the carbonyl carbons.

Using this specially adapted acid-chloride methodology circumvented the difficulties encountered during purification using the original HBTU mediated coupling procedure. In order to prove that no racemisation was occurring at either stereocenter during the reaction, the DL-valine equivalent of amide **38** was synthesized, and compared using NMR spectroscopy with the single diastereomeric products from the HBTU and acid chloride procedures. The NMR of epimeric amide clearly showed the two different diastereomers. This was most clear in the carbonyl region of the ^{13}C spectrum (Figure 2.6). It was possible to see peaks for both diastereomers for the carbamate carbon at 171 ppm, the valyl carbonyl at 157 ppm, and the imidazole C2 at 149 ppm, whereas only one epimer was visible for amide **38** obtained by using the reaction conditions developed. This evidence was considered conclusive showing that no significant racemization occurred during this reaction using the conditions proposed, and therefore it was deemed suitable for this synthetic route.
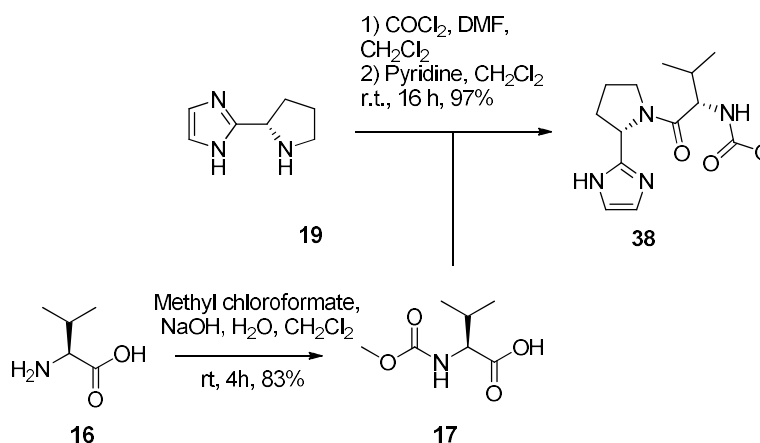
Scheme 2.13: Final conditions to reach amide **38**.

2.4.3 SEM Protection of Imidazole N-H to Form C-H Activation Reagents

Before the C-H activation of the imidazole could be attempted the imidazole N-H must be masked with a SEM group in accordance with Sames findings.[199,201] Both imidazole **42** with Boc-functionalisation, and fully functionalised imidazole **38** were protected to test the C-H activation conditions. Firstly the imidazole was deprotonated, using NaH in THF, then reacted with SEMCl, which was added dropwise. This method was used to produce compounds **43** (67 % yield) and **44** (47 % yield) (Scheme 2.14).
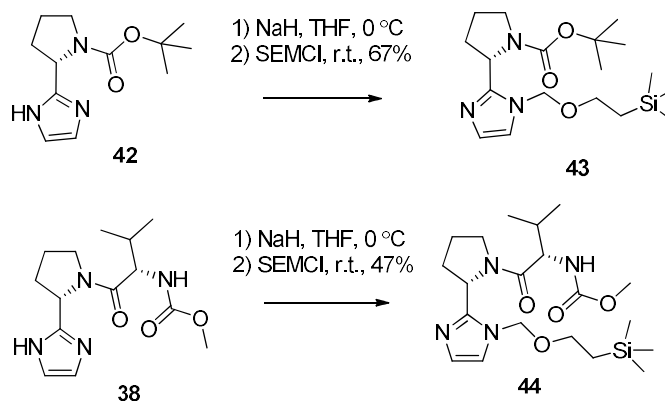
Scheme 2.14: SEM protection of imidazoles at the N-H position.

2.5 Optimisation of C-H Activation Conditions

With the completed synthesis of SEM-protected imidazoles **43** and **44**, the optimisation of C-H activation conditions was carried out. Initial optimisation was performed on the reaction between Boc-pyrrolidyl SEM-imidazole **43**. The initial conditions were based on Fagnou's broadly applicable conditions mentioned in section 2.3.1.2 where pivalic acid was used as an additive to facilitate the CMD transition state.[200]

With the completed synthesis of SEM-protected imidazoles **43** and **44**, the optimisation of C-H activation conditions was carried out. Initial optimisation was performed on the reaction between Boc-pyrrolidyl SEM-imidazole **43**. The initial conditions were based on Fagnou's broadly applicable conditions mentioned in section 2.3.1.2 where pivalic acid was used as an additive to facilitate the CMD transition state.[200]

The optimisation of the catalyst, base, additive, solvent and temperature was carried out. As expected, the absence of pivalic acid in experiments 6-8 lead to a significant drop in yields. There was little effect on changing the palladium ligands to the overall yields of reactions (experiments 2-5), although use of bis(diphenylphosphino)ferrocene (dppf) gave the best yield, and also reduced aryl bromide homo-coupling (see section 2.5.1.1). Changing the base did not affect the yield significantly either, although potassium carbonate was found to give the highest yield out of those used. Changing the solvent dramatically affected the yield, with DMA appearing the best of those tested. Interestingly dioxane, although giving a lower yield, appeared to significantly reduce aryldibromide homo-coupling (see section 2.5.1.1), and so was also explored further. 100 °C seemed to be the optimum temperature, providing the best balance between rate of reaction and side reactions occurring (experiments 19-22).

Table 2.4: Reaction data and yields for C-H activation under different conditions

Br—⟨benzene⟩—⟨benzene⟩—Br **18** + imidazole-SEM, R **43** → **45** (R = Boc-pyrrolidine)
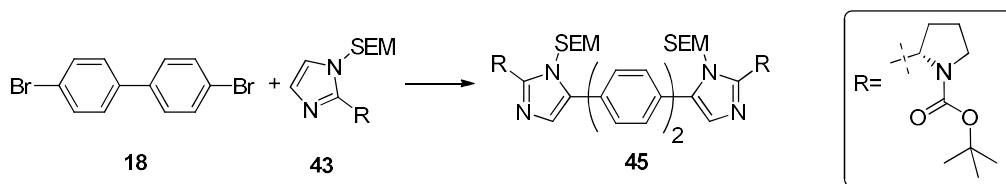
Experiment	Catalyst /Ligand	Base	Acid	Solvent	T (°C)	Yield (%)[a]
1	Pd(OAc)$_2$, HPCy$_3$BF$_4$	K$_2$CO$_3$	Pivalic Acid	DMA	140	61%
2	PdCl$_2$, HPCy$_3$BF$_4$	K$_2$CO$_3$	Pivalic Acid	DMA	140	49%
3	Pd(dppf)Cl$_2$	K$_2$CO$_3$	Pivalic Acid	DMA	140	65%
4	Pd(PCy$_3$)$_2$Cl$_2$	K$_2$CO$_3$	Pivalic Acid	DMA	140	57%
5	Pd(PPh$_3$)$_4$	K$_2$CO$_3$	Pivalic Acid	DMA	140	55%
6	Pd(OAc)$_2$, HPCy$_3$BF$_4$	Cs$_2$CO$_3$	-	DMA	140	42%
7	Pd(dppf)Cl$_2$	Cs$_2$CO$_3$	-	DMA	140	33%
8	Pd(PPh$_3$)$_4$	Cs$_2$CO$_3$	-	DMA	140	19%
9	Pd(dppf)Cl$_2$	Na$_2$CO$_3$	Pivalic Acid	DMA	140	34%
10	Pd(dppf)Cl$_2$	Cs$_2$CO$_3$	Pivalic Acid	DMA	140	60%
11	Pd(dppf)Cl$_2$	KOAc	Pivalic Acid	DMA	140	56%
12	Pd(dppf)Cl$_2$	K$_3$PO$_4$	Pivalic Acid	DMA	140	47%
13	Pd(dppf)Cl$_2$	K$_2$CO$_3$	Pivalic Acid	DMA	140	59%
14	Pd(dppf)Cl$_2$	K$_2$CO$_3$	Pivalic Acid	DMF	140	50%
15	Pd(dppf)Cl$_2$	K$_2$CO$_3$	Pivalic Acid	NMP	140	29%
16	Pd(dppf)Cl$_2$	K$_2$CO$_3$	Pivalic Acid	Dioxane	140	49%
17	Pd(dppf)Cl$_2$	K$_2$CO$_3$	Pivalic Acid	THF	140	50%
18	Pd(dppf)Cl$_2$	K$_2$CO$_3$	Pivalic Acid	IPA (wet)	140	9%
19	Pd(dppf)Cl$_2$	K$_2$CO$_3$	Pivalic Acid	DMA	100	69%
20	Pd(dppf)Cl$_2$	K$_2$CO$_3$	Pivalic Acid	DMA	120	61%
21	Pd(dppf)Cl$_2$	K$_2$CO$_3$	Pivalic Acid	DMA	160	53%
22	Pd(dppf)Cl$_2$	K$_2$CO$_3$	Pivalic Acid	DMA	80	64%
23	Pd(dppf)Cl$_2$	K$_2$CO$_3$	Pivalic Acid	Dioxane	100	60%
24	Pd(dppf)Cl$_2$	K$_2$CO$_3$	Pivalic Acid	Toluene	100	30%

[a]Yield of product determined by LCMS peak integration measured by adsorption at 254 nm.

In order to ensure that the LCMS profile of the reaction was indicative of the correct product forming, the reaction was carried out under the optimised conditions and the product was purified with a yield of 76% after purification. As the product contained Boc-protected pyrrolidine, it was difficult to fully characterise using ^{1}H NMR spectroscopy, due to the presence of rotamers (see section 2.4.1.1). Variable temperature NMR studies were carried out, and the peaks were found to fully converge at 70-80 °C (Figure 2.7), where the product was fully characterised as the compound **45** resulting from double C-H activation.
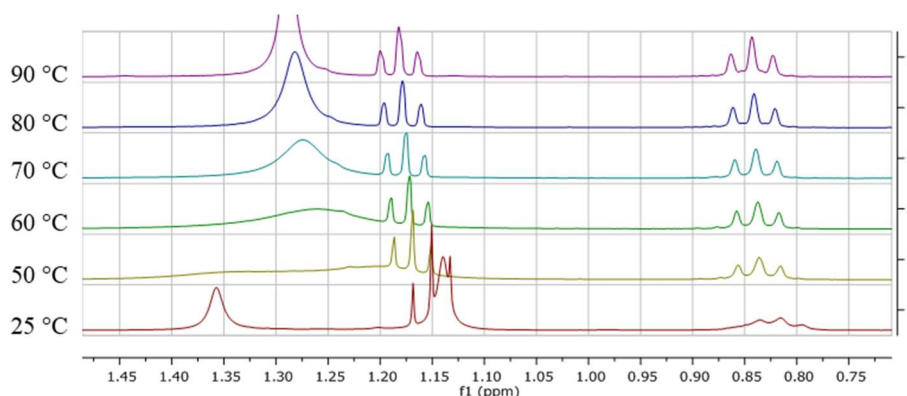

Figure 2.7: VT NMR of compound **45** to ensure proper assignment.

2.5.1.1 Homo-coupling as a competing reaction, lowering the yield

During the optimisation process of the C-H activation reaction, the major side-product was identified to be the product of homo-coupling between two dibromobisphenyl molecules which had then undergone homocoupling to form side product **47 (**Scheme 2.15**)**. The side product was clearly distinguishable from the product on the LCMS trace (Figure 2.8). Despite their distinct separation on the LCMS trace, the product and the side product of homo-coupling were found to completely co-elute on silica gel column chromatography. Because of this, lowering the amount of homo-coupling during the reaction was considered a priority during optimisation.
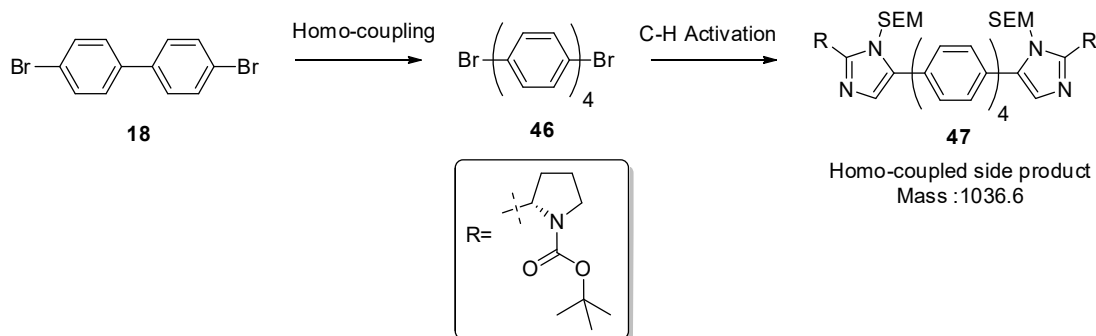
Scheme 2.15: Formation of side product through homo-coupling.

Use of the dppf ligand, as well as conducting the reaction at or below 100 °C, gave a notable reduction in the amount of homo-coupling (Figure 2.8, experiments 1 and 19). Interestingly, when 1,4-dioxane was used as a solvent, no mass ion for the homo-coupled product was detectable (Table 2.4, experiment 23), although there was a significant overall drop in yield, which meant that DMA was still the solvent of choice for the reaction.
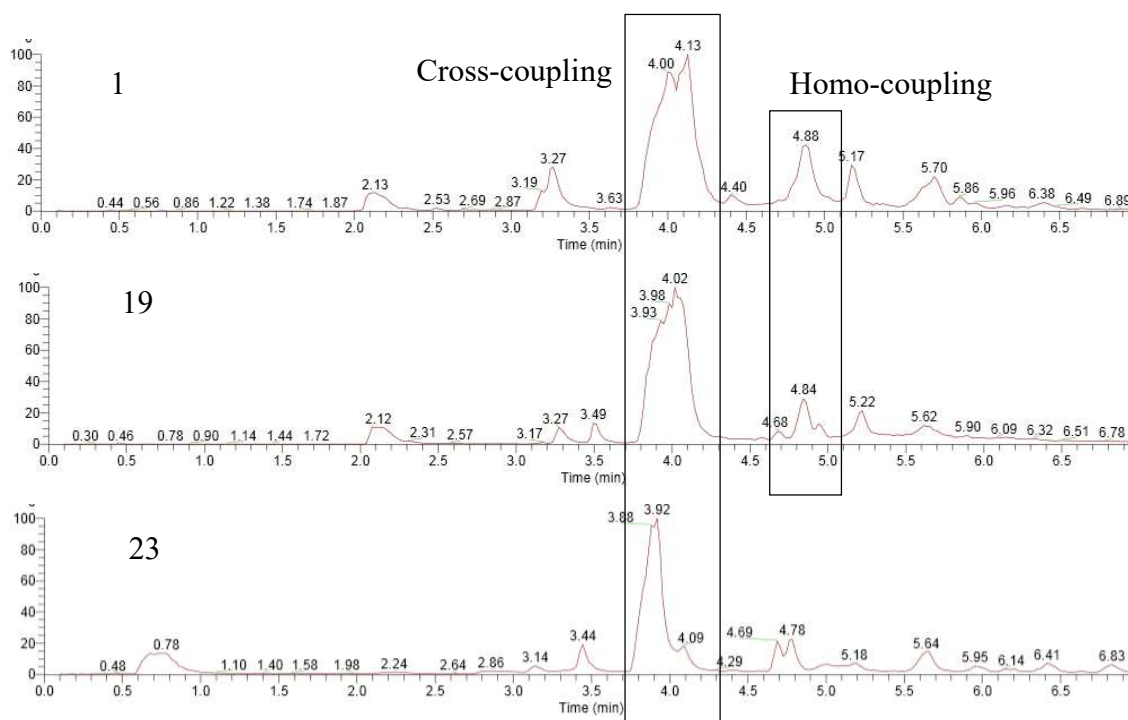

Figure 2.8: LCMS chromatogram for experiments 1, 19 and 23.

When the reaction was carried out in DMA at 80 °C (Table 2.4, entry 22), there was also less homo-coupling observed, however again there was a drop in yield. As such, the conditions used in experiment 19 were seemed to be the most effective overall, with a yield of 69% as measured by LCMS. This was found to equate to a yield of 76% of pure product after workup.

2.5.2 Confirming the Optimised Conditions for Fully Functionalised Imidazole 44

Once the conditions were optimised for imidazole **43**, a shorter optimisation process was carried out for imidazole **44** in order to ensure that the results observed previously regarding the best conditions held true. As the temperature, solvent and base were found to have the greatest effect on the reaction profile in the previous experiments, these were the only variables tested in this optimisation process (Table 2.5).

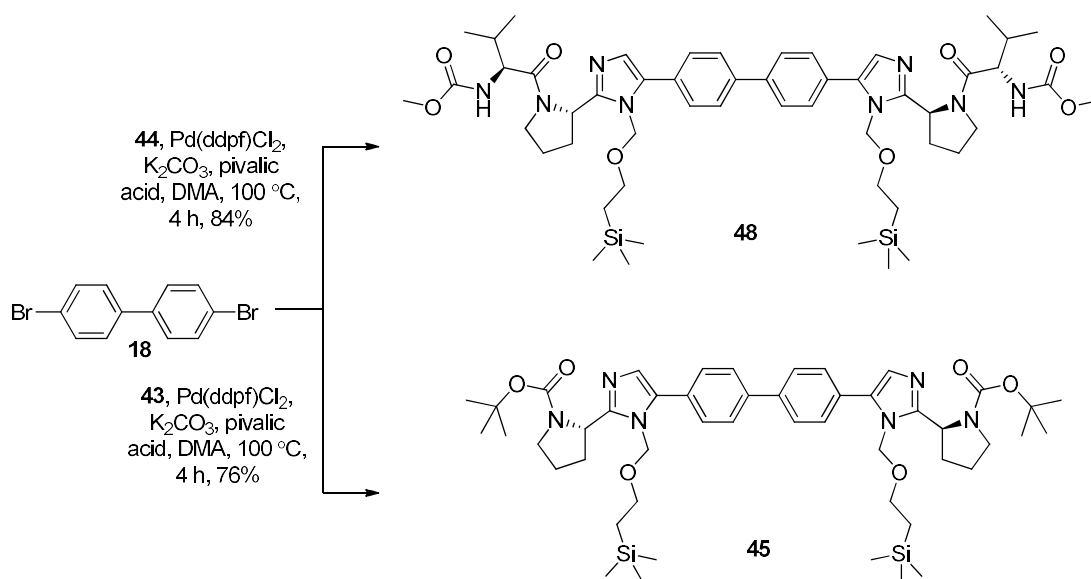**Table 2.5:** Optimisation conditions for the C-H Activation

Conditions: i) Pd(dppf)Cl₂, PivOH, base, 4 hours
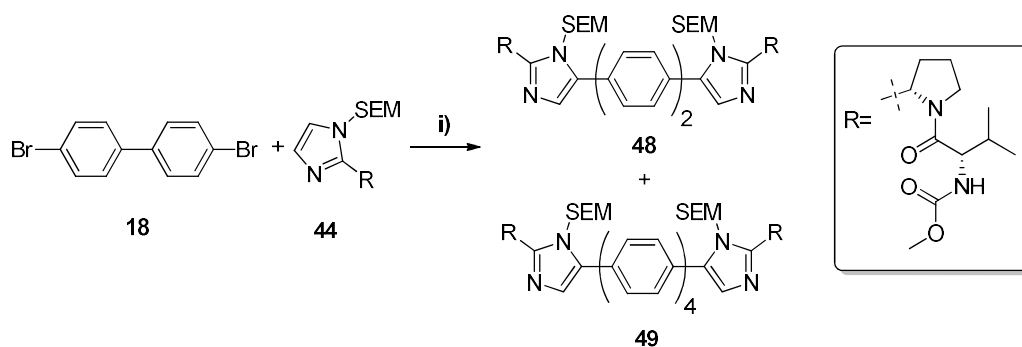
Entry	Base	Solvent	Temperature	Yield 48[a]	49:48[a]
1	K₂CO₃	DMF	100 °C	65%	0.13
2	K₂CO₃	dioxane	100 °C	10%	0
3	K₂CO₃	NMP	100 °C	38%	0.05
4	K₂CO₃	THF	100 °C	25%	0.11
5	K₂CO₃	DMA	100 °C	78%	0.03
6	K₂CO₃	DMA	60 °C	1%	0
7	K₂CO₃	DMA	80 °C	5%	0.15
8	K₂CO₃	DMA	120 °C	69%	0.07
9	Cs₂CO₃	DMA	100 °C	21%	0.10
10	KOAc	DMA	100 °C	3%	0.23
11	K₃PO₄	DMA	100 °C	72%	0.06

[a]Yield of **48** and ratio of **49:48** determined by LCMS peak integration measured by adsorption at 254 nm.

It was found that in general, the results obtained followed the same patterns for the yields of the product and by-product as when the reaction was previously optimised for imidazole **43**. K_2CO_3 was still the highest yielding base, and carrying out the reaction at 100 °C still seemed to be the best compromise between rate of reaction and side product formation. Gratifyingly, it was found that carrying out the reaction in DMA produced very little side product at all, whereas before, DMA enabled some homo-coupling of the aryldibromide to form side products. As a result, the overall yield by LCMS increased to 78 %.

As before, the optimised conditions were used on a larger scale to synthesise and purify the C-H activation product **48**, so that it could be fully characterised. Whereas the LCMS

predicted yield was measured as 78 %, the isolated yield was calculated as 84 % (Scheme 2.16). The lack of Boc groups also meant that no rotamers were visible by ^{1}H NMR.

Scheme 2.16: Isolated yields for optimised C-H activation reaction conditions

2.5.3   SEM-group Removal to Yield Daclatasvir

Once the C-H activation reaction had been optimised, conditions were tested for the removal of the SEM group from the product. Treatment of compound **48** with tetrabutlyammonium fluoride (TBAF) in ethyl acetate did not yield any product, even after heating at 60 °C for several days. More successful reaction conditions involved the use of trifluoroacetic acid (TFA) in dichloromethane, which showed quantitative conversion to the deprotected amine TFA salt by LCMS after eight hours. This material was then purified by preperative HPLC in order to obtain pure daclatasvir **10** to be fully characterised.
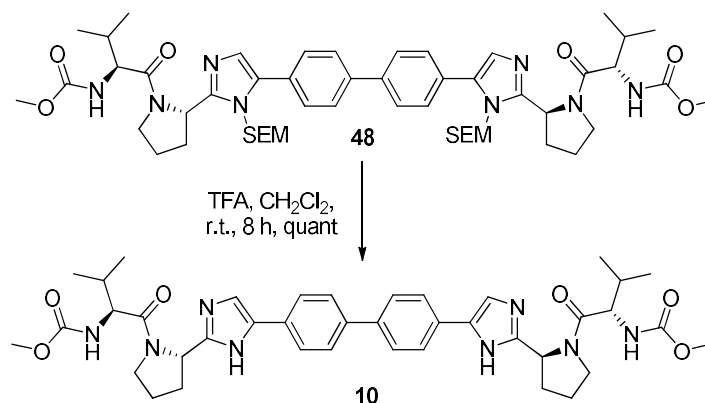
Scheme 2.17: The quantitative removal of the SEM protecting groups by TFA in CH_2Cl_2.

2.6 Testing the Scope of the C-H activation reaction on different aryl dibromides

With the conditions for the C-H activation and deprotection steps in hand, the scope of the reaction was tested. Aryl dibromide cores were selected from several examples of NS5A inhibitors in the literature,[187,214–216] as a means of proving the methodology could be used to synthesise pharmaceutically-relevant compounds (Figure 2.9). Secondly, other aryl dibromides were chosen to give a wide range of functionalities to test the scope of the reaction.
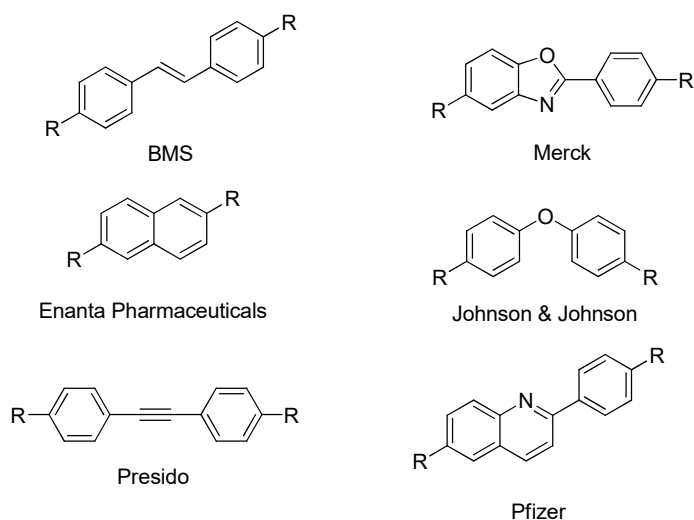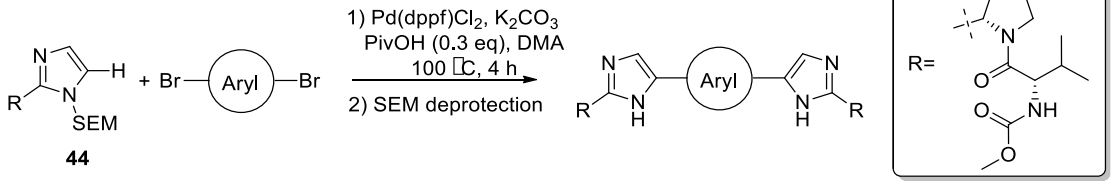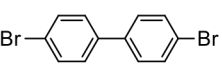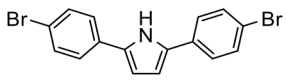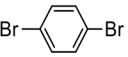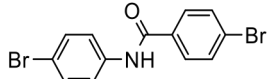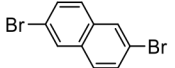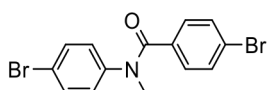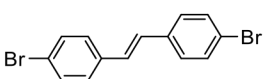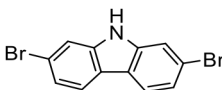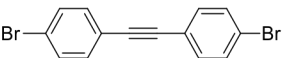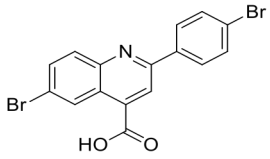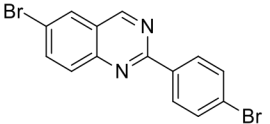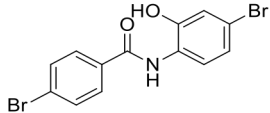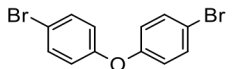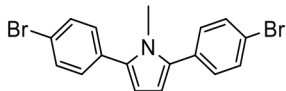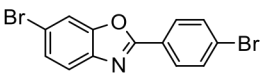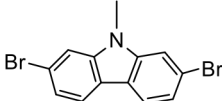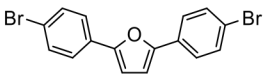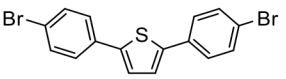

Figure 2.9: Cores selected for synthesis from published NS5A inhibitors.[187,214–216]

Overall, eighteen cores were either purchased or synthesised. The preparation of all synthetic intermediates is discussed in section 2.7. Of the 18 aryl dibromides tested under C-H activation, 11 were found to react. The results of these reactions can be seen in Table 2.6.

Overall, the reactions yielded between 42% and 80%. Generally, the larger and more complex the aryl dibromide, the lower the yield. Unless otherwise stated, the product of the C-H activation was fully characterised after removal of the SEM group by TFA in CH_2Cl_2. There are several functionalities that seem to be incompatible with the reaction, which are discussed below.

Table 2.6: Successful C-H activation reactions, with the yields of reaction.

Entry	Aryl Dibromide	Yield[a]	HD:P[b]	Entry	Aryl Dibromide	Yield[a]	HD:P[b]
1		78%	0.13	11		0%	-
2		80%	0	12		0%	-
3		64%	0.37	13		67%	0
4		37%	0.72	14		0%	-
5		0%	-	15		0%	-
6		0%	-	16		0%	-
7		42%	0.68	17[d]		49%	0
8		67%	0.12	18[d]		42%	0.31
9[c]		45%	0.26				
10[c]		58%	0.23				

[a]Yield of product determined by LCMS peak integration measured by adsorption at 254 nm. Characterisation was carried out on the fully deprotected product (see experimental section); [b]Ratio of homo-dimer to product; [c]Product decomposed under deprotection and characterisation was carried out on the SEM-protected product. [d]Product detected by LCMS but purification failed.

2.6.1 Incompatibility with alkynyl residues.

One core that was found to be incompatible with the palladium-catalysed C-H activation was the alkynyl biphenyl core **53**. There was found to be no observable change to the LCMS reaction profile when the reaction was followed during heating. This could be

explained by the affinity of palladium for binding to alkyne moieties, whereby it is known to be able to carry out a range of reactions.[217,218] For example, under similar conditions to those used in this reaction, the double addition of aryl-halides across the carbon-carbon triple bond has been observed (Scheme 2.18).[219]
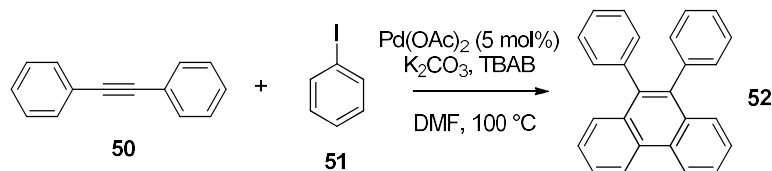
Scheme 2.18: Reactivity of biphenylalkyne reagent under similar conditions to those used in this work.[219]

When the alkyne-containing reagent was added to the solution, there was a significant colour change from light orange to deep red, which supports the theory that the co-ordination sphere of palladium may have changed. This is likely to affect the catalyst's ability to co-ordinate with the imidazole in order to carry out the C-H activation and thus prevent the reaction from occurring (Scheme 2.19).
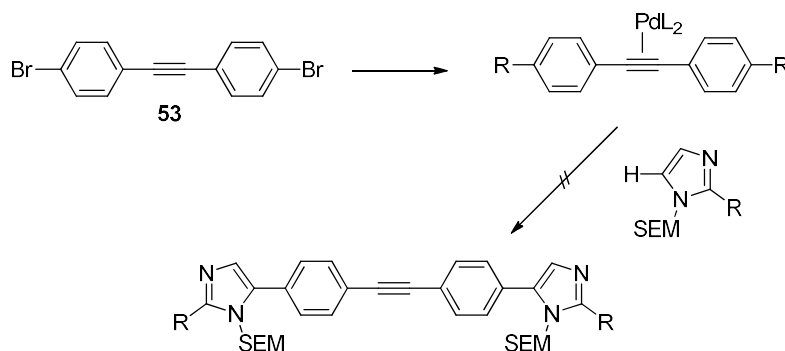
Scheme 2.19: Possible co-ordination of palladium to alkyne **53**, preventing the C-H reaction from proceeding.

2.6.2 Incompatibility with N-H residues
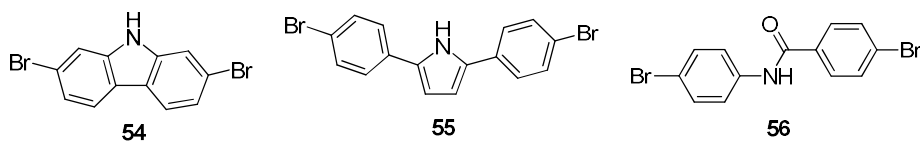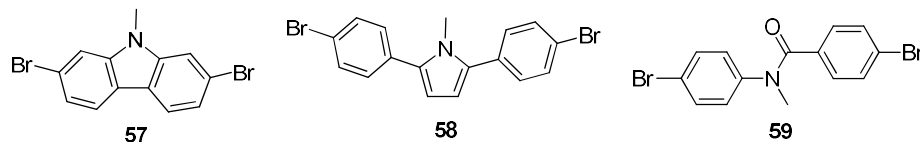
Figure 2.10: Methylation of N-H bonds instils compatibility with C-H activation.

When screening different cores against the C-H activation chemistry, it was found that three cores that were synthesised containing an N-H bond did not react. In order to test

whether it was the N-H bond which was the cause of unreactivity, each core was *N*-methylated. In each case, this resulted in a successful reaction (Figure 2.10).

This observation was supported by wider precedent in the literature where a C-H activation would not occur in the presence of an N-H, but the reaction proceeded after masking the N-H bond. For example the intramolecular imidazole C-H activation carried out by Suzuki and coworkers, where alkylation of an amide with a butyl group allowed the reaction to proceed (Figure 2.11; **A**).[197] The examples of reactions with an N-H present are nearly always intramolecular, for example Maes' work on isocryptolepine synthesis (Figure 2.11; **B**),[220] and Queiroz's thienocarbazole synthesis (Figure 2.11; **C**). The entropic favourability of the intramolecular reaction seems to counterbalance the loss reactivity that results from having an N-H present in the molecule.
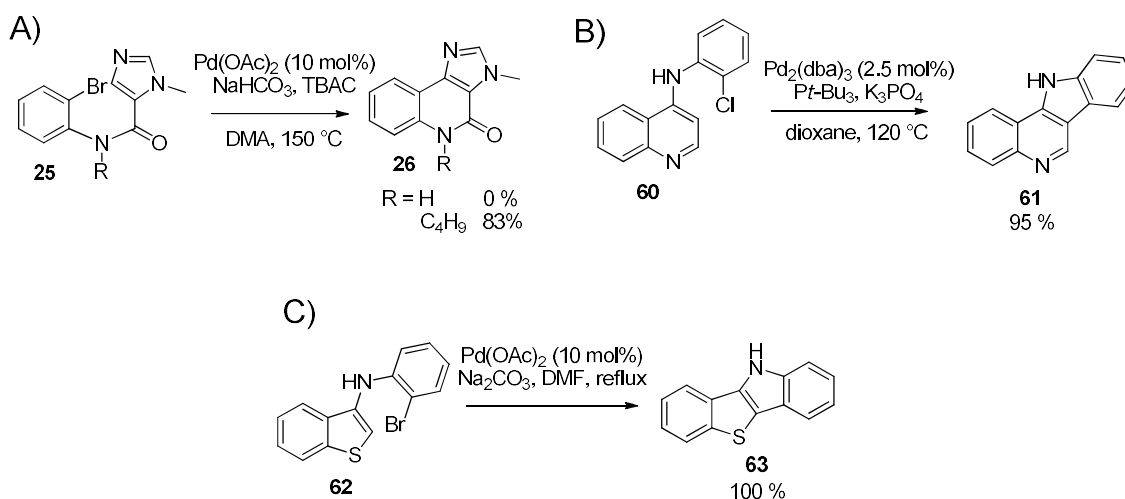
Figure 2.11: Literature examples of palladium catalysed C-H activation: A) Example of C-H activation not compatible with presence of NH bond;[197] B) Intramolecular pyridyl C-H activation, containing NH bond;[220] C) Intromolecular thiophene C-H activation, tolerating the presence of NH functionality.[221]

It therefore seems that the observation that N-H bonds are incompatible with this reaction is consistent with what has been reported in previous examples in the literature.

2.6.3 Incompatibility of pyridyl functionality

The final functionality which appeared to be incompatible with the C-H activation reaction was the pyridyl functionality, which was found in cores **64** and **65** (Figure 2.12). Aryl dibromide **64** contains a carboxylate functionality, although, as the presence of pivalic acid has been shown to be beneficial to the reaction, it seems unlikely that this is the cause of incompatibility in this case.
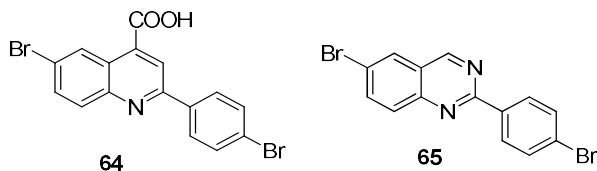
Figure 2.12: Cores with pyridyl functionality, unsuccessful at the C-H activation reaction.

Interestingly, there was some precedent in a patent published by Pfizer that quinazoline **65** should be compatible with the C-H activation chemistry.[216] Because of this, the core was investigated further. The C-H activation reaction was carried out using a different imidazole reagent, containing a Boc functionality in place of the valyl-carbamate (Scheme 2.20). In this case, the reaction was found to be successful.
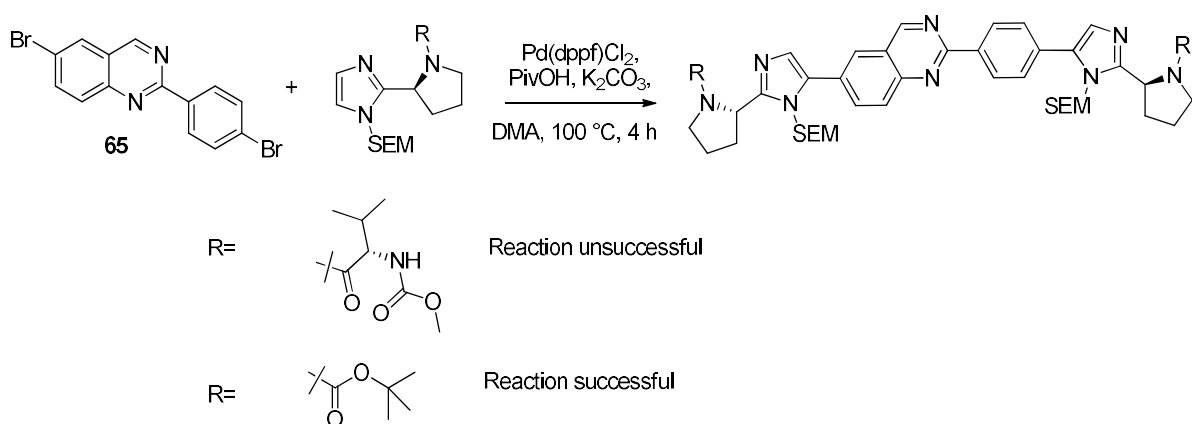
Scheme 2.20: The C-H activation reaction was successful with a Boc, but not with a valyl carbamate

In an attempt to explain this, both reactions were followed by LCMS analysis at regular intervals. In the unsuccessful reaction between quinazoline **65** and the valyl-carbamate imidazole reagent, the quinazoline starting material was found to be consumed, despite no product being formed. Instead, one predominant peak emerged as a side product. This side product could not be isolated and characterised, however its mass spectrum seemed to indicate a polymeric compound. When the equivalent reaction between quinazoline **65** and the Boc containing imidazole was carried out, the formation of the same polymeric by-product was observed, however the C-H activation reaction was much quicker and the formation of the desired product became the dominant one. This evidence suggested that the larger valyl-carbamate peptide functionality on the original reaction was affecting the rate of C-H activation due to presence of an *N*-H bond, meaning that the side reaction, which occurred in both reactions, became the dominant product.

Although it was not possible to identify the by-product, there is some precedent in the literature for potential reactions that may have been facilitated by the ability of the

quinazoline to act as a palladium ligand. The pyridine nitrogen has been shown to be a directing group for the C-H activation on aryl groups. Melanie Sanford and co-workers have done much of the work in this area. For example, it was shown that a pyridine group would direct the addition of a phenyl with *ortho*-regioselectivity to an 88% yield (Scheme 2.21).[192] Although normally an oxidising reagent is used for the C-H activation of arenes directed by pyridyl ligands, it demonstrates that the quinazoline may be directing C-H activation onto the phenyl group on **65**.
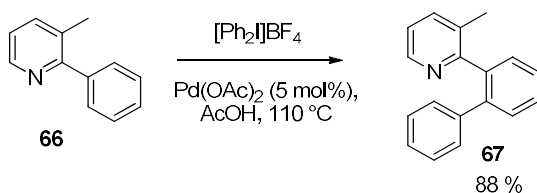
Scheme 2.21: Pyridyl group acting as a directing group for C-H activation.[192]

2.7 Synthesis of different cores for C-H activation

Many of the aryl-dibromides used to test the scope of the C-H activation reaction were not commercially available. The syntheses of these cores are described below.

2.7.1 Synthesis of bisphenylalkyne 53
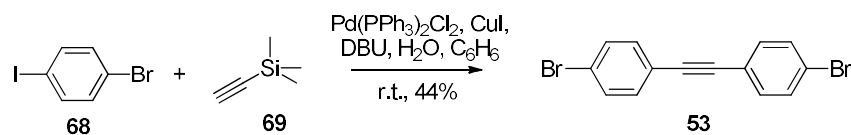
Scheme 2.22: Synthesis of bisphenylalkyne **53**

Bisphenylalkyne **53** was synthesized in one step from 1-bromo-4-iodobenzene (**68**), using methodology developed by Brisbois.[222] The reaction is a modified Sonogashira reaction, where trimethylsilyl acetylene is coupled to an aryl iodide, co-catalysed by palladium and copper. The presence of DBU and water then causes the labile trimethylsilyl group to cleave, liberating the second ethynyl hydrogen, which can undergo a second Sonogashira coupling. This reaction was successful with a yield of 44% after purification.

2.7.2 Synthesis of Dibromobiphenylether 71
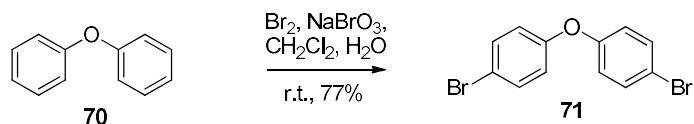
Scheme 2.23: Synthesis of dibromobiphenyl ether **71.**

Dibromobiphenylether **71** was produced using a simple bromination reaction, using bromine and sodium bromate as oxidizing agents. The reaction was carried out in a biphasic system of water and dichloromethane, and gave a 77% yield of the product.

2.7.3 Synthesis of Benzoxazole 75
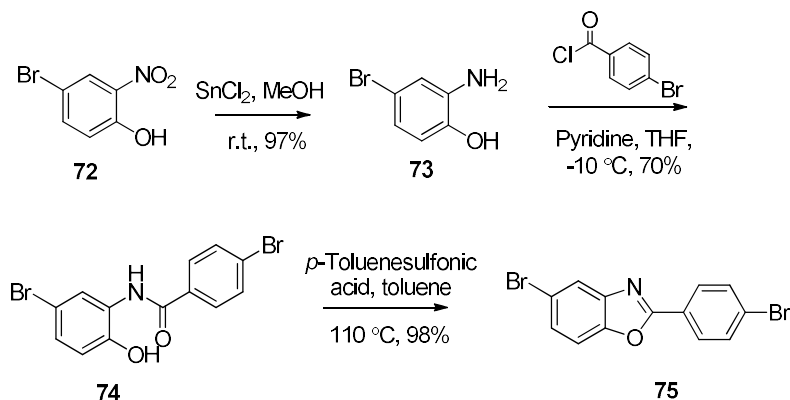
Scheme 2.24: Synthesis of benzoxazole **75**.

The first step of this synthesis involved tin (II) chloride reduction of 4-bromo-2-nitrophenol to the resultant aniline **73**.[223] This reaction gave a 97% yield after overnight stirring in methanol. Originally, the aminophenol **73** was reacted directly with 4-bromobenzoic acid in the presence of by propylphosphonic anhydride[224] to give benzoxazole **75** in one-pot. However, this reaction resulted in an intractable crude material, with no major product. Instead, the synthetic route summarised in Scheme 2.24 was undertaken. 4-Bromobenzoic acid was transformed to the benzoyl chloride derivative using oxalyl chloride and *N,N'*-dimethylformamide (DMF) in dichloromethane. This was then reacted with aminophenol **73** to give amide **74**. The temperature during the reaction was kept at -10 °C to control the selective formation of the amide over the less favoured ester. After all the starting material was consumed, any benzoic ester that had formed was saponified by stirring the crude material in aqueous lithium hydroxide, and then the resulting side products removed using aqueous work up. As a result, the product was obtained in 70 % yield. Amide **74** was then reacted with *p*-toluenesulfonic acid in refluxing toluene to obtain the benzoxazole **75** in 98% yield.[225]

2.7.4 Synthesis of amide 56 and *N*-methyl amide 59
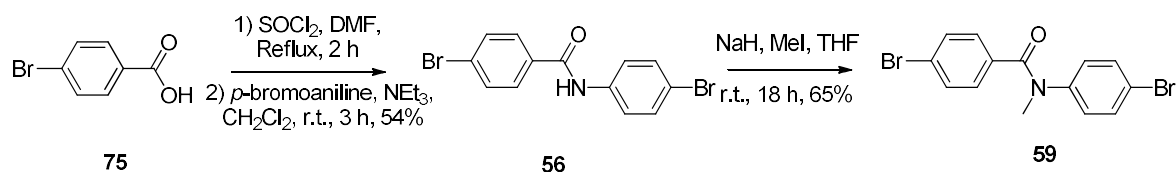
Scheme 2.25: Synthetic route to amide **56** and *N*-methylamide **59**.

In order to synthesize amide **56**, *p*-bromobenzoic acid was firstly transformed into *p*-bromobenzoyl chloride by refluxing in thionyl chloride for 2 hours, with catalytic DMF to catalyse the reaction through the Vilsmeier-Haack intermediate. This was then reacted with *p*-bromoaniline in the presence of triethylamine in dichloromethane at room temperature to give amide **56** to 54% yield. The amide was then *N*-methylated using sodium hydride and methyl iodide in THF, which after stirring for 18 hours at room temperature, afforded *N*-methylamide **59** in a 65% yield.

2.7.5 Synthesis of carbazole 54 and *N*-methylcarbazole 57
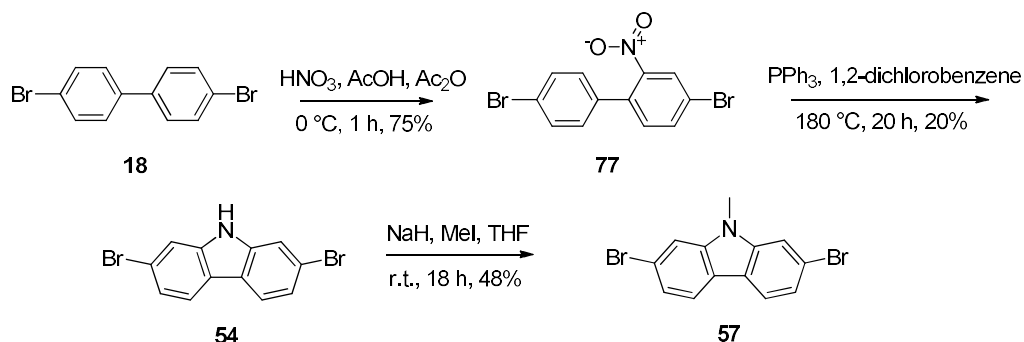
Scheme 2.26: Synthetic route to carbazole **54** and *N*-methylcarbazole **57**

Using a modified version of the conditions developed by Goto,[226] the nitration of 4,4'-dibromobiphenyl to compound **77** was successful, with a yield of 75%. Compound **77** was then reduced and cyclised to the corresponding carbazole **54** by refluxing in 1,2-dichlorobenzene in the presence of triphenylphosphine. Finally, carbazole **54** was methylated by reacting with sodium hydride, followed by methyl iodide, in THF, giving *N*-methylcarbazole **57** in 48 % yield.

The mechanism for this reaction is proposed to proceed by the triphenylphosphine mediated transformation of the nitro group to a nitroso.[227] Further reaction with triphenylphosphine allows cyclization of the carbazole ring with further production of triphenylphosphine oxide. Rearomatisation of this compound gives the carbazole product (Scheme 2.27).
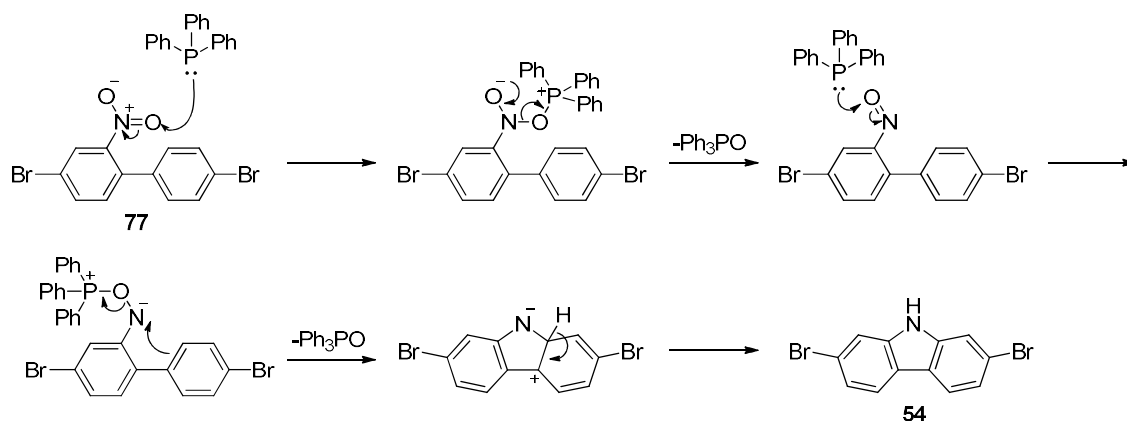
Scheme 2.27: Proposed mechanism of carbazole **54** formation using triphenylphosphine.[227]

2.7.6 Synthesis of Paal-Knorr products: furan 81, thiophene 82, and pyrroles 55 and 58

a) HCl (g), MeOH, CH$_2$Cl$_2$, 0 °C, r.t., 3 h, 74%; b) Lawesson's reagent, toluene, reflux, 3 h, 75 %; c)NH$_4$OAc, AcOH, reflux, 21 h, 62%
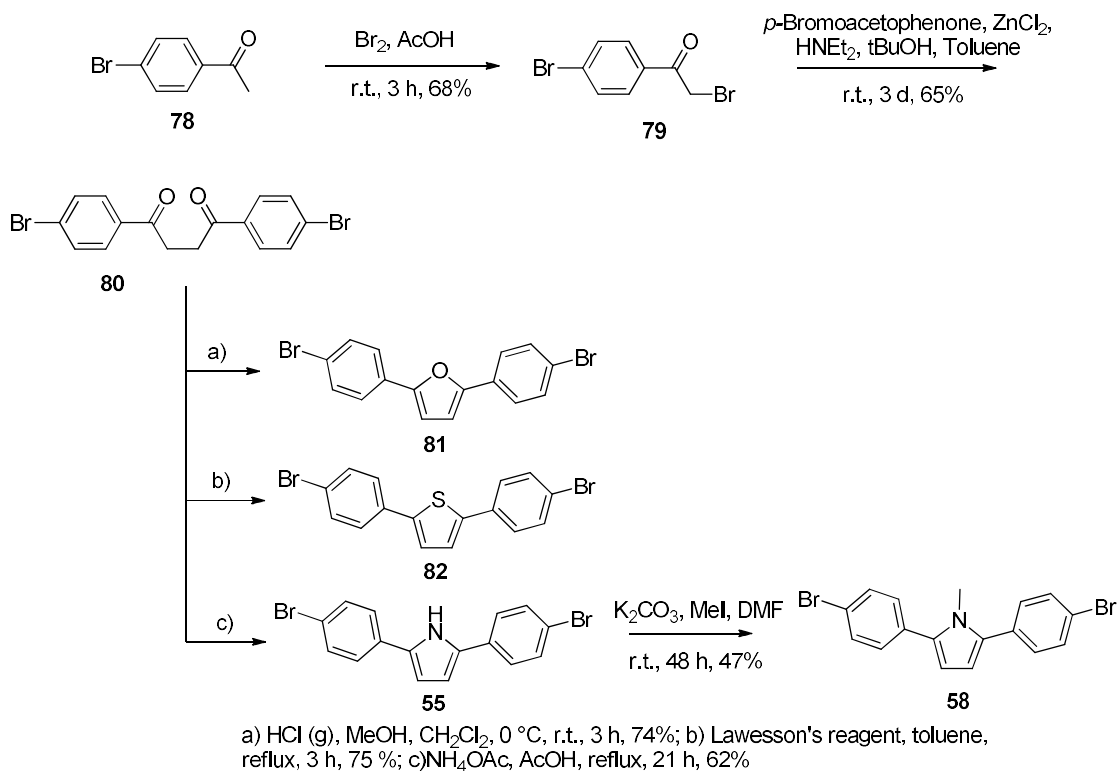
Scheme 2.28: Synthetic route involving Paal-Knorr synthesis of various heterocycles

The common intermediate for the preparation of these 4 heterocyclic compounds was the 1,4-diketone **80** (Scheme 2.28). In order to synthesize 1,4-diketone **80**, *p*-bromoacetophenone was brominated using elemental bromine in acetic acid, to give the α-bromo ketone **79** in 68% yield. The dimerization of this compound to compound **80** was first attempted with a Wurtz-type metal-radical reaction. This was based on work done by Ceylan, where α-bromoacetophenones were dimerised by treatment with elemental zinc and catalytic iodine.[228] These conditions proved unsuccessful, however,

and were substituted for an unusual set of conditions developed by Kulinkovich and coworkers,[229] for the synthesis of 1,4-diketones. Under these conditions, a mixture of 4-bromoacetophenone **78** and α-bromo ketone **79** was treated with a solution of zinc(II) chloride, *tert*-butanol and diethylamine. The proposed mechanism for this reaction is shown in Scheme 2.29. Zinc(II) chloride and *tert*-butanol form the active ClZnO*t*Bu species in solution. This facilitates the reaction between the methylketone and α-bromomethylketone. The use of the zinc reagent as a base results in this reaction being entirely selective to the desired product and prevents epoxide formation occurring as a side reaction. When treated with diethylamine, the product rearranges to the desired 1,4-diketone **80**, through a cyclopropyl intermediate.[230]
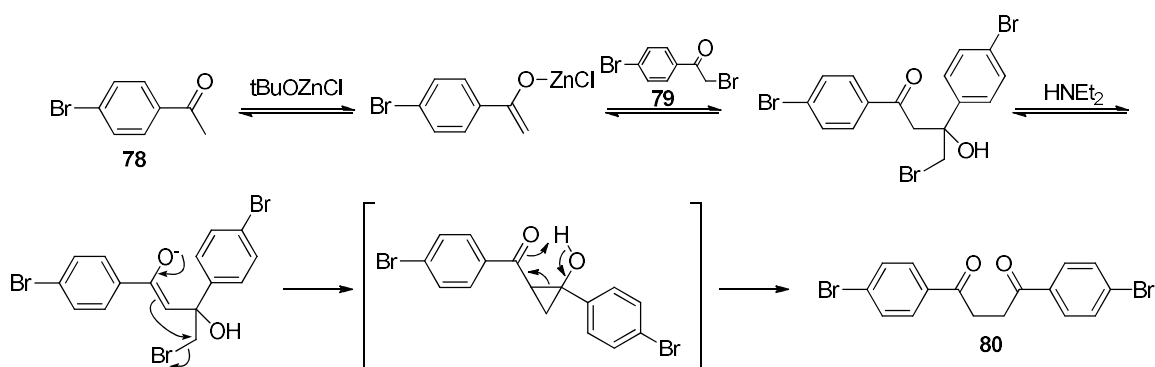
Scheme 2.29: Proposed mechanism for Kulinkovich's 1,4-diketone formation.[230]

Using these conditions, 1,4-diketone **80** was obtained in a 65% yield. Once this was achieved, the different Paal-Knorr conditions could be employed to achieve the three heterocyclic compounds. Furan **81** was obtained in a 74% yield by cyclocondensation obtained in a solution of methanol and dichloromethane saturated with HCl gas.[231] Interestingly, this reaction was found to be dependent on the temperature of the solution being kept below 0 °C during HCl gas saturation. This could be as a result of the higher solubility of HCl at lower temperatures allowing increased saturation of the solution. Thiophene **82** was synthesised by refluxing 1,4-diketone **80** in the presence of Lawesson's reagent in toluene for 3 hours, giving the product in 75% yield.[232] Finally, pyrrole **55** was synthesised in 62% yield by refluxing 1,4-diketone **80** in acetic acid, in the presence of ammonium acetate.[233] Methylation of **55** was carried out with potassium carbonate and methyl iodide in DMF at room temperature, to give *N*-methylpyrrole **58** in 47% yield.

2.7.7 Synthesis of quinazoline 65
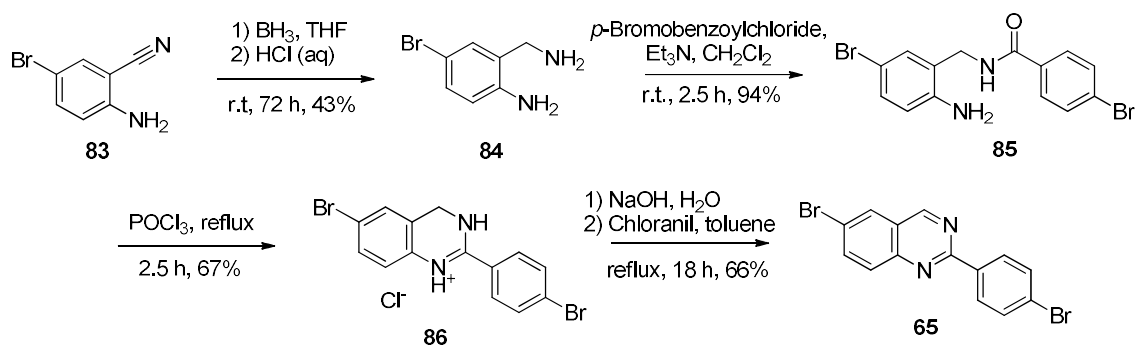
Scheme 2.30: Synthetic route to quinazoline **65**.

Quinazoline **71** was synthesized in a four step process reported by Pfizer.[187,216] Firstly, 2-amino-5-bromobenzonitrile was reduced to the corresponding primary amine **84** using a solution of borane in THF stirred over 3 days. After this, the resulting boron-nitrogen adducts were cleaved by HCl mediated hydrolysis to give the product in a 43% yield. 4-Bromobenzoyl chloride, prepared from the corresponding acid and sulfuryl chloride, was then reacted with amine **84** to give amide **85** in excellent yield. The reduced nucleophilicity of the aniline due to delocalisation of the nitrogen lone pair was found to be sufficient so that none of the aniline-amide regioisomer was observed in the crude material. **85** was then transformed into dihydroquinazolinyl chloride **86** by refluxing in phosphoryl chloride, to give the desired product in 67% yield. Finally, intermediate **86** was transformed into quinazoline **65** after two steps; liberation of the free base by treatment with aqueous sodium hydroxide and aromatisation in the presence of chloranil to give the desired aryl bromide core in a 66% yield.

2.8 Conclusion

11 successful out of 18 examples
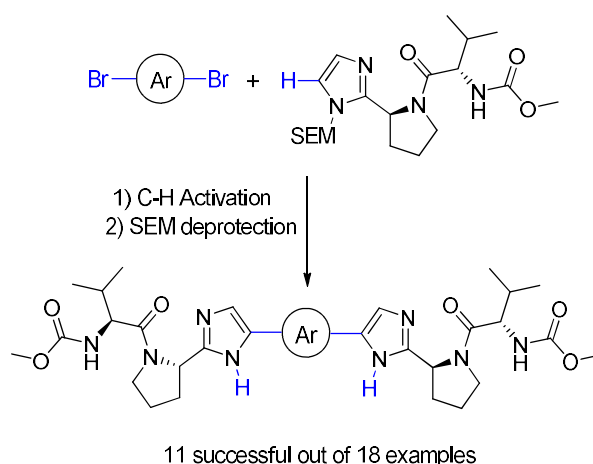
Scheme 2.31: Overview of the C-H activation methodology developed.

In conclusion, a new methodology has been developed in order to functionalise aryl dibromides with 'daclatasvir-like' side chains using C-H activation. The scope of the reaction has been tested, and chemical functionalities that are incompatible with the reaction conditions have been identified. Care has been taken throughout the synthetic route to ensure that racemisation of stereogenic centres did not occur, by monitoring the optical rotation of the molecules, as well as identifying diastereomers by ^{13}C NMR. This procedure has great potential for the development of new NS5A inhibitors and efficiently explore SAR, as the convergent nature of the synthesis means that the core region of the molecule can be diversified at a late stage. The scope of the reaction was tested, and out of 18 aryl dibromides tested, 11 were successful in the reaction.

The methodology was also developed with the mind set of potentially improving the physical properties of highly lipophilic α-helical mimetics, such as Hamiltons *i*, *i*+3/4, *i*+7 terphenyl compounds, by functionalisation with 'daclatasvir-like' solubilising side chains.[149] The C-H activation condition optimisation with the scope of reaction have been published under peer review.[234]

3   Targeting PALB2 with Computational-Based Design of Inhibitors

3.1   Chapter Summary

In this chapter computational work was carried out to design potential inhibitors of the partner and locator of BRCA2 (PALB2)/breast cancer 2 protein (BRCA2) PPI. The therapeutic potential of inhibiting the PALB2/BRCA2 interaction as a cancer treatment through synthetic lethality is outlined. The process by which small molecule mimetics were designed is described, followed by the computational alignment scoring and docking techniques used to predict which small molecule BRCA2 mimetics are most likely to inhibit the PALB2/BRCA2 interaction. The results of this work have been able to guide the design of BRCA2 mimetics for synthesis.

Miniproteins mimicking the BRCA2 WFXXL motif were then designed. During this process a large number of sequences were designed based on proteins determined to be structurally stable in the PDB, and then homology modelling and alignment techniques were used to select the best candidates. Ten different sequences were purchased, and the α-helicity of one sequence determined to be 95% by CD spectroscopy.

3.2   PALB2/BRCA2 as A Potential Target for Cancer Therapeutics

Large scale PPI drug discovery programmes in pharmaceutical companies tend to focus on well understood interactions such as p53/Hdm2[169,235,236] and the Bcl/BH3 family that have associated strong clinical target validation.[103,171,237,238] For these targets, the large amount of existing research detailing fragment and small-molecule binding activity provides an attractive starting point for the chemical leads required. Academic research groups can struggle to compete with such drug discovery programmes, as they do not often have large research budgets or access to the equipment necessary for a chemistry led SAR study. However, small-scale innovation from academic groups has provided much of the early-stage work that has enabled large scale PPI inhibitor development.[146,147,239–241] With this in mind, the PALB2/BRCA2 interaction was selected as an ideal target for an academic group. It offers potential as a cancer therapeutic target but there is a lack of known inhibitors in the literature.

PALB2 is a protein essential to the correct functioning of BRCA2 mediated repair of DNA double strand breaks by homologous recombination. It interacts with free BRCA2 within the cell and localises it to the DNA damage site.[242] The PPI occurs between *N*-terminal amino acids 10-30 of BRCA2 and a *C*-terminal hydrophobic binding cleft on one side of a W40-type β-propeller of PALB2 (Figure 3.1).[243] The binding region of BRCA2 forms a short α-helix that interacts with the PALB2 binding cleft predominantly through residues Trp31, Phe32 and Leu35, which exist in an *i*, *i*+1, *i*+4 configuration. Trp31 was identified as a hot-spot residue by mutation studies, which showed that BRCA2$_{(10-40)}$ sequences with W31R and W31C mutations exhibit a significant loss in activity.[242]
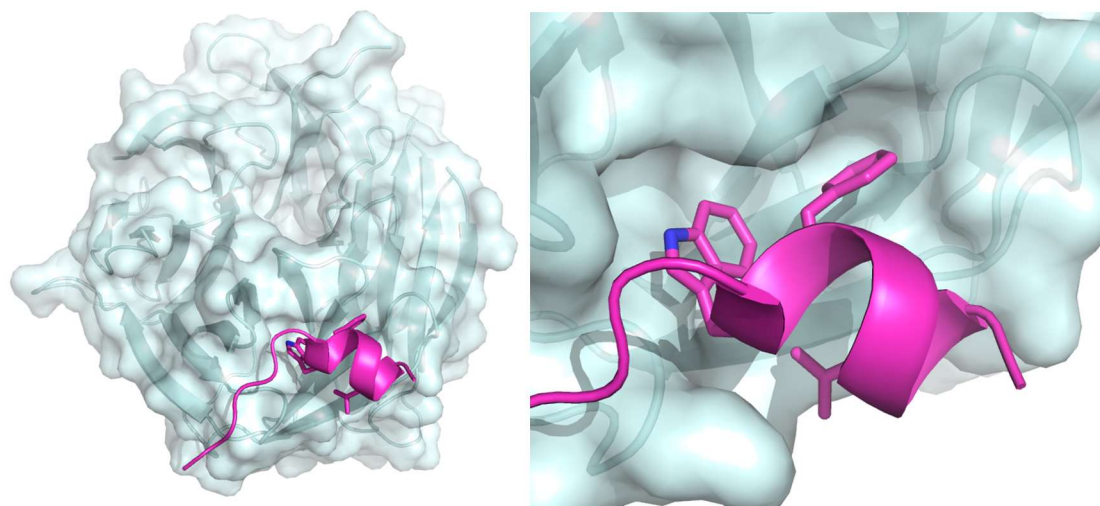
Figure 3.1: Structure of BRCA2 bound to PALB2 (PDB code: 3EU7). The BRCA2 α-helix is shown in magenta, displaying the key residues Trp31, Phe32 and Leu35. The PALB2 β-propeller structure is shown in light blue.

Although the exact role and function of PALB2 is yet to be fully understood, it is known that its interactions with BRCA2 and breast cancer 1 protein (BRCA1), another binding partner, are key to the correct functioning of homologous recombination within cells.[244] It seems that PALB2 acts as a structural link between BRCA1 and BRCA2, forming a multi-protein complex which is able to locate double-strand breaks and instigate homologous recombination.[245,246]

Loss of function mutations of BRCA1 and BRCA2, which impair homologous recombination within cells, have long been known to result in susceptibility to certain cancers.[247,248] As such it has been hypothesised that mutant forms of PALB2 will also

result in cancer susceptibility. Indeed, it was recently found that PALB2 mutations lead to a genetic predisposition to pancreatic cancers.[249] Mutations in PALB2, as in BRCA1 and BRCA2, also afford an increased risk of developing breast-cancer.[250] As a result, there is increased interest in identifying patients at high risk of developing tumours by establishing whether they have mutations in PALB2 using DNA sequencing.

3.2.1 Treating Cancer through Synthetic Lethality

Synthetic lethality is the genetic interaction of two mutations such that neither mutation alone affects cell viability, but that in combination they result in cell death.[251] This can be exploited in drug discovery to selectively target cancer cells over healthy human cells. The effect of one of the mutations can be mimicked using a drug to result in cell death only when the second mutation is already present in the cell.

The best-known example of this principle is in the development of PARP inhibitors.[252] The poly ADP ribose polymerase (PARP) family of enzymes are predominantly involved in DNA single-strand break repair. By inhibiting them with a drug, the cell's ability to repair DNA single-strand breaks is impaired, which generates an unsurmountable pressure in cancer cells already dysfunctional in DNA damage response pathways.

PALB2, BRCA2 and BRCA1 are all proteins important for forming the Holliday junctions that enable homologous recombination. Some people suffer with genetically inherited mutation of one of these three genes, which impairs their ability to repair double-stranded breaks effectively. This increases the risk that they develop certain cancers. In fact, carriers of a BRCA1 mutation have a greater than 87% chance of developing breast cancer within an average life time and overall approximately 5% of breast cancers can be traced to this abnormality.[253]

It was shown that sub-optimal PARP activity exhibited synthetic lethality with BRCA1 or BRCA2 mutations.[254] When either one of single-strand repair (PARP) or double-strand repair (BRCA1/2) are ineffective cells remain viable, albeit with an increased likelihood of generating cancerous mutations. However, when both DNA pathways are ineffective the cell is under sufficient stress to trigger apoptosis. This inspired a therapeutic strategy whereby cancer patients with defective BRCA containing tumours are treated with PARP inhibitors, which specifically target the cancer cells.

Olaparib (**87**), a PARP1 inhibitor developed by Astra-Zeneca was approved for the treatment of BRCA mutated ovarian cancer. It is currently undergoing phase II trials for the treatment of breast, prostate and colorectal cancers.[255,256] There are several more compounds undergoing phase III trials, such as the intravenously administered rucaparib (**88**) and the orally administered niriparib (**89**), highlighting the potential utility of drugs that harness the effects of synthetic lethality.[257]
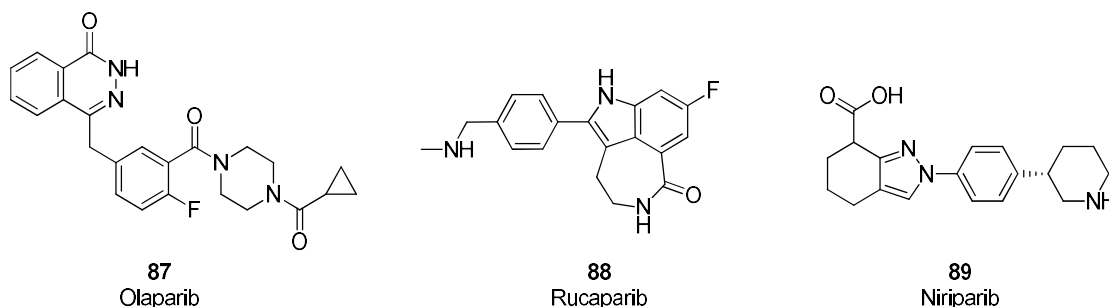
Figure 3.2: Structures of various PARP inhibitors.

3.2.2 Does PALB2 Have Therapeutic Potential?

Considering its involvement in homologous recombination, it is reasonable to assume that inhibition of the PALB2/BRCA2 interaction will impair repair of double-strand DNA breaks by disrupting Holliday junction formation. Considering what is now understood about synthetic lethality, and the proven therapeutic use of PARP inhibitors (section 3.2.1), inhibition of PALB2/BRCA2 may also prove synthetically lethal with other mutations which cause defects in alternative DNA damage repair pathways, such as in cases of defective single strand break (SSB) repair. Currently there are no known tool-compounds with which this hypothesis can be tested. It is therefore important to develop a competitive inhibitor of the PALB2/BRCA2 interaction with which to probe cell viability upon loss of PALB2 function. In this work computational techniques were used to test the potential of small molecules and miniproteins to mimic BRCA2 and bind to the surface of PALB2, before the synthesis of potential inhibitors carried out.

3.2.3 The Role of Bio- and Cheminformatics in Modern Drug Discovery

Computational techniques relevant to drug discovery tend to be split into two fields; bioinformatics and chemoinformatics.[258] The term bioinformatics covers computational methods that concern the structures of proteins and their interaction networks within the cell.[259] Chemoinformatics examines small molecules, predicting their physiochemical properties, as well as mining the vast amounts of data generated in biological assays.[260–

[263] The predictions made by computational calculations can help to guide the drug discovery process, reducing the number of compounds that need to be tested in order to achieve desirable activity against a drug target. This helps reduce the cost of drug discovery projects, particularly as the cost of computing decreases and the advancement of chemical models continues.

3.2.3.1   A Summary of Bioinformatics

The primary use of bioinformatics is in predicting the structure of proteins when there is no crystal structure available. The most accurate way of doing this is by homology modelling.[264] This requires identification of a sequence with high homology (usually >50%) to the query sequence, for which there is an existing crystal structure. The two sequences are aligned and the known structure used as a template to build up the structure of the modelled sequence. The model can then be refined by estimating the position of any loops that do not fit within the original template using molecular dynamics calculations, or by searching for known loops with similar structures that can be employed as a template. Using this technique it is estimated that with the current diversity of crystal structures in the protein data bank it is possible to predict the structure of the majority of proteins with a reasonable level of accuracy.[265]

When there are no crystal structures available with good homology to the query sequence it is possible to use protein threading to predict protein structure.[266,267] This technique relies on the fact that there are a limited number of folds that a protein can adopt, and thus for each fold within the structure of the query sequence there is likely to be a homologous fold in the PDB. Each amino acid is individually placed into a position on a structure in the PDB and energetically analysed to determine how well it fits the template. A best-fit template is selected from which a model of the sequence structure can be built.[267]

It is also theoretically possible to use *ab initio* calculations to predict protein secondary structure and folding based entirely on the protein sequence.[268] This uses huge amounts of computational power, however, and as such, is only possible to low accuracy on short protein sequences.[269]

The ability to process the large volume of data generated about proteins is key in bioinformatics. Large protein interaction network maps have been generated which help

us to understand how the loss of activity of one protein might affect others. A very popular service that will generate protein interaction networks is STRING.[270] This exists as an online database which contains information on 184 million different protein-protein interactions.

Bioinformatics can also make use of the structural properties and interaction networks of proteins in order to assess druggability.[17,271] This is essential for helping identify promising targets for drug discovery programmes. A more detailed analysis on predicting protein druggability was carried out in the thesis introduction (Section 1.2).

3.2.3.2 A Summary of Cheminformatics

Whereas bioinformatics helps us predict information regarding proteins, cheminformatics is the field of computational calculations used to make predictions about small molecules.[260] As information is generated, such as inhibitory activity against proteins, it is increasingly difficult to handle the accumulating data. Open databases such as ChEMBL[272] contain over 1.5 million distinct compounds, with nearly 14 million protein activities recorded. The ability to effectively data mine this information to find the relevant entries is essential, and is a primary focus of cheminformatics. Procedural flow-based programming environments such as Konstanz Information Miner (KNIME)[273] have proven to be of great use when adapted to enable processing of chemical databases.[274] As such, the ability to search chemical structures has become an important part of cheminformatics.

Cheminformatics has also led to the development of software able to predict the chemical properties of molecules based on their structures with reasonable accuracy.[275] Software suites such as MarvinSketch and ACDLabs have calculators that can predict properties such as molecular weight, molecular conformations, logP and pK_a, facilitating both the design and characterisation of drug-like molecules.[276,277]

Use of cheminformatics can also help us predict how well a small molecule ligand can bind to a protein.[278–280] Algorithms have been designed which model the chemical force fields generated by both ligand and protein, such as the Merck molecular force field (MMFF94)[281] and the assisted model building with energy refinement (AMBER)[282] force field. These enable prediction of lipophilic and charged interactions such as hydrogen

bonding between the ligand and the protein. In doing so, it is possible to generate a score indicating how well a particular ligand binds when computationally placed into a pocket on the protein. This allows for the 'virtual screening' of compounds without any need for chemical synthesis.[283] Although this is only a prediction and does not always perfectly reflect the results obtained later from biochemical screening, it enables medicinal chemists to direct the design of potential inhibitors, helping to reduce the number of compounds that must be synthesised in order to achieve high potency. Through the use of pharmacophores, which detail specific interactions and structural features necessary for good binding of lead compounds to proteins, it is possible to screen huge databases of compounds relatively quickly, to identify molecules that may bind in a similar way.

Overall, computational chemistry has proven invaluable to the drug design process. As such, here a series of computational experiments were carried out in order to facilitate the *ab initio* design of BRCA2 mimetics to potentially inhibit the PALB2/BRCA2 interaction.

3.3 Computational Design of Small molecule BRCA2 Mimetics

In order to design small molecule mimetics of BRCA2, a process was developed which used several different software packages to measure and refine structures (Figure 3.3). Firstly, instantJchem (ChemAxon, http://www.chemaxon.com) [284] was used to construct a database of potential mimetics based on existing mimetics in the literature. Conformers were then generated in KNIME[273] using RDKit[285] nodes, before being aligned to the BRCA2 helix from the BRCA2/PALB2 crystal structure (PDB code: 3EU7)[243] and ranked by score. Molecular Operating Environment (MOE, Chemical Computing Group, Montreal, Quebec, Canada, H3A 2R7, https://www.chemcomp.com/)[286,287] was used to check the strain energy (E_Strain) of the conformation, to remove any that were unlikely to form when binding. The top ranking structures could then be docked into PALB2 under structural restraints using MOE[287], to provide further information to help predict their effectiveness as PALB2 binders. The results of this sequence of computational processes was then either fed back into the design of a new database of potential mimetics to repeat the process, or used to finalise the design of structures of candidates for synthesis. Each stage of the process will be subsequently described in greater detail.
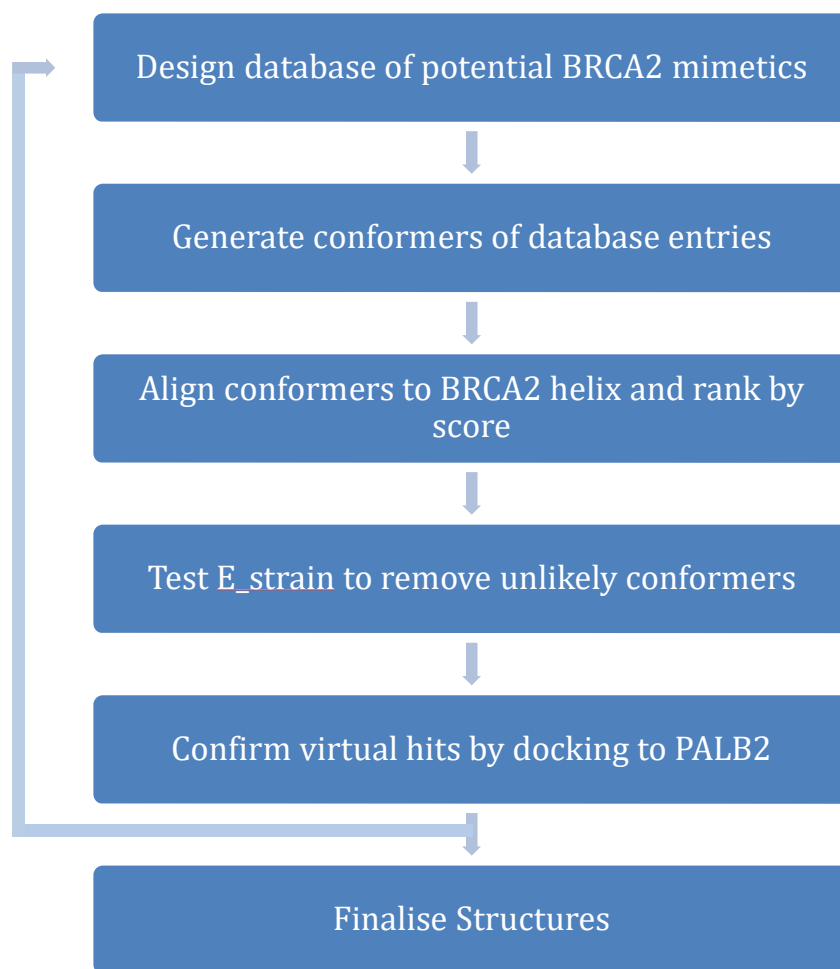
Figure 3.3: The computational design sequence followed for small molecule BRCA2 mimetics.

3.3.1 Designing a Database of Small Molecule BRCA2 Mimetics

As outlined in section 03.2, the BRCA2$_{10-40}$ sequence binds to PALB2 through an α-helix. Structural evidence (PDB code : 3EU7)[243] suggests that the key residues are Trp31, Phe32 and Leu35, which exist in an i, $i+1$, $i+4$ conformation. Examples of molecular scaffolds that were designed to mimic different configurations of amino acids on an α-helix were discussed earlier. These existing examples tend to follow the linear i, $i+3/4$, $i+7$ configuration that facilitates mimicry of PPIs such as p53/Hdm2 and Bcl-2/BH3. There are also examples of a few i, $i+1$ mimetics, such as the indane mimetics developed by Horwell.[153] In order to design a set of i, $i+1$, $i+4$ mimetic scaffolds, functionalities which enable i to $i+1$, and i to $i+4$ spacing between residues were identified, then combined to create hybrid structures.

3.3.1.1 Identifying i, i + 1 small molecule linkers

Few examples of i to $i+1$ scaffolds could be identified from the literature. The indane series of α-helical mimetics contain two modes by which an i to $i+1$ link can be made

(Figure 3.4, **a** and **b**).[148] The decision was also made to incorporate the unconstrained phenyl ring **c** into the design of mimetics as well. Finally, the azabicyclooctane moiety utilised by Spivey and co-workers as part of their 5-residue mimetic scaffold[288] was incorporated into potential *i, i*+1, *i*+4 mimetic designs.
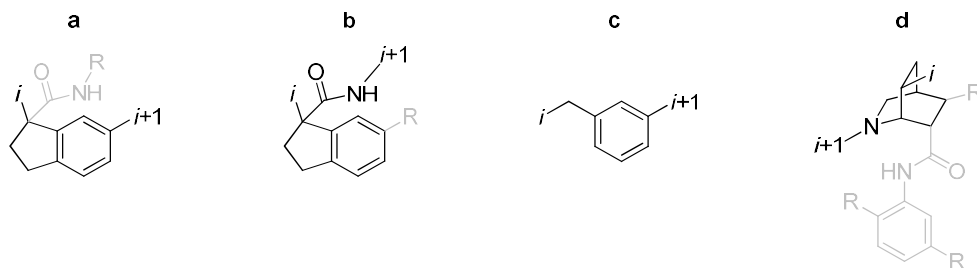
Figure 3.4: Examples of *i, i* + 1 scaffolds from published literature, including the indane series (**a** and **b**),[148] a ring opened indane (**c**), and an azabicyclooctane (**d**).[288]

3.3.1.2   Identifying i, i + 4 small molecule linkers

Many of the α-helical mimetic scaffolds outlined earlier had the potential to be incorporated into the design of the *i* to *i*+4 linker required for the BRCA2 mimetics (Figure 3.5). The *ortho*-substituted phenyl group from Hamilton's terphenyl scaffolds had the potential to be adapted into an *i* to *i*+4 linker (Figure 3.5, **a**). Oligobenzamides developed by Wilson, such as the *N*-functionalised (**b**)[151]   and *ortho*-phenol (**c**)[150] functionalised derivatives could also be incorporated into the design. Other scaffolds that were incorporated into *i, i*+1, *i*+4 mimetics in the database are benzoylurea (**d**),[156] 5-6-5 imidazole-phenyl-thiazole (**e** and **f**),[152] terephthalamide (**g**),[155] and the azabicyclooctane based scaffold (**h**).[288]
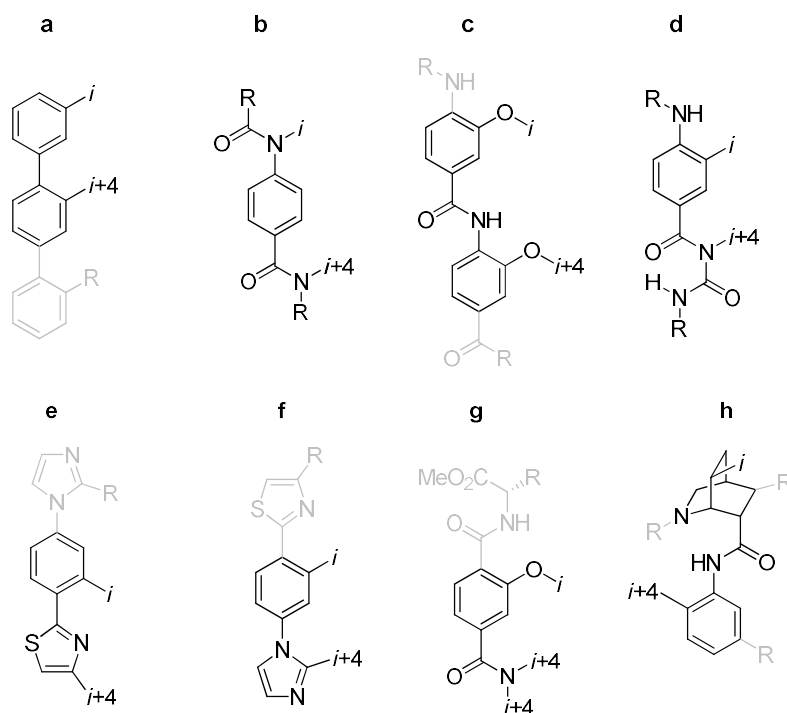
Figure 3.5: Examples of *i*, *i* + 4 scaffolds from published literature, including the terphenyl scaffold (**a**),[149] *N*-functionalised oligobenzamide (**b**),[151] ortho-functionalised oligobenzamide (**c**)[150], benzoylurea (**d**),[156] 5-6-5 imidazole-phenyl-thiazole (**e** and **f**)[152], terephthalamide (**g**),[155] and the azabicyclooctane based scaffold (**h**).[288]

3.3.1.3 Design of Databases Using Known Scaffold Moieties and Markush Enumeration

With the identification of reported *i* to *i*+1 and *i* to *i*+4 linkers, hybrid structures were designed in order to test computationally as BRCA2 mimetics. Initially, databases of structures were curated manually in instantJchem,[284] using structure editors to input each design, resulting in structures such as those shown in Figure 3.6.
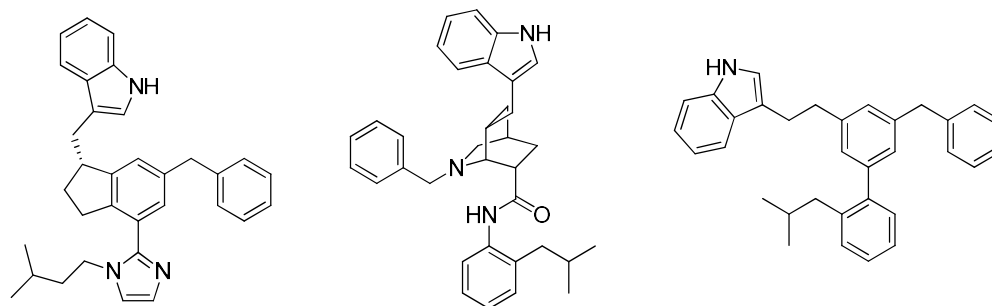
Figure 3.6: Examples of hybrid *i*, *i*+1, *i*+4 structures designed.

After alignment studies on the early sets of designed structures had been performed and the best BRCA2 mimetics established from these results, the design of sets of structures with smaller, iterative structural modifications was carried out. By doing so, it was hoped that the best scoring structures thus far could be optimised by introducing heteroatoms, as well as with modifications to the aliphatic residue linker lengths. During this process,

Markush enumeration software packages were used to generate databases of potential compounds.
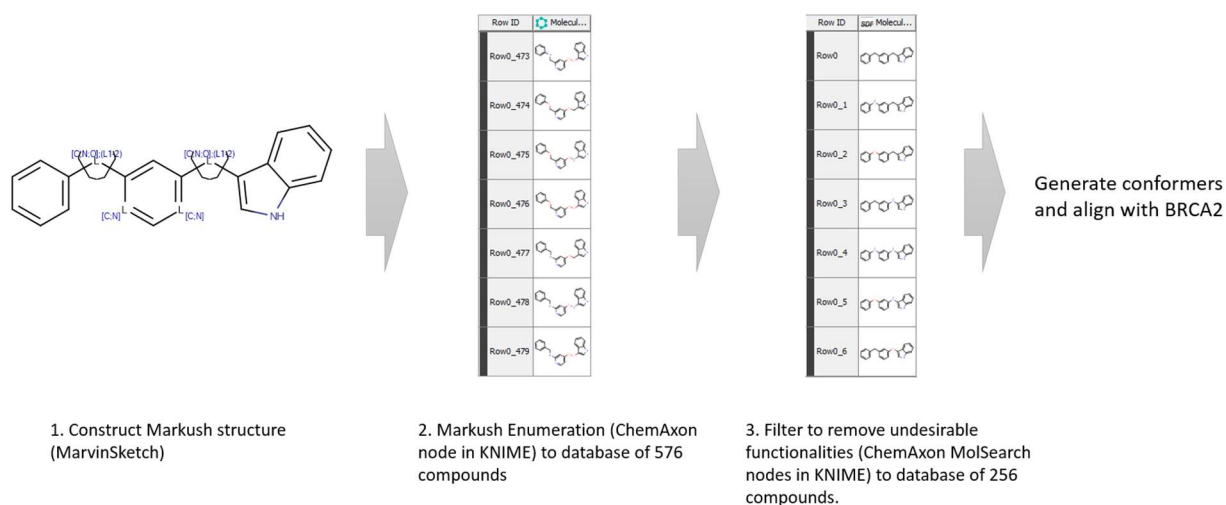
Figure 3.7: Use of ChemAxon nodes in KNIME to generate databases of designed BRCA2 mimetics from Markush structures.

Markush structures are best known as the diagrams used within patents in order to describe a set of similar compounds. Variation between a large number of compounds in the set is represented using a small number of structural diagrams. Variables such as different chain lengths, different ring sizes, use of different heteroatoms at particular positions, and use of different functional groups are often described. Using ChemAxon[284]MarvinBeans software package, it is possible to construct a Markush structure, and then have the software enumerate the structure into a database of every possible compound that is described by the Markush variables. This process was completed in KNIME[273] using ChemAxon nodes. By doing so the enumerated databases could be filtered to remove any structures with undesirable functionalities such as peroxide bonds, and concatenated with the output of other Markush enumerations to create a large database for analysis by alignment studies (Figure 3.7).

3.3.2 Generating Conformers of BRCA2 Mimetics and Aligning Them to The BRCA2 Helix Using RDKit

Once a set of potential BRCA2 mimetics had been designed, they were computationally analysed to determine how closely they mimicked the hotspot residues in BRCA2. This process was carried out in two steps. Firstly, each designed BRCA2 mimicking structure was used to generate a large number of rotational conformers to simulate how the shapes

they may potentially take in solution. These were then aligned with the hotspot residues of BRCA2.

RDKit[285] has been shown to be one of the most efficient and accurate freely available conformer generation tools.[289] It also features KNIME[273] integration, with nodes freely available for a number of functions on the community repository. It was therefore chosen as the conformer generation tool for this study. As the computational power required for conformer generation was relatively low, the RDKit tool was configured to generate a maximum of 5,000 non-duplicate conformers per structure. Although this number is higher than what was later found necessary to obtain good results for the alignment study, it still allowed completion of the analysis of a set of designed structures overnight.

Full α-helix (Source: 3EU7)   Reduced α-helix containing hot-spot residues   Aligned small molecule BRCA2 mimetic
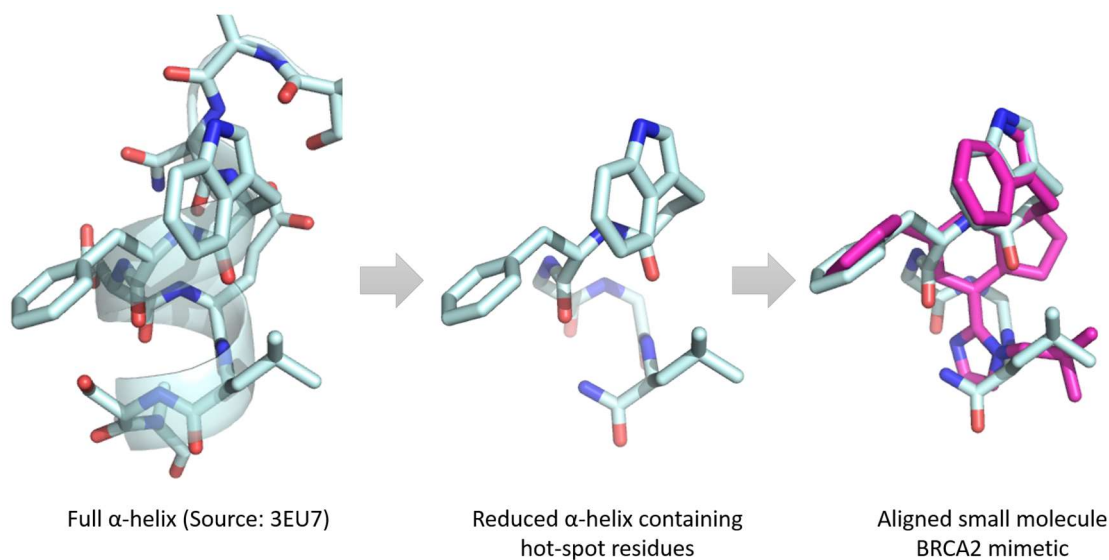
Figure 3.8: The process of reducing the BRCA2 helix to the key residues in PyMOL[290]before using RDKit[285] Open 3D Alignment to align with BRCA2 mimetic conformers.

PyMOL[290] was used to prepare the BRCA2 helix (PDB source: 3EU7)[243] by removing all atoms except the peptide backbone and the Trp31, Phe32 and Leu35 residues. This modified structure was then used as the template for the RDKit[285] Open 3D Alignment tool, which also has integration as a freely available node in KNIME.[273] This generated a score for the alignment of the mimetic with BRCA2. By using a KNIME workflow for both processes, they could be carried out in sequence, and repeated for each newly designed set of BRCA2 mimetics easily.

3.3.3 Using MOE to Remove Conformers with High E_Strain

Once the RDKit[285] nodes in KNIME[273] had been utilised to identify the scaffolds which best aligned with BRCA2, MOE[287] was used to test the conformational strain of the different conformers. Although the RDKit conformer generation tool generally generates low energy conformers, it was found that some of the conformers exhibited significant strain, and were deemed unlikely to form under binding conditions. The E_Strain tool in MOE calculates the strain that a molecular conformation is under within the chosen forcefield environment, as a measure of the difference in energy between a chosen conformation and its local rotational minima. The values are calculated in $kCal.mol^{-1}$. AMBER10[282] and MMFF94[281] are two of the most commonly used forcefield environments for electrostatic calculations, and both can operate within MOE. Of these, MMFF94 is generally considered better for small molecules, so was used in this case. The calculated E_Strain for the BRCA2 mimetic conformers typically ranged from 7 to 17 $kCal.mol^{-1}$. Of the best aligning mimetics, conformers with an E_Strain higher than 12 $kCal.mol^{-1}$ were removed, roughly halving the number of structures.

3.3.4 Using MOE to Dock the Best Conformers into PALB2 to Reinforce Predicted Activity

Once the best aligning conformers had been identified, and those under high strain removed, docking studies in MOE were carried out in an attempt to reinforce the prediction that the structures would bind to the PALB2-BRCA2 interface. The docking was carried out using the crystal structure of BRCA2 bound to PALB2 (PDB code: 3EU7).[243]

The docking studies were carried out to help understand how well the different structures were able to behave as BRCA2 mimetics. Because of this, restrictions were applied to the docking calculations, to encourage them to bind with the relevant functionalities in the Trp31, Phe32 and Leu35 positions. This was done by generating pharmacophores using the BRCA2 helix that requires indole, phenyl and isopropyl groups to be near the positions they would be in BRCA2. However, this approach still proved to be problematic as the software identified interactions with the core scaffold section of the structures, as well as the *i*, *i*+1 and *i*+4 functionalities. The most common binding mode produced was for the tryptophan-mimicking indole to be rotated by 180°, causing the aliphatic linker in

the 3- position to pick up lipophilic interactions with the rear of the protein pocket (Figure 3.9).
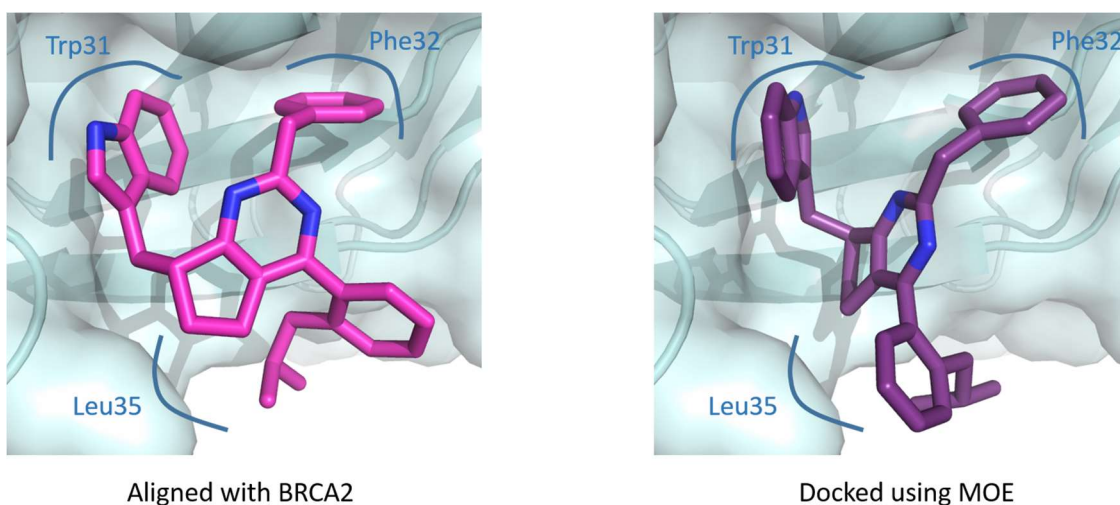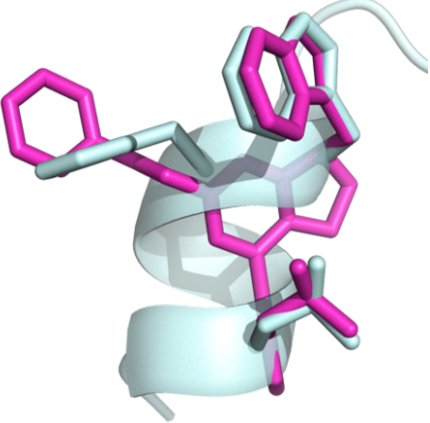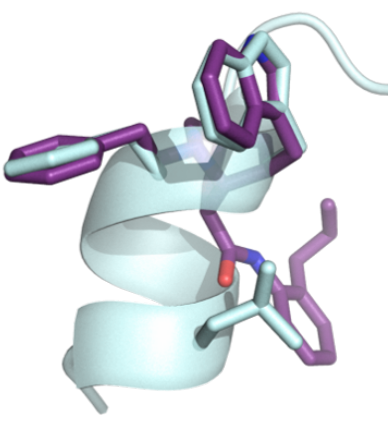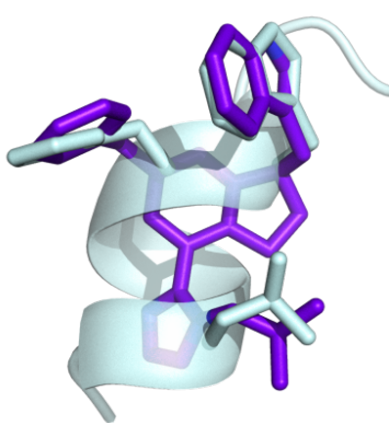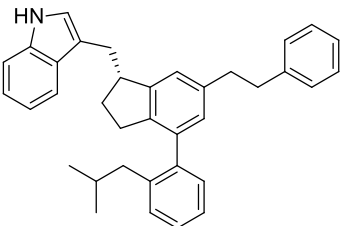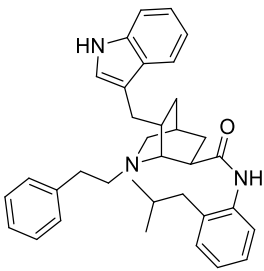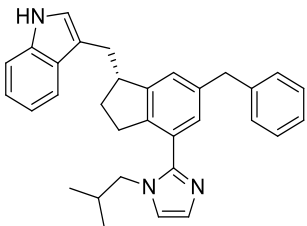
Figure 3.9: Comparison of results between a mimetic aligned with BRCA2, and docked into PALB2 using MOE. The binding sites of the BRCA hotspot residues have been labelled.

By increasing the binding restrictions by optimising the number of pharmacophores, it was eventually possible to dock the mimetics in the 'correct' configuration. However, because of the amount of restrictions applied when docking, the results became less and less accurate as a prediction. Consequently, little weighting was given to the docking scores when choosing candidates for synthesis.

3.3.5   Results of Alignment and Docking Studies
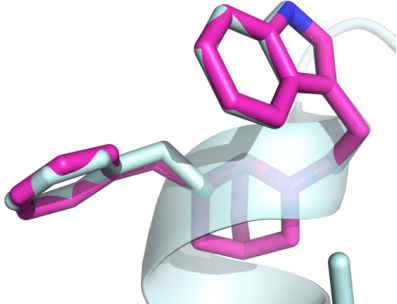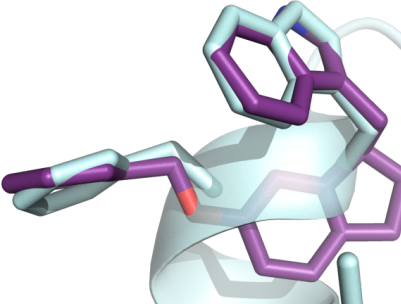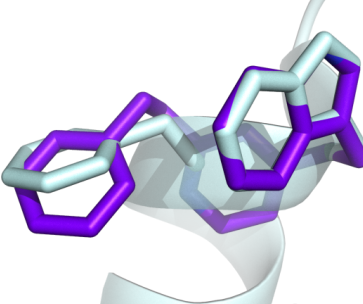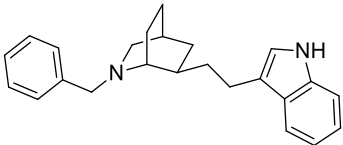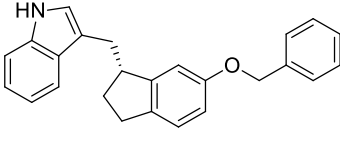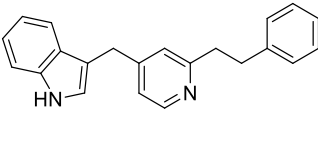
From the results of the alignment and docking studies, several different structures were found to have potential as *i*, *i*+1, *i*+4 BRCA2 mimetics (Table 3.1). The length of the *i*+1 linker to the phenyl group was found to be well tolerated with either one or two carbons, as each allowed alignment at different low energy conformations. The *i*+4 position was also found to not align as well with the BRCA2 helix in general. In a few cases (e.g. Table 3.1, example 1.1), good alignment was achieved in the *i*+4 position, although nearly always with reduced alignment in the *i*+1 position. These examples tended to have lower alignment scores as well. The azabicyclooctane containing structures (e.g. Table 3.1, example 1.2) in general had the best *i* to *i*+1 alignment, but in no cases were able to align with all three residues, and often had high calculated E_Strain. In some cases (e.g. Table 3.1, example 1.3), good overlap was achieved in each of the *i*, *i*+1 and *i*+4 positions, however this always occurred in conformations with E_Strain over the desired threshold.

Table 3.1: Examples of high scoring aligned *i*, *i*+1, *i*+4 structures.

Example 1.1	Example 1.2	Example 1.3

Alignment Score: 138.5	Alignment Score: 143.4	Alignment Score: 147.1
E_Strain: 8.7 kCal.mol^{-1}	E_Strain: 14.5 kCal.mol^{-1}	E_Strain: 14.4 kCal.mol^{-1}

Given the high E_Strain measured in the best aligning *i*, *i*+1, *i*+4 mimetics, and the tendency for the *i*+4 position to align poorly with BRCA2, *i*, *i*+1 mimetics were investigated as an alternative. In general, these were found to align well with the BRCA2 helix. As before, the azabicyclooctane (Table 3.2, example 2.1) containing structures aligned best, with low E_Strain values calculated. Indane and phenyl-based *i*, *i*+1 also showed good alignment with BRCA2, giving alignment scores comparable to the *i*, *i*+1, *i*+4 mimetics, despite the lower atom count. More importantly, the best aligning structures also were calculated to have E_Strain values nearly always under the chosen threshold of 12 (Table 3.2, examples 2.2 and 2.3). This put them into consideration as an alternative to the full *i*, *i*+1, *i*+4 mimetics when selecting candidates for synthesis.

Table 3.2: Examples of high scoring aligned *i*, *i*+1 structures.

Example 2.1	Example 2.2	Example 2.3

Alignment Score: 153.0	Alignment Score: 139.1	Alignment Score: 134.0
E_Strain: 9.0 kCal.mol^{-1}	E_Strain: 11.3 kCal.mol^{-1}	E_Strain: 10.8 kCal.mol^{-1}

The potential of *i*, *i*+4 mimetics to align with BRCA2 was also tested. Unsurprisingly given the observations of *i*, *i*+1, *i*+4 alignment, these were found to produce much lower scores from the RDKit Open 3D Align algorithm than other structures tested (Table 3.3, example 3.1 and 3.2) and not pursued further.

In all cases, it was found that the presence of heteroatoms in the core of the molecule, as well as hydrocarbon linkers was well tolerated. This could be of great use when designing candidates for synthesis, as it may improve the inherent aqueous solubility of an all-hydrocarbon mimetic.

Table 3.3: Examples of high scoring aligned *i*, *i+4* structures.

Example 3.1	Example 3.2

Alignment Score: 111.2	Alignment Score: 108.8
E_Strain: 12.0 kCal.mol^{-1}	E_Strain: 8.8 kCal.mol^{-1}
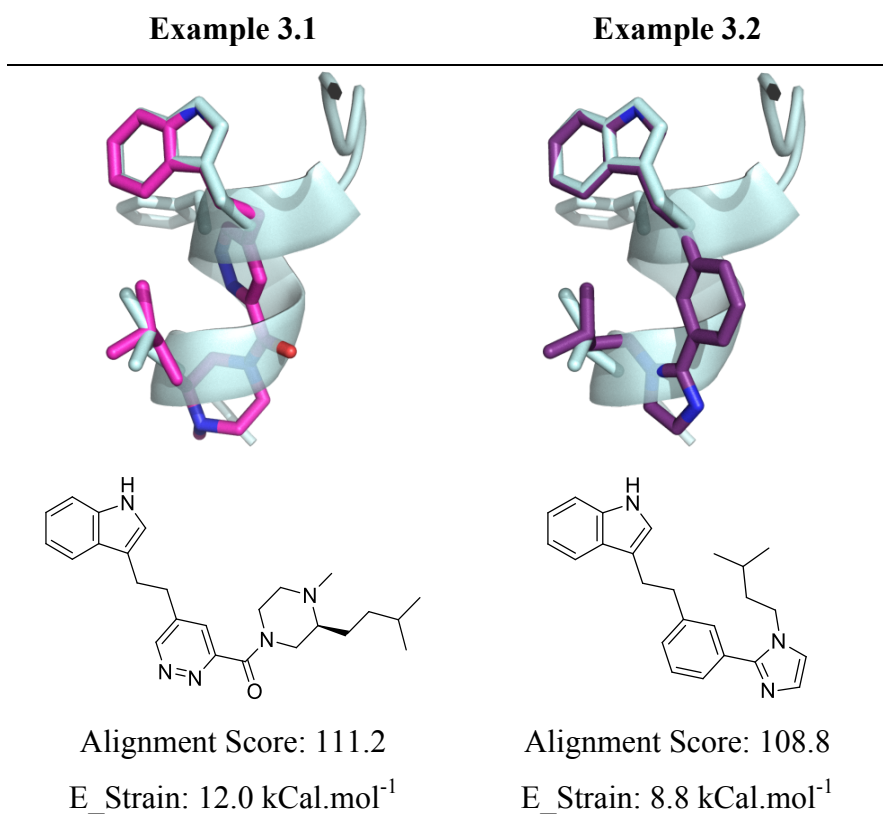
3.3.6 Searching For Alternative Routes to the *i+4* Position Using Docking Studies

Given that the *i+4* leucine position seemed to be the most difficult to align to with a small molecule mimetic, other potential options to reach this position were explored. From the PALB2/BRCA2 crystal structure, it seemed that the Trp31 indole 4-position points towards the Leu35 position. To explore this, indole analogues functionalised with aliphatic moieties were docked into the PALB2 structure.
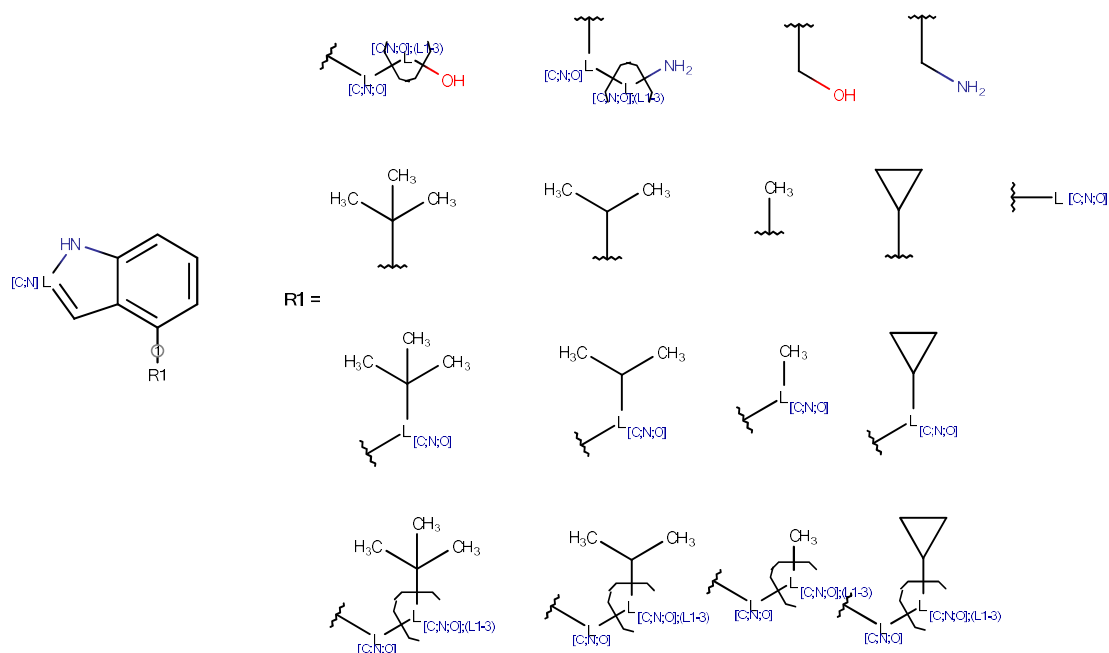
Figure 3.10: A Markush structure indicating the indoles docked into PALB2.

A database containing a series of different structures was created by using ChemAxon Markush enumeration (Figure 3.10), using filters as described before (section 3.3.1.3) to remove unwanted structures. A docking experiment was then set up in MOE, where pharmacophores were generated to ensure that the indole was locked into position in the Trp31 pocket. It was found by trial and error analysis that a total of 11 pharmacophores describing the indole atom placement was enough to lock the docked compounds in place. Once the methodology had been established, the functionalised indoles were docked into PALB2. The results demonstrated that the 4- position can indeed give access to the Leu35 pocket (Figure 3.11). A chain length of 3-5 atoms with a terminal lipophilic group seemed to be best tolerated for reaching the pocket. In some cases, a terminal amino or alcohol group was able to pick up hydrogen bonds on the far side of the Leu35 pocket. This seemed to support the hypothesis that indole functionalisation could prove as an alternative to mimicking the *i*+4 position, subject to synthetic accessibility.
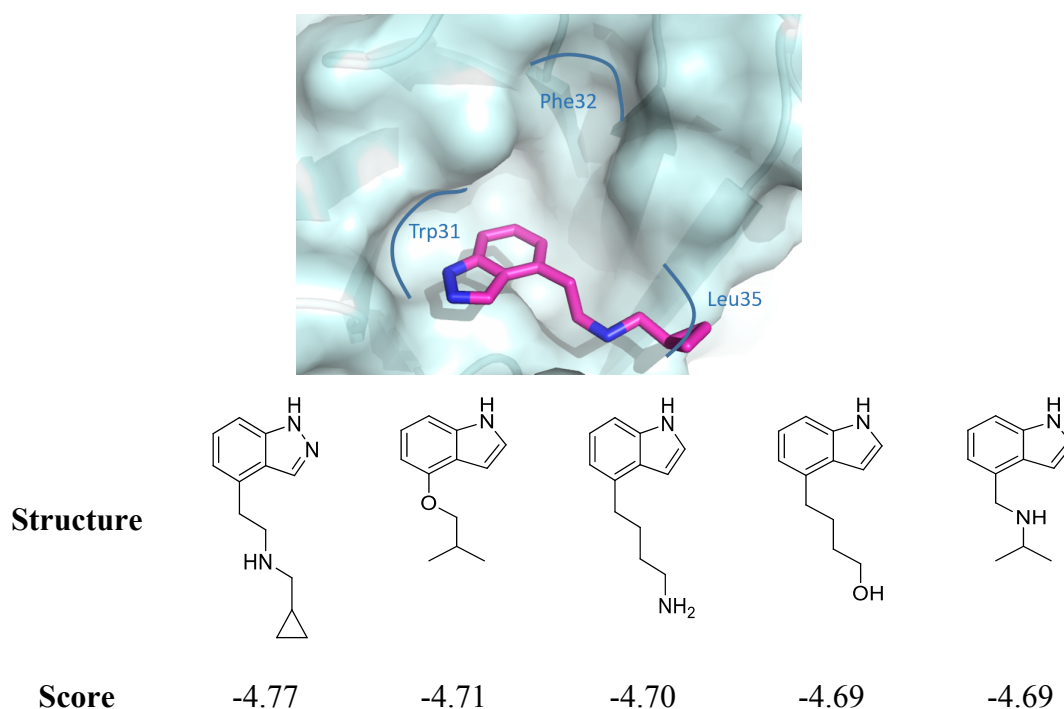
Figure 3.11: The top hits from the docking study in MOE showed that aliphatic functionalisation in the indole 4-position can comfortably occupy the position of Leu35 on BRCA2.

3.3.7 Choosing Candidates for Synthesis

Given the results from the computational work, the subjectively best results were incorporated into Markush structures (Figure 3.12) as potential candidates for synthesis. Given the tolerance of different scaffold-to-residue linkers and potentially solubilising heteroatoms in the alignment and docking studies, they were included in the designs. The phenyl group and N-substituted imidazole were by far the most effective *i*+4 linkers, so were chosen as candidates. As the *i*, *i*+1 mimetics seemed very promising in the alignment and docking studies, they were also chosen as candidates for synthesis, however, it was decided not to pursue the *i*, *i*+4 mimetics further, due to their lower alignment scores, and poor overlap with BRCA2.
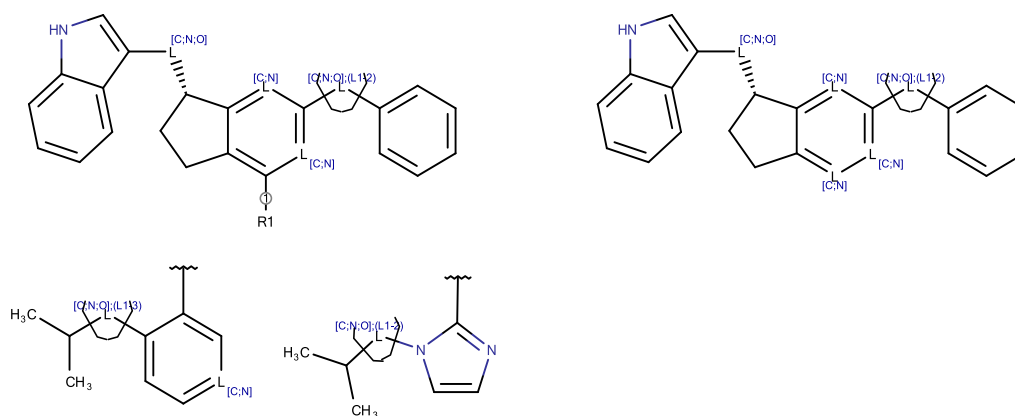
Figure 3.12: Markush structure showing the structures to be analysed for synthetic accessibility.

3.4 Design of Miniproteins to Mimic BRCA2

As an alternative approach to target the BRCA2/PALB2 PPI, a series of miniproteins were developed to mimic the BRCA2 helix. To achieve this, protein sequences were computationally analysed to select the best candidates to act as mimetics (Figure 3.13). An in depth review on the development of miniproteins was discussed in the thesis introduction (section 1.3.2).

Search for structurally stable miniproteins on the PDB

↓

Analyse each structure for potential α-helical *i*, *i*+1, *i*+4 mutation sites

↓

Make point mutations to incorporate WFXXL residues in each sequence

↓

Align and create homology models of the sequences based on the parent miniprotein structure

↓

Check for steric clashes when 'docked' in place on PALB2 and eliminate unsuitable structures

↓

Choose best miniproteins for synthesis

Figure 3.13: The design sequence followed for the development of BRCA2 miniprotein mimetics.

3.4.1 Searching for Structurally Stable Miniproteins with Known Structures

In order to design miniproteins, sequences must be found with structural information available. In order to achieve this, the PDB advance search function was used to find structures that matched chosen criteria. The criteria selected for were for a disulfide bond count > 1, a chain length < 31, the number of α-helices > 0, and the number of chains = 1. This returned a total of 87 different protein structures. To these was added the aPP sequence used by Schepartz as the parent structure for the design of many different miniproteins.[70,75] APP did not match the search criteria due to its lack of disulfide bonds, but had been proven to be a good template for miniprotein design in several examples. This gave a total of 94 different structures that were analysed for their suitability to adopt

the BRCA2 WFXXL motif. The PDB search contained all other structures used for the design of miniproteins that were discussed in section 1.3.2.

3.4.2 Analysing Structures for Suitability to Adopt the BRCA2 WFXXL Motif, and Designing Mutant Structures

Once the 94 different structures had been obtained, they were individually subjected to analysis in PyMOL[290] by manually visualising each structure to determine whether or not they could adopt the WFXXL motif. For this to be allowed, possible mutations had to be made in the solvent exposed face of an α-helix in an i, $i+1$, $i+4$ configuration. None of these mutations could be at the position of a disulfide-forming cysteine residue, in order to have as limited effect on the protein secondary structure as possible. In each case when it was possible to mutate in an i, $i+1$, $i+4$ configuration, the new structure was recorded. This resulted in a total of 53 different sequences being designed, using a total of 43 different miniprotein templates obtained from the PDB.

3.4.3 Constructing Homology Models of Designed Miniprotein Sequences
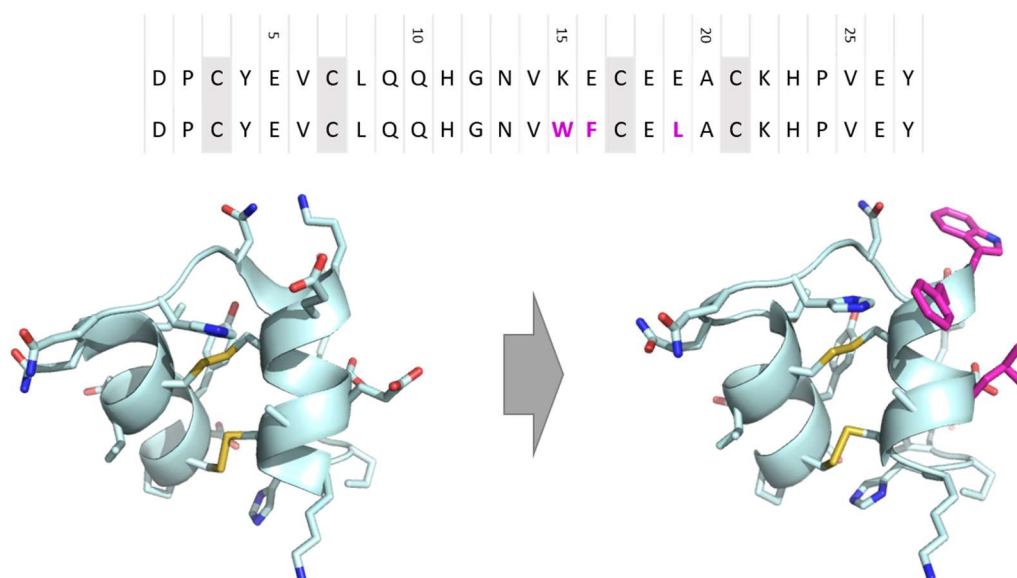
Figure 3.14: Creating a BRCA2 miniprotein mimetic from Om-toxin (Pdb Code: 1WQD)[291] using homology modelling.

Once the 53 different miniprotein sequences containing the WFXXL motif had been designed, homology modelling was used to construct the predicted 3D structures. This was carried out using homology modelling tools in MOE (Figure 3.14).[287] Firstly the

modified sequences were imported into MOE and aligned to the sequence from the structure of the parent miniprotein. The homology modelling tool was then used under Amber10,[282] with a maximum of 10 iterations. The best model of each was then exported for further analysis.

3.4.4 Superposition of Miniproteins with BRCA2 to Test for Steric Incompatibilities

Once the 53 homology models had been constructed, they were tested for steric incompatibilities with PALB2. For this to be done, firstly the WFXXL motif in each homology model was aligned with the equivalent residues in the BRCA2 helix. Each miniprotein was then superposed to BRCA2, locked to the position of the Trp31, Phe32 and Leu35 residues.
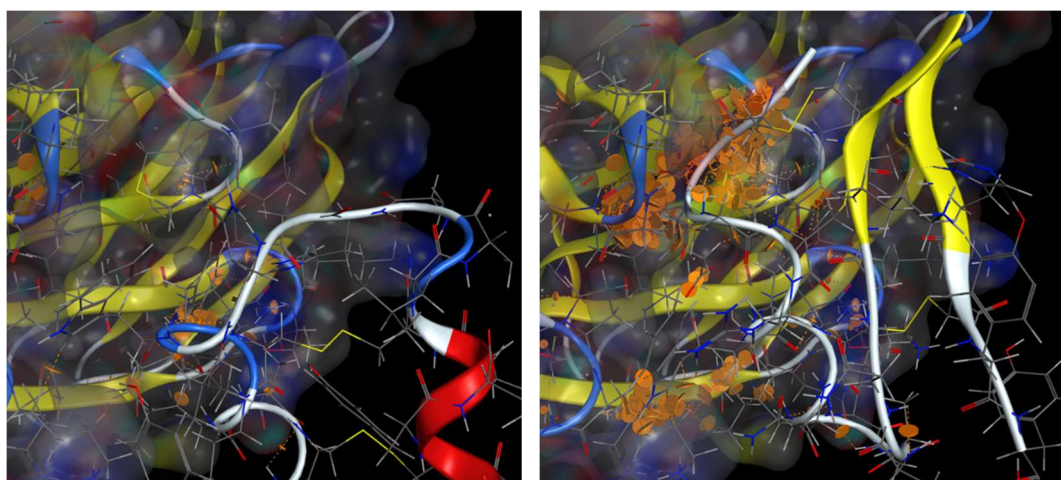
Figure 3.15: Left: Superposed WFXXL miniprotein based on 1WQD[291], with minimal steric clashing. Right: Superposed WFXXL miniprotein based on 2RTZ, showing significant steric clashing denoted by orange disks.

The 'Contacts' tool in MOE was then used to search for residues that were involved in steric clashes with PALB2 (Figure 3.15). Areas of significant steric clash were indicated by the clustering of orange 'steric-clash' disks. If a miniprotein was involved in a large amount of steric clashing when it was superposed to the BRCA2 position, then it was discarded as an option for synthesis. Using this technique, 22 of the 53 structures tested appeared to have little to no steric clashing with PALB2.

3.4.5 Selecting 10 Sequences for Synthesis

The set of 22 sequences that were able to fit in the PALB2 pocket without a significant steric clash were filtered down to a set of 10 sequences to be synthesised. Firstly, any sequence where a mutation had been made to a proline residue for part of the WFXXL motif was taken out of consideration as loss of a proline is likely to affect the way that

the protein folded. By doing this, the number of sequences was reduced from 22 to 15. Of these 15 peptides, 10 were selected based on the diversity of the different structures. For example, in any case where more than one miniprotein had been designed from the same parent sequence, only one was selected. The result was a set of 10 distinct miniproteins, all bearing the WFXXL motif.
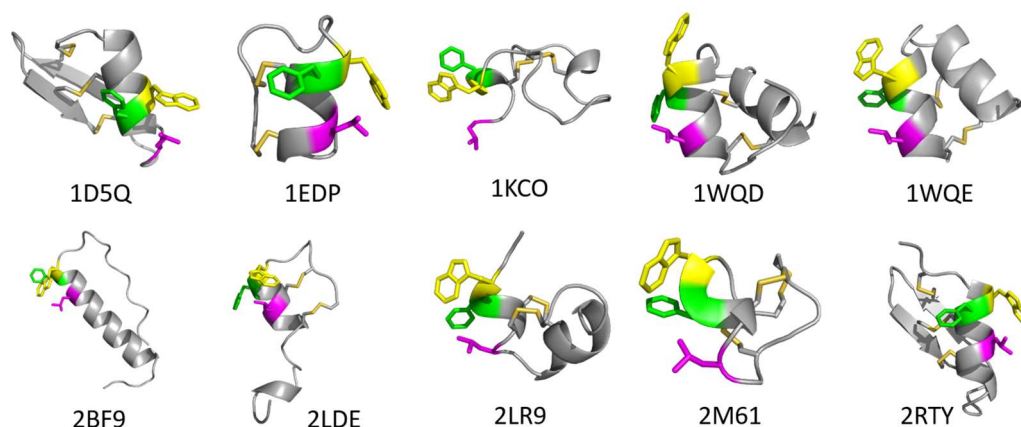
Figure 3.16: Structures of the 10 miniproteins selected for synthesis.

Of the miniproteins selected, the amino acid count was between 14 and 35. Of these, two were based on parent structures had had previously been used for the design of miniproteins. Avian pancreatic peptide (aPP, PDB code: 2BF9) has been used by Schepartz and co-workers to design miniproteins for a series of different targets.[75] Om-toxin3 (Pdb code: 1WQE) has been used by Brunsveld and co-workers to develop binders of oestrogen receptors.[88,89]

Table 3.4: The sequences of the 10 miniproteins selected for synthesis

PDB Code	Miniprotein Sequence	Amino acid count	Disulfide count
1D5Q	CNLWFCQLSCKSLGLKGGCQGSFCTCG	27	3
1EDP	CSCSSLMDWFCVLFCHL	17	2
1KCO	VQCPHFCYELDYELCPWFCYLX	21	2
1WQD	DPCYEVCLQQHGNVWFCELACKHPVEY	27	2
2LDE	CSCTDMSWFECLNFCHKDVIWVNRN	25	2
1WQE	NDPCEEVCIQHTGDVWFCELACQ	23	2
2BF9	GPSQPTYPGDDAPWFDLLRFYNDLQQYLNVVTRHRX	35	0
2LR9	FNWFCCLIPACRRNHKKFCX	19	2
2M61	CCRWFCGLGCHPCCX	14	3
2RTY	FSCDHWFCALRCLAQRRKGGKCKNGDCVCR	30	3

The other eight sequences are based on unique templates. Most of the examples contain two disulfide bonds, with three different examples of structures stabilised by three disulfide bonds, and one (aPP) with no disulfide bonds at all. The sequences were bought from PeptideSynthetics (Peptide Protein Research Ltd.) ready to be tested.

3.4.6 Testing the Miniprotein Sequences for Helicity

Once the peptide sequences had been bought, the helicity was assessed by circular dichroism to ensure that they had folded correctly before their affinity for PALB2 was tested in a biochemical assay. For the aPP based miniprotein which contained no disulfide bonds, this could be done using the purified sequence dissolved in water. The other nine proteins were submitted to oxidative folding conditions to ensure that they had formed disulfide bonds correctly, before they could be tested under circular dichroism.

The CD spectrum of the aPP (2BF9) based miniprotein showed a characteristically α-helical profile, with a peak at 192 nm, and two troughs at 209 and 222 nm respectively (Figure 3.17). The K2D3 online CD prediction service[292] predicted the miniprotein to have 95% helicity based on the recorded spectrum. This suggests that the miniprotein was folding in solution as hoped, and thus could act as a potential α-helical mimetic of BRCA2.
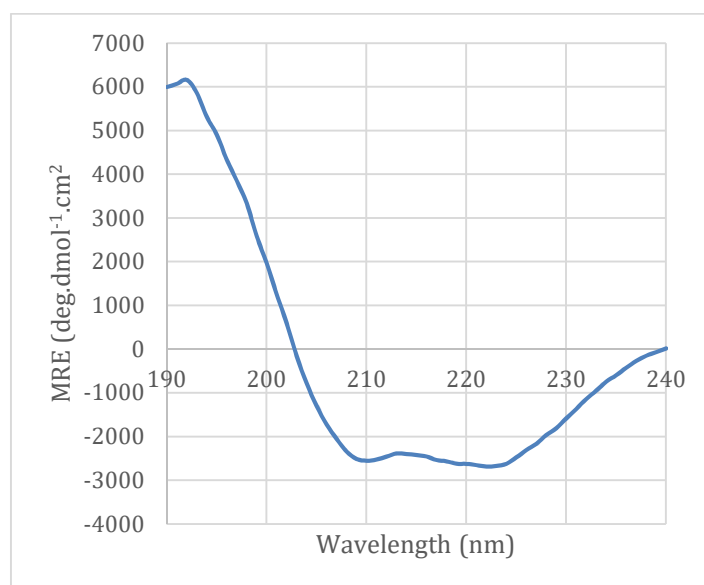
Figure 3.17: CD Spectrum of the aPP based miniprotein (2BF9), showing a characteristically α-helical profile.

The 9 miniproteins with disulfide bridges were then treated with tris(2-carboxyethyl)phosphine (TCEP) in a 2:1 mixture of phosphate buffer (pH 8.0) and trifluoroethanol to reduce any pre-existing disulfide bonds. Dimethylsulfoxide (DMSO)

was then added to quench the remaining TCEP, and the proteins were stirred in air for 48 hours to allow disulfide reformation. Unfortunately, the folded proteins were not purified due to time constraints, and thus the resulting CD spectra not reported in this thesis.

3.5 Conclusion

In conclusion, computational chemistry has been used for the design of both small molecule mimetics of BRCA2, and miniprotein mimetics containing the WFXXL binding motif. The small molecule mimetics designs were primarily scored by their ability to align with the BRCA2 helix, using RDKit nodes in KNIME to carry out many calculations in sequence. Docking studies were also used to evaluate the structures, but with less effect. This resulted in an understanding of which small molecules could be synthesised in order to obtain efficacy towards PALB2, allowing the progression of synthetic chemistry towards achieving this goal.

Miniproteins were also designed computationally, and analysed by use of homology modelling in MOE. This allowed a set of 10 miniproteins to be chosen from 94 initial candidates taken from the PDB. The 10 selected miniproteins were bought, and circular dichroism spectroscopy of one miniprotein showed that significant helicity was forming in solution. This suggests that the hot-spot residues are held at the correct configuration to have efficacy towards PALB2 in a biochemical assay.

4   The Synthesis of Small Molecule BRCA2 Mimetics

4.1   Chapter Summary

This chapter outlines the synthetic steps taken towards the synthesis of BRCA2 small molecule α-helical mimetics. The structures identified as potential mimetics in computational work outlined in chapter 3 are retrosynthetically analysed. Due to the synthetic complexity of the *i*, *i*+1, *i*+4 mimetics, the synthesis of *i*, *i*+1 mimetics is carried out. Initial synthesis towards a pyridazine containing scaffold design is unsuccessful due to difficulties experienced in selective pyridazine ring formation. Synthesis of all-hydrocarbon indane-based *i*, *i*+1 mimetics is more successful. The late-stage functionalisation of the *i* position is carried out using Wittig chemistry, and the *i*, *i*+1 position with the Sonogashira reaction and Ullman condensation. Overall, one *i*, *i*+1 mimetic is obtained, and two further mimetics are within two synthetic steps of a complete synthesis.

4.2   Retrosynthetic Analysis of Target Cores from Computational Results

Computational studies outlined in section 3.3 investigated the ability for different small molecule scaffolds to align with the BRCA2 hotspot residues that interact with PALB2. Small molecules that align well may be able to act as competitive ligands, inhibiting the PALB2/BRCA2 interaction. In this chapter, retrosynthetic analysis of the optimal structures from the computational study will be carried out. The progress towards synthesis of the most structurally accessible of these structures will then be described.

Before synthesis was attempted, retrosynthetic analysis was carried out on the results of the computational studies. This was performed with the intention of selecting cores which could be accessed synthetically in the fewest number of steps, with easy diversification of the *i*, *i*+1 or *i*+4 positions. Compounds containing stereogenic centres were initially synthesised in racemic form with the ultimate aim of separating *via* chiral HPLC methods.
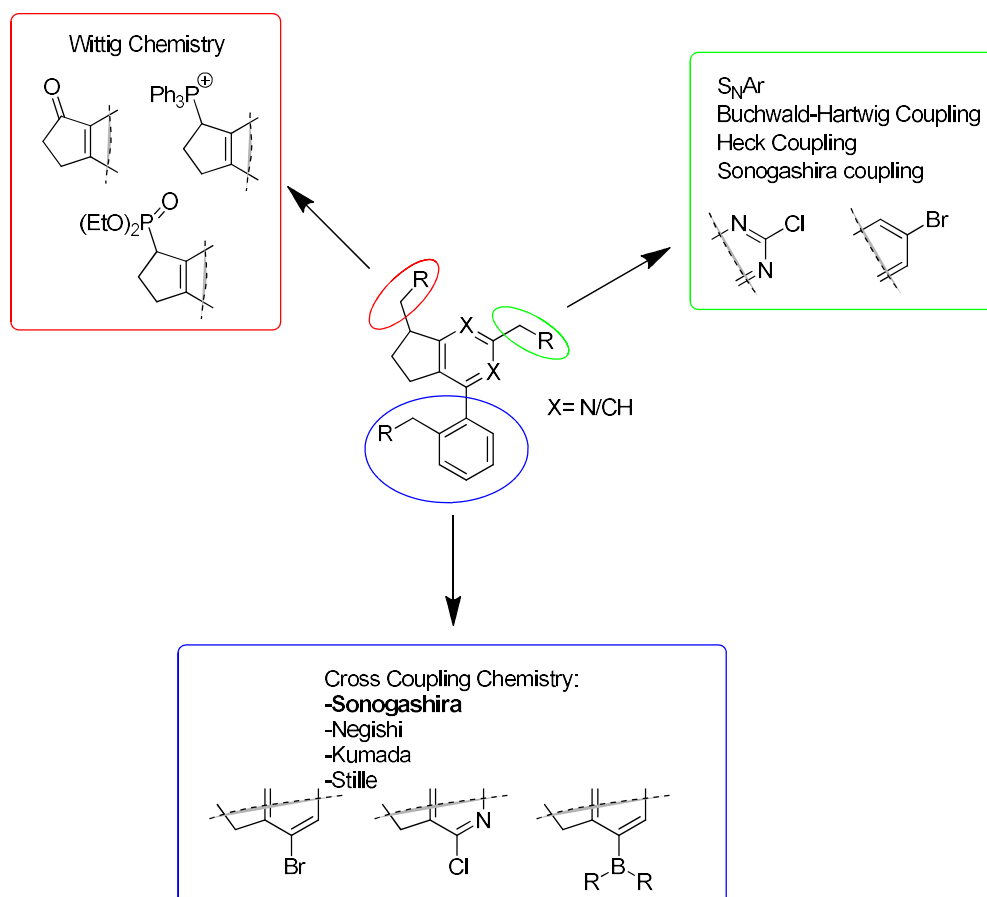
Figure 4.1: Potential methodology for functionalisation of an indane based scaffold in the *i*, *i*+1 and *i*+4 positions.

4.2.1 Retrosynthetic Analysis to Functionalise the *i*, *i*+1 and *i*+4 Positions

The results of computational alignment studies between small molecule mimetics and BRCA2 (section 3.3) suggested that by functionalising an indane in the 1, 4 and 6 positions, would give optimal overlay with the BRCA2 *i*, *i*+1 and *i*+4 hotspot positions. This structure was analysed to identify the best chemical methodology that could be used in order to selectively functionalise at these positions (Figure 4.1)

Functionalisation of the indane 1-position had previously been carried out by Horwell and co-workers using Wittig chemistry, followed by hydrogenation of the subsequent double bond.[153] This methodology was considered for the *i*, *i*+1, *i*+4 mimetics (Scheme 4.1).
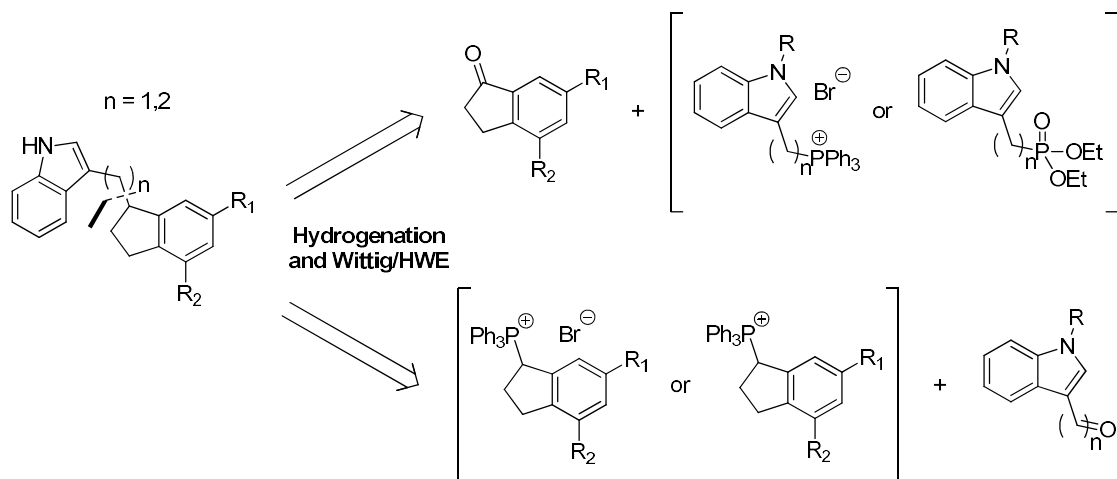
Scheme 4.1: Retrosynthetic cutting of the indole linking bond in the *i* position of the scaffold.

In the synthetic route described by Horwell,[153] the *i* position was functionalised with a phenyl group through Wittig chemistry,[293] to mimic phenylalanine. Application to indoles presented a challenge. The high acidity of the N-H bond, (pKa calculated using MarvinSketch[284] to be as low as 13.5 if the indole was electron deficient) meant it would deprotonate under the basic conditions of a Wittig[294,295] or Horner-Wadsworth-Emmons (HWE) reaction,[296,297] possibly affecting the result of the reaction. Because of this, protecting group chemistry would likely have to be employed. The stability of the phosphonium reagents, as well as the organohalide intermediates towards their synthesis, was also likely to be an issue. In the case where the chain length between the indole ring and the phosphorus is a single carbon atom, there is potential for elimination. This led to the possibility of switching the functional groups in the Wittig reaction, resulting in an indane phosphonium salt, and a carbaldehyde functionalised indole. By doing this, the indole reagent is likely to have increased stability.
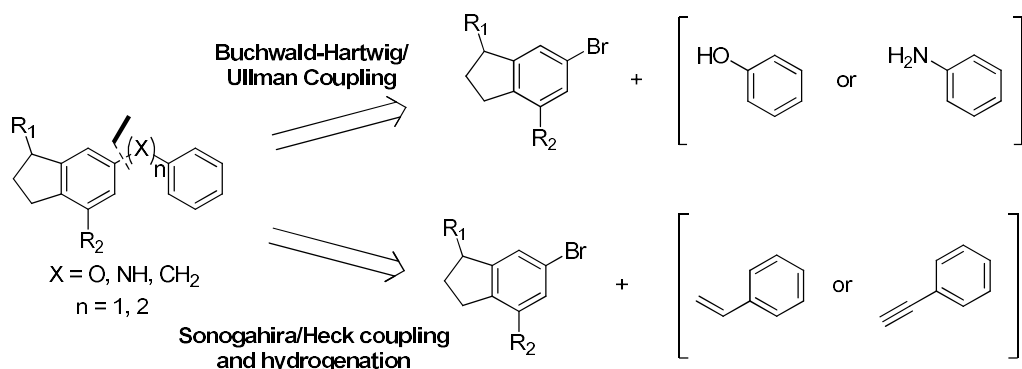
Scheme 4.2: Retrosynthetic cutting of the phenyl group linking chain in the *i*+1 position of the scaffold.

Retrosynthetic analysis of the *i*+1 phenyl group was then carried out (Scheme 4.2). The computational studies indicated that a chain length of one or two atoms between the phenyl group and the central scaffold provided the best overlap with BRCA2.

Synthesising compounds with a hydrocarbon chain length of one carbon atom appeared to be challenging. There are limited examples of transition metal-catalysed cross-coupling reactions at sp^3 hybridised carbon centres. Often the reagents required, such as benzylboronic acid derivatives, exhibit poor reactivity or are unstable. There are some examples in which this has been solved by using less reactive and thus more stable reagents, such as alkyltrifluoroboranes.[298] However, due to their lower reactivity, highly specific conditions are required for their reaction, such as expensive ruthenium based catalysts in the presence of ultraviolet light.[299,300] Because of this, one-atom linkers containing a heteroatom were considered because of more established reaction conditions.

The synthesis of biphenylethers is very well precedented using Ullmann condensation conditions.[301–303] This involves heating an aryl-halide with a phenol to high temperatures in the presence of a copper catalyst. These conditions can also be used to form C-N bonds, with the potential to produce a one atom aniline linker in the i+1 position. More recently, the development of the Buchwald-Hartwig reaction has allowed the formation of C-N and C-O bonds using milder conditions under palladium catalysis.[304–306] Both of these reactions would be suitable for the synthesis of single heteroatom linkers needed for BRCA2 mimetics.

For two atom linkers, it is possible to form this bond with an unsaturated two-carbon chain using palladium chemistry, which can then be hydrogenated to form the desired functionality. The Heck reaction can be used to cross-couple aryl-halides with terminal alkenes, such as styrene.[307] This involves heating the two reagents to high temperatures in the presence of a palladium catalyst and a base. The milder Sonogashira reaction[308,309] relies on the acidity of a terminal alkyne C-H bond. A copper/palladium co-catalytic system is employed, whereby the alkyne forms a cupric intermediate, facilitating transmetallation to the palladium during the catalytic cycle. This allows the reaction to be carried out at lower temperatures than the Heck reaction. The products of the reactions could then be hydrogenated to give the saturated ethylene as the desired two-atom linker.
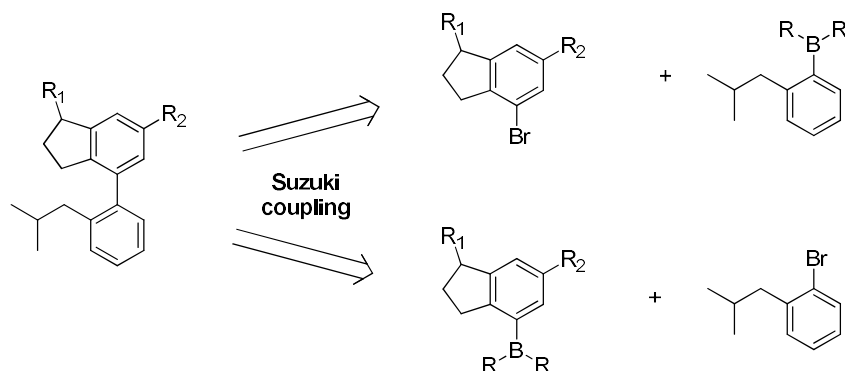
Scheme 4.3: Retrosynthetic cutting of the isopropyl group linking chain in the *i*+4 position of the scaffold.

Retrosynthetic analysis of the *i*+4 position suggested that disconnection at the aryl-aryl bond would be the best strategy, as it gave access to cross coupling reactions (Scheme 4.3). In particular, the Suzuki-Miyaura reaction allows the coupling between aryl halides and phenylboronic acid derivatives.[310,311] This reaction allows for the stereoselective cross-coupling of aryl rings using a palladium catalyst. As the aryl boronic acids required for i+4 functionalisation are commercially available, the Suzuki-Miyaura reaction was the preferred choice.

4.2.2 The Retrosynthetic Analysis of *i*, *i*+1 and *i*+4 Functionalisable Cores Proved Synthetically Challenging

Given the analysis of methodology which would allow functionalisation of the *i*, *i*+1 and *i*+4 positions during the synthesis of BRCA2 mimetics, the synthetic accessibility of different scaffold cores was analysed (Scheme 4.4). It was important to design the scaffolds so that there could be differentiation in the regioselectivity of cross coupling reactions to the *i*+1 and *i*+4 positions. One solution to this could be the use of *N*-methyliminodiacetyl (MIDA) boronates.[312] The MIDA group attenuates the reactivity of boron towards palladium, allowing other palladium catalysed reactions to be performed elsewhere in the molecule without reacting at this position. When the MIDA group is cleaved, the boron becomes reactive towards palladium, thus allowing iterative cross coupling reactions to take place.
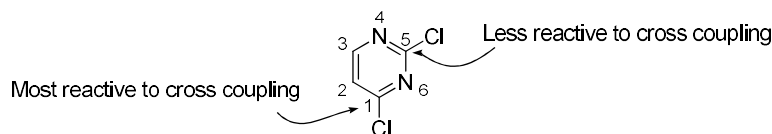
Figure 4.2: The reactivity of the 1 and 5 positions in dichloropyrimidines.

If a pyrimidine containing scaffold were used, it should be possible to differentiate between the $i+1$ and $i+4$ positions due to the differing reactivity of chlorides in this configuration (Figure 4.2).[313] The $i+4$ position should be the most reactive, and thus the first to react under cross coupling conditions.
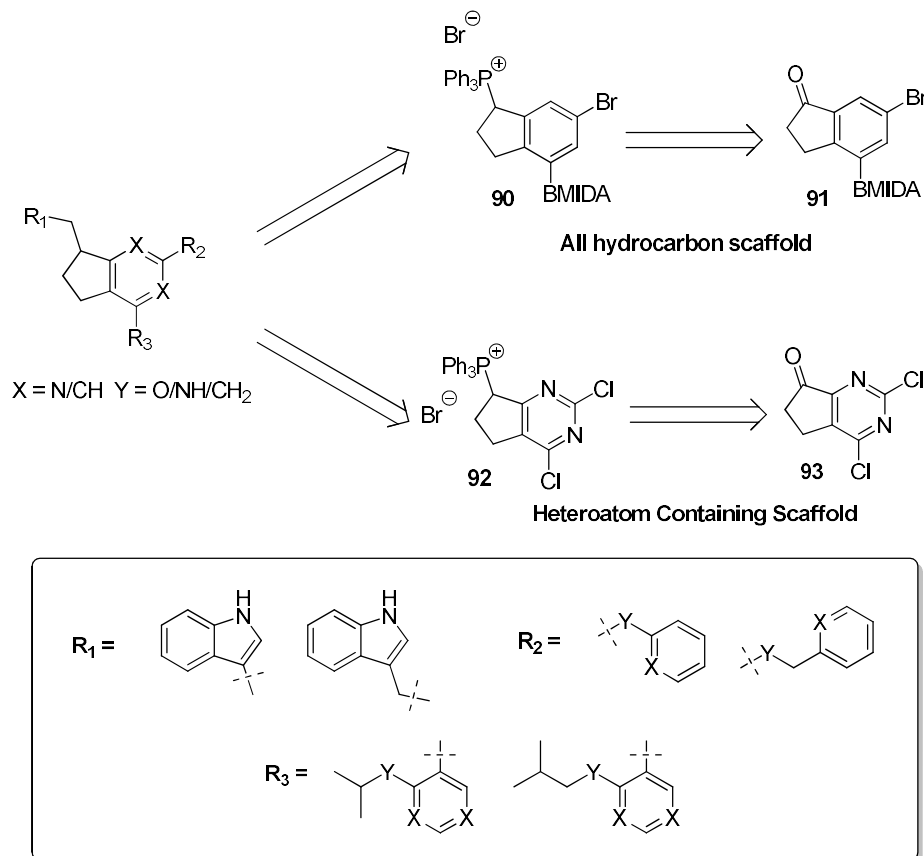
Scheme 4.4: Retrosynthetic analysis of i, $i+1$, $i+4$ scaffolds to produce cores with the potential for late stage diversification in each position.

Unfortunately, no straightforward or cost effective route that enabled the synthesis of these reagents could be devised. The standard route to form the indanone ring structure is *via* an intramolecular Friedel-Crafts acylation from a 3-phenylpropanoic acid. In the case of the all-hydrocarbon MIDA boronate containing indanone, it was difficult to buy starting materials that would lead to this structure within a small number of synthetic steps. There was also no literature precedent that suggested the electron poor pyrimidine ring would be able to undergo Friedel-Crafts acylation to form the ring. It also did not appear retrosynthetically that it would be possible to construct the pyrimidine ring onto a preformed cyclopentone based reagent. Due to the synthetic inaccessibility of the i, $i+1$, $i+4$ cores, the simpler i, $i+1$ scaffold was considered for synthesis.

4.2.3 Retrosynthetic Analysis of the *i, i*+1 Scaffold as a Synthetically Accessible Alternative
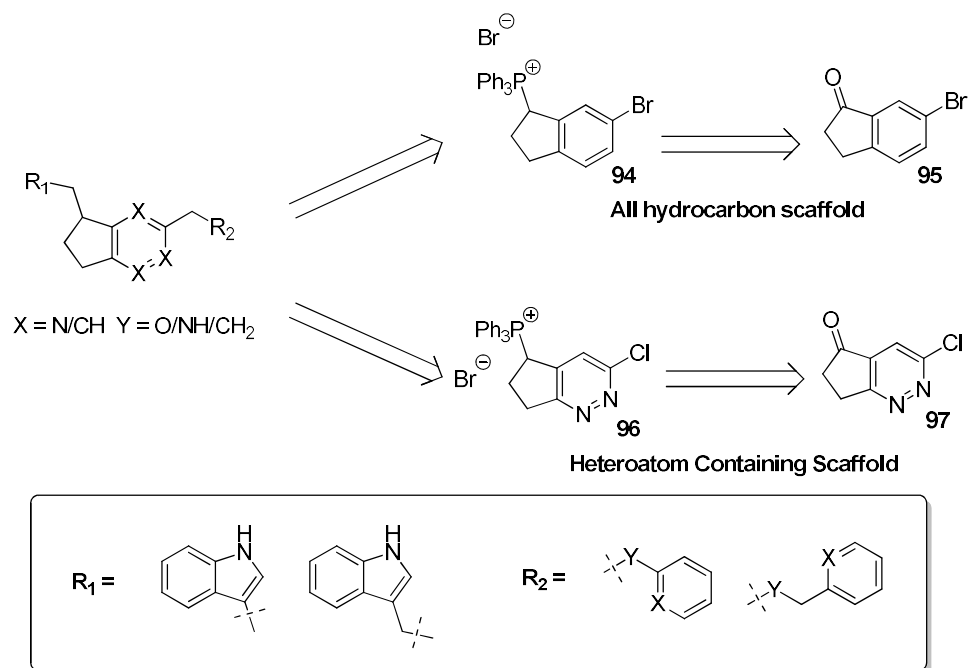
Scheme 4.5: Retrosynthetic analysis of *i, i*+1 scaffolds to produce cores with the potential for late stage diversification in each position.

The simplified structure of the *i, i*+1 scaffold appeared much easier to access synthetically based on retrosynthetic analysis. For a heteroatom containing scaffold, a pyridazine was chosen. In this configuration, the heteroatoms would be pointing away from the lipophilic pocket on PALB2, allowing them to have a positive influence on the physicochemical properties of the compound without having a negative effect on electrostatic interactions with the protein.
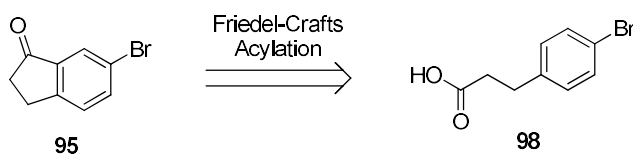
Scheme 4.6: Indanone **95** can be accessed synthetically in one step from commercially available starting materials.

Retrosynthetic analysis suggested it was possible to reach the all hydrocarbon bromoindanone intermediate **95** in one synthetic step from commercially available 3-(4-bromophenyl)propanoic acid (**98**) in one step, using Friedel-Crafts chemistry.
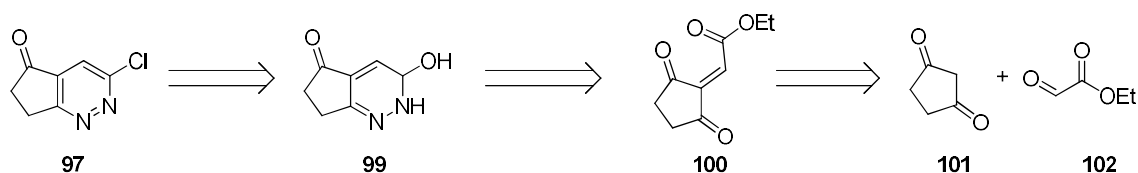
Scheme 4.7: Retrosynthetic analysis of pyridazine **97**

There was far less precedent for the synthesis of the pyridazine-based heteroatom containing intermediate **87**. As stated previously, there was no literature evidence to suggest that the ketone ring could be formed using an intramolecular Friedel-Crafts reaction on heteroatom containing rings, such as a pyridazine. Because of this, a strategy was developed to form the pyridazine ring from a pre-constructed cyclopentone (Scheme 4.7). It was predicted that by functionalising cyclopentadione **101** through an aldol condensation, and carefully reacting the intermediate with hydrazine, pyridazinone **99** may be produced. This could then be transformed to the chlorinated compound **97**, ready to be functionalised in the *i* and *i*+1 positions.

The *i*, *i*+1 mimetics were predicted to have much greater synthetic accessibility than their *i*, *i*+1, *i*+4 counterparts. The computational alignment studies described in chapter 3 did not highlight that a significant gain was made in terms of binding potential from the *i*+4 position in comparison to the *i* and *i*+1 positions. On top of this, significant gains in molecular weight had to be made as part of incorporating an *i*+4 group into the mimetic design, potentially reducing solubility in aqueous media. This led to the decision to pursue the synthesis of *i*, *i*+1 mimetics as a priority, in the hope that it would provide quicker results on the efficacy of small molecule mimetics towards PALB2.

4.3 Challenges towards the Synthesis of a Pyridazine Based *i*, *i*+1 Scaffold

Compared to other nitrogen containing heterocycles such as pyridine and pyrimidine, there is less literature precedent for synthetic routes to pyridazines. This may be influenced by the lack of naturally occurring pyridazines, and possibly by the toxicity of hydrazine as a reagent.

4.3.1 Review of Published Synthetic Routes to Pyridazines

The first example of a pyridazine synthesis was published in 1886 by Emile Fischer.[314] The most common route to a pyridazine is through the reaction of hydrazine with a 1,4-diketone moiety (Scheme 4.8).[315] If an unsaturated 1,4-diketone is reacted with hydrazine, it will be transformed directly into a pyridazine with the loss of two molecules of water, *via* a double condensation reaction (Scheme 4.8, **A**). However, often this reaction route can be difficult, due to the reliance on the *cis* conformation in order to form the ring. In cases where it may be difficult to selectively obtain the *cis* isomer, it is possible to synthesise the dihydropyridazine by reacting hydrazine with a saturated 1,4-diketone (Scheme 4.8, **B**). The resulting compound may then spontaneously oxidise to the aromatised pyridazine. If this does not occur, then an oxidant such as chloranil can be used in order to induce aromatisation.
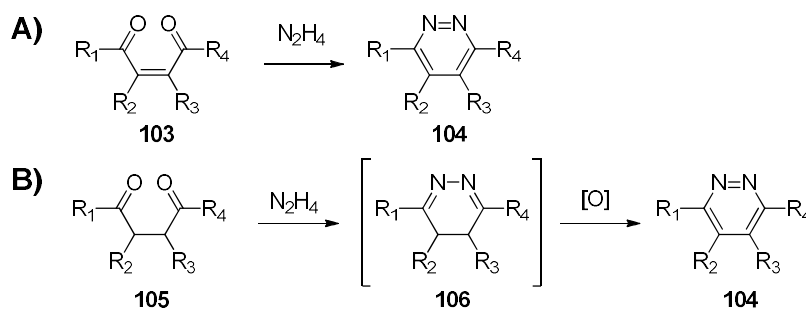
Scheme 4.8: General synthesises of pyridazines from saturated and unsaturated 1,4-diketone compounds.[315]

There are several other methods that have been published which offer variations on the standard pyridazine synthesis. The Schmidt pyridazinone synthesis involves the three-component reaction between an α-dicarbonyl compound, a malonate ester and hydrazine (Scheme 4.9).[316] Although there are many pathways that this reaction could follow, it is likely that the malonate ester reacts with the α-dicarbonyl compound *via* an aldol condensation, eliminating water and forming intermediate **109**. This can then react with hydrazine to form a hydrazone at the ketone position, followed by addition elimination with the adjacent ester to form a pyridazinone ring (**110**).
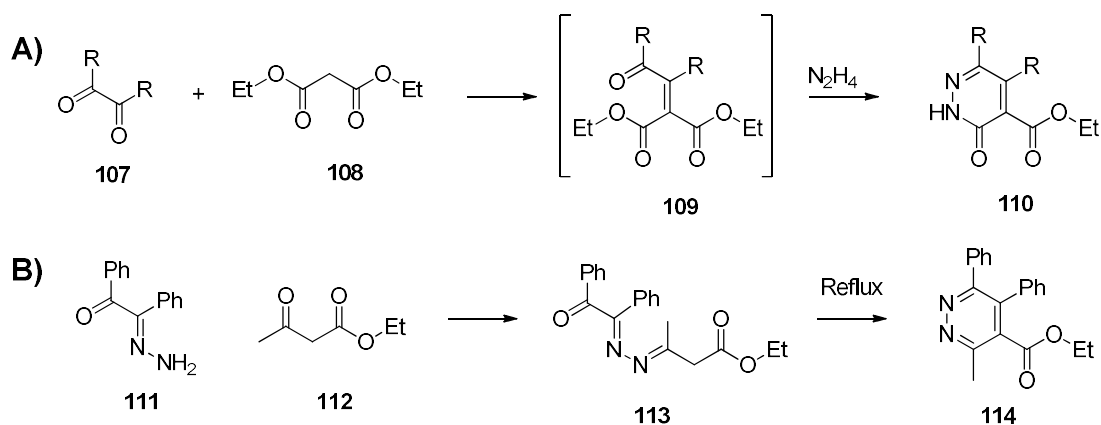
Scheme 4.9: A) Schmidt pyridazinone synthesis from α-dicarbonyl compounds and malonate esters.[316] B) Modified step-wise Schmidt synthesis by Evans and Schweizer allowing direct synthesis of pyrizadines.[317]

Evans and Schweizer developed a modified version of the Schmidt pyridazine synthesis, where they used a step-wise approach for the direct synthesis of pyridazines (Scheme 4.9, B).[317] They first prepared and isolated the biphenyl α-diketomonohydrazone **111**. They were then able to react this with 1,3-dicarbonyl **112** to form a bishydrazone. This was then transformed into the pyridazine on heating; the final step proceeds through an aldol condensation and the subsequent elimination of water to give **114**.
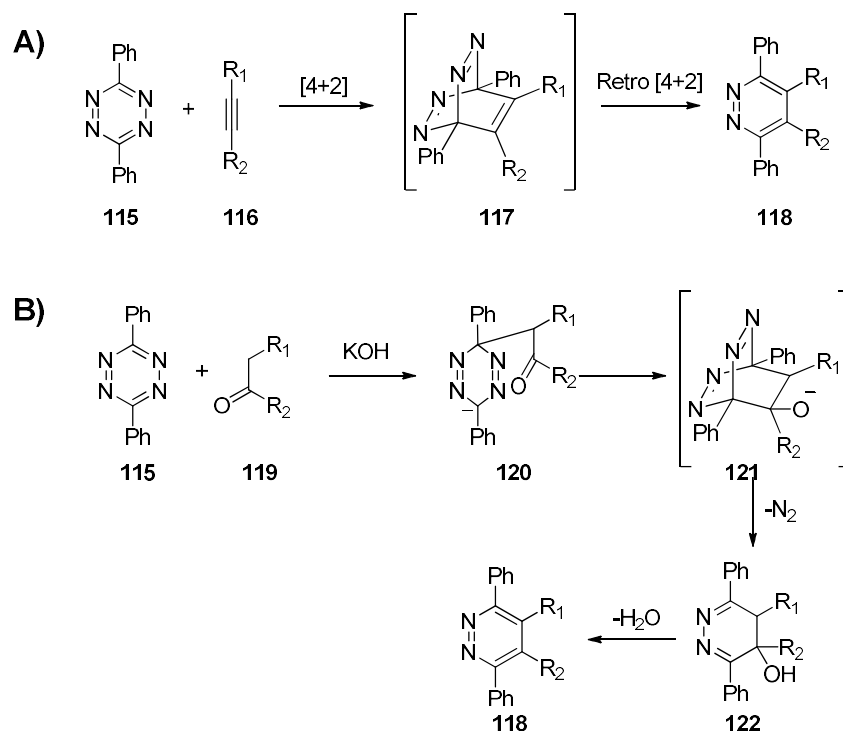
Scheme 4.10: A) Carboni-Lindsey reaction to form pyridazines from 2,3,4,5-tetrazines and alkynes.[318] B) Modification of the Carboni-Lindsey reaction to proceed using a keto-enolate by Haddadin and co-workers.[319]

The Carboni-Lindsey reaction forms a pyridazine through a very different reaction path. A 1,4-substituted 2,3,5,6-tetrazine is reacted with an alkyne through a 4+2 cycloaddition

reaction, forming the adduct **117**. A second pericyclic rearrangement then occurs, causing the loss of nitrogen gas, and resulting in a tetrasubstituted pyridazine ring (Scheme 4.10, A).[318]

This reaction was modified by Haddadin and co-workers, by deprotonating a ketone and then reacting with 2,3,5,6-tetrazine.[319] The enolate reacts at the 2,3,5,6-tetrazine 1-position, forming the intermediate **120**. The negative charge then acts as a nucleophile at the carbonyl position, forming a ring system **121**, similar to the intermediate in the Carboni-Lindsey reaction. Once this is formed, nitrogen is eliminated as before through a 4+2 sigmatropic rearrangement, forming a dihydropyridazine ring. Water is then eliminated, forming the pyridazine product **118**.
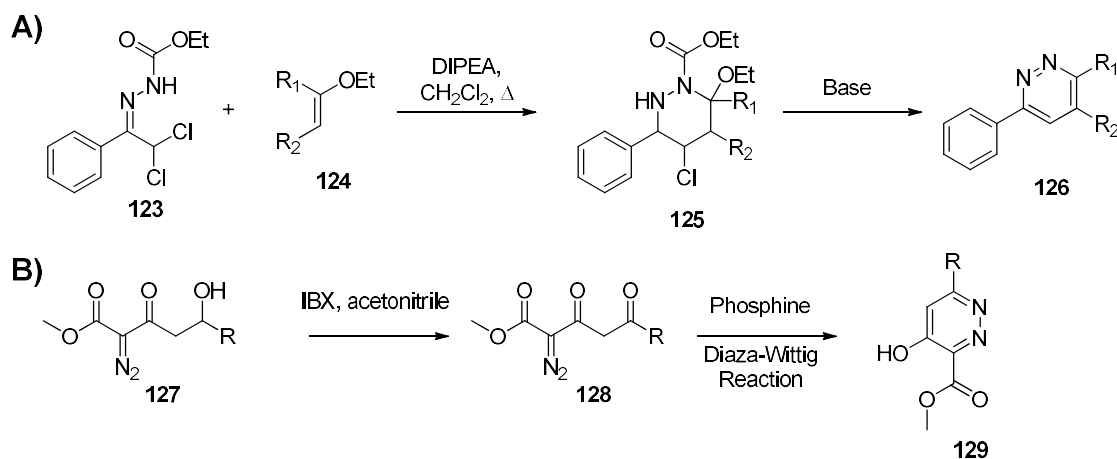
Scheme 4.11: A) South and co-workers pyridazine formation from 1-carbethoxy-3-phenyl-4-haloazodienes and electron rich olephins.[320] B) Strategy for the synthesis of pyridazines developed by Herdewijn and coworkers using diaza-Wittig chemistry.[321,322]

South and co-workers developed a technique for the synthesis of pyridazines using 1-carbethoxy-3-phenyl-4-haloazodienes (Scheme 4.11, A).[320] By reacting them with an electron rich alkene, such as the *O*-alkylated enol **124**, they found it was possible to form the hexahydropyridazine containing compound **125**. This compound then had the potential to form a pyridazine when treated with a base, through the elimination of hydrogen chloride, ethanol, carbon dioxide and ethane.

Herdewijn and coworkers developed methodology to form pyridazine rings using diaza-Wittig chemistry (Scheme 4.11, **B**).[321,322] They outlined a synthetic route to reach aldol **127**, where the functionality R could be diversified by the aldol reaction with an aldehyde. This could then be oxidised using 2-iodoxybenzoic acid (IBX) to the corresponding aldehyde. Treatment of this intermediate with a phosphine causes the aza-Wittig reaction

to proceed, driven by the formation of the phosphine oxide. The resulting compound is a 4-hydroxypyridazine.
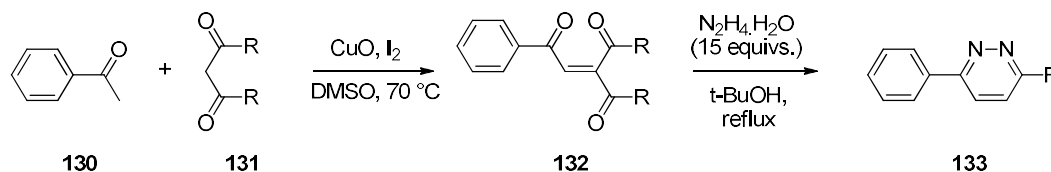
Scheme 4.12: Unexpected functional group cleavage observed by Wu and coworkers during pyridazine synthesis when using a large excess of hydrazine.[323]

Wu and co-workers observed an unusual functional group cleavage when they were synthesising pyridazines using unsaturated 1,4-dicarbonyl compounds. They had formed the 1,4-dicarbonyl precursors to pyridazines (**132**) by reacting benzophenone with 1,3-dicarbonyl compounds (**131**) in the presence of iodine and copper (II) oxide. However, they found that when they reacted this intermediate with a large excess (15 equivalents) of hydrazine in *tert*-butanol, rather than forming the pyridazine ring with a 3-substituted hydrazone, the functional group was cleaved, forming the disubstituted pyridazine **133**.

Given the different synthetic routes available to pyridazine rings, the synthesis of pyridazine-containing BRCA2 mimetic scaffolds was attempted.

4.3.2 Synthetic Steps Attempted Towards a Pyridazine Based BRCA2 Mimetic

Several different unsuccessful attempts were carried out to synthesise a pyridazine that could be functionalised into an *i*, *i*+1 mimetic. The unsuccessful synthetic routes are described below.

4.3.2.1 Attempted Route 1
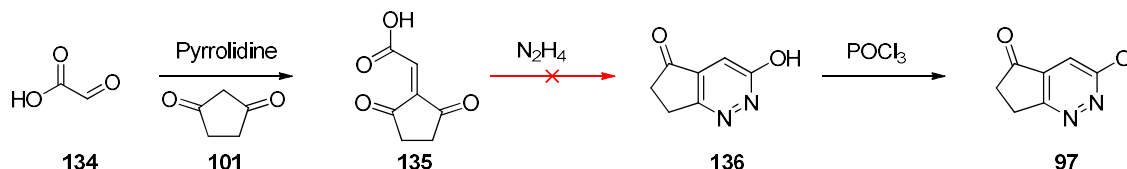
Scheme 4.13: The first attempted synthetic route to pyridazine **97**.

The first attempt at synthesising a pyridazine scaffold targeted chloro-substituted pyridazine **97**, from which substitution could be carried out in the *i* and *i*+1 positions as discussed previously (Scheme 4.13). It was hoped that after the condensation reaction between 1,3-cyclopentadione and glyoxylic acid, the resulting compound **135** could be treated with one equivalent of hydrazine to give the quinazolinone **136**. Treatment of this

with phosphorus oxychloride (POCl$_3$) would then give the desired compound **97**. The initial aldol condensation to form **135** did not give any product; however, when catalytic pyrrolidine was used in order to allow the reaction to proceed through a Knoevenagel condensation, the product was clearly visible by LCMS. Full purification and characterisation was not possible due to the hydrophilicity of the product, which precluded its extraction from aqueous media with organic solvents. Because of this, the product was treated with hydrazine without purification. However, no product was obtained, and instead a polymeric compound was observed by mass spectroscopy, and the route was abandoned.

4.3.2.2 Attempted Route 2
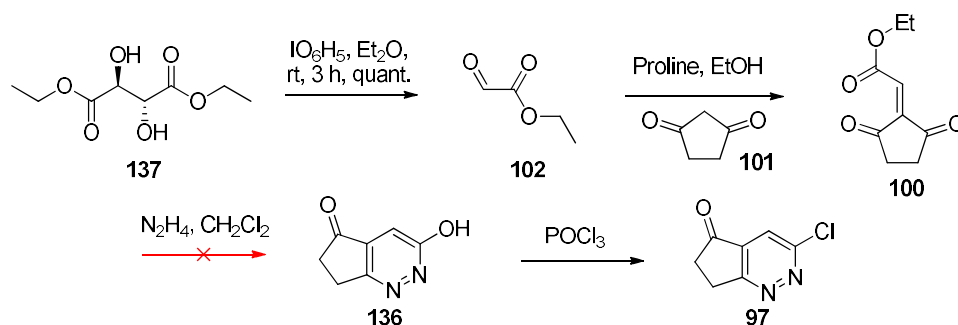
Scheme 4.14 The second attempted synthetic route to pyridazine **97**.

A new route was designed using ethyl glyoxylate to produce compound **100 (**Scheme 4.14**)**. It was hoped that the increased reactivity of esters towards hydrazine would counterbalance the problems faced in the previous attempt. The ozonolysis of diethyl fumarate was initially carried out to prepare ethylglyoxylate, however, no product was obtained, possibly due to the loss of product on workup due to its low boiling point of 110 °C.[324] Instead, the oxidative diol cleavage of diethyl L-(+)-tartrate by reaction with periodic acid was performed. To avoid evaporation of product, crude material was taken onto the next step without workup. A single peak with the mass of the hydroxyl acetal could be observed by LCMS. This was reacted under Knoevenagel conditions with 1,3-cyclopentadione in the presence of proline, to give the product **100**, which was directly reacted with one equivalent of hydrazine. However, a mixture of unidentified compounds was formed. This could potentially be due to the instability of unsaturated 1,4-dicarbonyl compounds for forming pyridazines observed by Wu and co-workers[323] (section 4.3.1), as well as the problem of selectivity with multiple carbonyl groups in one molecule. This route was not taken further.

4.3.2.3 Attempted Route 3
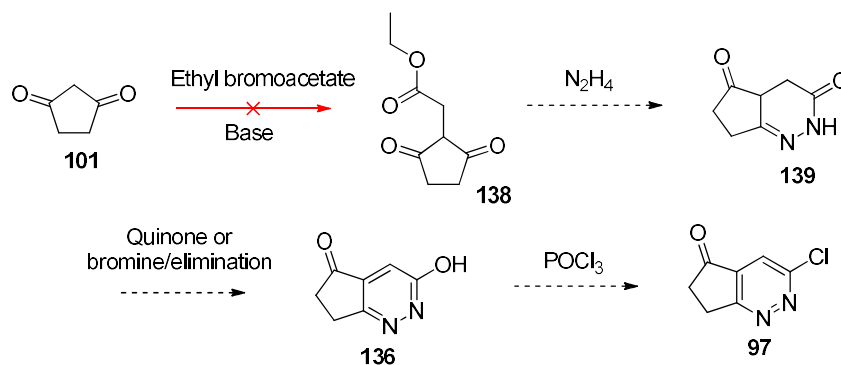
Scheme 4.15: Third first attempted synthetic route to pyridazine **97**.

In order to counterbalance the problem of selectivity, synthesis of the saturated compound **138** was attempted (Scheme 4.15). It was hoped that by doing so, the pyridazine could form after the formation of a hydrazone at either carbonyl, due to free rotation around the central bond. This may allow formation of the dihydropyridazinone, which could then be oxidised to the respective pyridazine **136**. However, problems were experienced in the alkylation of 1,3-cyclopentadione in the α-carbon 2-position.
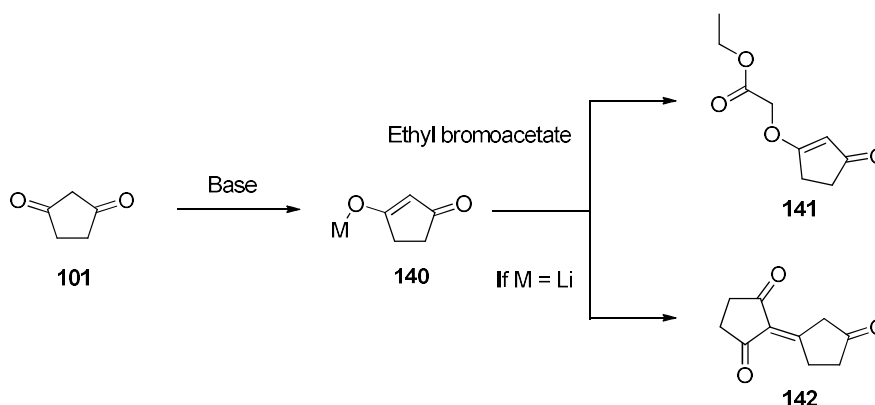
Scheme 4.16: The different products achieved from the attempted substitution reaction between cyclopentadione and ethyl bromoacetate, depending on the metal counterion present.

The reaction was attempted with potassium carbonate, potassium hydroxide, lithium hydroxide, lithium methoxide and trimethylamine as different bases. In most cases, it was found that the product from the reaction was the undesired enol ether **141** of 1,3-cyclopentadione and ethyl bromoacetate, resulting from *O*-alkylation, rather than the *C*-alkylated product (Scheme 4.16). Interestingly, when a lithium salt was used as a base, self-condensation of cyclopentadione occurred producing dimer **142**. This was potentially due to the ability of lithium ions to act as a Lewis acid, increasing the electrophilic nature of the carbonyl group. As the desired product **138** was not obtained, the synthetic route was abandoned.

4.3.2.4   Attempted Route 4
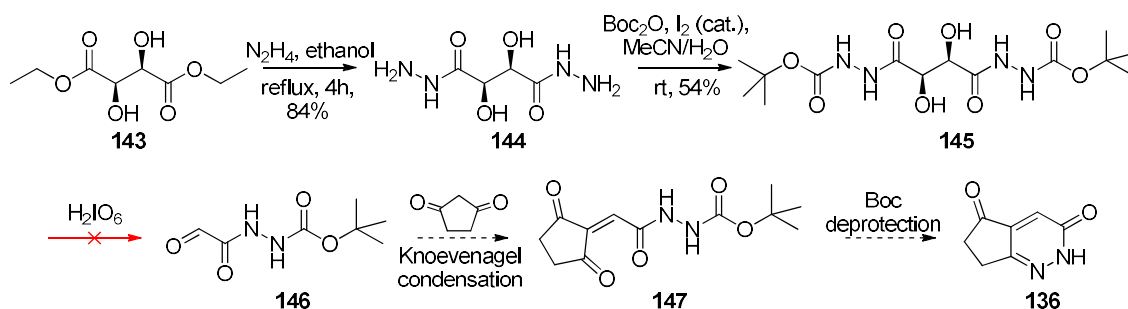
Scheme 4.17: The fourth attempted synthetic route employing a more step-wise approach to pyridazinone **136**.

In the light of the difficulties encountered in synthesis of pyridazines, a more step-wise approach was adopted (Scheme 4.17), in order to try and gain control over the selective addition of hydrazine to the carbonyl group required for pyridazine formation. This was attempted by the use of a Boc-protected hydrazide already in place on the glyoxylate reagent **146**. To synthesize the key intermediate, initially diethyl L-(+)-tartrate was transformed into the dihydrazide **144** by refluxing in ethanol in the presence of hydrazine. Some difficulty was encountered in functionalising this compound with Boc-protecting groups. Direct reaction of the dihydrazide **144** with Boc-anhydride in organic and in aqueous solvents under heating did not result in the Boc-protected product. The use of pyridine and of trimethylamine also yielded nothing. Using catalytic iodine as a Lewis acid was found to allow the reaction to proceed, giving the product in 54 % yield.

However, when the product was treated with periodic acid in diethyl ether to attempt oxidative diol cleavage, decomposition occurred, and effervescence was observed. When the reaction mixture was concentrated, only a small amount of unidentifiable material remained, suggesting that the starting material had decomposed to gaseous side products. Analysing the pKa of Boc-protected hydrazide **145** in MarvinSketch possibly helped develop an understanding of what had happened (Scheme 4.18).
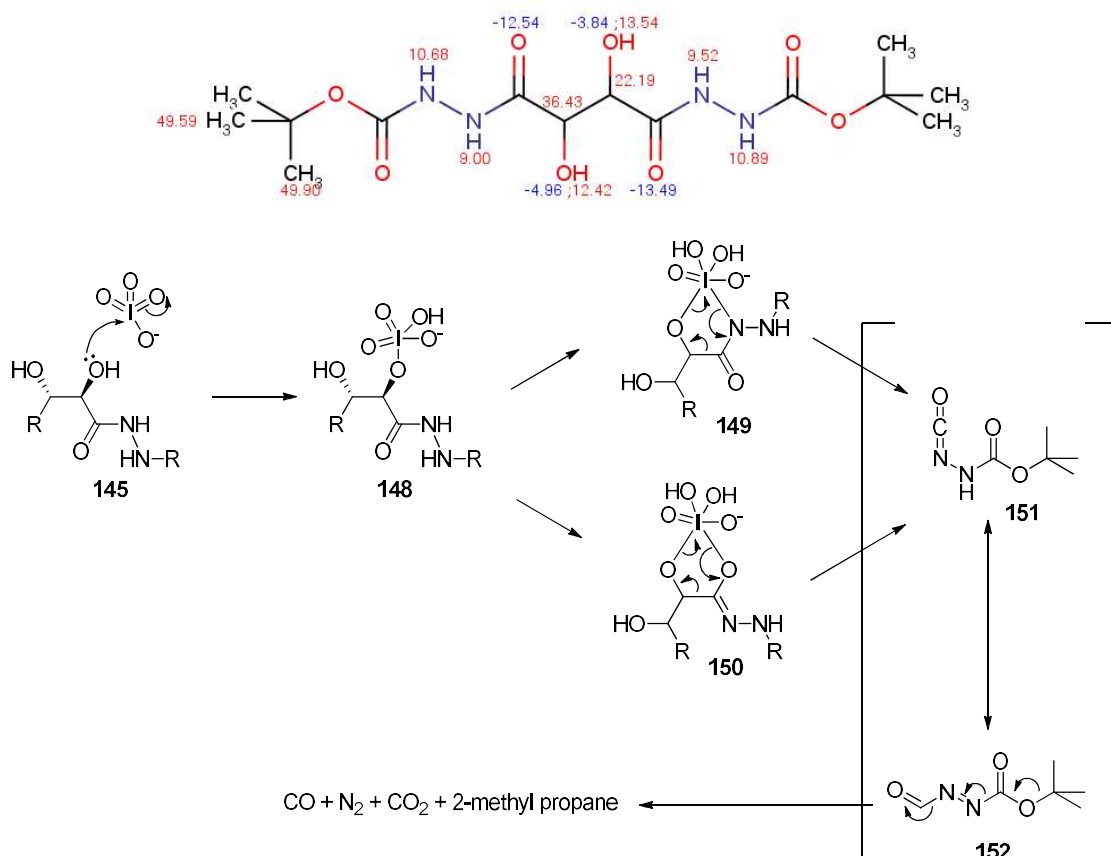
Scheme 4.18: Proposed mechanism for the decomposition of compound **145** observed when reacted with periodate species, with the pKa of different positions of Boc-hydrazide **145** as predicted by MarvinSketch.[284]

According to MarvinSketch, the point on the molecule with the lowest pKa is a hydrazide nitrogen, which has a predicted pKa of 9.0. Considering this, it is highly possible that during the reaction with periodate, the cyclic intermediate can form between one of the alcohol groups, and the hydrazide nitrogen, giving **149**. When the pericyclic oxidation occurs, this results in the cyanate containing compound **151**, which would tautomerise to compound **152**. This could then decompose to carbon monoxide, nitrogen, carbon dioxide and 2-methyl propane, as indicated in Scheme 4.18.

In an attempt to prevent this from happening, the reaction was attempted with sodium periodate in water, and silica-adsorbed sodium periodate in dichloromethane. However, in each case, the same result was observed. Because of this, this synthetic route was abandoned.

4.3.2.5 Attempted Route 5
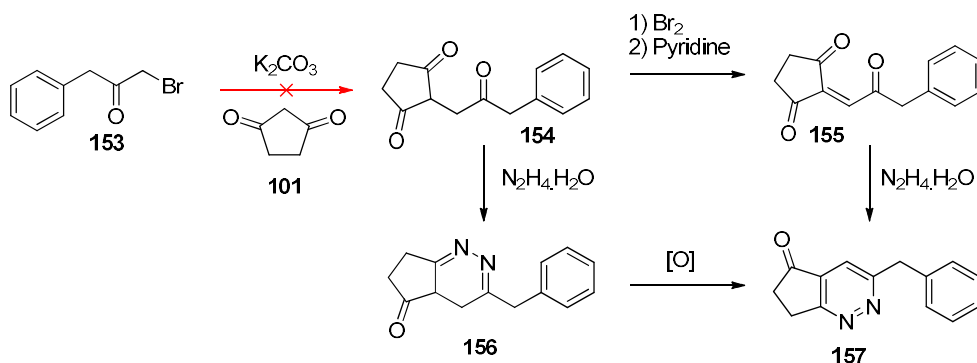
Scheme 4.19: The fifth attempted synthetic route to a pyridazine core, with early functionalisation of the *i*+1 position.

In a final attempt to synthesize a pyridazine containing *i*, *i*+1 mimetic, a synthetic route was designed that had a pre-functionalised *i*+1 position, containing a benzyl moiety. It was hoped that this would affect the formation of the pyridazine, as the ketone would be much more reactive than the esters from previous routes. However, as before, the alkylation of 1,3-cyclopentadione **101** was not successful due to the formation of the enol ether, and the synthetic route abandoned.

The inclusion of heteroatoms in the *i*, *i*+1 mimetic scaffold was desirable for effective tool compounds due to the expected increase in aqueous solubility, whilst still being compatible with the lipophilic PALB2 binding site as predicted by computational work detailed in chapter three. However, the synthesis of pyridazine-based *i*, *i*+1 scaffolds had proven problematic, with no fully synthesised mimetics produced after many different approaches were attempted. The synthesis of all-hydrocarbon indane-containing cores was more established in the literature. In order to achieve testable compounds, the focus was moved onto the synthesis of all-hydrocarbon *i*, *i*+1 mimetics, with the potential of returning to heteroatom containing cores as part of later SAR studies.

4.4 Synthesis of All-Hydrocarbon Indane-Based Wittig Reagents

The synthesis of all-hydrocarbon indanes to be used as the scaffold of *i*, *i*+1 mimetics was carried out, with the aim of functionalising the *i* position using Wittig or HWE chemistry (Scheme 4.20). The indanone functionality required for this was synthesised by intramolecular Friedel-Crafts acylation of a 3-phenylpropionic acid. The resulting ketone had the potential to react as the electrophile in a Wittig/HWE reaction. The ketone was

also converted into the respective phosphonium compound *via* the alcohol, allowing it to act as the nucleophile in the Wittig/HWE reaction too.
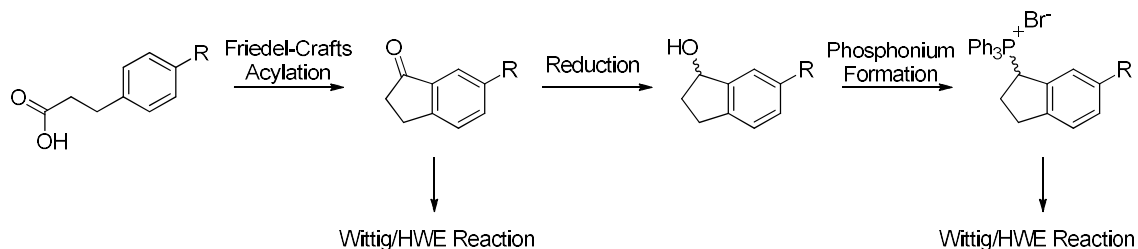
Scheme 4.20: General scheme for the synthesis of ketones and phosphonium salts for use in Wittig chemistry.

4.4.1 Intramolecular Friedel-Crafts Acylation Reactions.

Intramolecular Friedel-Crafts acylation reactions were carried out on three different 3-phenylpropionic acids, with different functionalities in the phenyl of the phenyl ring, to act as synthetic handles for *i*+1 functionalisation. Two propionic acids, 3-(4-bromophenyl)propionic acid (**98**) and 3-(4-methoxyphenyl)propionic acid (**161**) were obtained from commercial suppliers. The third acid, 3-(4-carboxyphenyl)propionic acid (**160**) was synthesised in two steps (Scheme 4.21).
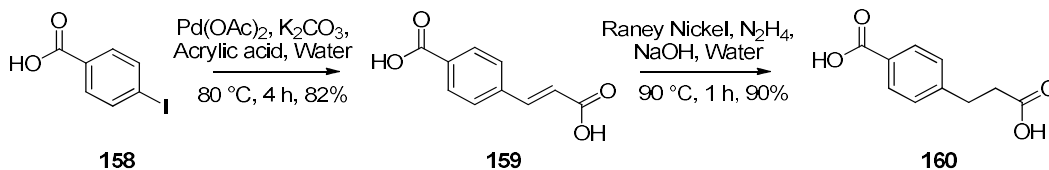
Scheme 4.21: Preparation of dicarboxylic acid **160** from 4-iodobenzoic acid.

4-Iodobenzoic acid (**158**) was coupled with acrylic acid under a palladium-catalysed Heck reaction in water. The product **159** was precipitated from the water by acidification, and obtained in 82% yield. The resulting carbon-carbon double bond from the Heck reaction was hydrogenated using Raney nickel in water, with hydrazine as a source of hydrogen. This gave the desired propionic acid **160** after 1 hour in 90% yield.

Once the three propionic acids had been obtained, they were subjected to Friedel-Crafts acylation conditions. 3-(4-Carboxyphenyl)propionic acid (**160**) was transformed into the indanone **163** by following published methodology.[153] It was reacted with aluminium trichloride in a melt with sodium chloride for 2 hours, giving the product in 83 % yield after workup.
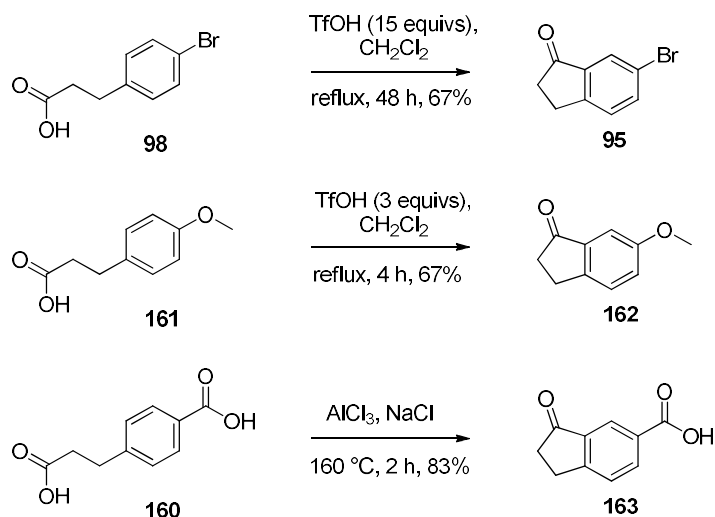
Scheme 4.22: The optimised conditions for three different Friedel-Crafts reactions to form indanones.

When 3-(4-bromophenyl)propionic acid (**98**) was reacted under these conditions, no product was obtained. However, using recently published conditions[325] and employing triflic acid as a catalyst to generate the carbonium ion required for acylation, the desired reaction occurred giving the product in 67% yield. The ease of acylation was heavily influenced by the electron density within the aromatic ring. For the electron poor bromophenyl derivative (**98**), 15 equivalents of triflic acid and forcing conditions (48 hours heating) were required to push the reaction to completion. However, for the electron rich anisole **161**, only three equivalents of triflic acid and four hours of heating were required to achieve the same yield.

4.4.2 Functionalisation of 6-Methoxyindanone (162) in the i+1 Position Prior to Wittig Functionalisation
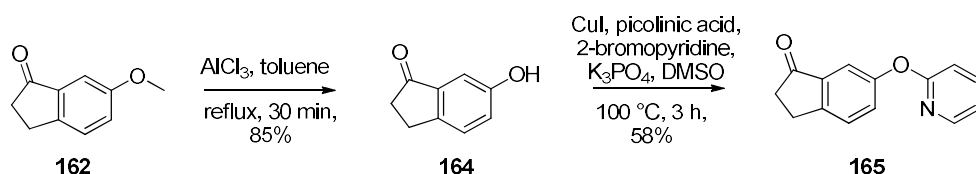
Scheme 4.23: The synthesis of *i*+1 functionalised indanone **165**, from 6-methoxyindanone **162**.

In the case of 6-methoxyindanone **162**, the decision was made to functionalise in the *i*+1 position before carrying out *i* functionalisation using Wittig chemistry. There was concern over whether the indole *N*-H would interfere with the Ullman conditions employed for aryl ether synthesis. Protecting group chemistry was considered, but there was no immediate solution which would not require multiple steps, and the exchange of protecting groups mid synthesis.

6-Methoxyindanone **162** was functionalised in two steps. Firstly, the anisole group was demethylated by refluxing with aluminium trichloride in toluene for 30 minutes. This gave the phenol **164** in 85% yield after purification. This was then coupled with 2-bromopyridine using an Ullmann condensation reaction. Picolinic acid has been shown to be a broadly applicable ligand for use in copper catalysed Ullmann condensation.[326,327] Use of these conditions gave the desired product **165** in 58% yield after 3 hours.
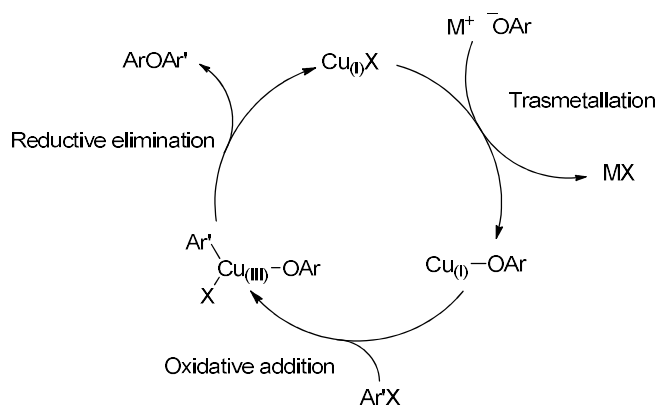
Scheme 4.24: The proposed catalytic cycle for an Ullman condensation to form an aromatic ether, involving an unusual Cu(III) oxidation state.[328]

The Ullmann condensation was first reported in 1905, and involves the use of catalytic copper to form aryl-heteroatom bonds.[301,302] Originally used for the formation of aryl ethers, it has since been developed to be used for a range of C-O and C-N bonds. The proposed mechanism of the reaction is similar to that of palladium catalytic cycles (Scheme 4.24).[328] Initially, a deprotonated heteroatom salt becomes co-ordinated to a copper(I) species, through a transmetallation reaction. The oxidative addition of an aryl halide to the copper can now occur, producing a highly unusual copper(III) species. Reductive elimination of both organic ligands forms a carbon-heteroatom bond, giving the product and completing the catalytic cycle.

4.4.3 The Reduction of Indanones to Alcohols
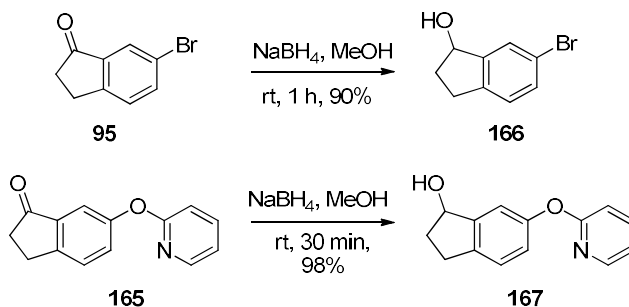
Scheme 4.25: The reduction of two different indanones to their respective alcohols using sodium borohydride.

The indanones synthesised were reduced to the corresponding alcohol by portion-wise addition of sodium borohydride in methanol, with no purification required. This was carried out on two different indanones. 6-Bromoindanone **95** was reduced to racemic 6-bromoindanol **166** in a 90 % yield after one hour. The 2-pyridyl functionalised indanone **165** was also reduced to alcohol **167**, in a 98% yield after 30 minutes.

4.4.4   The Conversion of Indanols to Phosphonium Salts for Wittig Chemistry
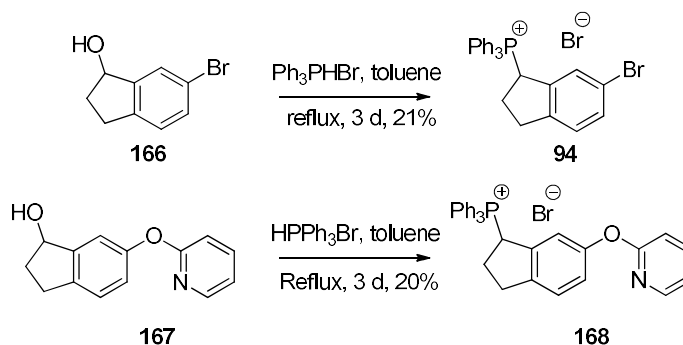
Scheme 4.26: One-step conversion of alcohols to their respective phosphonium salt, using triphenylphosphonium hydrobromide.

Once the indanols **166** and **167** had been synthesised, they were transformed into the respective triphenylphosphonium salts. Initially, this was attempted in one step, using triphenylphosphonium hydrobromide (Scheme 4.26).[329] This reagent acts as a source of hydrogen bromide (HBr), to convert the alcohol to the respective bromide. The bromide can then undergo a condensation reaction with triphenylphosphine, forming the phosphonium bromide salt as the product. The reaction gave the desired phosphonium salts at 21% and 20% yields for **94** and **168,** respectively. The reaction was found to only be effective on a large scale, with smaller scale (approximately 200 mg) reactions only yielding 3-4%. Although the yield of 20% was adequate to get enough material for the next step, a stepwise approach was tested in an attempt to improve the yield (Scheme 4.27).

In the step-wise approach, the indanol was reacted to the respective 1-bromoindane using phosphorus tribromide in dichloromethane. These conditions allowed the conversion of bromoindanol **166** to the corresponding bromide in a 94% yield after two hours. When the bromide was heated with triphenylphosphine in toluene at reflux, the phosphonium salt **94** was obtained in 57% yield. This yield, which equates to 54% over two steps, is an improvement on the 21% achieved for the one step method using triphenylphosphonium hydrobromide on the same indanol. Conversion of 1,6-dibromoindane to the diethyl

phosphonate ester was attempted using a Michaelis-Arbuzov reaction.[330] None of the desired product could be detected in the crude mixture by LCMS or ^{1}H NMR. However, the less sterically hindered benzylbromide **171** could be converted to the phosphonium salt **172** by reacting with triphenylphosphine in toluene under reflux with a higher yield of 92%.
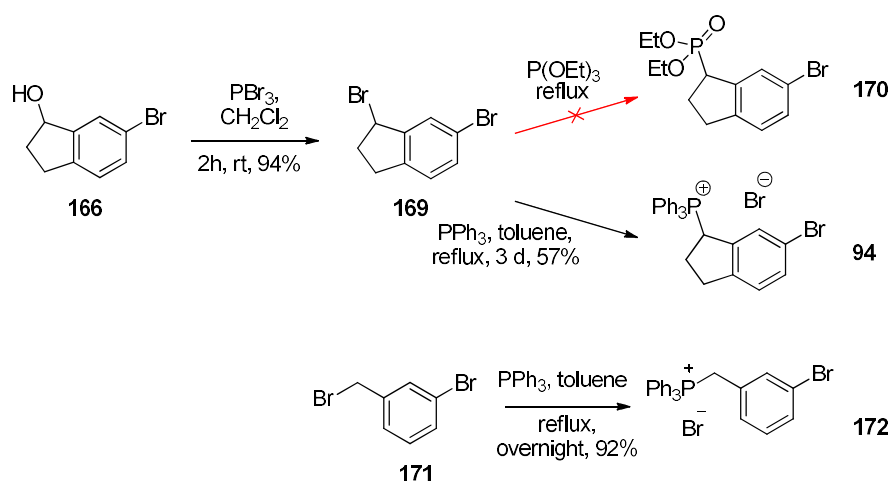
Scheme 4.27: Stepwise conversion of alcohols to the respective phosphonium salt, by conversion to the bromide, then reaction with triphenylphosphine.

The bromination of indanol **166** and **167** resulted in several problems (Scheme 4.28). Depending on the conditions used, three different brominated products were obtained. Initially, Appel reaction conditions were used in an attempt to form the bromoindane. However, the predominant component after column chromatography was identified as the tetrabromoindole **174** by ^{1}H NMR. The coupling constants of the three indane protons identified as being in an AAB system.

The use of phosphorus tribromide as an alternative to Appel conditions appeared to solve this problem, producing the desired product **169** in 96% without purification required. However, it was found that when the reaction was repeated on a day with high levels of sunshine in the lab, the only isolated product was the tribromoindane **173**. ^{1}H NMR analysis of the methylene protons at the indane 2-position suggested that only the (*R,R*) or (*S,S*) *trans* diastereomers were present. This was potentially due to benzyl radicals produced in the presence of light energy.

Synthesis of the *i*+1 functionalised bromoindole **175** was not achieved, and so the single-step method was used in this case.
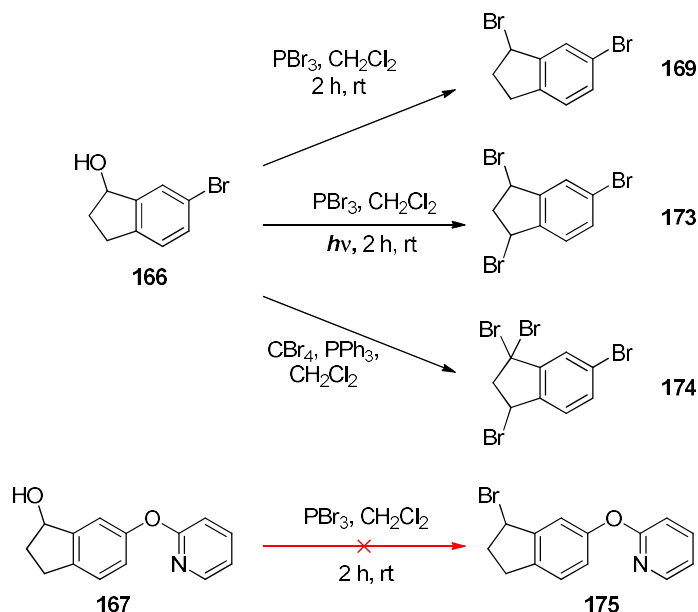
Scheme 4.28: The effect of reaction conditions on the isolated product of indanol bromination.

4.5 Using the Wittig Reaction to Functionalise the *i*+1 position

With the synthesis of the indanone and indane phosphonium salts complete, Wittig chemistry could be employed to functionalise the scaffolds in the *i* position with an indole moiety.
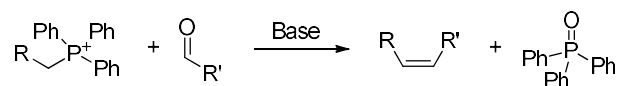
Scheme 4.29: General reaction scheme of a Wittig reaction.

Conditions were explored for the functionalisation of the *i* position of the scaffolds synthesised, using Wittig chemistry. For these reactions, the phosphonium and carbonyl synthetic handles could be placed in one of two different pairings (Scheme 4.30). For the mimetics with a 1-carbon linker, it was considered best to react the indane as a phosphonium salt. This is because of the high risk of elimination of triphenylphosphine from 3-triphenylphosphomethyl indole.
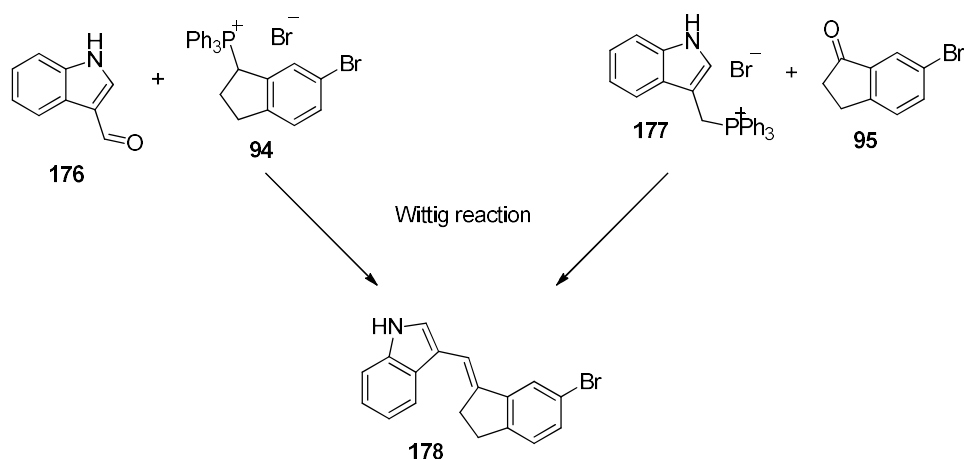
Scheme 4.30: The synthetic handles can be placed in two different pairings for the Wittig reaction.

4.5.1 Using the Wittig Reaction to Synthesise Reagent 178

Initial attempts to perform a Wittig reaction on phosphoindane **94** were unsuccessful (Scheme 4.31). Phosphoindane **94** was reacted with indole-3-carbaldehyde **176**, using potassium carbonate with 18-crown-6 in dichloromethane, however only starting materials could be detected in the crude material after being stirred overnight.
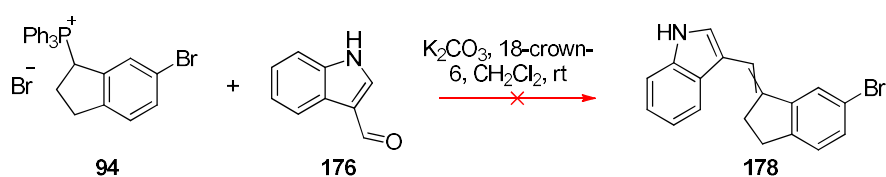
Scheme 4.31: The failed Wittig reaction in the presence of the indole N-H bond.

It was hypothesised that the indole N-H bond was interfering with the reaction. With a pKa calculated to be 13.5 (MarvinSketch),[284] it is the most acidic proton in both reagents. Because the negative charge would be delocalised to the aldehyde, the deprotonated species would be significantly less electrophilic. In order to circumvent this problem, the N-H bond was masked with an electron withdrawing carboxybenzyl (Cbz) protecting group.
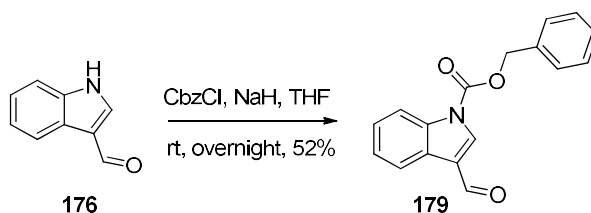
Scheme 4.32: The protection of the indole N-H bond with a Cbz group.

To protect the indole-3-carboxaldehyde, it was deprotonated with sodium hydride, then reacted with carboxybenzyl chloride (CbzCl) in tetrahydrofuran overnight, giving the product **179** in 52% yield.

Table 4.1: Different conditions tested for the Wittig reaction between compounds **94** and **179**.

	Conditions	Yield
1	K$_2$CO$_3$, 18-crown-6, CH$_2$Cl$_2$, r.t, 3 days	10%
2	BuLi, THF, -78 °C to r.t, 6 hours	Trace by TLC, not isolatable
3	DBU, THF, reflux, overnight	0%
4	DBU, ethanol, r.t., overnight	Trace by TLC, not isolatable
5	LiHMDS, THF, 0 °C to r.t., overnight	Trace by TLC, not isolatable
6	NaH, THF, r.t., overnight	22%

Once the protected indole had been synthesised, it was submitted to the same Wittig conditions as used before, and the desired product was obtained in a 10% yield. In an attempt to increase the yield, six different sets of alternative reaction conditions were screened (Table 4.1). Of these, the only other set of conditions that gave product was the use of sodium hydride in THF. This could be due to the large amount of steric hindrance around the benzylic position, which had less effect on the smaller bases such as sodium hydride.
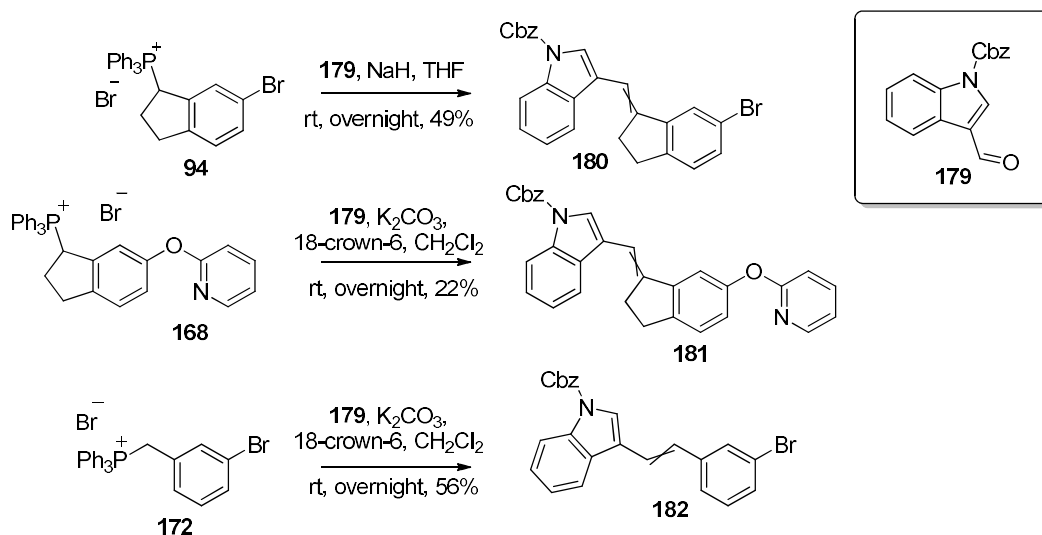
Scheme 4.33: Different successful Wittig reactions to functionalise the *i* position of BRCA2 mimetics.

On scale-up of the optimal Wittig conditions in Table 4.1, the yield was found to improve from 22% to 49%. The Wittig reactions of *i*+1 functionalised phosphonium salt **168**, and

benzylphosphonium salt **172** were also attempted. In both of these cases, the reaction conditions with sodium hydride were found to be ineffective. Instead, the reaction was carried out using potassium carbonate and 18-crown-6 and was found to be successful in both cases. The *i*+1 functionalised phosphonium salt was reacted overnight at room temperature with *N*-Cbz protected indole **179** to give the product **181** in 22% yield. The lower yield could be due to the increased electron density provided by the aryl ether increasing the pKa and causing destabilisation of the phosphonium ylid. For the benzylic phosphonium bromide **172**, reaction with *N*-Cbz protected indole **179** afforded the product as approximately a 1:1 mixture of both *cis* and *trans* isomers in 56% yield. Although it was difficult to separate the isomers by column chromatography, small fractions of each isomer were obtained and fully characterised. As the stereochemistry of each double bond was to be subsequently lost at a later stage by hydrogenation, determination of the exact stereochemistry of other Wittig products was not attempted.

4.5.2   Synthesis of Reagents for the Two-Carbon Linker i Functionalisation
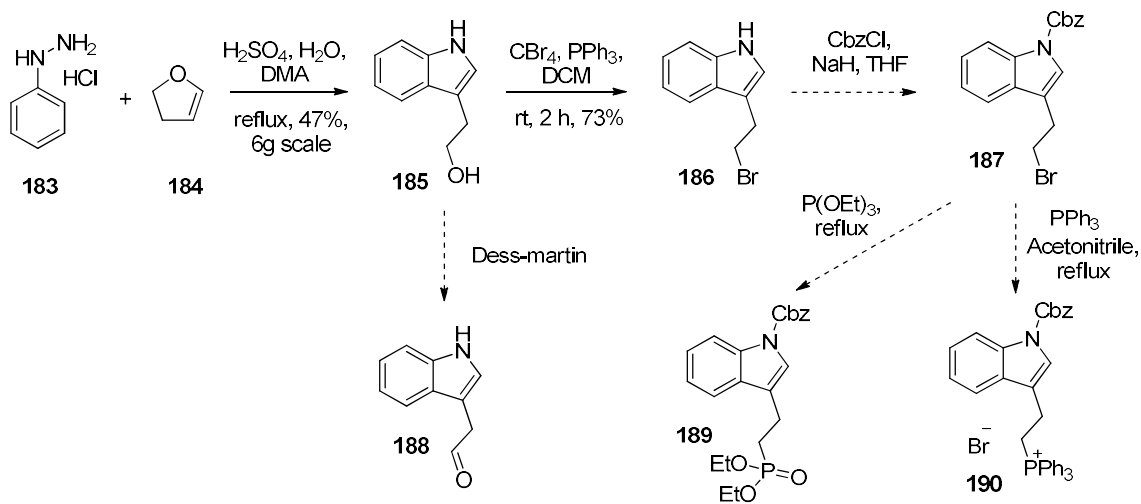
Scheme 4.34: Synthesis towards reagents for a 2 chain linker in the *i* position, with reversed synthetic handles.

It was possible to predict the most stable pairing for the Wittig synthetic handles when designing the synthetic route for the one carbon linker due to possible elimination of the indole phosphonium salt. For the two carbon linker, however, it was less clear which pairing would be the most effective. As a result, the synthesis of an indole functionalised with an aldehyde and with a phosphonium salt were planned, so that both could be tested (Scheme 4.34). Indole **185** was synthesised from 1,2-dihydrofuran using a Fischer indole synthesis, giving the product in 47% yield. The alcohol group was then transformed to

the corresponding bromide by using Appel conditions with carbon tetrabromide and triphenylphosphine in dichloromethane. This gave the product **186** in 73% yield after purification. Unfortunately, due to time constraints, the completion of the synthesis of both aldehyde **188** and phosphonium **190** were not completed.

4.6 Using Heck and Sonogashira Chemistry to Functionalise the i+1 Position, and Hydrogenation to Give Final Compounds

Once the *i* position had been functionalised, the Heck reaction and Sonogashira coupling reactions were investigated as a means of functionalising in the *i*+1 position.
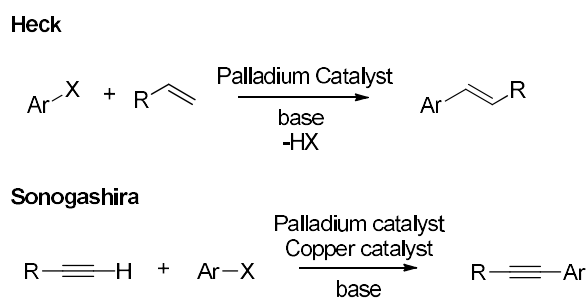
Scheme 4.35: A general reaction scheme for a Heck reaction[307,331] and Sonogashira reaction.[308]

Initially, the Heck reaction was used for the attempted functionalisation of the *i*+1 position (Scheme 4.36). The reaction was attempted with both styrene and 4-vinylpyridine. Palladium(*o*-tolyl)$_3$ was used as it has been shown to be a widely applicable ligand for the Heck reaction. Unfortunately, no product was detectable in the crude reaction mixture, despite the consumption of starting material. Instead, a complicated mixture of unidentifiable side products was produced in both cases.
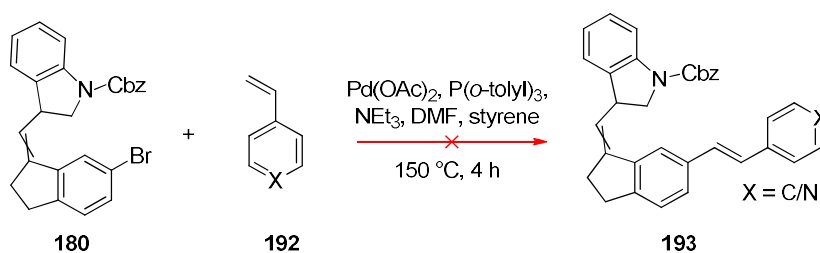
Scheme 4.36: Failed heck reaction conditions using styrene and 4-vinylpyridine to form heck product **193**.

The Sonogashira reaction was used as an alternative to the Heck reaction (Scheme 4.37). Comparatively forcing conditions were chosen. When the reaction was carried out at room temperature no reaction was observed, probably due to the insolubility of the starting material. However, when the reaction mixture was heated to 80 °C, the starting material began to dissolve, allowing the reaction to occur. As a result, the product **194** was obtained in 24% yield.
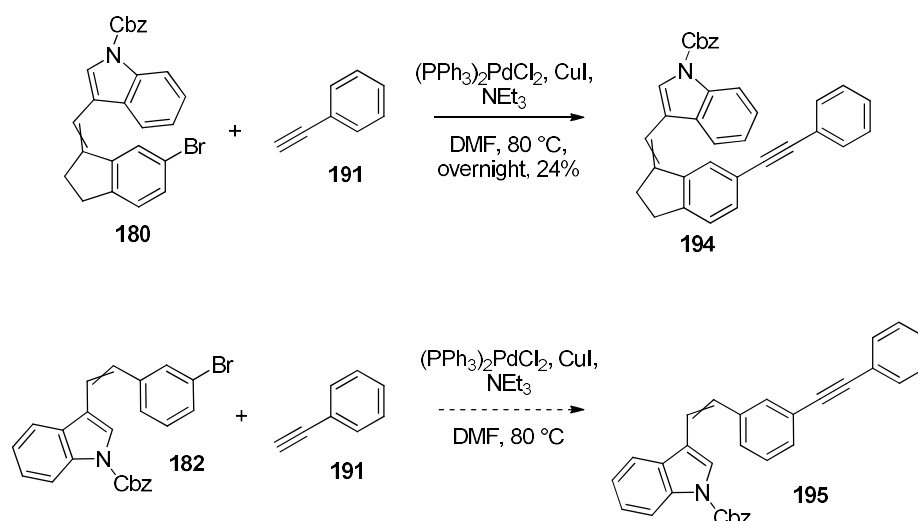
Scheme 4.37: The successful Sonogashira reaction between bromoindane **180** and ethynyl benzene to give **194**, as well as the reaction between phenylbromide **182** and ethynyl benzene that was not carried out due to time constraints.

After the *i*+1 functionalisation by Sonogashira coupling, the final stage of the synthesis was palladium catalysed hydrogenation. This would potentially deprotect the *N*-Cbz group, as well has saturate the *i* alkene bond and the *i*+1 alkyne bond. Compound **194** was stirred in THF in the presence of palladium on charcoal, with a hydrogen atmosphere overnight. The crude material was then purified by preparative HPLC, to give the product **196** in a yield of 18%.
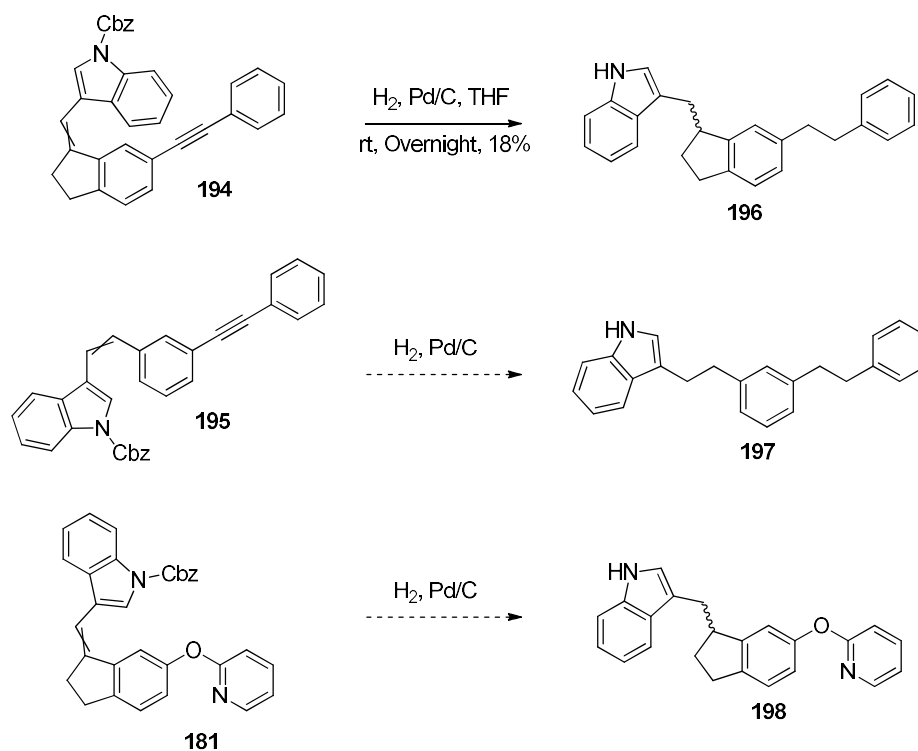
Scheme 4.38: Prospective hydrogenations to give BRCA2 mimetics.

4.7 Conclusions

In conclusion, synthetic routes were designed for the synthesis of small molecule BRCA2 mimetics, based on the results of computational studies. Due to the difficulty in designing a short synthetic route of i, $i+1$, $i+4$ mimetics, the synthesis of the simplified i, $i+1$ mimetics was performed instead.

Difficulty was encountered with the synthesis of pyridazine containing scaffolds to serve as a backbone for the i, $i+1$ mimetics. The problems seemed to arise from the high reactivity of hydrazine rendering selective reaction with the desired carbonyl groups unachievable. A step-wise approach was attempted, but abandoned when the reaction was unsuccessful. After many failed synthetic routes, the synthesis of pyridazine containing cores was abandoned.

The synthesis of an all-hydrocarbon i, $i+1$ scaffold was more successful, and several indanes were formed using an intramolecular Friedel-Crafts acylation, followed by subsequent conversion to the triphenylphosphonium salt *via* the alcohol. This allowed for late-stage functionalisation in the i and $i+1$ positions.

Functionalisation in the i position was carried out by Wittig chemistry. By protecting indole-3-carboxaldehyde with an *N*-Cbz group, the aldehyde group was rendered electrophilic enough to undergo a Wittig reaction with a triphenylphosphonium salt. It was subsequently reacted to form three different i, $i+1$ mimetic scaffolds with tryptophan (indole) functionalisation in the i position.

Functionalisation in the $i+1$ position was carried out using a Sonogashira coupling reaction. This was performed on one scaffold, and subsequently hydrogenated to produce the fully synthesised BRCA2 i, $i+1$ mimetic, ready for testing. Two other mimetics also are within two synthetic steps of the end of the synthetic route, but were not able to be synthesised due to time constraints.

5   Conclusion and Further Work

5.1   Conclusions

This thesis has investigated the processes by which protein α-helices can be mimicked by designed structures.

In the introduction the identification of PPI targets for inhibitor design, and in particular how modern computational methods have been able to predict the druggability of a PPI with increasing accuracy, was covered. The different methods that have been used to design α-helical mimetics were then described. This covered: using unnatural amino acid residues to create protein-mimicking *foldamers*, using naturally occurring and highly stable short peptide structures as surrogates for the design of *miniproteins*, placing a covalent linker at a set position in a peptide sequence to form a *stapled helix*, and by the design of *small molecule* scaffolds that are able to be functionalised to mimic the position of amino acid residues in an α-helix. In particular, the use of *miniproteins* to inhibit PPIs was reviewed in depth due to a lack of existing published reviews in the literature. This helped develop an understanding of the ways in which an inhibitor for a novel PPI target could be developed. It also highlighted aspects of α-helical mimetic research, in particular the solubility and cell permeability of PPI inhibitors, which were in need of further scientific development.

11 successful out of 18 examples
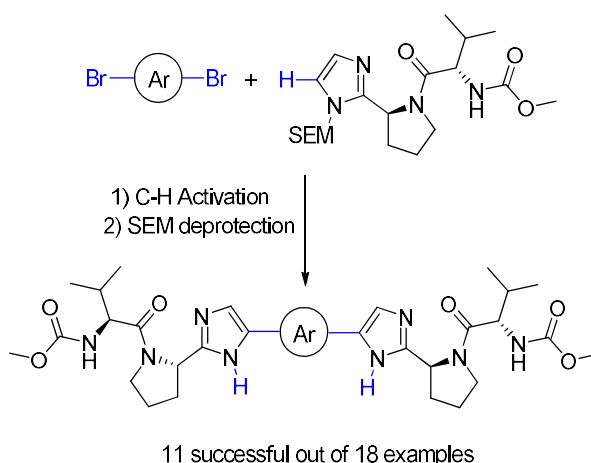
Scheme 5.1: Overview of the C-H activation methodology developed.

In unusual cases, approved drug compounds are able to exhibit excellent pharmacological (PK) profiles despite having unfavourable physical properties. It was hypothesised that it may be possible to incorporate structural features from these compounds into PPI inhibitors in order for them to also adopt better PK profiles. This hypothesis was explored

in chapter 2. The HCV NS5A inhibitor, daclatasvir, was identified as having good oral bioavailability alongside a high molecular weight of 739 Daltons. This property was attributed to a pair of symmetrical polar side chains within the structure of daclatasvir. Methodology which would allow the functionalisation of small molecule PPI inhibitors with poor PK profiles with 'daclatasvir-like' side chains could be of use for the development of membrane permeable PPI inhibitors. C-H activation was identified as potential methodology for the formation of the required C-C bonds. This reaction was attempted and proved successful, and the conditions optimised in order to reduce the amount of homo-coupling leading to inseparable side products. The scope of the optimised conditions was tested and of 18 different aryl dibromide cores investigated, 11 were able to undergo C-H activation (Scheme 5.1). This methodology was shown to have the potential to synthesise a range of 'daclatasvir-like' compounds as potential NS5A inhibitors, resulting in a peer-reviewed publication.[234] It can be used to functionalise a wide range of aryl bromides with solubilising side chains, and could potentially help towards synthesis of PPI inhibitors with good PK profiles.
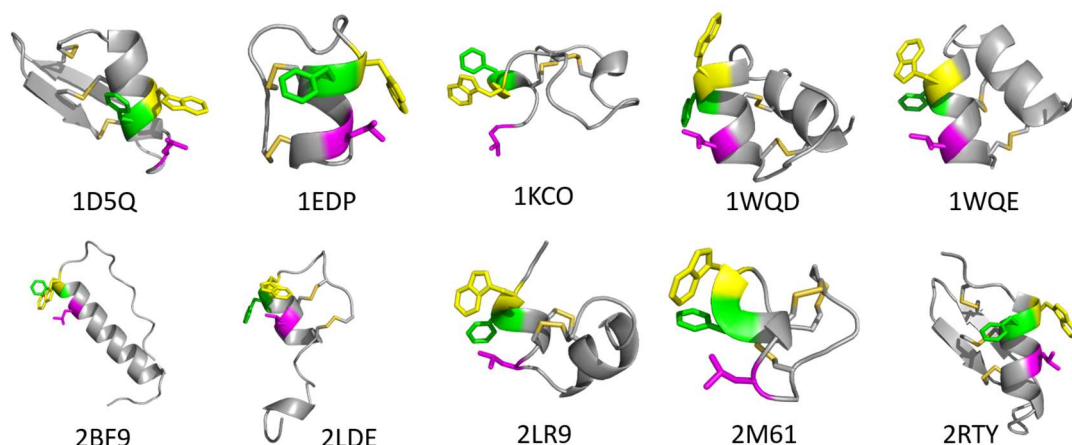
Figure 5.1: Structures of the 10 miniproteins selected for synthesis.

The PALB2/BRCA2 PPI was identified as a potential candidate as a target of cancer therapeutics through synthetic lethality with damaged DNA break repair mutations. However, there are no published inhibitors of the PALB2/BRCA2 interaction. It was hypothesised that the methods reviewed to mimic α-helical secondary structure could be used to develop tool compounds to probe the therapeutic value of a PALB2/BRCA2 inhibitor in cell-based assays. A series of miniproteins containing the WFXXL hot-spot motif presented by BRCA2 were designed using computational techniques to test for compatibility when binding to PALB2 (Figure 5.1). Upon purchasing the protein

sequences, CD spectroscopy of one of the sequences suggested that the WFXXL motif had not disrupted α-helicity, which was calculated to be 95% in solution. The miniprotein sequences are now ready for testing in a biochemical assay.
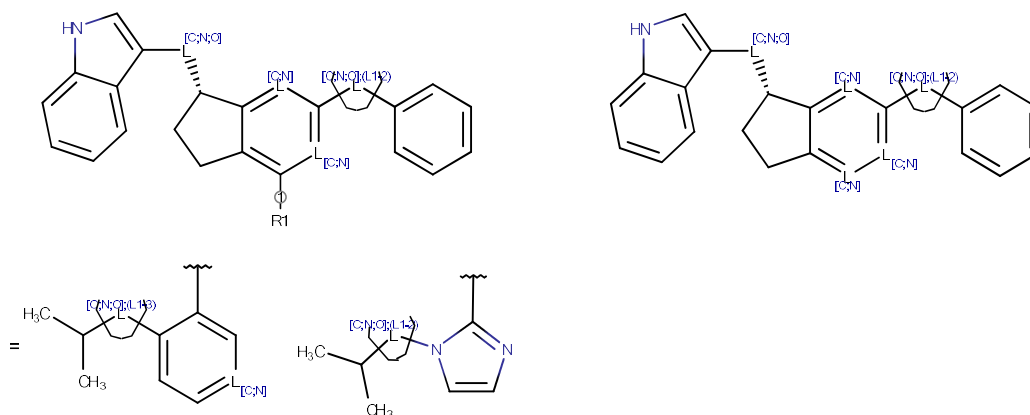
Figure 5.2: Markush structure showing the structures to be analysed for synthetic accessibility.

Small molecule *i*, *i*+1, *i*+4 and *i*, *i*+1 were designed by using a series of computational studies to analyse the ability of small molecule scaffolds to align with BRCA2. This gave an understanding of which small molecules were the most effective BRCA2 mimetics, and therefore be able to inhibit the BRCA2/PALB2 interaction (Figure 3.12). Due to the synthetic complexity, higher molecular weight, and less 'drug-like' structures of *i*, *i*+1, *i*+4, the synthesis of *i*, *i*+1 mimetics was carried out. Although the synthesis of pyridazine based *i*, *i*+1 mimetics was unsuccessful, the synthesis of hydrocarbon scaffolds had more success and allowed the development of methodology to functionalise in the *i* and *i*+1 positions. This resulted in the complete synthesis of one all-hydrocarbon BRCA2 mimetic, as well as progress towards the synthesis of several others. These compounds can then be tested in a biochemical assay in order to determine whether or not mimicking the BRCA2 helix has resulted in efficacy towards PALB2.
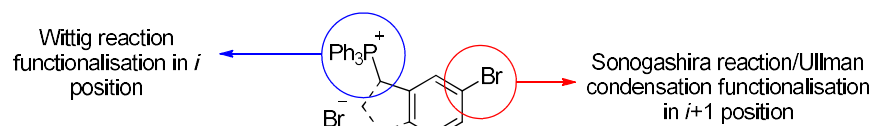
Figure 5.3: Functionalisation of hydrocarbon *i*, *i*+1 α-helical mimetic scaffolds was carried out using Wittig, Sonogashira and Ullman chemistry.

5.2   Further Work

This thesis has led to a number of areas where further work can be done. The synthesis of BRCA2 *i*, *i*+1 mimetics with a one-carbon *i* linker was not completed, and three synthetic steps are required in order to synthesise two more mimetics (Figure 5.4),

including one with a one-atom *i*+1 linker. This would be of value for understanding the potential SAR of these molecules based on the determined efficacy towards PALB2.
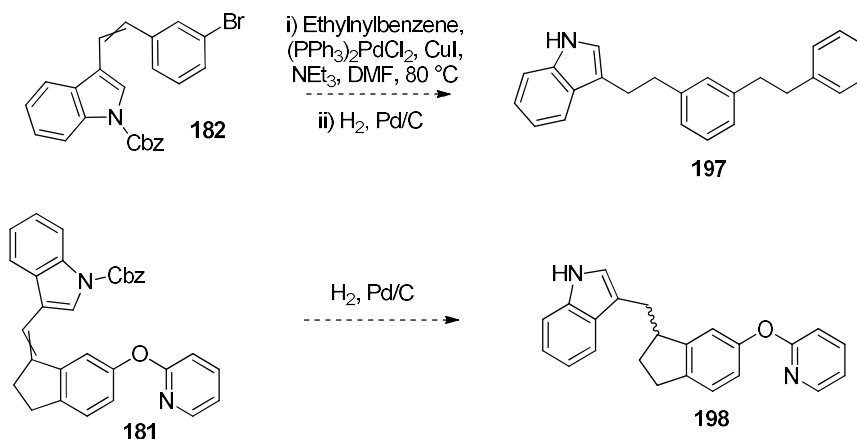
Figure 5.4: Final steps to the completion of mimetics with a one-carbon linker in the *i* position.

The synthesis and activity of BRCA2 mimetics with a two-atom *i* linker has not been explored. The synthetic route to these compounds from 3-(2-bromoethyl)-indole derivatives (**187**) has been designed, however the synthesis was not carried out (Figure 5.5). This synthetic route should hopefully be compatible with the established methodology and give further understanding of the SAR of these compounds.
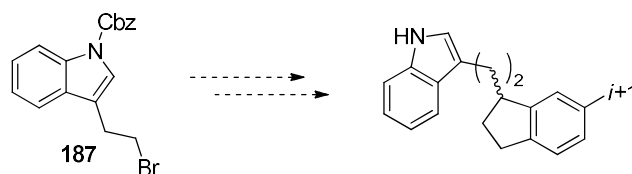
Figure 5.5: Synthesis of mimetics with a two-carbon linker in the *i* position.

Once the activity of the BRCA2 miniproteins has been established, it could be optimised by sequence modifications. This could be done either by the rational mutation of certain residues that have potential for picking up further interaction at the binding site,[89] or by using phage display methods on a library of generated sequences.[91] It is also possible to regulate the cellular uptake of miniproteins by incorporating basic residues such as arginine,[332] facilitating active transport into cells. This could be explored in the case of any miniprotein which demonstrates good activity against PALB2, and lead towards the development of a tool compound.

Finally, the hypothesis of whether incorporation of 'daclatasvir-like' side chains into small molecule mimetics would increase the solubility and bioavailability is yet to be tested. Mimetics with known activity against PPIs, such as Hamilton's terphenyl

compounds,[149] could be synthesised with daclatasvir-like side chains incorporated and PK profiles determined in order to fully test this hypothesis.

6   Experimental Information

6.1   General Information

All reactions were conducted under an atmosphere of nitrogen unless otherwise stated. Anhydrous solvents were used as purchased or were purified under nitrogen as follows using activated molecular sieves.

Thin layer chromatography was performed on glass plates pre-coated with Merck silica gel 60 F254. Visualisation was achieved with U.V. fluorescence (254 nm) or by staining with a phosphomolybdic acid dip or a potassium permanganate dip. Flash column chromatography was carried out using pre-packed columns filled with Aldrich silica gel (40-63 µm) on an ISCO Combiflash Rf, or a Biotage Isolera Prime.

Proton nuclear magnetic resonance spectra were recorded at 500 MHz on a Varian 500 spectrometer (at 30 °C), using residual isotopic solvent (CHCl$_3$, δ_H = 7.27 ppm, DMSO δ_H = 2.50 ppm, 3.33 ppm (H$_2$O), MeOH δ_H = 3.31 ppm, 4.87 ppm (H$_2$O)) as an internal reference. Chemical shifts are quoted in parts per million (ppm). Coupling constants (J) are recorded in Hertz (Hz).

Carbon nuclear magnetic resonance spectra were recorded at 125 MHz on a Varian 500 spectrometer and are proton decoupled, using residual isotopic solvent (CHCl$_3$, δ_C = 77.00 ppm, DMSO δ_C = 39.52 ppm, MeOH δ_C = 49.00 ppm) as an internal reference. Carbon spectra assignments are supported by DEPT editing and chemical shifts (δ_C) are quoted in ppm.

Infrared spectra were recorded on a Perkin Elmer FT-IR One spectrometer as either an evaporated film or liquid film on sodium chloride plates. Absorption maxima are reported in wave numbers (cm^{-1}). Only significant absorptions are presented in the data, with key stretches identified in brackets.

LCMS data was recorded on a Waters 2695 HPLC using a Waters 2487 UV detector and a Thermo LCQ ESI-MS. Samples were eluted through a Phenomenex Lunar 3µ C18 50 mm × 4.6 mm column, using acetonitrile and water (3:7 to 7:3) acidified by 0.01% formic acid.

HPLC analysis and purification was performed on an Agilent 1100 series HPLC spectrometer, using a Phenominex Luna 10 μ C18 150 mm × 15 mm column, eluted using acetonitrile and water (3: 7 to 7: 3)

High resolution mass spectrometry (HRMS) spectra were recorded on Bruker Daltonics Apex III ESI-MS, with an Apollo ESI probe using a methanol spray. Only molecular ions, fractions from molecular ions and other major peaks are reported as mass/charge (m/z) ratios.

Microwave reactions were performed using a Biotage Initiator 8+ microwave reactor.

6.1.1 General Software Packages

Calculation of molecular weight, pKa, logP, logD and NMR spectra predictions were carried out using MarvinSketch (ChemAxon Kft, 1031 Budapest, Hungary, https://www.chemaxon.com).

Drawing of chemical structures was carried out using ChemDraw Prime 15.1.0.144 (PerkinElmer Informatics Inc, Massachusetts 02451, USA, http://www.cambridgesoft.com/).

Visualisation and rendering of three-dimensional protein and small-molecule images was carried out using PyMOL 1.7.6.6 (Schrodinger, LLC, Cambridge, Cambridgeshire, CB1 2JD, https://www.pymol.org).

6.2 Chemistry Experimental Procedures

A. General Procedure for C-H activated cross coupling for compound without deprotection

The appropriate aryl dibromide (0.10 mmol), (*S*)-*tert*-butyl 2-(1-((2-(trimethylsilyl)ethoxy)methyl)-1*H*-imidazol-2-yl)pyrrolidine-1-carboxylate (**44**, 100 mg, 0.24 mmol), pivalic acid (6.1 mg, 0.06 mmol), K_2CO_3 (41 mg, 0.30 mmol) and Pd(ddpf)Cl_2 (7.3 mg, 0.01 mmol) were placed in a sealed reaction vial. The reaction vial was placed under reduced pressure then purged with nitrogen, repeating the process 5 times. The content of the sealed vial was dissolved in dimethylacetamide (1 mL), and the solvent was deoxygenated by bubbling with nitrogen for 20 minutes. The solution was then irradiated in a microwave at 100 °C for 4 hours. The resulting solution was quenched with water (15 mL), and extracted with EtOAc (2 × 10 mL). The organic phase was

washed with concentrated sodium bicarbonate solution (2 × 15 mL) water (2 × 15 mL), and brine (2 × 15 mL), before being concentrated under reduced pressure to afford a solid which was purified by mass-directed auto purification (30% to 70% acetonitrile in water over 25 minutes).

Dimethyl (2*S*,2'*S*)-1,1'-((2*S*,2'*S*)-2,2'-(5,5'-(biphenyl-4,4'-diyl)bis(1-((2-(trimethylsilyl) ethoxy)methyl)-1*H*-imidazole-5,2-diyl))bis(pyrrolidine-2,1-diyl))bis(3-methyl-1-oxobutane-2,1-diyl)dicarbamate (48)
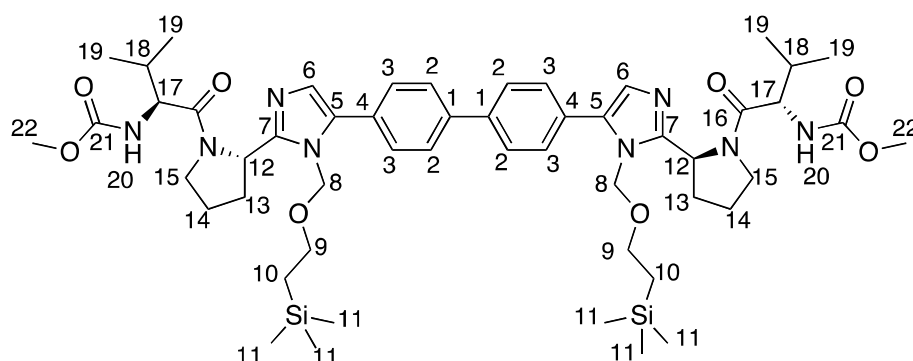
¹H NMR (500 MHz, CD₃OD) 7.79 (4H, d, J = 7.8 Hz, 3-C*H*), 7.59 (4H, d, J = 7.8 Hz, 2-C*H*), 7.04 (2H, s, 6-C*H*), 5.79 (2H, d, J = 10.9 Hz, 8-C*H*$_A$), 5.32 (2H, d, J = 10.9 Hz, 8-C*H*$_B$), 5.24 (2H, s, 12-C*H*), 4.22 (2H, d, J = 7.5 Hz, 17-C*H*), 4.11 − 4.03 (2H, m, 15-C*H*$_A$), 3.95 − 3.86 (2H, m, 15-C*H*$_B$), 3.66 (6H, s, 22-C*H*$_3$), 3.61 − 3.44 (4H, m, 10-C*H*$_2$), 2.37 (4H, m, 13-C*H*$_A$, 14-C*H*$_A$), 2.21 − 1.96 (6H, m, 13-C*H*$_B$, 14-C*H*$_B$, 18-C*H*), 0.90 (16H, m, 19-C*H*$_3$, 9-C*H*$_2$), -0.03 (18H, s, 11-Si(C*H*$_3$)$_3$); **¹³C NMR** (126 MHz, CD₃OD) 171.7 (16-*C*O), 158.0 (21-*C*O), 150.9 (7-N*C*N), 140.0 (1-*C*), 133.1 (5-*C*), 129.4 (2-*C*H), 128.8 (4-*C*), 127.0 (3-*C*H), 125.4 (6-*C*H), 72.1 (8-*C*H$_2$), 65.6 (10-*C*H$_2$), 58.3 (17-*C*H), 53.2 (12-*C*H), 51.3 (22-*C*H$_3$), 47.6 (15-*C*H$_2$), 31.6 (13-*C*H$_2$), 30.1 (18-*C*H), 24.7 (14-*C*H$_2$), 18.4 (19-*C*H$_3$), 17.30 (9-*C*H$_2$), -2.84 (11-Si(*C*H$_3$)$_3$); **IR** (neat, ν$_{max}$, cm^{-1}) 3276 (brw), 2959 (m), 1719 (s), 1631 (s), 1507 (m), 1441 (s), 1248 (s), 1189 (m), 1178 (s), 1033 (m), 820.21 (s), 693 (m); **HRMS (ESI):** calcd. for C₅₂H₇₈N₈NaO₈Si₂ [M+H]$^+$ 1021.5368, found 1012.5373; **[α]$_D$** -39.4 ° (c 0.6 in CHCl₃, T 21.0 °C)

Dimethyl ((2*S*,2'*S*)-((2*S*,2'*S*)-2,2'-(5,5'-(furan-2,5-diylbis(4,1-phenylene))bis(1-((2-(trimethylsilyl)ethoxy)methyl)-1*H*-imidazole-5,2-diyl))bis(pyrrolidine-2,1-diyl))bis(3-methyl-1-oxobutane-2,1-diyl))dicarbamate (199)
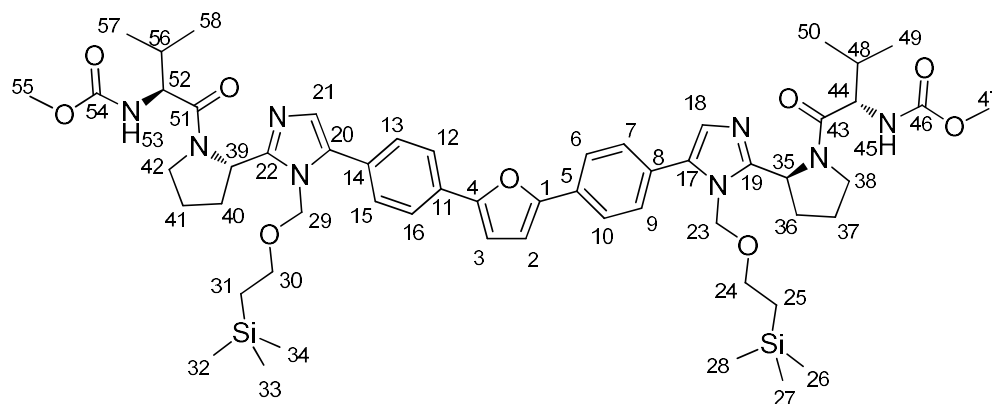
¹H NMR (500 MHz, CDCl₃) δ 7.80 (d, J = 8.0 Hz, 4H, 7/9/13/15-C*H*), 7.49 (d, J = 8.0 Hz, 4H, 6/10/12/16-C*H*), 7.05 (s, 2H, 2/3-C*H*), 6.82 (s, 2H, 18/21-C*H*), 5.79 (d, J = 10.6 Hz, 2H, 29/23-C*H*H), 5.29 (d, J = 9.1 Hz, 2H, 45/53-N*H*), 5.24 – 5.16 (m, 4H, 23/29-CH*H*, 35/39-C*H*), 4.35 (dd, J = 9.2, 6.4 Hz, 2H, 44/52-C*H*), 3.98 – 3.87 (m, 4H, 38/42-C*H₂*), 3.67 (s, 6H, 47/55-C*H₃*), 3.59 – 3.46 (m, 4H, 24/30-C*H₂*), 2.55 – 2.42 (m, 2H, 37/41-C*H*H), 2.39 – 2.26 (m, 4H, 36/40-C*H₂*), 2.15 – 1.91 (m, 4H, 37/41-CH*H*, 48/56-C*H*), 1.01 – 0.87 (m, 10H, 50/58-C*H₃*, 25/31-C*H₂*), 0.80 (d, J = 6.7 Hz, 6H, 49/57-C*H₃*), 0.00 (s, 18H, 26/27/28/32/33/34-CH₃); **¹³C NMR** (126 MHz, CDCl₃) δ 170.9 (43/51-*C*O), 157.1 (46/54-*C*O), 153.3 (8/14-*C*), 151.6 (1/4-*C*), 150.7 (19/22-*C*), 133.8 (5/11-*C*), 132.8 (17/20-*C*), 129.3 (7/9/13/15-*C*H), 124.0 (6/10/13/15-*C*H), 108.0 (18/21-*C*H), 88.3 (2/3-*C*H), 65.9 (24/30-*C*H₂), 57.6 (44/52-*C*H), 52.9 (35/39-*C*H), 52.2 (47/55-*C*H₃), 47.6 (38/42-*C*H₂), 31.8 (36/40-*C*H₂), 31.1 (48/56-*C*H), 25.2 (37/41-*C*H₂), 19.3 (49/57-*C*H₃), 18.0 (25/31-*C*H₂), 17.3 (50/58-*C*H₃), -1.4 (26/27/28/32/33/34-CH₃); **IR** (neat, ν$_{max}$, cm⁻¹) 2955 (m), 1719 (s), 1640 (s), 1503 (m), 1440 (m), 1249 (s), 1188 (m), 1082 (s), 1030 (w), 837 (s), 697 (w); **HRMS (ESI):** calcd. for C₅₆H₈₁N₈O₇Si₂ [M+H]⁺ 1065.5660, found 1065.5642; **[α]$_D$** -102.1 ° (c 0.3 in CHCl₃, T 22.5 °C)

Dimethyl ((2S,2'S)-((2S,2'S)-2,2'-(5,5'-(thiophene-2,5-diylbis(4,1-phenylene))bis(1-((2-(trimethylsilyl)ethoxy)methyl)-1*H*-imidazole-5,2-diyl))bis(pyrrolidine-2,1-diyl))bis(3-methyl-1-oxobutane-2,1-diyl))dicarbamate (200)
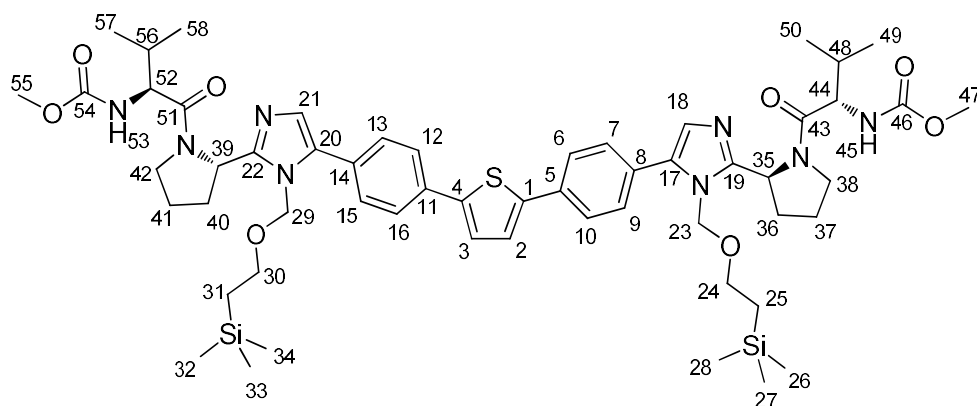
¹H NMR (500 MHz, CDCl₃) δ 7.71 (d, J = 7.9 Hz, 4H, 7/9/13/15-C*H*), 7.47 (d, J = 7.9 Hz, 4H, 6/10/12/16-C*H*), 7.38 (s, 2H, 2/3-C*H*), 7.12 (s, 2H, 18/21-C*H*), 5.81 (d, J = 10.6 Hz, 2H, 23/29-C*H*H), 5.29 – 5.14 (m, 6H, 23/29-CH*H*, 35/39-C*H*, 45/53-N*H*), 4.39 – 4.31 (m, 2H, 44/52-C*H*), 4.21 – 4.05 (m, 2H, 38/42-C*H*H), 3.99 – 3.84 (m, 2H, 38/42-CH*H*), 3.66 (s, 6H, 47/55-C*H₃*), 3.59 – 3.43 (m, 4H, 24/30-C*H₂*), 2.58 – 2.28 (m, 6H, 37/41-C*H*H, 36/40-C*H₂*), 2.17 – 1.93 (m, 4H, 37/41-CH*H*, 48/56-C*H*), 1.01 – 0.85 (m, 10H, 50/58-C*H₃*, 25/31-C*H₂*), 0.79 (d, J = 6.7 Hz, 6H, 49/57-C*H₃*), 0.00 (s, 18H, 26/27/28/32/33/34-C*H₃*); **¹³C NMR** (126 MHz, CDCl₃) δ 171.4 (43/51-*C*O), 157.1 (46/44-*C*O), 150.3 (19/22-*C*), 143.2 (1/4-*C*), 134.5 (5/11-C), 134.1 (8/14-*C*), 132.9 (17/20-*C*), 129.7 (6/10/12/16-*C*H), 125.9 (7/9/13/15-*C*H), 124.8 (2/3-*C*H), 118.4 (18/21-*C*H), 72.9 (23/29-*C*H₂), 66.4 (24/30-*C*H₂), 57.7 (44/52-*C*H), 52.8 (35/39-*C*H), 52.2 (47/55-*C*H₃), 47.7 (38/42-*C*H₂), 31.7 (36/40-*C*H₂), 30.6 (48/56-*C*H), 25.4 (37/41-*C*H₂), 19.4 (50/58-*C*H₃), 18.0 (25/31-*C*H₂), 17.2 (49/57-*C*H3), -1.4 (26/27/28/32/33/34-*C*H₃); **IR** (neat, ν$_{max}$, cm⁻¹) 3681 (brw), 2967 (m), 2844 (m), 1718 (m), 1637 (m), 1508 (w), 1455 (m), 1429 (m), 1055 (s), 1033 (s), 1015 (s), 838 (m); **HRMS (ESI):** calcd. for C₅₆H₈₁N₈O₆Si₂S [M+H]⁺ 1081.5431, found 1081.5449; **[α]$_D$** -69.1 ° (c 1.2 in CHCl₃, T 22.7 °C)

B. General Procedure for C-H activated cross coupling and subsequent SEM-removal

The appropriate aryl dibromide (0.10 mmol), (*S*)-*tert*-butyl 2-(1-((2-(trimethylsilyl)ethoxy)methyl)-1*H*-imidazol-2-yl)pyrrolidine-1-carboxylate (**7**, 100 mg, 0.24 mmol), pivalic acid (6.1 mg, 0.06 mmol), K₂CO₃ (41 mg, 0.30 mmol) and

Pd(ddpf)Cl$_2$ (7.3 mg, 0.01 mmol) were placed in a sealed reaction vial. The reaction vial was placed under reduced pressure then purged with nitrogen, repeating the process 5 times. The contents of the sealed vial were dissolved in dimethylacetamide (1 mL), and the solvent was deoxygenated by bubbling with nitrogen for 20 minutes. The solution was then irradiated in a microwave at 100 °C for 4 hours. The resulting solution was quenched with water (15 mL), and extracted with EtOAc (2 × 10 mL). The organic phase was washed with saturated sodium bicarbonate solution (2 × 15mL) water (2 × 15 mL), and brine (2 × 15 mL), before being concentrated under reduced pressure. The resulting solid material was then redissolved in dichloromethane (0.5 mL) before addition of trifluoroacetic acid (0.50 mL, 6.5 mmol). The solution was stirred under an inert atmosphere for 4 hours, before being diluted with dichloromethane (15 mL). The solution was then washed with saturated sodium bicarbonate solution (2 × 30 mL), with care taken to release any overpressure from gas formation. The organic phase was concentrated under reduced pressure to afford a solid which was purified by mass-directed auto purification.

Dimethyl ((2S,2'S)-((2S,2'S)-2,2'-(5,5'-([1,1'-biphenyl]-4,4'-diyl)bis(1H-imidazole-5,2-diyl))bis(pyrrolidine-2,1-diyl))bis(3-methyl-1-oxobutane-2,1-diyl))dicarbamate (10)
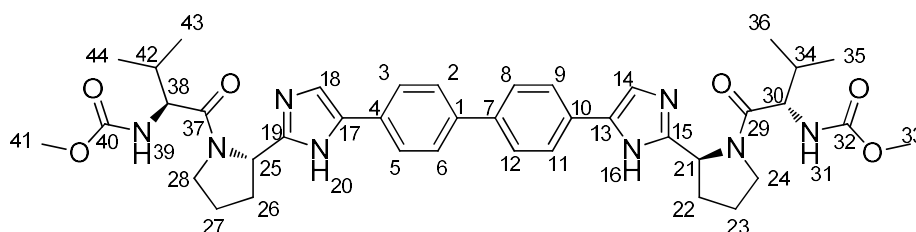
¹H NMR (500 MHz, CD$_3$OD) 7.73 (4H, d, J = 8.1 Hz, 3/5/9/11-C*H*), 7.66 (4H, d, J = 8.1 Hz, 2/6/8/12-C*H*), 7.37 (2H, s, 14/18-C*H*), 5.19 (2H, dd, J = 8.0, 6.3 Hz, 21/25-C*H*), 4.24 (2H, d, J = 7.4 Hz, 30/38-C*H*), 4.05 – 3.95 (2H, m, 24/28-C*H$_A$*), 3.92 – 3.84 (2H, m, 24/28-C*H$_B$*), 3.66 (6H, s, 33/41-C*H$_3$*), 2.41 – 2.17 (6H, m, 22/26-C*H$_2$*, 23/27-C*H$_A$*),), 2.11 – 2.04 (4H, m, 23/27-C*H$_A$*, 34/42-C*H*), 0.93 (12H, dd, J = 22.0, 6.8 Hz, 35/36/43/44-C*H$_3$*); **¹³C NMR** (126 MHz, CD$_3$OD) 172.0 (32/40-CO), 158.0 (15/19-C), 149.6 (32/40-CO), 139.0 (4/10-C), 136.1 (17/13-C), 130.9 (1/7-C), 126.6 (2/6/8/12-CH), 125.0 (3/5/9/11-CH), 116.2 (14/18-CH), 58.4 (21/25-CH), 54.8 (30/38-CH), 51.3 (33/41-CH$_3$), 47.4 (24/28-CH$_2$), 30.9 (22/26-CH$_2$), 30.1 (34/42-CH), 24.7 (23/27-CH$_2$), 18.4 (35/36/43/44-CH$_3$), 17.0 (35/36/43/44-CH$_3$); **IR** (neat, ν$_{max}$, cm^{-1}) 2964 (w), 2875 (w),

1705 (s), 1632 (s), 1446 (s), 1393 (s), 1258 (w), 1035 (m), 829 (m), 773 (m); **HRMS (ESI):** calcd. for $C_{40}H_{51}N_8O_6$ [M+H]$^+$ 739.3926, found 739.3937; **[α]$_D$** -170.0 ° (c 0.6 in CHCl$_3$, T 22.2 °C)

Dimethyl ((2*S*,2'*S*)-((2*S*,2'*S*)-2,2'-(5,5'-(1,4-phenylene)bis(1*H*-imidazole-5,2-diyl))bis(pyrrolidine-2,1-diyl))bis(3-methyl-1-oxobutane-2,1-diyl))dicarbamate (201)
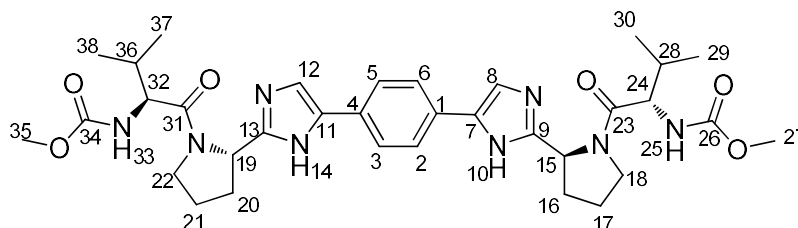
¹H NMR (500 MHz, CD$_3$OD) 7.88 (4H, s, 2/3/5/6-C*H*), 7.48 (2H, s, 8/12-C*H*), 5.29 – 5.19 (2H, m, 15/19-C*H*), 4.29 – 4.17 (2H, m, 24/32-C*H*), 4.14 – 3.97 (2H, m, 18/22-C*H$_A$*), 3.93 – 3.77 (2H, m, 18/22-C*H$_B$*), 3.66 (6H, s, 27/35-C*H$_3$*), 2.60 – 2.46 (16/20-C*H$_A$*, m, 2H), 2.32 – 1.95 (8H, m, 16/20-C*H$_B$*, 17/21-C*H$_2$*, 28/36-C*H*), 0.96 – 0.85 (12H, m, 29/30/37/38-C*H$_3$*); **¹³C NMR** (126 MHz, CD$_3$OD) 172.7 (23/31-*C*O), 158.1 (9/13-*C*), 149.4 (23/34-*C*O), 147.5 (7/11-*C*), 126.0 (1/4-*C*), 118.7 (2/3/5/6-*C*H), 115.3 (8/12-*C*H), 58.5 (24/32-*C*H), 53.7 (15/19-*C*H), 53.2 (18/22-*C*H$_2$*), 51.4 (27/35-*C*H$_3$*), 29.8 (28/36-*C*H), 24.9 (16/20-*C*H$_2$*), 18.4 (17/21-*C*H$_2$*), 16.7 (29/30/37/38-*C*H$_3$*); **IR** (neat, ν$_{max}$, cm^{-1}) 3284 (brw), 2962 (w), 1711 (s), 1625 (s), 1521 (s), 1444 (s), 1370 (w), 1264 (s), 1192 (m), 1098 (m), 1037 (m), 846 (m), 734 (s); **HRMS (ESI):** calcd. for $C_{34}H_{47}N_8O_6$ [M+H]$^+$ 663.3613, found 663.3608; **[α]$_D$** -288.6 ° (c 0.3 in CHCl$_3$, T 22.6 °C)

Dimethyl ((2*S*,2'*S*)-((2*S*,2'*S*)-2,2'-(5,5'-((*E*)-ethene-1,2-diylbis(4,1-phenylene))bis(1*H*-imidazole-5,2-diyl))bis(pyrrolidine-2,1-diyl))bis(3-methyl-1-oxobutane-2,1-diyl))dicarbamate (202)
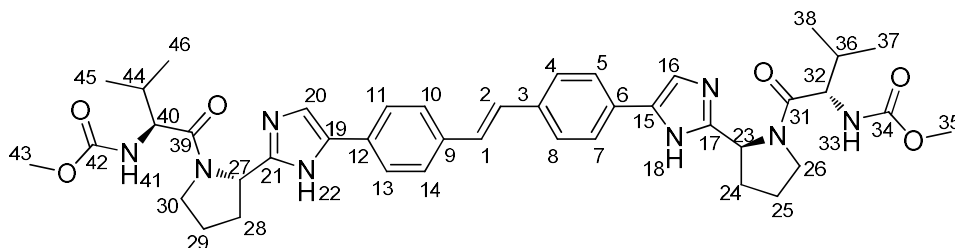
¹H NMR (500 MHz, CD$_3$OD) 8.16 (2H, s, 1/2-C*H*), 7.92 (4H, d, J = 8.4 Hz, 5/7/11/13-C*H*), 7.82 (4H, d, J = 8.4 Hz, 4/8/10/14-C*H*), 7.66 (2H, s, 16/20-C*H*), 5.24 (2H, dd, J =

8.1, 6.2 Hz, 23/27-C*H*), 4.26 (2H, d, J = 7.3 Hz, 32/40-C*H*), 4.09 – 4.02 (2H, m, 26/30-C*H*$_A$), 3.97 – 3.81 (2H, m, 26/30-C*H*$_B$), 3.67 (6H, s, 35/43-C*H*$_3$), 2.52 – 2.41 (2H, m, 24/28-C*H*$_A$), 2.32 – 2.02 (8H, m, 24/28-C*H*$_B$, 25/29-C*H*$_2$, 36/44-C*H*), 0.98 – 0.86 (12H, m, 37/38/45/46-C*H*$_3$); ^{13}**C NMR** (126 MHz, CD$_3$OD) 172.4 (31/39-*C*O), 163.2 (17/19-*C*), 158.1 (15/19-*C*), 149.6 (34/42-*C*O), 133.0 (3/9-*C*), 128.5 (6/12-*C*), 125.0 (1/2-*C*H), 123.7 (5/7/11/13-*C*H), 123.0 (4/8/10/14-*C*H), 115.8 (16/20-*C*H), 58.5 (23/27-*C*H), 54.4 (32/40-*C*H), 51.4 (35/43-*C*H$_3$), 47.5 (26/30-*C*H$_2$), 30.9 (24/28-*C*H$_2$), 30.0 (36/44-*C*H), 24.8 (25/29-*C*H$_2$), 18.4 (37/38/45/46-*C*H$_3$), 16.9 (37/38/45/46-*C*H$_3$); **IR** (neat, ν_{max}, cm^{-1}) 2966 (w), 2876 (w), 1701 (s), 1631 (s), 1521 (m), 1443 (m), 1264 (w), 1201 (s), 1132 (m), 1033 (w), 799 (w), 719 (m); **HRMS (ESI):** calcd. for C$_{42}$H$_{53}$N$_8$O$_6$ [M+H]$^+$ 765.4083, found 765.4118; **[α]$_D$** -538.3 ° (c 0.3 in CHCl$_3$, T 22.6 °C)

Dimethyl ((2*S*,2'*S*)-((2*S*,2'*S*)-2,2'-(5,5'-(naphthalene-2,6-diyl)bis(1*H*-imidazole-5,2-diyl))bis(pyrrolidine-2,1-diyl))bis(3-methyl-1-oxobutane-2,1-diyl))dicarbamate (203)
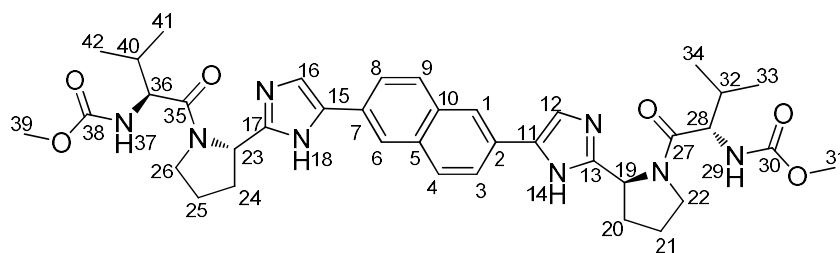
^{1}**H NMR** (500 MHz, CD$_3$OD) 7.64 – 7.54 (4H, m, 1/4/6/9-C*H*), 7.41 (2H, s, 12/16-C*H*), 7.20 – 7.13 (2H, m, 3/8-C*H*), 5.19 (2H, t, J = 7.0 Hz, 19/23-C*H*), 4.24 (2H, d, J = 7.3 Hz, 28/36-C*H*), 4.06 – 3.97 (2H, m, 22/26-C*H*$_A$), 3.92 – 3.83 (2H, m, 22/26-C*H*$_B$), 3.66 (6H, s, 31/39-C*H*$_3$), 2.43 – 2.32 (2H, m, 20/24-C*H*$_A$), 2.31 – 1.97 (8H, m, 20/24-C*H*$_B$, 21/25-C*H*$_2$, 32/40-C*H*), 0.97 – 0.85 (12H, m, 33/34/41/42-C*H*$_3$); ^{13}**C NMR** (126 MHz, CD$_3$OD) 172.2 (27/35-*C*O), 158.0 (13/17-*C*), 149.4 (30/38-*C*O), 136.7 (5/10-*C*), 135.6 (11/15-*C*), 130.0 (2/7-*C*), 127.8 (3/8-*C*H), 126.7 (1/6-*C*H), 124.8 (4/9-*C*H), 115.9 (12/16-*C*H), 58.4 (19/23-*C*H), 54.6 (28/36-*C*H), 51.4 (31/39-*C*H$_3$), 47.5 (22/26-*C*H$_2$), 30.9 (20/24-*C*H$_2$), 30.1 (32/40-*C*H), 24.7 (21/25-*C*H$_2$), 18.4 (33/34/41/42-*C*H$_3$), 16.9 (33/34/41/42-*C*H$_3$); **IR** (neat, ν_{max}, cm^{-1}) 3287 (brw), 2962 (w), 2875 (w), 1711 (s), 1628 (w), 1517 (m) 1443 (m), 1370 (w), 1264 (m), 1181 (m), 1131 (m), 1035 (m), 947 (w), 841 (m), 775 (m), 732 (s); **HRMS (ESI):** calcd. for C$_{38}$H$_{49}$N$_8$O$_6$ [M+H]$^+$ 713.3770, found 713.3779; **[α]$_D$** -222.1 ° (c 0.6 in CHCl$_3$, T 22.1 °C)

Dimethyl ((2*S*,2'*S*)-((2*S*,2'*S*)-2,2'-(5,5'-(oxybis(4,1-phenylene))bis(1*H*-imidazole-5,2-diyl))bis(pyrrolidine-2,1-diyl))bis(3-methyl-1-oxobutane-2,1-diyl))dicarbamate (204)
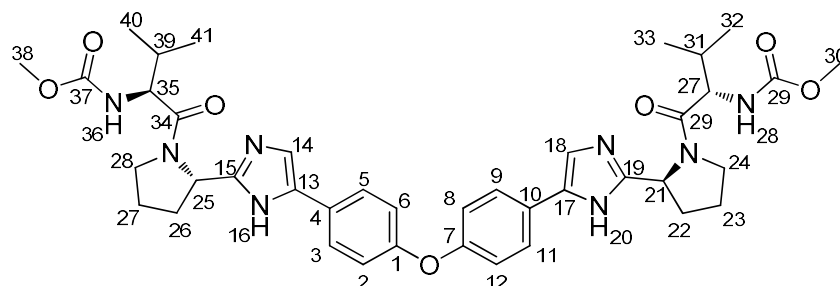
**¹H NMR** (500 MHz, CD₃OD) 8.13 (2H, s, 20/36-N*H*), 7.68 (4H, d, J = 8.0 Hz, 3/5/9/11-C*H*), 7.48 (2H, s, 14/18-C*H*), 7.08 (4H, d, J = 8.0 Hz, 2/6/8/12-C*H*), 5.20 (2H, t, J = 7.1 Hz, 21/25-C*H*), 4.24 (2H, d, J = 7.2 Hz, 27/35-C*H*), 4.09 – 3.99 (2H, m, 24/28-C*H*ₐ), 3.92 – 3.82 (2H, m, 24/28-C*H*ʙ), 3.66 (6H, s, 30/38-C*H*₃), 2.50 – 2.35 (2H, m, 22/26-C*H*ₐ), 2.31 – 1.97 (8H, m, 22/26-C*H*ʙ, 23/27-C*H*₂, 31/39-C*H*)), 0.97 – 0.87 (12H, m, 32/33/40/41-C*H*₃); **¹³C NMR** (126 MHz, CD₃OD) 172.3 (29/34-*C*O), 163.5 (15/19-*C*), 158.1 (29/37-*C*O), 156.9 (1/7-*C*), 149.1 (13/17-*C*), 134.9 (4/10-*C*), 126.6 (3/5/9/11-*C*H), 118.9 (2/6/8/12-*C*H), 114.8 (14/16-*C*H), 58.4 (21/25-*C*H), 54.3 (27/35-*C*H), 51.4 (30/38-*C*H₃), 47.5 (24/28-*C*H₂), 30.9 (22/26-*C*H₂), 30.0 (31/39-*C*H), 24.8 (23/27-*C*H₂), 18.4 (32/33/40/41-*C*H₃), 16.8 (32/33/40/41-*C*H₃); **IR** (neat, νₘₐₓ, cm⁻¹) 3294 (brw), 2962 (w), 2875 (w), 1712 (m), 1627 (s), 1497 (m), 1424 (m), 1237 (s), 1199 (m), 1198 9m), 1130 (m), 1033 (m), 798 (m), 733 (s), 701 (m); **HRMS (ESI):** calcd. for C₄₀H₅₁N₈O₇ [M+H]⁺ 755.3875, found 755.3882; **[α]ᴅ** -114.3 ° (c 0.6 in CHCl₃, T 22.0 °C)

Methyl *N*-[(2S)-1-[(2S)-2-{5-[4-(5-{2-[(2S)-1-[(2S)-2-[(methoxycarbonyl)amino]-3-methylbutanoyl]pyrrolidin-2-yl]-1*H*-imidazol-5-yl}-1,3-benzoxazol-2-yl)phenyl]-1H-imidazol-2-yl}pyrrolidin-1-yl]-3-methyl-1-oxobutan-2-yl]carbamate (205)
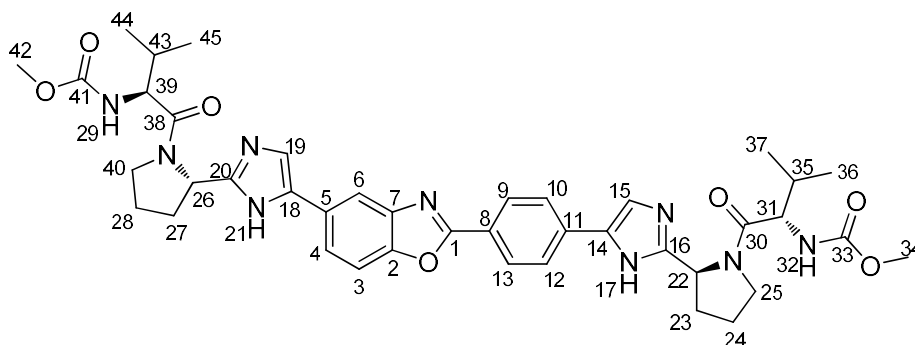
¹H NMR (500 MHz, CD₃OD) δ 8.33 – 8.20 (2H, m, 10/12-C*H*), 8.02 (1H, d, J = 8.3 Hz, 6-C*H*), 7.94 – 7.84 (2H, m, 9/13-C*H*), 7.83 – 7.70 (2H, m, 3-C*H*, 4-C*H*), 7.72 – 7.59 (2H, m, 15-C*H*, 19-C*H*), 5.27 – 5.18 (2H, m, 22/26-C*H*), 4.25 (2H, dd, J = 7.3, 7.3 Hz, 31/39-C*H*), 4.13 – 3.98 (2H, m, 25/40-C*H*H), 3.95 – 3.84 (2H, m, 25/40-CH*H*), 3.66 (6H, s, 34/42-C*H₃*), 2.57 – 2.35 (2H, m, 23/27-C*H*H), 2.32 – 2.02 (8H, m, 23/27-CH*H*, 24/28-C*H₂*, 35/43-C*H*), 0.99 – 0.89 (12H, m, 36/37/44/45-C*H₃*); **¹³C NMR** (126 MHz, CD₃OD₃) δ 173.7 (30/38-*C*O), 165.0 (1-*C*), 159.5 (33/41-*C*O), 152.9 (2-*C*), 152.4 (16/20-*C*), 152.0 (7-*C*), 151.5 (14/18-*C*), 150.3 (5-*C*), 136.6 (8-*C*), 136.0 (11-*C*), 129.3 (10/12-*C*H), 126.5 (9/13-*C*H), 124.5 (4-*C*H), 118.0 (19-*C*H), 117.5 (6-*C*H), 116.4 (15-*C*H), 112.5 (3-*C*H), 59.9 (31/39-*C*H), 55.4 (22/44-*C*H), 52.8 (34/42-*C*H₃), 49.0 (25/40-*C*H₂), 32.4 (24/28-*C*H₂), 31.5 (35/43-*C*H), 26.3 (23/27-*C*H₂), 19.8 (36/44-*C*H₃), 18.3 (37/45-*C*H₃); **IR** (neat, ν_{max}, cm^{-1}) 3284 (brw), 2966 (w), 2875 (w), 1699 (m), 1628 (s), 1517 (m), 1424 (m), 1368 (w), 1265 (m), 1180 (s), 1132 (s), 1039 (m), 825 (m), 798 (m), 732 (s); **HRMS (ESI):** calcd. for $C_{41}H_{50}N_9O_7$ [M+H]$^+$ 780.3828, found 780.3804; **[α]ᴅ** -446.6 ° (c 0.3 in CHCl₃, T 22.4 °C)

Methyl *N*-[(2S)-1-[(2S)-2-(5-{4-[(4-{2-[(2S)-1-[(2S)-2-[(methoxycarbonyl)amino]-3-methylbutanoyl]pyrrolidin-2-yl]-1*H*-imidazol-5-yl}phenyl)(methyl)carbamoyl]phenyl}-1H-imidazol-2-yl)pyrrolidin-1-yl]-3-methyl-1-oxobutan-2-yl]carbamate (206)
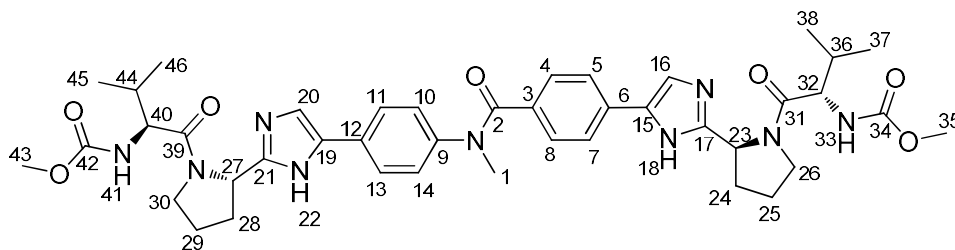
¹H NMR (500 MHz, CDCl₃) δ 7.51 – 7.39 (m, 2H, 5/7-C*H*), 7.35 – 7.26 (m, 2H, 10/14-C*H*), 7.26 – 7.20 (m, 2H, 4/8-C*H*), 7.15 – 7.04 (m, 2H, 16/20-C*H*), 6.98 (d, J = 8.2 Hz, 2H, 11/13-C*H*), 5.58 (s, 2H, 33/41-N*H*), 5.18 (s, 2H, 23/27-C*H*), 4.30 (dd, J = 8.1, 8.1 Hz, 2H, 32/40-C*H*), 3.90 – 3.77 (m, 2H, 26/30-C*H*H), 3.74 – 3.63 (m, 8H, 26/30-CH*H*, 35/43-C*H₃*), 3.46 (s, 3H, 1-C*H₃*), 2.35 – 2.11 (m, 4H, 24/25/28/29-C*H*H), 2.09 – 1.85 (m, 6H, 24/25/28/29-CH*H*, 36/44-C*H*), 0.88 – 0.77 (m, 12H, 37/38/45/46-C*H₃*); **¹³C NMR** (100 MHz, CDCl₃) δ 172.6 (31/39-*C*O), 170.2 (2-*C*O), 157.1 (34/42-*C*NO), 148.5 (17/21-*C*N₂), 143.6 (9-*C*), 138.5 (15/19-*C*N), 134.3 (3-*C*), 132.3 (6-*C*), 129.3 (4/8-*C*H), 127.1 (11/13-*C*H), 125.5 (5/7-*C*H), 124.6 (12-*C*), 123.9 (10/14-*C*H), 115.3 (16/20-*C*H), 57.7

(32/40-*CH*), 54.2 (23/27-*CH*), 52.3 (35/43-*CH*$_3$), 47.8 (26/30-*CH*$_2$), 38.4 (1-*CH*$_3$), 30.8 (36/44-*CH*), 28.4 (24/28-*CH*$_2$), 25.2 (25/29-*CH*$_2$), 19.2 (38/46-*CH*$_3$), 17.437/45-*CH*$_3$); **IR** (neat, v_{max}, cm^{-1}) 3227 (brw), 2958 (w), 1698 (s), 1630 (s), 1532 (m), 1445 (m), 1396 (m), 1203 (m), 1137 (m), 1038 (w), 847 (w); **HRMS (ESI):** calcd. for C$_{41}$H$_{50}$N$_9$O$_7$ [M+H]$^+$ 796.4141, found 796.4240; **[α]$_D$** -243.3 ° (c 0.6 in CHCl$_3$, T 22.4 °C)

(S)-2-(Methoxycarbonylamino)-3-methylbutanoic acid (17)[203]
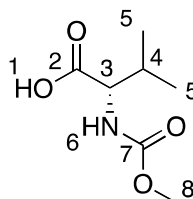
To a stirred solution of L-valine (20.0 g, 171 mmol) in water (150 mL) and toluene (150 mL) was added NaOH (13.7 g, 341 mmol). The solution was cooled to 0 °C, and methyl chloroformate (14.5 mL, 188 mmol) was added dropwise over 10 minutes. The solution was allowed to warm to room temperature and stirred for 16 hours. The solution was diluted with EtOAc (200 mL) to dissolve any precipitate, and the organic phase was separated. The aqueous phase was acidified with aqueous HCl (1 M, 100 mL), and extracted with EtOAc (2 × 200 mL). The combined organic phases were concentrated under reduced pressure, and the resultant white paste dissolved in warm water (500 mL). The solution was freeze dried, and the resulting white powder dried under reduced pressure at 70 °C for 16 hours to give the product as a white powder (25.7 g, 84%).

^{1}H NMR (500 MHz, DMSO-*d*$_6$) 12.54 (1H, brs, 1-COO*H*), 7.27 (1H, d J = 8.4 Hz, 6-N*H*), 3.84 (1H, dd, J = 8.4, 5.9 Hz, 3-C*H*), 3.53 (3H, s, 8-C*H*$_3$), 2.08 – 1.96 (1H, m, 4-C*H*), 0.97 – 0.77 (6H, m, 5-CH(C*H*$_3$)$_2$); **^{13}C NMR** (126 MHz, DMSO-*d*$_6$) 173.7 (2-*C*O), 157.4 (7-*C*O), 60.0 (3-*C*H), 51.8 (8-*C*H$_3$), 30.0 (4-*C*H), 19.6 (5-CH(*C*H$_3$)$_2$), 18.4 (5-CH(*C*H$_3$)$_2$); **IR** (neat, v_{max}, cm^{-1}) 3397 (m), 2965 (m), 1732 (s), 1664 (s), 1551 (s), 1395 (m), 1306 (s), 1255 (s), 1179 (s), 1104 (s), 1019 (s), 892 (s), 778 (s), 735 (m); **HRMS (ESI):** calcd. for C$_7$H$_{13}$NNaO$_4$ [M+Na]$^+$ 198.0737, found 198.0738

(*S*)-2-(Pyrrolidin-2-yl)-1H-imidazole (19)[203]
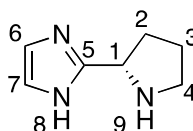
To a solution of (*S*)-*tert*-butyl 2-(1*H*-imidazol-2-yl)pyrrolidine-1-carboxylate (**42**, 10.0 g, 42.1 mmol) in dichloromethane (200 mL) at 0 °C, was added dropwise trifluoroacetic acid (24 mL, 316 mmol). The solution was allowed to warm to room temperature, and stirred for a further 16 hours, before being concentrated under reduced pressure. The crude material was then purified by adsorption onto an ion exchange column in methanol (SCX-2, 2 × 50 g silica), followed by elution by ammonia in methanol (2.0 M), and concentration under reduced pressure to give amine **19** as a brown oil (5.6 g, 97%).

¹H NMR (500 MHz, CDCl₃) 7.44 (1H, s, 9-N*H*), 6.89 (2H, s, 6,7-C*H*), 4.25 (1H, t, J = 7.4 Hz, 1-C*H*), 3.02-2.95 (1H, m, 4-C*H*$_A$), 2.93-2.86 (1H, m, 4-C*H*$_B$), 2.17-2.06 (1H, m, 2-C*H*$_A$), 2.03-1.94 (1H, m, 2-C*H*$_B$), 1.86-1.69 (2H, m, 3-C*H*$_2$); **¹³C NMR** (126 MHz, CDCl₃) 150.1 (5-*C*), 121.7 (6,7-*C*H), 56.1 (1-*C*H), 46.5 (4-*C*H$_2$), 32.0 (2-*C*H$_2$), 25.5 (3-*C*H$_2$); **IR** (neat, ν$_{max}$, cm^{-1}) 2971 (brm), 1648 (m), 1531 (m), 1408 (s), 1342 (m), 1065 (m), 748 (s); **HRMS (ESI):** calcd. for C$_7$H$_{12}$N$_3$ [M+H]$^+$ 138.1026, found 138.1026

Methyl (*S*)-1-(((*S*)-2-(1*H*-imidazol-2-yl)pyrrolidin-1-yl)-3-methyl-1-oxobutan-2-ylcarbamate (38)
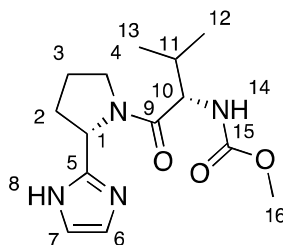
To a solution of (*S*)-2-(methoxycarbonylamino)-3-methylbutanoic acid (**17**, 1.25 g, 6.98 mmol) in dichloromethane (20 mL) and a drop of dimethylformamide at 0 °C was added dropwise oxalyl chloride (0.67 mL, 7.0 mmol). The solution was allowed to warm to room temperature, and stirred for 3 hours. This solution was then added dropwise to a solution of (*S*)-2-(pyrrolidin-2-yl)-1*H*-imidazole **19** (870 mg, 6.35 mmol) and pyridine (1.9 mL, 19 mmol) in dichloromethane (30 mL) at 0 °C. The solution was allowed to warm to room temperature and stir for 16 hours. The solution was then neutralized by the slow addition of ammonia (2 M) in isopropanol (20 mL) dropwise. The resultant precipitate was removed by vacuum filtration, and the filtrate was concentrated under reduced pressure to give a brown solid, which was lyophilized from ice to remove residual water and pyridine, and afforded the product as a brown, glass foam (1.82 g, 97%).

¹H NMR (500 MHz, CDCl₃) 11.50 (1H, s, 8-N*H*), 7.21 (1H, d, J = 8.4 Hz, 14-N*H*), 6.92 (1H, s, 6-C*H*), 6.72 (1H, s, 7-C*H*), 5.05 (1H, dd, J = 7.7, 3.5 Hz, 1-C*H*), 4.04-3.98 (1H, m, 10-C*H*), 3.80-3.64 (2H, m, 4-C*H₂*), 3.53 (3H, s, 16-C*H₃*), 2.19-1.80 (5H, m, 2-C*H₂*, 3-C*H₂*, 11-C*H*), 0.90-0.75 (6H, m, 12-C*H₃*, 13-C*H₃*); **¹³C NMR** (500 MHz, CDCl₃) 170.8 (9-*C*O), 157.2 (15-*C*O), 148.9 (5-*C*H), 127.3 (6-*C*H), 115.9 (7-*C*H), 58.4 (10-*C*H), 54.4 (1-*C*H), 51.9 (16-*C*H₃), 47.2 (4-*C*H₂), 30.9 (3-*C*H₂), 30.2 (11-*C*H), 24.7 (2.*C*H₂), 19.4 (12-*C*H₃, 13-*C*H₃), 18.9 (12-*C*H₃, 13-*C*H₃); **IR** (neat, v_{max}, cm⁻¹) 3151 (brw), 2970 (w), 1703 (m), 1632 (m), 1520 (m), 1446 (m), 1362 (m), 1520 (m), 1446 (m), 1362 (m), 1261 (m), 1183 (m), 1099 (m), 835 (s), 741 (s); **HRMS (ESI):** calcd. for $C_{14}H_{23}N_4O_3$ [M+H]⁺ 295.1765, found 295.1761

(*S*)-*tert*-Butyl 2-(hydroxymethyl)pyrrolidine-1-carboxylate (40)[206]
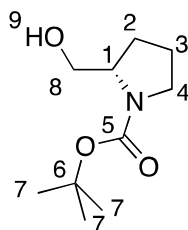
To a solution of (*S*)-1-(3,3-dimethylbutanoyl)pyrrolidine-2-carboxylic acid **37** (20.0 g, 92.9 mmol) in THF (200 mL) at -10 °C was added dropwise BH₃.SMe₂ solution (2.0 M in THF, 60.4 mL). The solution was allowed to warm to room temperature and stirred for a further 16 hours. After cooling to 0 °C, MeOH (80 mL) was added dropwise to quench the reaction. The solution was then concentrated under reduced pressure, and the resulting yellow oil was dissolved in EtOAc (150 mL), washed with saturated aqueous NaHCO₃ solution (2 × 100 mL) and brine (2 × 100mL). The organic phase was then dried over magnesium sulfate and concentrated under reduced pressure, to give alcohol **40** as a pale yellow oil (18.2 g, 98%).

¹H NMR (500 MHz, CDCl₃) 4.69 (1H, s, 9-O*H*), 3.91 (1H, s, 1-C*H*), 3.66 (2H, s, 8-C*H₂*), 3.44-3.35 (1H, m, 4-C*H₄*), 3.31-3.22 (1H, m, 4-C*H₈*), 2.00-1.90 (1H, m, 2-C*H₄*), 1.88-1.68 (2H, m, 3- C*H₂*), 1.58-1.48 (1H, m, 2-C*H₈*), 1.42 (9H, s, 7-C*H₃*); **¹³C NMR** (126 MHz, CDCl₃) 156.9 (5-*C*O), 80.0 (6-*C*(CH₃)₃), 67.3 (8-*C*H₂), 60.1 (1-*C*H), 47.5 (4-*C*H₂), 28.6 (2-*C*H₂), 28.4 (7-C(*C*H₃)₃) 23.7 (3-*C*H₂). **IR** (neat, v_{max}, cm⁻¹) 3405 (brm), 2974 (m), 2877 (w), 1666 (s), 1391 (s), 1252 (w), 1163 (s), 1105 (m), 1049 (m), 772 (m); **HRMS**

(ESI): calcd. for $C_{10}H_{19}NNaO_3$ [M+Na]$^+$ 224.1257, found 224.1258; **[α]$_D$** -49.9 ° (*c* 1.3 in CHCl$_3$, T 21.0 °C)

(*S*)-*tert*-Butyl 2-formylpyrrolidine-1-carboxylate (41)[206]
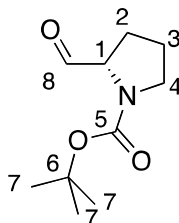
Swern Oxidation:

To a solution of oxalyl chloride (8 mL, 92.6 mmol) in dry dichloromethane (150 mL) at -78 °C under nitrogen was added dropwise DMSO (13.2 mL, 185.2 mmol). To this solution was added dropwise (*S*)-*tert*-butyl 2-(hydroxymethyl)pyrrolidine-1-carboxylate **40** (17.4 g, 84.2 mmol), and allowed to stir for 30 minutes. Triethylamine (60 mL, 421 mmol) was added to the solution dropwise over 30 minutes, before being allowed to warm to room temperature. The reaction was then quenched with deionized water (150 mL), and stirred for a further 15 minutes. The layers were allowed to separate, before extracting the aqueous phase with dichloromethane (150 mL). The combined organic phases were washed with saturated aqueous NaHCO$_3$ solution (2 × 150 mL) and brine (2 × 150 mL), before being dried over magnesium sulfate and concentrated under reduced pressure to yield aldehyde **41** as a pale orange oil (16.7 g, 99% yield).

Aerobic Oxidation:

To a stirred solution of *tert*-butyl (2*S*)-2-(hydroxymethyl)pyrrolidine-1-carboxylate **40** (6.06 g, 30.1 mmol) in acetonitrile (100 mL) at was added (2,2,6,6-tetramethylpiperidin-1-yl)oxidanyl (TEMPO, 235 mg, 1.51 mmol), 2,2'-bipyridyl (235 mg, 1.51 mmol), and tetrakisacetonitrile copper(I) triflate (567 mg, 1.51 mmol). 1-methylimidazole (0.24 mL, 3.01 mmol) was then added to the solution, and stirred for 36 hours in an open flask, replacing solvent as it evaporated. The solution was then diluted with water (150 mL) and extracted with diethyl ether (3 × 150 mL). The combined organic phases were washed with brine (2 × 100 mL) and dried over MgSO$_4$, before being concentrated to dryness under reduced pressure. The crude material was then passed through a silica plug (approx. 30 g silica), eluted with a 1:1 mixture of petrol ether:diethyl ether, and concentrated under reduced pressure to give the product as a light orange oil (5.40 g, 90%).

¹H NMR (500 MHz, CDCl₃) 9.53 (0.4H, s, 8-C*H*O), 9.44 (0.6H, s, 8-C*H*O), 4.17 (0.4H, t, J = 5.6 Hz, 1- C*H*), 4.03 (0.6H, t, J = 5.6 Hz, 1- C*H*), 3.63-3.37 (2H, m, 4-C*H₂*), 2.15-1.73 (4H, m, 2-C*H₂*, 3-C*H₂*), 1.49-1.36 (9H, m, 7-C(C*H₃*)₃); **¹³C NMR** (126 MHz, CDCl₃) 200.5 (8-*C*HO), 200.3 (8-*C*HO), 154.9 (5-*C*O), 153.9 (5-*C*O), 80.5 (7-*C*(CH₃)₃), 80.1 (7-*C*(CH₃)₃), 65.0 (1-*C*H), 64.8 (1-*C*H), 46.8 (4-*C*H₂), 46.7 (4-*C*H₂), 28.2 (C(*C*H₃)₃), 27.9 (2-*C*H₂), 26.7 (2-*C*H₂), 24.6 (4-*C*H₂), 23.9 (4-*C*H₂); **IR** (neat, v_max, cm⁻¹) 2976 (m), 2880 (w), 1735 (m), (1687 (s), 1390 (s), 1256 (m), 1160 (s), 1119 (m), 911 (w), 858 (w), 772 (m); **[α]_D** -94.5 ° (*c* 0.6 in CHCl₃, T 24.2 °C)

(S)-*tert*-Butyl 2-(1*H*-imidazol-2-yl)pyrrolidine-1-carboxylate (42)[203]
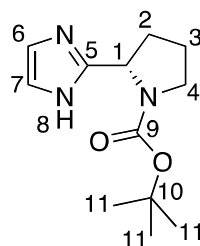
To a solution of (S)-*tert*-butyl 2-formylpyrrolidine-1-carboxylate **41** (16.7 g, 84.2 mmol) in methanol (150 mL) and glyoxal solution (40 % wt. in water, 35.6 mL, 353.0 mmol) at 0 °C was added dropwise ammonia solution (35 % wt. in water, 31.0 mL, 555 mmol). The solution was allowed to warm to room temperature and stirred for 15 hours. The solution was concentrated under reduced pressure, and the resultant residue was extracted with EtOAc (2 × 150 mL). Any precipitate was removed by filtration, and the organic phase was washed with concentrated aqueous NaHCO₃ solution (2 × 100 mL), and brine (2 × 100 mL). The solution was then dried over magnesium sulfate and concentrated under reduced pressure to give imidazole **42** as a brown solid (19.4 g, 97%).

¹H NMR (500 MHz, CDCl₃) 6.95 (2H, s, 6/7-C*H*), 4.94 (1H, dd, J = 7.6, 2.6 Hz, 1-C*H*), 3.43 - 3.34 (1H, m, 3-C*H₂*), 2.95 - 2.88 (1H, m, 2-C*H_A*), 2.17 - 2.08 (2H, m, 2-C*H_B*, 4-C*H_A*), 1.95 (1H, ddd, J = 8.9, 6.0, 4.2 Hz, 4-C*H_B*), 1.48 (9H, s, 10-C(C*H₃*)₃); **¹³C NMR** (126 MHz, CDCl₃) 156.4 (8-*C*O), 148.9 (5-N*C*NH), 80.2 (9-*C*(CH₃)₃), 53.9 (1-*C*H), 47.2 (3-*C*H₂), 28.5 (10-C(*C*H₃)₃), 28.1 (2-*C*H₂), 24.8 (4-*C*H₂); **IR** (neat, v_max, cm⁻¹) 2977 (w), 2877 (w), 1687 (s), 1390 (s), 1161 (s), 1101 (s), 966 (w), 890 (m), 760 (m), 694 (w); **HRMS (ESI):** calcd. for C₁₂H₂₀N₃O₂ [M+H]⁺ 238.1550, found 238.1551; **[α]_D** -159.4 ° (*c* 1.4 in CHCl₃, T 24.6 °C)

(*S*)-*tert*-Butyl 2-(1-((2-(trimethylsilyl)ethoxy)methyl)-1*H*-imidazol-2-yl)pyrrolidine-1-carboxylate (43)[216]
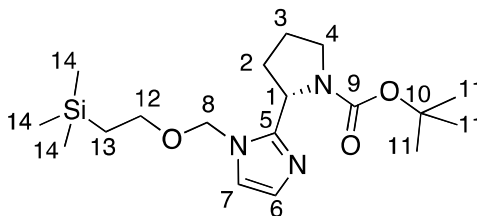
To a solution of (*S*)-*tert*-butyl 2-(1*H*-imidazol-2-yl)pyrrolidine-1-carboxylate (**42**, 2.50 g, 10.5 mmol) in dry THF (100 mL) at 0 °C under an atmosphere of nitrogen was added portion-wise NaH (60% in mineral oil, 0.50 g, 12 mmol), and stirred for one hour. 2-(Trimethylsilyl)ethoxymethyl chloride (2.1 mL, 12 mmol) was added to the solution dropwise, which was then allowed to warm to room temperature and stir for a further 22 hours. The resulting solution was quenched with deionized water (150 mL) and extracted with EtOAc (2 × 150 mL). The combined organic phases were washed with water (150 mL), and brine (2 × 150 mL), before being dried over MgSO₄ and concentrated under reduced pressure to give a brown oil. The crude material was purified by flash chromatography (50 g silica, EtOAc:Pet. Ether, 25% to 100%) to give the protected imidazole **43** as a pale brown oil (2.61 g, 67%).

¹H NMR (500 MHz, CDCl₃) 7.18 - 7.09 (1H, m, 7-C*H*), 6.78 (1H, d, J = 5.7 Hz, 6-C*H*), 5.60 (0.4H, d, J = 10.9 Hz, 8-C*H₂*), 5.37 (0.6H, d, J = 10.9 Hz, 8-C*H₂*), 5.25 (1H, d, J = 10.9 Hz, 8-C*H₂*), 4.97 - 4.82 (1H, m, C*H*), 3.50 - 3.32 (4H, m, 4-C*H₂*, 12-C*H₂*), 2.28 - 2.02 (2H, m, 2-C*H*$_A$, 3-C*H*$_A$), 1.90 - 1.75 (2H, m, 2-C*H*$_B$, 3-C*H*$_B$), 1.36 (4H, s, 11-C*H₃*), 1.12 (5H, s, 11-C*H₃*), 0.91 – 0.77 (2H, m, 13-C*H₂*), -0.04 (9H, s, 14-C*H₃*); **¹³C NMR** (126 MHz, CDCl₃) 153.5 (9-*C*O), 150.1 (5-*C*), 126.9 (6-CH), 120.4 (7-*C*H), 78.5 (10-*C*), 74.4 (8-C*H₂*), 65.5 (12-C*H₂*), 52.5 (1-*C*H), 46.7 (4-CH), 32.8 (3-*C*H₂), 28.4 (11-*C*H₃), 23.5 (2-*C*H₃), 17.7 (13-*C*H₂), -0.95 (14-*C*H₃); **IR** (neat, ν$_{max}$, cm⁻¹) 2975 (brw), 1683 (s), 1478 (w), 1386 (s), 1248 (m), 1159 (m), 1103 (s), 986 (w), 913 (m), 883 (s), 761 (m), 729 (m), 697 (m), 664 (w).

Methyl (*S*)-3-methyl-1-oxo-1-(((*S*)-2-(1-((2-(trimethylsilyl)ethoxy)methyl)-1*H*-imidazol-2-yl)pyrrolidin-1-yl)butan-2-ylcarbamate (44)
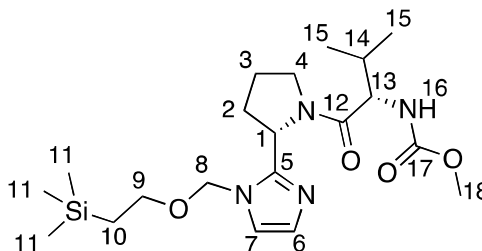
To a stirred suspension of NaH (305 mg, 7.63 mmol) in dry THF (100 mL) at 0 °C, was added dropwise a solution of methyl (*S*)-1-(((*S*)-2-(1*H*-imidazol-2-yl)pyrrolidin-1-yl)-3-methyl-1-oxobutan-2-ylcarbamate (**38**, 1.35 g, 4.54 mmol) in dry THF (20 mL), and the resulting solution was allowed to stir for one hour. To this solution was added dropwise 2-(trimethylsilyl)ethoxymethyl chloride (1.35 mL, 7.63 mmol) in dry THF (5 mL), before warming to room temperature and stirring for 16 hours. The reaction was quenched with deionized water (100 mL), and extracted with EtOAc (2 × 100 mL). The combined organic phases were washed with water (100 mL) and brine (2 × 100 mL), before being dried over $MgSO_4$ and concentrated under reduced pressure to give an orange oil (2.03 g). The crude material was purified by flash chromatography (25 g silica, EtOAc:petroleum ether, 30% to 100%), to give protected imidazole **44** (0.91 g, 47%) as a pale brown oil.

^{1}H NMR (500 MHz, DMSO-d_6) 7.17 (1H, s, 16-N*H*), 7.10 (1H, d, J = 1.2 Hz, 7-C*H*), 6.77 (1H, d, J = 1.2 Hz, 6-C*H*), 5.71 (1H, d, J = 10.9 Hz, 8-C*H$_A$*), 5.24 (1H, d, J = 10.9 Hz, 8-C*H$_B$*), 5.11 (1H, dd, J = 8.2, 3.8 Hz, 1-C*H*), 3.99 (1H, dd, J = 8.4, 8.4 Hz, 13-C*H*), 3.88 - 3.80 (1H, m, 4-C*H$_A$*), 3.74 – 3.66 (1H, m, 4-C*H$_B$*), 3.56 - 3.45 (5H, m, 9-C*H$_2$*, 18-C*H$_3$*), 2.37 - 2.27 (1H, m, 3-C*H$_A$*), 2.14 - 2.05 (1H, m, 2-C*H$_A$*), 1.98 - 1.88 (2H, m, 2-C*H$_A$*, 3-C*H$_B$*), 1.84 - 1.73 (1H, m, 14-C*H*), 0.92 - 0.69 (8H, m, 10-C*H$_2$*, 15-C*H$_3$*), -0.04 (9H, s, 11-C*H$_3$*); **^{13}C NMR** (126 MHz, DMSO d_6) 170.4 (12-*C*O), 157.2 (17-*C*O), 149.3 (5-*C*), 127.0 (6-*C*H), 120.3 (7-*C*H), 74.6 (8-*C*H$_2$), 65.4 (9-*C*H$_2$), 58.3 (13-*C*H), 51.9 (1-*C*H), 51.8 (18-*C*H$_3$), 47.2 (4-*C*H$_2$), 31.5 (2-*C*H$_2$), 30.3 (14-*C*H), 24.9 (3-*C*H$_2$), 19.1 (15-*C*H$_3$), 17.7 (10-*C*H$_2$), -0.9 (11-*C*H$_3$); **IR** (neat, v_{max}, cm^{-1}) 3261 (w), 2955 (w), 1716 (s), 1619 (s), 1522 (m), 1444 (m), 1248 (m), 1091 (s), 1033 (m), 834 (s), 725 (m); **HRMS (ESI):** calcd. for $C_{20}H_{36}N_4NaO_4$ [M+H]$^+$ 447.2398, found 447.2395; **[α]$_D$** 32.1 (c 0.6 in CHCl$_3$, T 20.5 °C).

(2*S*,2'*S*)-*tert*-Butyl   2,2'-(5,5'-(biphenyl-4,4'-diyl)bis(1-((2-(trimethylsilyl)ethoxy)-methyl)-1*H*-imidazole-5,2-diyl))dipyrrolidine-1-carboxylate (45)
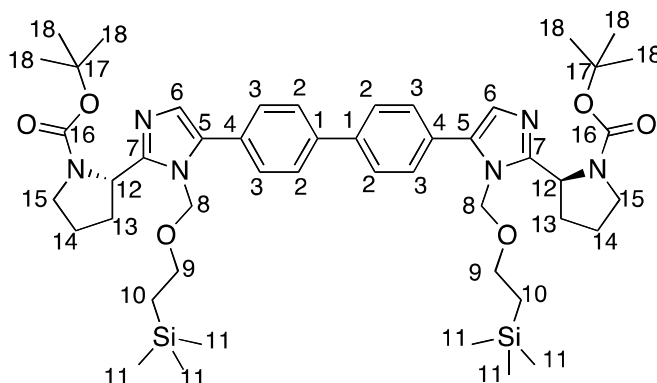
4,4'-Dibromobiphenyl (**18**, 178 mg, 0.57 mmol), (*S*)-*tert*-butyl 2-(1-((2-(trimethylsilyl)ethoxy)methyl)-1*H*-imidazol-2-yl)pyrrolidine-1-carboxylate (**43**, 500 mg, 1.36 mmol), pivalic acid (35 mg, 0.34 mmol), K_2CO_3 (236 mg, 1.71 mmol) and Pd(ddpf)Cl$_2$ (42 mg, 0.057 mmol) were placed in a sealed reaction vial. The reaction vial was placed under reduced pressure then purged with nitrogen, repeating the process 5 times. The content of the sealed vial was dissolved in dimethylacetamide (5 mL), and the solvent was deoxygenated by bubbling with nitrogen for 20 minutes. The solution was then irradiated in a microwave reactor at 100 °C for 4 hours. The resulting solution was quenched with water (50 mL), and extracted with EtOAc (2 × 35 mL). The organic phase was washed with brine (3 × 50 mL), then passed through a silica plug, eluted with MeOH:CH$_2$Cl$_2$ (1:9). The filtrate was concentrated under reduced pressure to give a dark brown oil (0.62 g). The crude material was purified by flash chromatography (25 g silica, MeOH:CH$_2$Cl$_2$, 1% to 8%) to give compound **45** as a brown solid (382 mg, 76%).

^{1}H NMR (500 MHz, CDCl$_3$) 7.75 - 7.63 (4H, m, 3-C*H*), 7.59 - 7.47 (4H, m, 2-C*H*), 7.13 - 7.05 (2H, m, 6-C*H*), 5.75 (1H, d, J = 10.8 Hz, 8-C*H$_2$*), 5.38 (1H, d, J = 10.8 Hz, 8-C*H$_2$*), 5.25 - 5.18 (2H, m, 8-C*H$_2$*), 5.18 - 4.84 (2H, m, 12-C*H*), 3.83 - 3.35 (8H, m, 9-C*H$_2$*, 15-C*H$_2$*), 2.54 - 1.82 (8H, m, 13-C*H$_2$*, 14-C*H$_2$*), 1.42 (10H, s, 18-C*H$_3$*), 1.24 (8H, s, 18-C*H$_3$*), 0.99 - 0.79 (4H, m, 10-C*H$_2$*), -0.01 (18, s, 11-C*H$_3$*); **^{13}C NMR** (126 MHz, CDCl$_3$) 154.6 (16-*C*O), 151.5 (7-*C*), 139.8 (1-*C*), 139.5 (5-*C*), 132.8 (4-*C*), 129.5 (2-*C*H), 127.1 (3-*C*H), 126.7 (6-*C*H), 79.4 (17-*C*), 72.3 (8-*C*H$_2$), 65.6 (9-*C*H$_2$), 52.7 (12-*C*H), 46.8 (15-*C*H$_2$), 32.4 (14-*C*H$_2$), 28.5 (18-*C*H$_3$), 24.0 (13-*C*H$_2$), 17.9 (10-*C*H$_2$), -1.44 (11-*C*H$_3$); **IR** (neat, ν$_{max}$, cm^{-1}) 2951 (brw), 1697 (s), 1390 (s), 1248 (m), 1163 (m), 1079 (s), 832 (s),

761 (m), 694 (m); **HRMS (ESI):** calcd. for $C_{48}H_{73}N_6O_6Si_2$ [M+H]$^+$ 885.5125, found 885.5125

1,2-Bis(4-Bromophenyl)ethyne (53)[222]
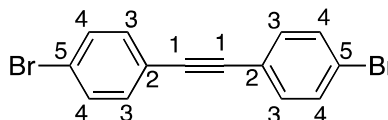
To a solution of 1-bromo-4-iodo benzene **68** (1.0 g, 3.5 mmol), Pd(PPh$_3$)Cl$_2$ (74 mg, 0.21 mmol), and CuI (67 mg, 0.35 mmol) in benzene under N$_2$ gas, was added dropwise DBU (3.2 mL, 21 mmol), trimethylsilyl acetylene (0.25 mL, 1.8 mmol), and H$_2$O (25 μL, 1.41 mmol). The solution was stirred at room temperature for 20 hours, before being diluted with water (75 mL) and diethyl ether (75 mL). The phases were separated, and the organic phase was washed with aqueous HCl (1M, 3 × 50 mL), before being dried over MgSO$_4$, and concentrated under reduced pressure. The crude material was purified by flash chromatography (10 g silica, petroleum ether, 100%), to give the product as a pale yellow solid (0.26 g, 44%).

^{1}H NMR (500 MHz, CDCl$_3$) 7.50 (4H, d, J = 8.4 Hz, 4-C*H*), 7.39 (4H, d, J = 8.4 Hz, 3-C*H*); **^{13}C NMR** (126 MHz, CDCl$_3$) 133.0 (3-CH), 131.7 (4-*C*H), 122.8 (2-*C*), 121.9 (5-*C*), 89.4 (1-*C*).

2,7-Dibromo-9H-carbazole (54)[333]
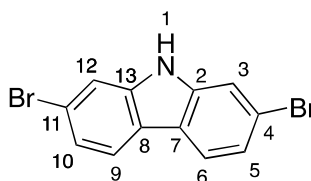
 A stirred solution of 4-bromo-1-(4-bromophenyl)-2-nitro-benzene (**77**, 1.5 g, 4.2 mmol) and triphenylphosphine (5.00 g, 19.1 mmol) in 1,2-dichlorobenzene (10 mL) was refluxed at 180 °C for 20 hours. The solution was allowed to cool to room temperature, and then concentrated under reduced pressure to form a dark oil. This was then washed with a mixture of ethyl acetate:petroleum ether (1:9, 100 mL) under stirring for 3 hours, forming a dark precipitate, which was removed by filtration. The filtrate was concentrated under reduced pressure, and purified by column chromatography (0 to 10% ethyl acetate: petroleum ether (40-60°C), followed by crystallization from chloroform, to give the

product as a white solid (278 mg, 20%).

^{1}H NMR (500 MHz, DMSO-d_6) 8.06 (1H, s, 1-N*H*), 7.88 (2H, d, J = 8.2 Hz, 6-C*H*, 9-C*H*), 7.37 (2H, dd, J = 8.2 Hz, 1.9, 5-C*H*, 10-C*H*), 7.27 (2H, d, J = 1.9, 3-C*H*, 12-C*H*); **^{13}C NMR** (126 MHz, DMSO-d_6) 140.3 (2-*C*, 13-*C*), 123.3 (5-*C*H, 10-*C*H), 121.8 (4-*C*Br, 11-*C*Br), 121.4 (6-*C*H, 9-*C*H), 119.7 (7-*C*, 8-*C*), 113.8 (3-*C*H, 12-*C*H).

2,5-Bis(4-bromophenyl)-1H-pyrrole (55)[334]
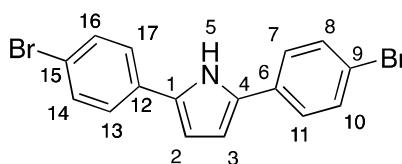
A stirred solution of 1,4-bis(4-bromophenyl)butane-1,4-dione (**80**, 500 mg, 1.28 mmol) and ammonium acetate (346 mg, 4.48 mmol) in acetic acid (10 mL) was allowed to reflux for 21 hours. The solution was then allowed to cool to room temperature, and concentrated under reduced pressure. The resulting brown solid was washed with concentrated aqueous potassium carbonate (30 mL), before being crystallized from acetic acid to give the product as a white, crystalline solid (301 mg, 62%).

^{1}H NMR (500 MHz, DMSO-d_6) 11.34 (1H, s, 5-N*H*), 7.72 (4H, d, J = 8.1 Hz, 7-C*H*, 11-C*H*, 13-C*H*, 17-C*H*), 7.55 (4H, d, J = 8.1, 8-C*H*, 10-C*H*, 14-C*H*, 16-C*H*), 6.65 (2H, s, 2-C*H*, 3-C*H*); **^{13}C NMR** (126 MHz, DMSO-d_6) 132.8 (1-*C*, 4-*C*), 132.0 (6-*C*, 12-*C*), 131.9 (8-*C*H, 10-*C*H, 14-*C*H, 16-*C*H), 126.4 (7-*C*H, 11-*C*H, 14-*C*H, 16-*C*H), 119.0 (9-*C*Br, 15-*C*Br), 109.0 (2-*C*H, 3-*C*H).

4-Bromo-N-(4-bromophenyl)benzamide (56)[335]
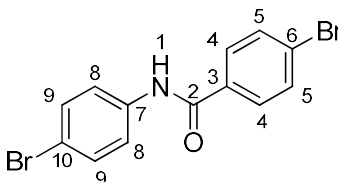
A suspension of 4-bromobenzoic acid (1.00 g, 4.97 mmol) in thionyl chloride (5.0 mL, 69 mmol) with a drop of DMF was heated under reflux under stirring. The solution was allowed to reflux for 2 hours, then concentrated under reduced pressure, to give a light orange oil, which solidified upon cooling to a crystalline solid (1.09 g)

To a stirred solution of 4-bromoaniline (777 mg, 4.52 mmol) and triethylamine (1.9 mL, 13.6 mmol) in dichloromethane (30 mL) at 0 °C was added dropwise a solution of the crystalline solid produced in the first part of the experiment in dichloromethane (10 mL). After addition, the solution was allowed to warm to room temperature, and allowed to stir for a further three hours. The solution was then cooled to −78 °C, and the resulting white precipitate was collected by vacuum filtration, and washed with cold dichloromethane (2 x 20 mL) to give the produce as a fine, white, crystalline powder (872 mg, 54%).

^{1}H NMR (500 MHz, DMSO-d_6) ^{1}H NMR (500 MHz, DMSO-d_6) δ 10.40 (1H, s, 1-N*H*), 7.90 (2H, d, J = 8.2 Hz, 4-C*H*), 7.75 (4H, d, J = 8.2 Hz, 5-C*H*, 8-C*H*), 7.53 (2H, d, J = 8.2 Hz, 9-C*H*). **^{13}C NMR** (126 MHz, DMSO-d_6) δ 165.1 (2-*C*O), 138.8 (7-*C*), 134.2 (3-*C*), 131.9 (9-CH), 131.9 (5-*C*H), 130.3 (4-*C*H), 125.9 (6-*C*Br), 122.8 (8-*C*H), 116.0 (10-*C*Br).

2,7-Dibromo-9-methyl-9*H*-carbazole (57)[336]
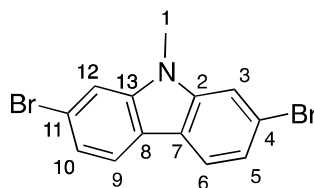
A solution of 2,7-dibromo-9*H*-carbazole (**54**, 200 mg, 0.62 mmol) in dry *N,N*-dimethylformamide (2 mL) at room temperature was added dropwise to sodium hydride (32 mg, 0.8 mmol) previously washed with ether. After 15 minutes of stirring, iodomethane (48 μL, 0.77 mmol*)* was added to the solution, and allowed to stir for a further 18 hours. The solution was quenched with water (20 mL), and extracted with ethyl acetate (2 × 20 mL). The combined organic phases were then washed with brine (2 × 20 mL), dried over magnesium sulfate, and concentrated under reduced pressure. The crude material was purified by flash column chromatography, eluted with petroleum ether and ethyl acetate (0 to 10 %), then recrystallized from a 1:1 mixture of petroleum ether and ethyl acetate, to give the product as a white, crystalline solid (101 mg, 48%).

^{1}H NMR (500 MHz, CDCl$_3$) 7.88 (2H, d, J = 8.2 Hz, 9-C*H*, 6-C*H*), 7.54 (2H, d, J = 1.7 Hz, 3-C*H*, 12-C*H*), 7.36 (2H, dd, J = 8.2, 1.7 Hz, 5-C*H*, 10-C*H*), 3.78 (3H, s, 1-C*H₃*); **^{13}C NMR** (126 MHz, CDCl$_3$) 141.9 (2-*C*, 13-*C*), 122.6 (6-*C*H, 9-*C*H), 121.4 (5-*C*H, 10-*C*H), 121.2 (4-*C*Br, 11-*C*Br), 119.7 (7-*C*, 8-*C*), 111.9 (3-CH, 12-*C*H), 29.3 (1-C*H₃*).

2,5-Bis(4-bromophenyl)-1-methyl-1*H*-pyrrole (58)[337]
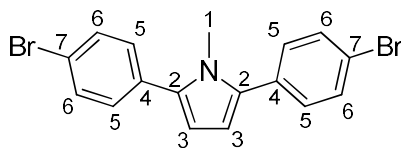
A solution of 2,5-bis(4-bromophenyl)-1*H*-pyrrole (**55**, 300 mg, 0.800 mmol), iodomethane (0.10 mL, 1.59 mmol), and potassium carbonate (330 mg, 2.39 mmol) in *N,N*-dimethylformamide (5 mL) was stirred under an inert atmosphere of nitrogen for 48 hours. The solution was then quenched with water (20 mL), and extracted with ethyl acetate (2 × 20 mL). The combined organic phases were washed with brine (2 × 30 mL), dried over magnesium sulfate, and concentrated under reduced pressure. The crude solid was purified by flash column chromatography eluting with Petroleum Ether: Ethyl Acetate (0 to 10 %), then recrystallized from ethyl acetate, to give the product as a white, crystalline solid (146 mg, 47%).

^{1}H NMR (500 MHz, CDCl$_3$) δ 7.56 (4H, d, J = 8.4 Hz, 5-C*H*), 7.34 (4H, d, J = 8.4 Hz, 6-C*H*), 6.32 (2H, s, 3-C*H*), 3.58 (3H, s, 1-C*H$_3$*); **^{13}C NMR** (126 MHz, CDCl$_3$) δ 136.1 (2-*C*), 132.2 (4-*C*), 131.6 (5-*C*H), 130.1 (6-*C*H), 121.0 (7-*C*Br), 109.2 (3-*C*H), 34.2 (1-*C*H$_3$).

4-Bromo-*N*-(4-bromophenyl)-*N*-methylbenzamide (59)[338]
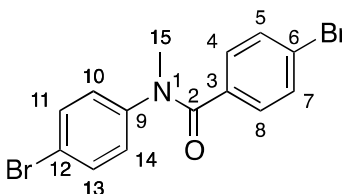
A solution of 4-bromo-*N*-(4-bromophenyl)benzamide (**56**, 250 mg, 0.700 mmol) in dry tetrahydrofuran (5 mL) was added to sodium hydride (42 mg, 1.1 mmol) washed with ether. After 90 minutes, iodomethane (0.26 mL, 4.2 mmol) was added to the solution, and allowed to stir for a further 18 hours. The solution then was quenched with water (30 mL), and extracted with diethyl ether (2 × 30 mL). The combined organic phase was dried over magnesium sulfate and concentrated under reduced pressure to give the crude material as an oil. The crude material was purified by column chromatography, eluted with petroleum ether and ethyl acetate (10 to 30 %) to give the product as a crystalline solid (168 mg, 65%).

¹H NMR (500 MHz, CDCl$_3$) 7.40 − 7.31 (4H, m, 4-C*H*, 8-C*H*, 10-C*H*, 14-C*H*), 7.17 (2H, d, J = 8.4 Hz, 11-C*H*, 13-C*H*), 6.90 (2H, d, J = 8.6 Hz, 5-C*H*, 7-C*H*), 3.46 (3H, s, 15-C*H₃*); **¹³C NMR** (126 MHz, CDCl$_3$) 169.4 (2-*C*O), 143.7 (9-*C*N), 134.3 (3-C), 132.5 (10-*C*H, 14-*C*H), 131.2 (11-*C*H, 13-*C*H), 130.3 (5-*C*H, 7-*C*H), 128.3 (4-*C*H, 8-*C*H), 124.4 (6-*C*Br), 120.3 (12-*C*Br), 38.4 (15-*C*H₃); **IR** (neat, v_{max}, cm⁻¹) 3059 (w), 2925 (w), 1640 (m), 1588 (m), 1484 (s), 1354 (m), 1296 (m), 1183 (m), 1068 (m), 1005 (m), 834 (s), 735 (s), 688 (m); **HRMS (ESI):** calcd. for C$_{14}$H$_{12}$Br$_2$NO [M+H]⁺ 369.9260, found 369.9276

7-Bromo-3-(4-bromophenyl)quinoline-4-carboxylic acid (64)[339]
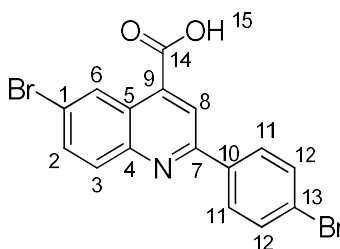
A solution of 5-bromoisatin (2.00 g, 8.85 mmol), 1-(4-bromophenyl)ethanone (1.76 g, 8.85 mmol) and potassium hydroxide (1.48 g, 26.6 mmol) in ethanol (100 mL) was heated at 60 °C for 15 hours. The solution was then concentrated under reduced pressure to give a red residue, which was redissolved in water (200 mL) and petroleum ether (150 mL). The precipitate was removed by filtration, and the filtrate separated. The aqueous phase was acidified with 1 M HCl, to give an orange precipitate. This precipitate was then triturated in a mixture of hot ethanol (90 mL) and water (10 mL) to give the product as a pale brown solid (621 mg, 17%).

¹H NMR (500 MHz, DMSO-*d₆*) δ 9.05 (1H, s, 8-C*H*), 8.42 (1H, s, 6-C*H*), 8.18 (2H, d, J = 8.1 Hz, 11-C*H*), 8.01 (1H, d, J = 8.9 Hz, 3-C*H*), 7.89 (1H, d, J = 8.9 Hz, 2-C*H*), 7.73 (2H, d, J = 8.1 Hz, 12-C*H*); **¹³C NMR** (126 MHz, DMSO-*d₆*) δ 155.6 (14-*C*OOH), 147.4 (7-*C*), 137.6 (4-*C*), 133.3 (2-*C*H), 132.4 (12-*C*H), 132.0 (3-*C*H), 129.6 (11-*C*H), 129.0 (8-*C*H), 125.9 (1-*C*Br), 124.2 (13-*C*Br), 120.9 (5-*C*), 119.7 (6-*C*H).

6-Bromo-2-(4-bromophenyl)quinazoline (65)[340]
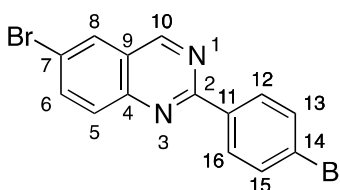
6-Bromo-2-(4-bromophenyl)-3,4-dihydroquinazolin-1-ium chloride (**86,** 569 mg, 1.42 mmol) was stirred in an emulsion of aqueous sodium hydroxide (30 mL) and dichloromethane (30 mL) at room temperature for 2 hours. The resulting solution was extracted with ethyl acetate (2 × 75 mL), and dried over magnesium sulfate, before being concentrated under reduced pressure. This residue was taken up in toluene (10 mL), and chloranil (350 mg, 1.42 mmol) was added. The solution was refluxed under an inert atmosphere for 18 hours, then concentrated under reduced pressure. The solid residue was taken up in dichloromethane (250 mL), and the remaining precipitate removed by vacuum filtration. The filtrate was washed with aqueous sodium hydroxide solution (3 M, 100 mL) and water (2 × 100 mL), and dried over magnesium sulfate, before being concentrated under reduced pressure to give quinazoline **65** as an off-white solid (341 mg, 66%)

^{1}H NMR (500 MHz, CDCl$_3$) 9.39 (1H, s, 10-C*H*), 8.50 (2H, d, J = 8.6 Hz, 12-C*H*, 16-C*H*), 8.10 (1H, d, J = 2.0 Hz, 8-C*H*), 8.02 − 7.93 (2H, m, 5-C*H*, 6-C*H*), 7.67 (2H, d, J = 8.6 Hz, 13-CH, 15-CH); **^{13}C NMR** (126 MHz, CDCl$_3$) 160.4 (2-*C*), 159.4 (10-*C*H), 149.4 (4-*C*), 137.8 (1-*C*H), 136.5 (9-*C*H), 131.9 (13-C*H*, 15-C*H*), 130.4 (5-*C*H), 130.2 (12-*C*H, 16-*C*H), 129.2 (8-*C*H), 125.8 (11-*C*), 124.5 (14-*C*Br), 121.0 (7-*C*Br).

4,4'-Oxybis(bromobenzene) (71)[341]
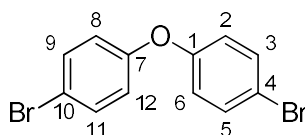
To a solution of diphenyl ether (0.50 mL, 3.2 mmol), and NaBrO$_3$ (158 mg, 1.05 mmol) in dichloromethane (2.5 mL) and water (2.5 mL), was added dropwise bromine (0.20 ml, 3.5 mmol). The solution was then stirred at room temperature for 2 hours, until no starting material was visible by TLC. The solution was diluted with dichloromethane (10 mL) and water (10 mL), and 1 M aqueous sodium thiosulfate solution was added until the solution had turned colourless. The organic phase was washed with water (2 × 10 mL) and brine (2 × 10 mL), before being dried over MgSO$_4$, and concentrated under reduced pressure to the product as an off-white solid (792 mg 77%).

¹H NMR (500 MHz, CDCl₃) 7.46 (4H, d, J = 8.8 Hz, 3/5/9/11-C*H*), 6.89 (4H, d, J = 8.8 Hz, 2/6/8/12-C*H*), **¹³C NMR** (126 MHz, CDCl₃) 156.0 (1/7-*C*O), 132.8 (3/5/9/11-*C*H), 120.6 (2/6/8/12-*C*H), 116.2 (4/10-*C*Br)

2-Amino-4-bromophenol (73)[342]
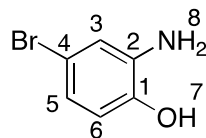
A suspension of tin(II) chloride (5.9 g, 23 mmol) in concentrated aqueous HCl (37 %, 11 mL) and MeOH (20 mL) was cooled to 0 °C. 4-bromo-2-nitrophenol (**72**, 1.0 g, 4.6 mmol) was then added to the solution in one portion, and allowed to stir for 6 hours, whereby the solution had gone colourless. The solution was then diluted with ethyl acetate (60 mL), and neutralised with concentrated aqueous NaHCO₃ solution. The white precipitate was then removed by filtration, and washed with ethyl acetate (120 mL). The filtrate was separated, and the aqueous phase extracted with ethyl acetate (3 × 50 mL). The combined organic phase was dried over MgSO₄, and concentrated under reduced pressure, to give the product as a light brown solid (836 mg, 97%).

¹H NMR (500 MHz, CDCl₃) 9.21 (1H, s, 7-O*H*), 6.71 (1H, d, J = 2.4 Hz, 3-C*H*), 6.55 (1H, d, J = 8.3 Hz, 6-C*H*), 6.48 (1H, dd, J = 8.3, 2.4 Hz, 5-C*H*), 4.76 (2H, s, 8-N*H₂*); **¹³C NMR** (126 MHz, CDCl₃) 143.7 (1-*C*OH), 139.3 (2-*C*NH₂), 118.5 (5-*C*H), 116.4 (3-*C*H), 116.1 (6-*C*H), 111.1 (4-*C*Br).

4-Bromo-*N*-(5-bromo-2-hydroxyphenyl)benzamide (74)
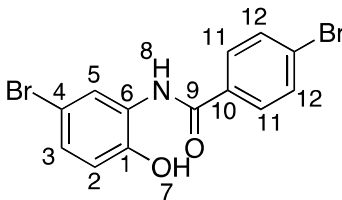
A solution of 4-bromobenzoic acid (1.0 g, 5 mmol), and 3 drops of dimethylformamide in sulfuryl chloride (7 mL) was refluxed at 80 °C for 3 hours. The solution was then concentrated under reduced pressure to yield a yellow, crystalline solid (1.03 g).

To a stirred solution of 2-amino-4-bromophenol (**73**, 568 mg, 3.02 mmol) and pyridine (0.8 mL, 10 mmol) in dry THF (20 mL) at −10 °C, was added dropwise a solution of the yellow solid (676 mg, 3.3 mmol) in dry THF (6 mL). The solution was allowed to stir at −10 °C for 6 hours, before being quenched with concentrated aqueous $NaHCO_3$ solution (30 mL), and diluted with ethyl acetate (40 mL). The aqueous phase was extracted with ethyl acetate (30 mL), and the combined organic phases were washed with HCl (1 M, 50 mL), and brine (50 mL), before being concentrated under reduced pressure to give a brown solid. This crude material was dissolved in a solution of NaOH (150 mg, 3.8 mmol) in water (10 mL) and THF (20 mL), and allowed to stir at room temperature for 20 hours. The resulting solution was diluted with ethyl acetate (100 mL), and the organic phase was washed with HCl (1 N, 2 × 150 mL), concentrated aqueous $NaHCO_3$ solution (2 × 150 mL), and brine (2 × 100 mL). The solution was dried over $MgSO_4$ and concentrated under reduced pressure, to give the product as a light brown, crystalline solid (787 mg, 70%).

^{1}H NMR (500 MHz, DMSO-d_6) 10.11 (1H, s, 8-N*H*), 9.57 (1H, s, 7-O*H*), 8.00 - 7.81 (3H, m, 5-C*H*, 11-C*H*), 7.74 (2H, d, J = 8.6 Hz, 12-C*H*), 7.19 (1H, dd, J = 8.6, 2.5 Hz, 3-C*H*), 6.88 (1H, d, J = 8.6 Hz, 2-C*H*); **^{13}C NMR** (126 MHz, DMSO-d_6) 164.9 (9-*C*O), 149.3 (1-*C*OH), 133.7 (3-*C*H), 132.0 (12-*C*H), 130.2 (11-*C*H), 128.8 (10-*C*), 128.3 (13-*C*Br), 127.7 (5-*C*H), 126.9 (6-*C*NH), 126.0 (2-*C*H), 109.9 (4-*C*Br); **IR** (neat, v_{max}, cm^{-1}) 3182 (brm), 1631 (s), 1587 (s), 1519 (s), 1481 (s), 1409 (s), 1320 (s), 1283 (s), 1104 (s), 1068 (s), 1009 (s), 930 (m), 870 (s), 844 (s), 812 (s), 752 (m), 723 (m), 688 (m); **HRMS (ESI):** calcd. for $C_{13}H_{10}Br_2N_1O_2$ [M+H]$^+$ 371.9053, found 371.9056.

5-bromo-2-(4-bromophenyl)benzo[*d*]oxazole (75)
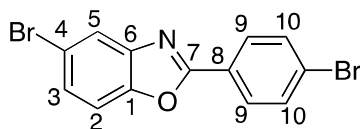
To a solution of *p*-toluenesulfonic acid monohydrate (80 mg, 0.42 mmol) in toluene (5 mL) was added 4-bromo-*N*-(5-bromo-2-hydroxyphenyl)benzamide (**74**, 50 mg, 0.13 mmol), and heated at 110 °C for 5 hours. The resulting solution was diluted with water (15 mL) and ethyl acetate (15 mL). The aqueous phase was extracted with ethyl acetate (15 mL), before the combined organic phase was washed with $NaHCO_3$ solution (2 × 20 mL) and brine (2 × 20 mL). The solution was then dried over $MgSO_4$, and concentrated under reduced pressure to give the product as a light brown solid (45 mg, 98%).

^{1}**H NMR** (500 MHz, CDCl$_3$) 8.11 (2H, d, J = 8.6 Hz, 10-C*H*), 7.91 (1H, d, J = 1.9 Hz, 5-C*H*), 7.69 (2H, d, J = 8.6 Hz, 9-C*H*), 7.54 - 7.42 (2H, m, 2-C*H*, 3-C*H*); ^{13}**C NMR** (126 MHz, CDCl$_3$) 163.2 (7-*C*), 149.8 (6-*C*), 143.6 (1-*C*), 132.3 (10-*C*H), 129.1 (9-*C*H), 128.4 (3-*C*H), 126.7 (8-*C*), 125.6 (11-*C*Br), 123.1 (5-*C*H), 117.5 (4-*C*Br), 111.8 (2-*C*H); **IR** (neat, v$_{max}$, cm^{-1}) 3084 (w), 2919 (w), 1587 (m), 1546 (m), 1474 (s), 1446 (s), 1423 (m), 1397 (s), 1260 (s), 1069 (s), 1008 (s), 891 (m), 830 (s), 797 (s), 750 (m), 726 (s), 679 (m); **HRMS (ESI):** calcd. for C$_{13}$H$_8$Br$_2$NO [M+H]$^+$ 351.8967, found 351.8967.

4,4'-Dibromo-2-nitro-1,1'-biphenyl (77)[333]
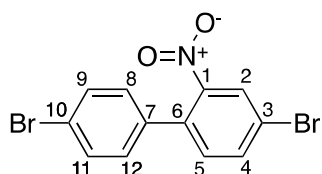
A mixture of nitric acid (68%, 5.0 mL, 110 mmol) and glacial acetic acid (3mL) was added dropwise to a stirred solution of 1-bromo-4-(4-bromophenyl)benzene (**18**, 2.00 g, 6.41 mmol) in acetic anhydride (20 mL) and dichloromethane (4 mL) at 0 °C. The solution was allowed to warm to room temperature and stir for 1 hour. The solution was then neutralised with aqueous NaOH (3 M) and extracted with dichloromethane (2 × 40 mL). The organic phase was dried over MgSO$_4$ and concentrated under reduced pressure. The resulting material was recrystallized from ethanol to yield the product as a light yellow solid (1.72 g, 75%).

^{1}**H NMR** (500 MHz, DMSO-d_6) 8.27 (1H, d, J = 2.0 Hz, 2-C*H*), 7.98 (1H, dd, J = 8.2, 2.0 Hz, 4-C*H*), 7.66 (2H, d, J = 8.1 Hz, 8-C*H*, 12-C*H*), 7.51 (1H, d, J = 8.2 Hz, 5-C*H*), 7.30 (2H, d, J = 8.1 Hz, 9-C*H*, 11-C*H*); ^{13}**C NMR** (126 MHz, DMSO-d_6) 149.5 (1-*C*NO$_2$), 136.2 (4-*C*H), 135.7 (6-*C*), 133.9 (5-*C*H), 133.5 (7-*C*), 132.2 (8-*C*H, 12-*C*H), 130.4 (9-*C*H, 11-*C*H), 127.2 (2-*C*H), 122.6 (10-*C*Br), 121.7 (3-*C*Br).

2-Bromo-1-(4-bromophenyl)ethanone (79)[343]
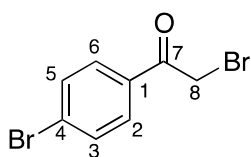
To a stirred solution of 4-bromoacetophenone (**78**, 6.00 g, 30.3 mmol) in acetic acid (45 mL) was added dropwise bromine (1.55 mL, 30.3 mmol). The solution was allowed to stir at room temperature for a further three hours, until a colour change from dark brown to yellow had occurred. The solution was then diluted with ethyl acetate (100 mL), and washed with sodium thiosulfate solution (2 M, 2 × 50 mL), sodium bicarbonate solution (saturated, 2 × 50 mL), and water (2 × 50 mL). The solution was then dried over magnesium sulfate and concentrated under reduced pressure to give a white solid. The crude material was then recrystallized from a 10:1 mixture of petrol ether and ethyl acetate to give the product as a white, crystalline solid (5.75 g, 68% yield).

^{1}H NMR (500 MHz, CDCl$_3$) 7.85 (2H, d, J = 8.6 Hz, 2-CH, 6-CH), 7.65 (2H, d, J = 8.6 Hz, 3-CH, 5-CH), 4.40 (2H, s, 8-CH_2); **^{13}C NMR** (126 MHz, CDCl$_3$) 190.4 (7-CO), 132.7 (1-C), 132.2 (3-CH, 5-CH), 130.4 (2-CH, 6-CH), 129.3 (4-CBr), 30.3 (8-CH$_2$).

1,4-Bis(4-bromophenyl)butane-1,4-dione (80)[228]
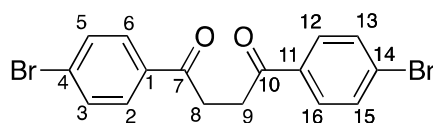
Zinc(II) chloride (3.91 g, 28.7 mmol) was dried by melting at 350 °C for 25 minutes, until gas evolution had ceased. After cooling to room temperature, the solidified material was then stirred in a solution of toluene (13.5 mL), diethylamine (2.2 mL, 22 mmol) and *tert*-butanol (2.1 mL, 22 mmol) for 2.5 hours, until all solid material had dissolved. To this solution was added 4-bromoacetophenone (**78**, 4.26 g, 21.5 mmol), followed by 2-bromo-1-(4-bromophenyl)ethanone (**79**, 4.00 g, 14.3 mmol), and the reaction vessel was sealed under an inert atmosphere and stirred for a further 3 hours. The solution was then allowed to stand for 3 days, causing a precipitate to form. The reaction mixture was then quenched with sulfuric acid (5 % in water, 20 mL), and the precipitate collected by vacuum filtration. The precipitate was washed with toluene (20 mL), water (20 mL) and methanol (20 mL) to give the product as a crystalline, white solid (3.67 g, 65%).

^{1}H NMR (500 MHz, CDCl$_3$) 7.90 (4H, d, J = 8.5 Hz, 2-CH, 6-CH, 12-CH, 16-CH), 7.63 (4H, d, J = 8.5 Hz, 3-CH, 5-CH, 12-CH, 16-CH), 3.41 (4H, s, 8-CH_2, 9-CH_2); **^{13}C NMR** (126 MHz, CDCl$_3$) 197.5 (7-CO, 10-CO), 135.4 (1-C, 11-C), 131.9 (3-CH, 5-CH, 12-CH,

16-C*H*), 129.6 (2-C*H*, 6-C*H*, 12-C*H*, 16-C*H*), 128.4 (4-*C*Br, 14-*C*Br), 32.4 (8-*C*H$_2$, 9-*C*H$_2$).

2,5-Bis(4-bromophenyl)furan (81)[344]
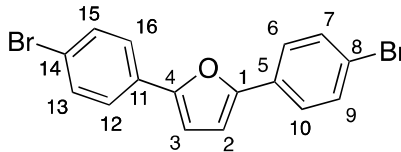
A stirred solution of 1,4-bis(4-bromophenyl)butane-1,4-dione (**80**, 250 mg, 0.635 mmol) in methanol (3.0 mL) and dichloromethane (7.0 mL) at 0 °C was saturated with HCl gas, produced from the reaction of concentrated sulfuric acid with sodium chloride. The solution was allowed to warm to room temperature and stirred for a further three hours. The solution was then concentrated under reduced pressure give a brown solid, which was washed with concentrated aqueous potassium carbonate solution (30 mL). The remaining crude material was then crystallized from a 3:7 mixture of chloroform and diethyl ether to give 2,5-bis(4-bromophenyl)furan (**81**) as a white, crystalline solid (177 mg, 74%).

¹H NMR (500 MHz, DMSO-*d*$_6$) 7.77 (4H, d, J = 8.6 Hz, 6-C*H*, 10-C*H*, 13-C*H*, 15-C*H*), 7.63 (4H, d, J = 8.6 Hz, 7-C*H*, 9-C*H*, 13-C*H*, 15-C*H*), 7.14 (2H, s, 2-C*H*, 3-C*H*); **¹³C NMR** (126 MHz, DMSO-*d*$_6$) 152.4 (1-*C*H, 4-*C*H), 132.3 (7-*C*H, 9-*C*H, 13-*C*H, 15-*C*H), 129.6 (5-*C*, 11-*C*), 125.9 (6-*C*H, 10-*C*H, 12-*C*H, 16-*C*H), 121.1 (8-*C*Br, 14-*C*Br), 109.6 (2-*C*H, 3-*C*H).

2,5-Bis(4-bromophenyl)thiophene (82)[345]
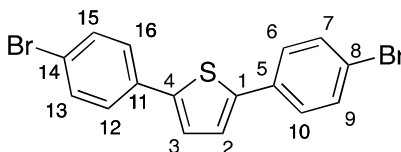
A stirred solution of 1,4-bis(4-bromophenyl)butane-1,4-dione (**80**, 250 mg, 0.64 mmol) and Lawesson's reagent (308 mg, 0.76 mmol) in toluene (5.0 mL) was allowed to reflux for 3 hours. The solution was then allowed to cool to room temperature, before adding diethyl ether (10.0 mL) and further cooling the solution to −78 °C. After allowing the solution to sit for 10 minutes, the resulting precipitate was collected by vacuum filtration,

and washed with chilled (−78 °C) diethyl ether (10 mL), to give 2,5-bis(4-bromophenyl)thiophene (**82**) as a white, crystalline solid (188 mg, 75%).

^{1}H NMR (500 MHz, CDCl$_3$) 7.55 - 7.46 (8H, m, 6-C*H*, 7-C*H*, 9-C*H*, 10-C*H*, 12-C*H*, 13-C*H*, 15-C*H*, 16-C*H*), 7.27 (2H, s, 2-C*H*, 3-C*H*); **^{13}C NMR** (126 MHz, CDCl$_3$) 142.7 (1-*C*, 4-*C*), 133.1 (5-*C*, 11-*C*), 132.0 (7-*C*H, 9-*C*H, 13-*C*H, 15-*C*H), 127.1 (6-*C*H, 10-*C*H, 12-*C*H, 16-*C*H), 124.5 (2-*C*H, 3-*C*H), 121.5 (8-*C*Br, 14-*C*Br).

2-(Aminomethyl)-4-bromoaniline (84)[340]
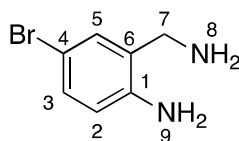
To a solution of 2-amino-5-bromobenzonitrile (**83**, 2.00 g, 10.2 mmol) in dry THF (20 mL) at 0 °C was added dropwise a solution of borane in THF (1 M, 12.5 mL, 12.5 mmol). The solution was then allowed to warm to room temperature, and stir for a further 72 hours. The solution was quenched dropwise with ethanol (10 mL), before hydrochloric acid (4 M, 10 mL) was added, and the solution was allowed to stir for a further 16 hours. The solution was then concentrated under reduced pressure, and *tert*-butyl acetate (20 mL) was added, before basifying the aqueous phase with sodium hydroxide tablets. The organic phase was then separated and washed with water (2 × 15 mL), before being dried over magnesium sulfate, and concentrated under reduced pressure to give a light brown solid. This crude material was then purified by column chromatography (25 g silica, MeOH:CH$_2$Cl$_2$, 0% to 15%) to give the product as a white, crystalline solid (874 mg, 43%).

^{1}H NMR (500 MHz, DMSO-*d$_6$*) 7.18 (1H, d, J = 2.4 Hz, 5- C*H*), 7.04 (1H, dd, J = 8.4, 2.4 Hz, 3-C*H*), 6.54 (1H, d, J = 8.4 Hz, 2-C*H*), 5.21 (2H, s, 9-N*H$_2$*), 3.56 (2H, s, 7-C*H$_2$*), 1.75 (2H, brs, 8-N*H$_2$*); **^{13}C NMR** (126 MHz, DMSO-*d$_6$*) 146.3 (1-*C*NH$_2$), 130.4 (5-*C*H), 129.7 (3-*C*H), 117.7 (6-*C*H), 116.7 (2-*C*H), 107.0 (4-*C*Br), 43.0 (7-*C*H$_2$).

N-(2-Amino-5-bromobenzyl)-4-bromobenzamide (85)[340]
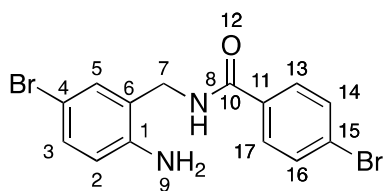
4-Bromobenzoic acid (1.00 g, 5.00 mmol) was suspended in thionyl chloride (5 mL), and a drop of DMF was added. The solution was heated to reflux, and stirred for 2 hours, until no more effervescence was observed. The solution was then cooled to room temperature and concentrated under reduced pressure to give an orange, crystalline solid (1.097 g).

To a stirred solution of 2-(aminomethyl)-4-bromoaniline **84** (730 mg, 3.63 mmol) and triethylamine (0.70 mL, 5.4 mmol) in dichloromethane (10 mL) at 0 °C was added dropwise a solution of the previously formed orange solid (800 mg, 3.64 mmol) in dichloromethane (5 mL). The solution was allowed to warm to room temperature and stir for a further 2.5 hours. The reaction was quenched with water (15 mL), causing precipitation of a white solid upon formation of an emulsion. The white solid was redissolved using ethyl acetate (100 mL), the phases separated, and the organic phase washed with potassium carbonate solution (conc., 2 × 50 mL) and water (2 × 50 mL). The solution was then dried over magnesium sulfate, and concentrated under reduced pressure to yield amide **85** as an oil, which later solidified to an off-white solid (1.31 g, 94%)

^{1}H NMR (500 MHz, DMSO-d_6) 9.01 (1H, s, 8-N*H*), 7.82 (2H, d, J = 8.4 Hz, 13-C*H*, 17-C*H*), 7.69 (2H, d, J = 8.4 Hz, 14-C*H*, 16-C*H*), 7.14 − 7.07 (2H, m, 3-C*H*, 5-C*H*), 6.59 (1H, d, J = 8.4 Hz, 3-CH), 5.31 (2H, s, 9-N*H2*), 4.29 (2H, d, J = 6.0 Hz, 7-C*H2*); **^{13}C NMR** (126 MHz, DMSO-d_6) 166.0 (10-*C*O), 146.0 (1-*C*NH$_2$), 133.5 (11-*C*), 131.9 (14-*C*H, 16-*C*H), 131.6 (5-*C*H), 130.7 (3-*C*H), 129.8 (13-*C*H, 17-*C*H), 125.6 (15-*C*Br), 124.7 (6-*C*), 117.0 (2-*C*H), 106.7 (4-*C*Br).

6-Bromo-2-(4-bromophenyl)-3,4-dihydroquinazolin-1-ium chloride (86)[340]
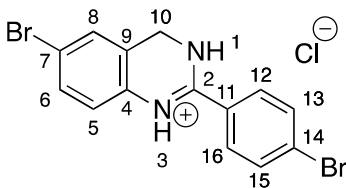
A solution of *N*-(2-amino-5-bromobenzyl)-4-bromobenzamide **85** (1090 mg, 2.84 mmol) in phosphoryl chloride (7 mL) was refluxed for 2.5 hours. The solution was then concentrated under reduced pressure, and the residue was neutralized by washing with concentrated aqueous sodium bicarbonate (30 mL). The solid material was extracted with ethyl acetate (50 mL), and the precipitate was removed by vacuum filtration. The filtrate was dried over magnesium sulfate, before being concentrated under reduced pressure to give the product as an off-white, crystalline solid (767 mg, 67%).

^{1}H NMR (500 MHz, DMSO-d_6) 12.64 (1H, s, 1-N*H*), 11.10 (1H, s, 3-N*H*), 7.98 − 7.83 (4H, m, 12-C*H*, 13-C*H*, 15-C*H*, 16-C*H*), 7.58 − 7.48 (2H, m, 6-C*H*, 8-C*H*), 7.40 (1H, d, J = 8.5 Hz, 5-C*H*), 4.85 (2H, s, 10-C*H₂*); **^{13}C NMR** (126 MHz, DMSO-d_6) 157.5 (2-*C*), 132.4 (13-*C*H, 15-*C*H), 132.0 (4-*C*), 131.9 (6-*C*H), 131.3 (12-*C*H, 16-*C*H), 129.8 (8-*C*H), 128.5 (11-*C*), 126.3 (9-*C*), 121.0 (14-*C*Br), 119.8 (5-*C*H), 119.2 (7-*C*Br), 42.1 (10-*C*H₂); **HRMS (ESI):** calcd. for $C_{14}H_{10}Br_2N_4$ [M+H]$^+$ 364.9283, found 364.9283.

(6-Bromo-2,3-dihydro-1*H*-inden-1-yl)triphenylphosphonium bromide (94)
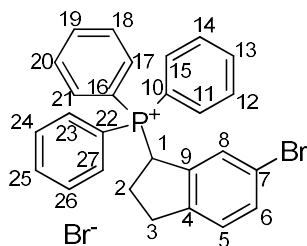
Single Step Method

A solution of 6-bromoindan-1-ol (**166**, 107 mg, 0.500 mmol) and triphenylphosphonium bromide (172 mg, 0.500 mmol) in toluene (5 mL) was refluxed overnight at 110 °C, forming a white precipitate. The solution was cooled to room temperature, and the white precipitate was collected by vacuum filtration. Recrystallization of this white precipitate from chloroform gave the product as a white solid (56 mg, 21%).

Two Step Method

To a solution of 6-bromoindan-1-ol (**166**, 2.00 g, 9.39 mmol) in dichloromethane (5 mL) at 0 °C was added dropwise a solution of phosphorus tribromide (1.76 mL, 18.8 mmol) in dichloromethane (8 mL). The solution was allowed to warm to room temperature and stir for a further 2 hours. This solution was then quenched with saturated NaHCO₃ solution, and the phases separated. The aqueous phase was extracted with

dichloromethane (2 x 5 mL), dried, and concentrated under reduced pressure to give the benzyl bromide as a light brown oil (2.45 g)

To a solution of 1,6-dibromoindane from the previous step (6.06 g, 22.0 mmol) in toluene (60 mL) was added triphenylphosphine (5.76 g, 22.0 mmol) and the solution was refluxed for 4 days. The solution was then cooled to room temperature and filtered, and the precipitate was washed with diethyl ether (2 x 30 mL), before being triturated in boiling chloroform. Drying in the vacuum oven for 20 hours yielded the product as a white solid (6.32 g, 51% over 2 steps).

^{1}H NMR (500 MHz, CDCl$_3$) δ 7.93 – 7.83 (6H, m, 12/14/18/20/24/26-C*H*), 7.81 – 7.75 (3H, m, 13/19/25-C*H*), 7.70 – 7.62 (6H, m, 11/15/17/21/23/27-C*H*), 7.34 – 7.28 (1H, m, 6-C*H*), 6.92 – 6.86 (2H, m, 5-C*H, *8-C*H*), 3.43 – 3.20 (1H, m, 3-C*H*H), 2.74 – 2.63 (1H, m, 3-CH*H*), 2.44 – 2.27 (1H, m, 2-C*H*H), 1.69 – 1.52 (1H, m, 2-CH*H*). **^{1}H NMR** (500 MHz, DMSO-d6) δ 7.91 (3H, t, J = 7.1 Hz, 13/19/25-C*H*), 7.85 – 7.68 (m, 12H, 11/15/17/21/23/27-C*H*, 12/14/18/20/24/26-C*H*), 7.45 (1H, dd, J = 8.3, 1.9 Hz, 6-C*H*), 7.12 (1H, d, J = 8.3 Hz, 5-C*H*), 6.76 (1H, d, J = 1.9, 8-C*H*), 6.33 – 6.17 (1H, m, 1-C*H*), 3.06 – 2.83 (1H, m, 3-C*H*H), 2.78 – 2.65 (1H, m, 3-CH*H*), 2.48 – 2.31 (1H, m, 2-C*H*H), 1.63 – 1.53 (1H, m, 2-CH*H*); **^{13}C NMR** (126 MHz, DMSO-d6) δ 145.9 (4-*C*), 145.8 (4-*C*), 137.3 (9-*C*), 135.7 (13/19/25-*C*H), 135.6 (13/19/25-*C*H), 134.4 (12/14/18/20/24/26-*C*H), 134.3 (12/14/18/20/24/26-*C*H), 132.4 (6-*C*H), 130.8 (11/15/17/21/23/27-*C*H), 130.8 (11/15/17/21/23/27-*C*H), 129.0 (8-*C*H), 128.9 (8-*C*H), 127.8 (5-*C*H), 127.7 (5-*C*H), 119.9 (7-*C*Br), 119.8 (7-*C*Br), 118.1 (10/16/22-*C*), 117.5 (10/16/22-*C*), 79.7 (1-*C*), 30.7 (3-*C*H$_2$), 27.4 (2-*C*H$_2$); **IR** (neat, v_{max}, cm^{-1}) 3317 (m), 2979 (w), 1754 (m), 1673 (s), 1493 (s), 1393 (w), 1369 (m), 1244 (s), 1156 (s), 1125 (s), 1172 (m), 878 (w), 754 (m); **HRMS (ESI):** calcd. for C$_{27}$H$_{23}$BrP [M]$^+$ 457.0715, found 457.0714.

6-Bromo-2,3-dihydro-1*H*-inden-1-one (95)[346]
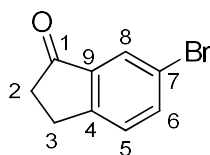
To a cooled suspension of 3-(4-bromophenyl)propanoic acid (2.29 g, 10.0 mmol) in dichloromethane (5 mL) at 0 °C was added dropwise trifluoromethanesulfonic acid (13.3 mL, 150 mmol). The solution was warmed to room temperature, then heated in the

microwave at 80 °C for 1 hour. The resulting brown solution was poured into ice (150 mL), causing a large exotherm when mixing. The ice/water solution was neutralised with saturated $NaHCO_3$ solution and then extracted with CH_2Cl_2 (3 x 70 mL). The combined organic phase was dried over $MgSO_4$ and concentrated under reduced pressure to yield an off-white powder (1.83 g) Purification by flash chromatography (PE:EtOAc, 0-20%, 20g silica) yielded the product as a crystalline white powder (182 mg, 70%).

^{1}H NMR: (500 MHz, CDCl$_3$) δ 7.89 (1H, d, J = 1.9 Hz, 8-C*H*), 7.69 (1H, dd, J = 8.2, 1.9 Hz, 6-C*H*), 7.37 (1H, d, J = 8.2 Hz, 5-C*H*), 3.13 – 3.07 (2H, m, 3-C*H$_2$*), 2.75 – 2.70 (2H, m, 2-C*H$_2$*); **^{13}C NMR**: (126 MHz, CDCl$_3$) δ 205.1 (1-*C*O), 153.5 (4-*C*), 138.9 (9-*C*), 137.3 (6-*C*H), 128.2 (5-*C*H), 126.7 (8-*C*H), 121.6 (7-*C*Br), 36.5 (2-*C*H$_2$), 25.5 (3-*C*H$_2$); **LCMS** (acetonitrile/water, 30-90%, 20 mins) Retention time: 15.28 min, m/z: 210.9 [M+H]$^+$.

(*2R,3R*)-2,3-Dihydroxysuccinohydrazide (144)[347]
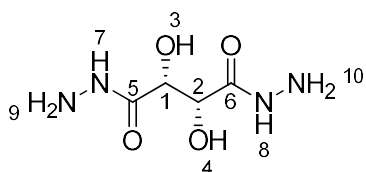
To a solution of diethyl (2*R*,3*R*)-2,3-dihydroxybutanedioate (0.83 mL, 4.85 mmol) in ethanol (10mL) was added hydrazine hydrate (0.94 mL, 19.4 mmol), and heated to reflux for 16 hours. The solution was cooled to room temperature and the precipitate collected by vacuum filtration, and dried under vacuum to give the product as a white solid (698 mg, 81%).

^{1}H NMR (500 MHz, DMSO-d6) δ 8.76 (2H, s, 7/8-N*H*), 5.36 (2H, s, 3/4-O*H*), 4.25 (2H, s, 1/2-C*H*); **LCMS** (acetonitrile/water, 10-90%, 10 mins) Retention time: 0.66 min, m/z: 366.1 [M+H]$^+$.

Di-tert-butyl 2,2'-((2R,3R)-2,3-dihydroxysuccinyl)bis(hydrazinecarboxylate) (145)
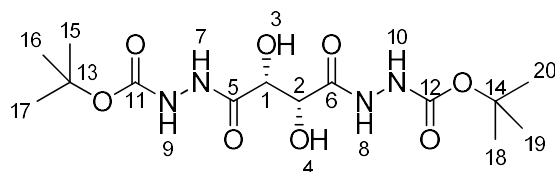
To a stirred suspension of *tert*-butoxycarbonyl *tert*-butyl carbonate (**144**, 2.45 g, 11.23 mmol) and (2R,3R)-2,3-dihydroxybutanedihydrazide (1.00 g, 5.61 mmol) in acetonitrile (40 mL) and water (10 mL) at room temperature was added a single crystal of iodine. The solution was allowed to stir overnight. If there was still undissolved starting material, several more crystals of iodine were added and the solution allowed to continue stirring. The solution was then concentrated under reduced pressure, giving a white slurry. This slurry was then triturated in dichloromethane (approx. 25 mL) by sonication. The solid was then collected by filtration, and dried under vacuum, giving the product as a white solid (989 mg, 47 %).

¹H NMR: (500 MHz, DMSO-d6) δ 9.36 (2H, s, 7/8-N*H*), 8.70 (2H, s, 9/10-N*H*), 5.37 (2H, d, J = 7.6 Hz, 3/4-O*H*), 4.30 (2H, d, J = 7.6 Hz, 1/2-C*H*), 1.39 (18H, s, 15/16/17/18/19/20-C*H₃*); **¹³C NMR** (126 MHz, DMSO-*d*6) δ 171.6 (5/6-*C*O, 155.5 (11/12-*C*O), 79.6 (13/14-*C*), 72.8 (1/2-*C*H), 28.5 (15/16/17/18/19/20-*C*H₃); **IR** (neat, v_{max}, cm^{-1}) 3332 (m), 2979 (w), 1754 (m), 1714 (m), 1674 (s), 1494 (s), 1393 (w), 1369 (m), 1296 (s), 1245 (s), 1158 (s), 1127 (s), 1173 (m), 879 (w), 794 (m), 755 (m), 676 (m); **HRMS (ESI):** calcd. for $C_{14}H_{26}N_4NaO_8$ [M+Na]$^+$ 401.1648, found 401.1643.

(*E*)-4-(2-Carboxyvinyl)benzoic acid (159)[348]
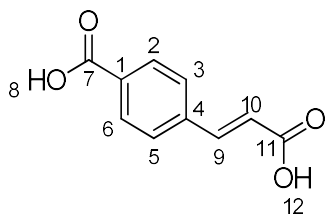
Acrylic acid (6.22 mL, 90.72 mmol) was added dropwise to a solution of potassium carbonate (16.7 g, 121 mmol) and palladium(II) acetate (136 mg, 0.600 mmol) in water (150 mL). To this solution was then added 4-iodobenzoic acid (15.0 g, 60.5 mmol), forming a slurry. This slurry was then heated at 80 °C for 4 hours under vigorous stirring, before being cooled to room temperature. The solution was filtered through Celite then acidified with 1 M HCl, causing formation of a fine, white precipitate. The precipitate was removed by vacuum filtration, and dried in a vacuum oven at 80 °C to give the pure product as a fine white powder (10.6 g, 82%).

^{1}**H NMR** (600 MHz, DMSO-d6) δ 12.82 (2H, s, 8-COO*H*, 12-COO*H*), 7.93 (2H, d, J = 8.3 Hz, 2H, 2/8-C*H*), 7.78 (2H, d, J = 8.3 Hz, 3/5-C*H*), 7.62 (1H, d, J = 16.0 Hz, 9-C*H*), 6.62 (1H, d, J = 16.0 Hz, 10-C*H*); **LCMS** (acetonitrile/water, 30-90%, 20 mins) Retention time: 7.44 min, m/z: 191.0 [M-H]$^-$.

4-(2-Carboxyethyl)benzoic acid (160)[349]
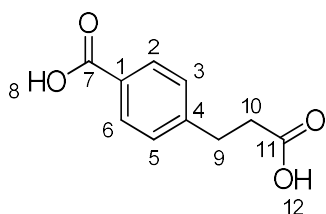
To a solution of sodium hydroxide (3.6 g, 90 mmol) and 4-[(*E*)-2-carboxyvinyl]benzoic acid (**159**, 8.65 g, 45.0 mmol) in water (250 mL) was added a small spatula of Raney nickel, and the solution was heated to 90 °C. Hydrazine hydrate (45.0 mL, 925 mmol) was then added to the solution dropwise, causing the evolution of gas. After 1 hour, gas evolution had ceased. The Raney nickel was removed by filtration, giving a colourless solution. The solution was then acidified by concentrated HCl, resulting in the formation of a white precipitate. The white precipitate was collected by vacuum filtration and washed with water (30 mL) and ethanol (20 mL). The product was then dried in the vacuum oven, giving the product as a white solid (7.82 g, 90%).

^{1}**H NMR** (500 MHz, DMSO-d6) δ 7.83 (2H, d, J = 8.0 Hz, 2/6-C*H*), 7.33 (2H, d, J = 8.0 Hz, 3/5-C*H*), 2.87 (2H, t, J = 7.5 Hz, 9-C*H$_2$*), 2.55 (2H, t, J = 7.5 Hz, 10-C*H$_2$*).

6-Methoxy-2,3-dihydro-1*H*-inden-1-one (162)[350]
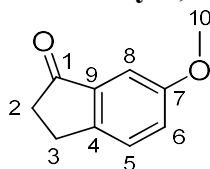
Trifluoromethanesulfonic acid (6.6 mL, 75 mmol) was added to a cooled (0 °C) solution of 3-(4-methoxyphenyl)propionic acid (2.7 g, 15 mmol) in dichloromethane (15mL) in a 20 mL microwave vial. The vial was sealed and the tube heated to 80 °C overnight. The solution was cooled to room temperature and poured into ice (approx. 60 mL). The mixture was extracted with dichloromethane (3 x 20 mL), and washed with brine (2 x 20 mL), before being dried over magnesium sulfate and concentrated under reduced

pressure. The crude material was purified by column chromatography (0 to 40%), and then recrystallized from hexanes to give the product as a light yellow, crystalline solid (856 mg, 35%).

¹H NMR (500 MHz, CDCl₃) δ 7.35 (1H, d, J = 7.7 Hz, 5-C*H*), 7.20 – 7.16 (2H, m, 6-C*H*, 8-C*H*), 3.83 (3H, s, 10-C*H₃*), 3.10 – 3.02 (2H, m, 3-C*H₂*), 2.74 – 2.67 (2H, m, 2-C*H₂*).

3-Oxo-2,3-dihydro-1*H*-indene-5-carboxylic acid (163)[153]
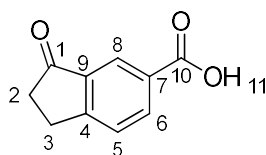
4-(2-Carboxyethyl)benzoic acid (**160**, 250 mg, 1.29mmol), aluminium trichloride (1.25 g, 9.37 mmol) and sodium chloride (250 mg, 4.28 mmol) were placed in a 25 mL round bottom flask fitted with a condenser, and heated to 180 °C. After one hour, a liquid had formed. The liquid was cooled to room temperature and ice water (7.5 mL) and 6M HCl (7.5 mL) was added, followed by ethyl acetate (20 mL). The layers were separated and the aqueous layer was washed with ethyl acetate (3 x 20 mL). The combined organic layers were then washed with 1 M HCl (20 mL), water (2 x 20 mL) and brine (20 mL), dried over MgSO₄, and concentrated under reduced pressure to give the product as a white solid (192 mg, 85%).

¹H NMR (500 MHz, DMSO-d6) δ 13.15 (1H, s, 11-COO*H*), 8.19 (1H, dd, J = 8.0, 1.6 Hz, 6-C*H*), 8.10 (1H, d, J = 1.6 Hz, 8-C*H*), 7.70 (1H, d, J = 8.0 Hz, 5-C*H*), 3.20 – 3.13 (2H, m, 3-C*H₂*), 2.73 – 2.66 (2H, m, 2-CH₂).

6-Hydroxy-2,3-dihydro-1*H*-inden-1-one (164)[351]
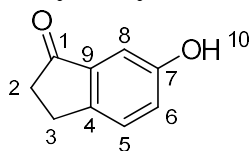
6-Methoxyindan-1-one (**162**, 5.0 g, 31 mmol) and aluminium trichloride (12.3 g, 92.0 mmol) were dissolved in toluene (100 mL) and heated to reflux, whereby a large about of gas was released. Once gas production had ceased (approx. 3 hours) the reaction was cooled to room temperature and poured into ice, forming a precipitate. The precipitate

was dissolved in ethyl acetate, and the organic phase separated, before being washed with water (2 x 30 mL), dried over $MgSO_4$, and concentrated under reduced pressure. The resulting residue was washed with CH_2Cl_2 to give the product as a brown, crystalline solid (4.04 g, 88%).

¹H NMR (500 MHz, $CDCl_3$) δ 7.35 (1H, d, J = 8.2, 5-C*H*), 7.22 (1H, d, J = 2.5 Hz, 8-C*H*), 7.16 (1H, dd, J = 8.3, 2.5 Hz, 6-C*H*), 5.98 (1H, s, 10-O*H*), 3.10 – 3.03 (2H, m, 3-C*H₂*), 2.75 – 2.69 (2H, m, C*H₂*).

6-(Pyridin-2-yloxy)-2,3-dihydro-1*H*-inden-1-one (165)
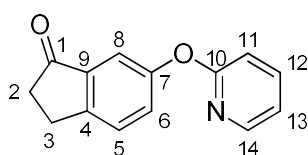
6-Hydroxyindan-1-one (**164**, 3.89 g, 26.3 mmol), 2-picolinic acid (650 mg, 5.2 mmol), copper iodide (500 mg, 2.6 mmol), and tribasic potassium phosphate (11.2 g, 52.5 mmol) were placed into a 25 mL round bottom flask and the air replaced with nitrogen. Dimethyl sulfoxide (8mL) and 2-bromopyridine (3.76 mL, 39.4 mmol) were added, and the solution heated to 100 °C under stirring. After 3 hours, the reaction mixture was poured into water (30 mL) and ethyl acetate (30 mL). The fractions were filtered through a celite plug, and the organic phase washed with water (2 x 50 mL), and brine (2 x 30 mL), before being dried over $MgSO_4$ and concentrated under reduced pressure. The crude material was purified by column chromatography (EtOAc, PE, 0% to 50%), to the product at a white solid (3.40 g, 58%)

¹H NMR (500 MHz, $CDCl_3$) δ 8.17 (1H, dd, J = 5.0, 1.5 Hz, 14-C*H*), 7.71 (1H, ddd, J = 8.2, 7.2, 2.0 Hz, 12-C*H*), 7.53 – 7.47 (2H, m, 5-C*H*, 8-C*H*), 7.40 (1H, dd, J = 8.2, 2.4 Hz, 6-C*H*), 7.01 (1H, ddd, J = 7.1, 4.9, 0.9 Hz, 13-C*H*), 6.97 (1H, dd, J = 8.3, 0.9 Hz, 11-C*H*), 3.14 (2H, t, J = 5.7 Hz, 3-C*H₂*), 2.74 (2H, t, J = 6.1 Hz, 2-C*H₂*); **¹³C NMR** (126 MHz, $CDCl_3$) δ 206.1 (1-*C*O), 163.3 (10-*C*), 153.8 (9-*C*), 151.0 (7-*C*), 147.6 (14-*C*H), 139.5 (12-*C*H), 138.5 (4-*C*), 128.2 (6-*C*H), 127.6 (5-*C*H), 118.8 (13-*C*H), 115.5 (8-*C*H), 111.9 (11-*C*H), 36.9 (2-*C*H₂), 25.4 (3-*C*H₂); **IR** (neat, v_{max}, cm⁻¹) 3054 (w), 2928 (w), 1698 (s), 1590 (m), 1465 (s), 1423 (s), 1266 (m), 1229 (s), 1175 (m), 990 (w), 913 (m), 861 (s), 782 (s), 740 (m), 697 (m); **MS (EI):** *m/z* 78 (13%), 168 (7 %), 169 (12 %), 196 (11 %),

197 (26 %), 224 (100%), 225 (65%, mass ion), 226 (22%); **LCMS** (acetonitrile/water, 30-70%, 4 mins) Retention time: 2.05 min, m/z: 226.2 [M+H]$^+$.

6-Bromo-2,3-dihydro-1*H*-inden-1-ol (166)[352]
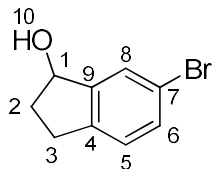
To a suspension of 6-bromoindan-1-one (**95**, 2.53 g, 12.0 mmol) in methyl alcohol (30 mL) was added portion-wise sodium borohydride (454 mg, 12.0 mmol), keeping the level of effervescence under control. After 1 hour, the effervescence had ceased, and the reaction mixture was concentrated under reduced pressure. The residue was taken up in ethyl acetate (30 mL) and water (30 mL), and the phases separated. The aqueous phase was extracted with a further 30 mL of ethyl acetate, before the combined organic phases were dried over MgSO$_4$, and concentrated under reduced pressure to give the product as a white, crystalline solid (2.39 g, 93%).

^{1}H NMR (500 MHz, CDCl$_3$) δ 7.54 (1H, d, J = 1.8 Hz, 8-C*H*), 7.37 (1H, dd, J = 8.0, 1.8 Hz, 6-C*H*), 7.12 (1H, d, J = 8.0 Hz, 5-C*H*), 5.22 (1H, t, J = 6.3 Hz, 10-O*H*), 3.05 – 2.93 (1H, m , 3-C*H*H), 2.84 – 2.70 (1H, m, 3-CH*H*), 2.59 – 2.43 (1H, m 2-C*H*H), 2.02 – 1.89 (1H, m, CH*H*).

6-(Pyridin-2-yloxy)-2,3-dihydro-1*H*-inden-1-ol (167)
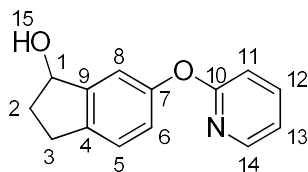
To a suspension of 6-(2-pyridyloxy)indan-1-one (**165**, 3.00 g, 13.3 mmol) in methanol (25 mL) was added portion-wise sodium borohydride (504 mg, 13.3 mmol), keeping the level of effervescence under control. The solution was allowed to stir for 45 minutes, by which point effervescence had ceased, and concentrated under reduced pressure. The residue was taken up in ethyl acetate (30 mL) and water (30 mL), and the phases separated. The aqueous phase was extracted with a further 30 mL of ethyl acetate, before the combined organic phases were washed with brine (30 mL) dried over MgSO$_4$, and

concentrated under reduced pressure to give the product as a light brown oil (2.97 g, 98%).

¹H NMR (500 MHz, CDCl₃) δ 8.15 (1H, dd, J = 5.0, 2.1 Hz, 14-C*H*), 7.66 (1H, ddd, J = 8.3, 7.2, 2.0 Hz, 12-C*H*), 7.23 (1H, d, J = 8.1 Hz, 5-C*H*), 7.15 (1H, d, J = 2.2 Hz, 8-C*H*), 7.01 (1H, dd, J = 8.1, 2.3 Hz, 6-C*H*), 6.96 (1H, ddd, J = 7.2, 5.0, 0.9 Hz, 13-C*H*), 6.89 (1H, d, J = 8.3 Hz, 11-C*H*), 5.19 (1H, t, J = 6.3 Hz, 1-C*H*OH), 3.05 – 2.97 (1H, m, 3-C*H*H), 2.84 – 2.73 (1H, m, 3-CH*H*), 2.54 – 2.46 (1H, m, 2-C*H*H), 2.32 (1H, s, 15-O*H*), 2.00 – 1.90 (1H, m, 2-CH*H*); **¹³C NMR** (126 MHz, CDCl₃) δ 164.0 (10-*C*), 153.1 (7-*C*), 147.6 (14-*C*H), 146.8 (9-*C*), 139.4 (4-*C*), 139.3 (12-*C*H), 125.7 (5-*C*H), 121.3 (6-*C*H), 118.2 (13-*C*H), 117.1 (8-*C*H), 111.5 (11-*C*H), 76.3 (1-*C*HOH), 36.4 (2-*C*H₂), 29.3 (3-*C*H₂); **IR** (neat, ν_max, cm⁻¹) 3323 (brw), 2935 (w), 2851 (w). 1590 (m), 1466 (s), 1424 (s), 1243 (s), 1167 (m), 1143 (m), 1048 (m), 921 (w), 857 (m), 822 (w) 776 (s); **MS (EI):** *m/z* 78 (23%), 198 (23%), 199 (17%), 226 (99%), 227 (100%, mass ion) 228 (17%); **LCMS** (acetonitrile/water, 30-70%, 4 mins) Retention time: 1.64 min, m/z: 228.1 [M+H]⁺.

Triphenyl(6-(pyridin-2-yloxy)-2,3-dihydro-1*H*-inden-1-yl)phosphonium   bromide (168)
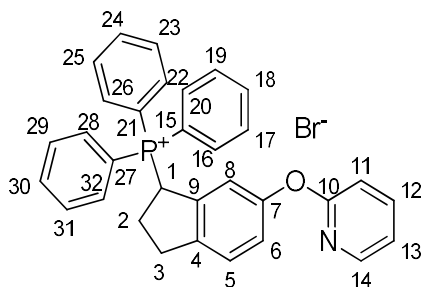
A solution of 6-(2-pyridyloxy)indan-1-ol (**167**, 1.91 g, 8.39 mmol) and triphenylphosphonium bromide (3.45 g, 10.1 mmol) in toluene (30 mL) was refluxed overnight at 110 °C, forming a white precipitate. The solution was cooled to room temperature, and the white precipitate was collected by vacuum filtration. The solid material was triturated in ether to give the product as a light pink solid (906 mg, 20%).

¹H NMR (500 MHz, CDCl₃) δ 8.06 (1H, d, J = 4.7 Hz, 14-C*H*), 7.96 – 7.80 (7H, m, 17/19/23/25/29/31-C*H*, 12-C*H*), 7.74 – 7.66 (3H, m, 18/24/30-C*H*), 7.64 – 7.54 (6H, m, 16/20/22/26/28/32-C*H*), 7.13 – 7.05 (1H, m, 13-C*H*), 7.05 – 6.84 (3H, m, 5-C*H*, 6-C*H*, 11-C*H*), 6.59 (1H, s, 8-C*H*), 3.43 – 3.20 (1H, m, 3-C*H*H), 2.82 – 2.65 (1H, m, 2-C*H*H),

2.50 – 2.32 (1H, m, 3-CH*H*), 1.78 – 1.63 (1H, m, 2-CH*H*).^{13}**C NMR** (126 MHz, CDCl$_3$) δ 162.1 (10-*C*), 152.6 (4-*C*), 145.1 (14-*C*H), 142.8 (9-*C*), 142.3 (12-*C*H), 134.8 (18/24/30-*C*H), 134.4 (17/19/23/25/29/31-*C*H), 134.3 (17/19/23/25/29/31-*C*H), 130.4 (16/20/22/26/28/32-*C*H), 130.3 (16/20/22/26/28/32-*C*H), 126.2 (5-*C*H), 121.9 (6-*C*H), 119.6 (8-*C*H), 118.8 (13-*C*H), 118.3 (15/21/27-*C*), 117.7 (15/21/27-*C*), 112.5 (11-*C*H), 39.9 (1-*C*), 39.6 (1-*C*), 30.6 (2-*C*H$_2$), 27.7 (3-*C*H$_2$); **IR** (neat, ν$_{max}$, cm^{-1}) 3442 (w), 2827 (w), 1733 (w), 1584 (m), 1486 (m), 1462 (m), 1423 (s), 1234 (s), 1110 (s), 993.4 (m), 948.6 (m), 883 (m), 842 (m), 784 (m), 706 (s), 752 (s), 725 (s), 693 (s); **HRMS (ESI):** calcd. for C$_{32}$H$_{27}$NOP [M]$^+$ 472.1825, found 472.1814; **LCMS** (acetonitrile/water, 30-70%, 4 mins) Retention time: 1.20 min, m/z: 472.1 [M]$^+$.

(3-Bromobenzyl)triphenylphosphonium bromide (172)
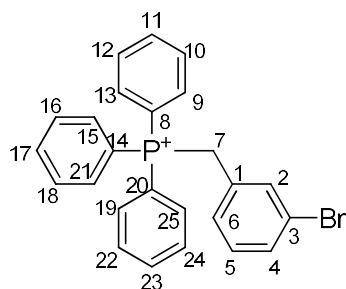
To a solution of 3-bromobenzyl bromide (1.00 g, 4.00 mmol) in toluene (13 mL) was added triphenylphosphine (1.05 g, 4.00 mmol) and the solution was refluxed for 16 hours, forming a white precipitate. The white precipitate was collected by vacuum filtration, and washed with diethyl ether (2 x 15 mL), and allowed to dry to give the product as a white, crystalline powder (1.89 g, 92 %).

^{1}**H NMR** (500 MHz, CDCl$_3$) δ 7.83 – 7.73 (9H, m, 10/12/16/18/22/24-C*H*, 11/17/23-C*H*), 7.67 – 7.59 (6H, m, 9/13/15/21/19/25-C*H*), 7.35 – 7.28 (2H, m, 4-C*H*, 6-C*H*), 7.02 – 6.94 (2H, m, 2-C*H*, 5-C*H*), 5.57 (2H, d, J = 14.7 Hz, 7-C*H$_2$*); ^{13}**C NMR** (126 MHz, CDCl$_3$) δ 135.0 (11/17/23-*C*H), 134.9 (11/17/23-*C*H), 134.5 (10/12/16/18/22/24-*C*H), 134.4 (10/12/16/18/22/24-*C*H), 134.0 (2-*C*H), 131.3 (6-*C*H), 130.6 (4-*C*H), 130.2 (5-*C*H), 130.2 (9/13/15/21/19/25-*C*H), 130.1 (9/13/15/21/19/25-*C*H), 129.8 (1-*C*), 122.2 (3-*C*Br), 118.0 (8/14/20-*C*), 117.3 (8/14/20-*C*), 30.5 (7-*C*H$_2$), 30.2 (7-*C*H$_2$); **IR** (neat, ν$_{max}$, cm^{-1}) 3054 (w), 2844 (w), 2777 (w), 1727 (s), 1456 (m), 1390 (s), 1308 (m), 1247 (s), 1217 (s), 1075 (m), 996 (m), 867 (s), 810 (m), 740 (s), 696 (s); **HRMS (ESI):** calcd. for C$_{25}$H$_{21}$BrP [M]$^+$ 431.0559, found 431.0553.

Benzyl 3-formyl-1*H*-indole-1-carboxylate (179)[353]
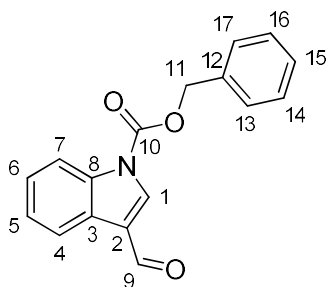
To a suspension of sodium hydride (220 mg, 16.5 mmol) in tetrahydrofuran (15 mL) at 0 °C was added portion-wise indole-3-carboxaldehyde (2.2 g, 15 mmol) over 20 minutes. The solution was allowed to warm to room temperature and stir for a further 20 minutes. The solution was then cooled to 0 °C and benzyl chloroformate (2.36 mL, 16.5 mmol) was added, before the solution was allowed to warm to room temperature and stir overnight. The solution was diluted with water (50 mL) and EtOAc (50 mL), and the aqueous phase was further extracted with ethyl acetate (2 x 50 mL). The combined organic phase was washed with saturated NaHCO$_3$ solution (40 mL) and brine (2 x 40 mL), before being dried over magnesium sulphate and concentrated under reduced pressure to give a brown oil (5.092 g). The crude material was recrystallized from to give the product as a white crystalline solid (2.311 g, 55%).

^{1}H NMR (500 MHz, Chloroform-d) δ 10.10 (1H, s, 9-C*O*), 8.30 (1H, d, J = 7.9 Hz, 4-C*H*), 8.26 (1H, s, 1-C*H*), 8.20 (1H, d, J = 8.1 Hz, 7-C*H*), 7.55 – 7.50 (2H, m, 13/17-C*H*), 7.48 – 7.37 (m, 5H, 5-C*H*, 6-C*H*, 14/16-C*H*, 15-C*H*), 5.52 (2H, s, 11-C*H$_2$*); **LCMS** (acetonitrile/water, 30-70%, 10 mins) Retention time: 2.62 min, m/z: 280.0 [M+H]$^+$.

(*E/Z*)-Benzyl 3-((6-bromo-2,3-dihydro-1*H*-inden-1-ylidene)methyl)-1*H*-indole-1-carboxylate (180)
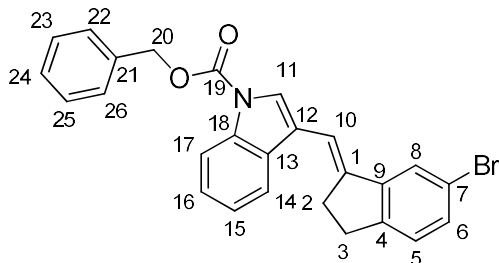
To a stirred suspension of (6-bromoindan-1-yl)-triphenyl-phosphonium bromide (**94**, 1.62 g, 3.00 mmol) in tetrahydrofuran (15 mL) at 0 °C was added sodium hydride (132 mg, 3.3 mmol). The solution was stirred for a further 20 minutes before adding benzyl 3-

formylindole-1-carboxylate (**179**, 920 mg, 3.3 mmol). The solution was then allowed to warm to room temperature and stir for a further 18 hours. The solution was then concentrated under reduced pressure, giving a light brown precipitate. This precipitate was washed with water (40 mL) and CH_2Cl_2 (30 mL) before triturating in ethyl acetate (30 mL) to give the product as a fine white powder (671 mg, 49 %).

^{1}H NMR (500 MHz, CDCl$_3$) δ 8.19 (1H, s, 14-C*H*), 7.82 – 7.74 (2H, m, 15-C*H*, 17-C*H*), 7.71 (1H, s, 11-C*H*) 7.55 – 7.50 (2H, m, 22/26-C*H*), 7.48 – 7.30 (6H, m, 23/25-C*H*, 16-C*H*, 24-C*H*, 8-C*H*, 6-C*H*), 7.16 (1H, d, J = 7.9 Hz, 5-C*H*), 7.08 (1H, s, 10-C*H*), 5.51 (2H, s, 20-C*H$_2$*), 3.21 – 2.77 (4H, m, 2-C*H$_2$*, 3-C*H$_2$*); **^{13}C NMR** (126 MHz, CDCl$_3$) δ 151.0 (19-*C*O), 145.0 (1-*C*), 144.4 (4-*C*), 143.3 (9-*C*), 135.2 (21-*C*), 135.0 (18-*C*), 130.7 (6-*C*H), 130.4 (13-*C*), 128.8 (23/25-*C*H), 128.7 (16-*C*H), 128.4 (22/26-*C*H), 126.7 (5-*C*H), 125.2 (24-*C*H), 123.2 (8-*C*H), 122.4 (11-*C*H), 120.8 (17-*C*H), 119.2 (7-*C*Br), 118.9 (15-*C*H), 117.7 (12-*C*), 115.3 (14-*C*H), 108.4 (10-*C*H), 68.8 (20-*C*H$_2$), 31.7 (3-*C*H$_2$), 30.2 (2-*C*H$_2$); **IR** (neat, $ν_{max}$, cm^{-1}) 3052 (w), 2919 (w), 1726 (s), 1455 (s), 1400 (s), 1366 (s), 1352 (s), 1309 (s), 1259 (s), 1250 (s), 1219 (s), 1173 (m), 1091 (m), 1034 (m), 961 (m), 867 (m), 795 (m), 758 (s), 741 (s), 691 (s); **HRMS (ESI):** calcd. for $C_{26}H_{20}BrNNaO_2$ [M+Na]$^+$ 480.0570, found 480.0572.

(*E/Z*)-Benzyl 3-((6-(pyridin-2-yloxy)-2,3-dihydro-1*H*-inden-1-ylidene)methyl)-1*H*-indole-1-carboxylate (181)
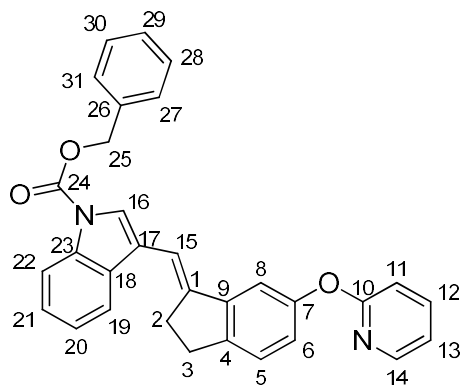
To a stirred suspension of triphenyl-[6-(2-pyridyloxy)indan-1-yl]phosphonium bromide (**168**, 750 mg, 1.36 mmol) in dichloromethane (4 mL) and 18-crown-6 (3.6 mg, 0.01 mmol) in a screw-top reactivial under nitrogen gas was added potassium carbonate (1.87 g, 13.6 mmol). The solution was stirred for half an hour, and benzyl 3-formylindole-1-carboxylate (**179**, 417 mg, 1.49 mmol) was added. The solution was then allowed to stir

room temperature overnight (approximately 18 hours). The solution was then diluted with water (20 mL) and dichloromethane (20 mL). The aqueous phase was washed with dichloromethane (20 mL), and the combined organic phases were washed with brine (20 mL) and dried over magnesium sulfate, filtered, and concentrated under reduced pressure. The resulting residue was purified by column chromatography (PE/EtOAc, 0% to 20%), causing a crystalline solid to form in the product fraction test tubes. This was collected by vacuum filtration, giving the product as a light yellow crystalline solid (132 mg, 21%).

^{1}H NMR (500 MHz, CDCl$_3$) δ 8.24 (1H, dd, J = 5.1, 2.0 Hz, 14-CH), 8.19 (1H, s, 19-CH), 7.76 – 7.67 (3H, m, 8-CH, 12-CH, 16-CH), 7.53 (2H, d, J = 7.3 Hz, 27/31-CH), 7.47 – 7.29 (7H, m, 22-CH, 28/30-CH, 21-CH, 29-CH, 5-CH, 20-CH), 7.06 – 6.98 (3H, m, 15-CH, 6-CH, 13-CH), 6.96 (1H, d, J = 8.3 Hz, 11-CH), 5.52 (2H, s, 25-CH_2), 3.18 – 3.13 (2H, m, 3-CH_2), 3.10 – 3.06 (2H, m, 2-CH_2); **^{13}C NMR** (126 MHz, CDCl$_3$) δ 164.2 (10-C), 153.4 (7-C), 151.0 (24-CO), 147.8 (14-CH), 144.3 (4-C), 143.8 (9-C), 142.7 (1-C), 139.3 (12-CH), 135.2 (26-C), 135.0 (23-C), 130.5 (18-C), 128.8 (28/30-CH), 128.7 (21-CH), 128.4 (27/31-CH), 126.2 (5-CH), 125.0 (29-CH), 123.1 (20-CH), 122.2 (16-CH), 121.5 (6-CH), 119.4 (17-C), 119.0 (8-CH), 118.2 (13-CH), 115.3 (19-CH), 112.8 (22-CH), 111.3 (11-CH), 107.9 15-CH), 68.8 (25-CH$_2$), 32.1 (2-CH$_2$), 30.2 (3-CH$_2$); **IR** (neat, ν$_{max}$, cm^{-1}) 3034 (w), 2929 (w), 1726 (s), 1591 (m), 1470 (s), 1458 (s), 1429 (s), 1399 (s), 1306 (m), 1240 (s), 1214 (s), 1073 (m), 1014 (m), 913 (w), 852 (m), 745 (s); **HRMS (ESI):** calcd. for C$_{62}$H$_{48}$N$_4$NaO$_6$ [2M+Na]$^+$ 967.3466, found 967.3432; **MS (EI):** m/z 91 (23%), 243 (76%), 337 (58%), 368 (12%), 428 (11%), 472 (70%).

(*E/Z*)-Benzyl 3-(3-bromostyryl)-1*H*-indole-1-carboxylate (182)
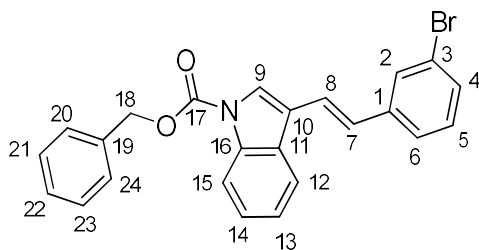
To a stirred suspension of (3-bromophenyl)methyl-triphenyl-phosphonium bromide (**172**, 102 mg, 0.200 mmol) in dichloromethane (0.4 mL) and 18-crown-6 (1 mg, 0.01 mmol) in a screw-top reactivial under nitrogen gas was added potassium carbonate (276 mg, 2.00 mmol). The solution was stirred for half an hour, before benzyl 3-formylindole-1-carboxylate (**179**, 61 mg, 0.22 mmol) was added. The solution was then allowed to stir

at ambient temperatures overnight (approx. 18 hours). The solution was then diluted with water (20 mL) and CH_2Cl_2 (20 mL). The aqueous phase was washed with CH_2Cl_2 (20 mL), and the combined organic phases were washed with brine (20 mL) and dried over $MgSO_4$, before being filtered and concentrated under reduced pressure. The resulting residue was purified by column chromatography (PE/EtOAc, 0% to 10%) to give *trans* and *cis* isomers of the product. The combined weight of both isomers was 44 mg, as a white crystalline solid (56% yield).

(*cis* isomer)

^{1}H NMR (500 MHz, CDCl$_3$) δ 8.14 (1H, d, J = 8.2 Hz, 28-C*H*), 7.46 (1H, s, 2-C*H*), 7.42 (1H, s, 9-C*H*), 7.40 – 7.25 (8H, m, 20/24-C*H*, 21/23-C*H*, 13-C*H*, 15-C*H*, 4-C*H*, 6-C*H*), 7.22 – 7.13 (2H, m, 22-C*H*, 14-C*H*), 7.03 (1H, t, J = 7.8 Hz, 5-C*H*), 6.66 – 6.58 (2H, m, 7-C*H*, 8-C*H*), 5.37 (2H, s, 18-C*H$_2$*); **^{13}C NMR** (126 MHz, CDCl$_3$) δ 150.3 (17-*C*O), 139.9 (1-*C*), 138.0 (16-*C*), 135.0 (19-*C*), 131.5 (2-*C*H), 130.2 (4-*C*H), 129.8 (5-*C*H), 129.8 (7-*C*H), 128.7 (21/23-*C*H), 128.6 (22-*C*H), 128.2 (20/24-*C*H), 127.1 (6-*C*H), 124.9 (15-*C*H), 123.7 (9-*C*H), 123.7 (11-*C*), 123.0 (14-*C*H), 122.4 (3-*C*Br), 121.1 (8-*C*H), 119.8 (13-*C*), 117.4 (10-*C*), 115.1 (12-*C*H), 68.7 (18-*C*H$_2$); **IR** (neat, ν$_{max}$, cm^{-1}) 3034 (w), 1739 (s), 1589 (w), 1566 (w), 1456 (s), 1395 (m), 1350 (m), 1255 (s), 1221 (s), 1074 (m), 1018 (w), 747 (m), 697 (m); **HRMS (ESI):** calcd. for $C_{24}H_{19}BrNO_2$ [M+H]$^+$ 434.0579, found 434.0574

(*trans* isomer)

^{1}H NMR (500 MHz, CDCl$_3$) δ 8.24 (1H, d, J = 8.0 Hz, 12-C*H*), 7.88 (1H, dd, J = 7.6, 1.4 Hz, 15-C*H*), 7.78 (1H, s, 9-C*H*), 7.67 (1H, d, J = 1.8 Hz, 2-C*H*), 7.50 (2H, d, J = 7.1 Hz, 20/24-C*H*), 7.46 – 7.29 (7H, m, 21/23-C*H*, 13-C*H*, 14-C*H*, 4-C*H*, 5-C*H*, 6-C*H*), 7.25 – 7.16 (2H, m, 7-C*H*, 22-C*H*), 7.11 (1H, d, J = 16.5 Hz, 8-C*H$_2$*), 5.47 (2H, s, 18-C*H$_2$*); **^{13}C NMR** (126 MHz, CDCl$_3$) δ 150.7 (17-*C*O), 139.8 (1-*C*), 136.2 (16-*C*), 135.0 (19-*C*), 130.2 (4-*C*H), 130.1 (5-*C*H), 129.8 (2-*C*H), 128.9 (22-*C*H), 128.8 (21/23-*C*H), 128.5 (20/24-*C*H), 127.7 (8-*C*H), 125.2 (14-*C*H), 124.8 (6-*C*H), 123.9 (15-*C*), 123.4 (9-*C*H), 123.0 (11-*C*), 122.9 (3-*C*Br), 121.2 (7-*C*H), 120.0 (13-*C*H), 119.5 (10-*C*H), 115.5 (12-*C*H), 68.9 (18-*C*H$_2$); **IR** (neat, ν$_{max}$, cm^{-1}) 3035 (w), 2983 (w), 1739 (s), 1455 (m), 1393 (m), 1356 (m), 1247 (s), 1095 (m), 1004 (m), 956 (m), 910 (m), 747 (m), 696 (s); **LRMS (EI):** *m/z* 91 (100%), 217 (34%), 387 (66%), 389 (65%), 431 (46%, mass ion), 433 (48%, mass ion).

2-(1*H*-Indol-3-yl)ethanol (185)[354]
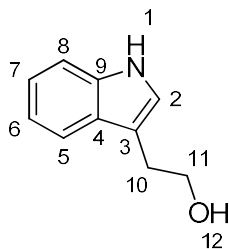
Phenylhydrazine (7.7 mL, 78 mmol) was added to 1M HCl solution (80 mL), and topped up with water (20 mL) and sulfuric acid (2.71 mL, 48.2 mmol). To this solution was added *N,N*-dimethylacetamide (100 mL), and heated to 100 °C. 2,3-dihydrofuran (6.04 mL, 78.3 mmol) was added dropwise over 5 mins to the solution. The solution was then stirred at 100 °C for 2 hours, developing a red colour. The solution was extracted with ethyl acetate (3 x 100 mL), and the combined organic phase was washed with 0.5 M LiCl solution (100 mL), water (2 x 100 mL) and brine (2 x 100 mL). The organic phase was then dried over magnesium sulfate, and concentrated under reduced pressure to give an orange oil. The orange oil was purified by column chromatography (PE:EtOAc, 0-65%, 100 g silica) to give the product as a light orange, crystalline solid (5.89 g, 47%).

¹H NMR (500 MHz, DMSO-*d₆*) δ 10.74 (1H, s, 1-N*H*), 7.50 (1H, d, *J* = 7.9 Hz, 5-C*H*), 7.32 (1H, d, *J* = 8.1 Hz, 8-C*H*), 7.14 – 7.10 (1H, m, 2-C*H*), 7.04 (1H, d, *J* = 7.7 Hz, 6-C*H*), 6.96 (1H, d, *J* = 7.4 Hz, 7-C*H*), 4.58 (1H, t, *J* = 5.4 Hz, 12-O*H*), 3.65 (2H, td, *J* = 7.3, 5.4 Hz, 11-C*H₂*), 2.84 (2H, d, *J* = 7.3 Hz, 10-C*H₂*).

3-(2-Bromoethyl)-1*H*-indole (186)[355]
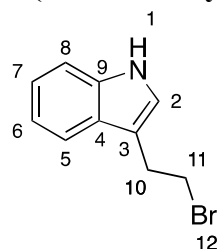
To a solution of triphenylphosphine (680 mg, 2.6 mmol) and 2-(1*H*-indol-3-yl)ethanol (**185**, 320 mg, 2.0 mmol) in dichloromethane (3 mL) at 0 °C was added dropwise a solution of tetrabromomethane (860 mg, 2.6 mmol) in dichloromethane (2 mL). The solution was allowed to warm to room temperature and stir for a further 3 hours. The solution was then concentrated under reduced pressure, giving a brown oil. This crude

mixture was purified by column chromatography (P:E/EtOAc, 0-25%), giving 329 mg of a light brown solid. The brown solid was then recrystallized from PE (100-120 °C fraction), to give the product as a crystalline, off-white solid (329 mg, 73%).

¹H NMR (500 MHz, Chloroform-d) δ 8.01 (1H, s, 1-N*H*), 7.62 (1H, d, J = 7.5 Hz, 5-C*H*), 7.39 (1H, d, J = 7.5 Hz, 8-C*H*), 7.24 (1H, dd, J = 7.5 Hz, 6-C*H*), 7.17 (1H, dd, J = 7.5 Hz, 7-C*H*), 7.10 (1H, s, 2-C*H*), 3.67 (2H, t, J = 7.7 Hz, 11-C*H₂*Br), 3.37 (2H, t, J = 7.7 Hz, 10-C*H₂*).

(*E/Z*)-Benzyl 3-((6-(phenylethynyl)-2,3-dihydro-1H-inden-1-ylidene)methyl)-1*H*-indole-1-carboxylate (194)
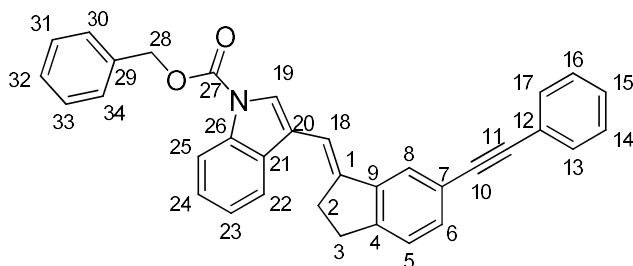
In a screw cap reaction vial with a septum was added a solution of bis(triphenylphosphine)palladium(II) dichloride (2.81 mg, 0.01 mmol), copper iodide (1.5 mg, 0.010 mmol), triethylamine (139 μL, 1.00 mmol) and benzyl 3-[(*E/Z*)-(6-bromoindan-1-ylidene)methyl]indole-1-carboxylate (**180**, 92 mg, 0.20 mmol), before being purged with nitrogen. Phenylacetylene (26 μL, 0.24mmol) was then added, and the sealed vial was heated to 70 °C. The solution was allowed to stir overnight before being cooled to room temperature. The solution was diluted with ethyl acetate (20 mL) and water (20 mL), and the aqueous phase further extracted with ethyl acetate (10 mL). The combined organic phase was then washed with saturated sodium bicarbonate solution (2 x 20 mL), 1M HCl solution (2 x 20 mL), and brine (20 mL), before being concentrated onto silica and purified by flash column chromatography (0% to 25% EtOAc:PE). Precipitate from the chromatographic elute were collected by vacuum filtration to give the product as a white crystalline solid (23 mg, 24%).

¹H NMR (500 MHz, Chloroform-*d*) δ 8.20 (1H, s, 22-C*H*), 7.85 (1H, s, 8-C*H*), 7.80 (1H, d, *J* = 7.8 Hz, 25-C*H*), 7.73 (1H, s, 19-C*H*), 7.58 (2H, d, *J* = 7.8 Hz, 13/17-C*H*), 7.53 (2H, d, *J* = 7.3 Hz, 30/34-C*H*), 7.48 – 7.33 (7*H*, m, 5-C*H*, 31/33-CH, 6-C*H*, 32-C*H*, 14/16-C*H*), 7.32 – 7.27 (1H, m, 15-C*H*), 7.18 – 7.31 (1H, m, 18-C*H*), 5.52 (2H, s, 28-C*H₂*), 3.20

– 3.14 (2H, m, 3-C*H₂*), 3.09 – 3.02 (2H, m, 2-C*H₂*); **¹³C NMR** (126 MHz, CDCl3) δ 151.0 (27-*C*O), 146.7 (4-*C*), 143.8 (9-*C*), 142.3 (1-*C*), 135.2 (29-*C*), 131.6 (13/17-*C*H), 131.3 (6-*C*H), 130.5 (26-*C*), 128.8 (31/33-*C*H), 128.7 (5-*C*H), 128.4 (14/16-*C*H), 128.3 (30/34-*C*H), 128.1 (24-*C*H), 125.3 (15-*C*H), 125.1 (32-*C*H), 123.4 (7-*C*), 123.2 (23-*C*H), 123.1 (8-*C*H), 122.2 (21-*C*), 121.6 (12-*C*), 119.4 (20-*C*), 119.0 (25-*C*H), 115.3 (22-*C*H), 107.9 (18-*C*H), 89.8 (10-*C*), 88.8 (11-*C*), 68.8 (28-*C*H₂), 31.6 (2-*C*H₂), 30.7 (3-*C*H₂); **IR** (neat, ν_{max}, cm^{-1}) 3049 (w), 2915 (w), 1727 (s), 1543 (w), 1455 (s), 1392 (s), 1352 (m), 1396 (m), 1249 (s), 1218 (s), 1076 (m), 1013 (m), 867 (m), 754 (s), 740 (s), 698 (m), 613 (m); **HRMS (ESI):** calcd. for $C_{68}H_{50}N_2NaO_4$ [2M+Na]$^+$ 981.3663, found 981.3656; **LRMS (EI):** *m/z* 91 (87%), 344 (100%), 355 (45%), 435 (13%), 436 (6%), 479 (51%, mass ion), 480 (19%, mass ion).

3-((6-Phenethyl-2,3-dihydro-1*H*-inden-1-yl)methyl)-1*H*-indole (196)
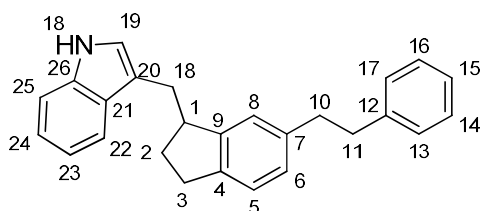
A stirred solution of benzyl 3-[(*E/Z*)-[6-(2-phenylethynyl)indan-1-ylidene]methyl]indole-1-carboxylate (**194**, 23 mg, 0.05 mmol) and palladium on charcoal (2 mg) in tetrahydrofuran (5 mL) was placed under an atmosphere of hydrogen gas. The solution was stirred until no starting material was present by TLC (approximately 42 hours). The hydrogen atmosphere was then replaced with nitrogen gas. The solution was filtered through a celite plug to remove any palladium on charcoal, and concentrated under reduced pressure. The crude material was purified using prep-scale HPLC (acetonitrile/water, 50% to 95%, 20 minutes), to give the product as a light brown solid (3 mg, 18%)

¹H NMR (500 MHz, CDCl₃) δ 7.92 (1H, s, 18-N*H*), 7.65 (1H, d, J = 7.9 Hz, 22-C*H*), 7.38 (1H, d, J = 8.1 Hz, 25-C*H*), 7.33 – 7.28 (2H, m, 13/17-C*H*), 7.23 – 7.18 (4H, m, 14/16-C*H*, 15-C*H*, 19-C*H*), 7.17 – 7.13 (2H, m, 6-C*H*, 24-C*H*), 7.03 (2H, s, 5-C*H*, 23-C*H*), 6.92 (1H, s, 8-C*H*), 3.60 – 3.50 (1H, m, 7-C*H*), 3.23 (1H, dd, J = 14.6, 5.5 Hz, 18-C*H*H), 2.91 – 2.73 (7H, m, 3-C*H₂*, 10-C*H₂*, 11-C*H₂*, 18-CH*H*), 2.25 – 2.17 (1H, m, 2-C*H*H), 1.89 – 1.78 (1H, m, 2-CH*H*); **¹³C NMR** (126 MHz, CDCl₃) δ 147.6 (4-*C*), 142.0

(12-*C*), 141.7 (9-*C*), 139.6 (7-*C*), 136.3 (26-*C*), 128.4 (14/16-*C*H), 128.2 (13/17-*C*H), 126.6 (21-*C*), 125.8 (15-*C*H), 124.2 (6-*C*H), 124.0 (5-*C*H), 121.9 (19-*C*H), 121.8 (8-*C*H), 119.2 (22-*C*H), 119.0 (24-*C*H), 115.3 (23-*C*H), 113.2 (20-*C*), 111.0 (25-*C*), 45.4 (1-*C*H), 38.2 (10-*C*H$_2$), 37.8 (11-*C*H$_2$), 32.4 (3-*C*H$_2$), 30.8 (2-*C*H$_2$, 18-*C*H$_2$); **IR** (neat, ν_{max}, cm^{-1}) 3641 (w), 2959 (s), 2926 (s), 2871 (m), 1724 (m), 1435 (m), 1383 (w), 1351 (w), 1273 (m), 1160 (m), 1121 (m), 1059 (w), 970 (w), 753 (w); **MS (EI):** *m/z* 91 (57%), 120 (37%), 129 (34%), 219 (100%), 243 (26%), 337 (24%), 352 (85%, mass ion).

6.3   Biochemical Procedures

6.3.1   Circular Dichroism

Circular dichroism was carried out on a JASCO J-715 series CD spectrometer. The protein was suspended as a 1.2 mg/mL solution in deionised water. This solution was diluted to a 0.3 mg/mL solution using deionised water, then placed in a 0.5 mm cuvette. The spectrum was recorded between 300 nm and 180 nm, and baseline-corrected by subtraction of a spectrum of deionised water. The data was then converted to deg. Cm^2.dmol^{-1} for analysis.

6.3.2   Miniprotein Refolding General Procedure

A lyophilised sample of 1.0 mg peptide sequence was suspended in phosphate buffer (3 mL, pH 8.0). This solution was mixed with trifluoroethanol (6 mL), and TCEP solution added (0.5 M, 30 μL, pH 7.0) before being allowed to stir for three hours. DMSO (1 mL) was then added to quench any remaining TCEP, and the solution stirred in oxygen for 48 hours to promote the formation of disulphide bridges. The resulting solution was concentrated under a stream of pressurised air at 30 °C, giving a solid residue ready to be desalted using Waters Sep-Pak C-18 200mg cartridges.

6.4   Computational Procedures

6.4.1   Small Molecule Mimetics: Markush Enumeration

Aim: To generate a database of structures designed to mimic BRCA2, with filters to remove any structurally unstable or unwanted functionalities generated.

Software Package(s): KNIME version 2.9.1 (KNIME AG, 8005 Zurich, Switzerland, https://www.knime.org), Infocom JchemExtentions 2.6.3.v0200 (Infocom Corp., Shibuya-ku, Tokyo, 150-0001, https://www.infocom.co.jp/bio/knime/)

General Procedure:
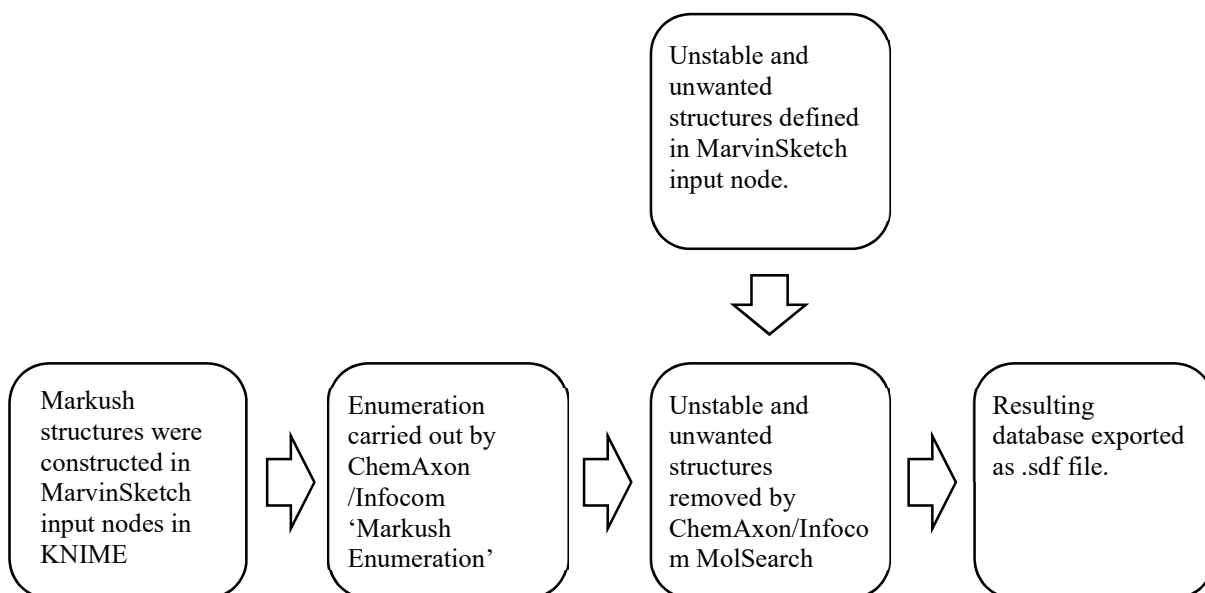
Details and Parameters:

Markush Structures

Markush structures were constructed in the ChemAxon 'Marvin Sketch' node. The 'atom list' and 'link node' tools were used to define parameters for each atomic centre. The information was output as the ChemAxon 'MrvCell' format. An example of a typical Markush structure is shown below.
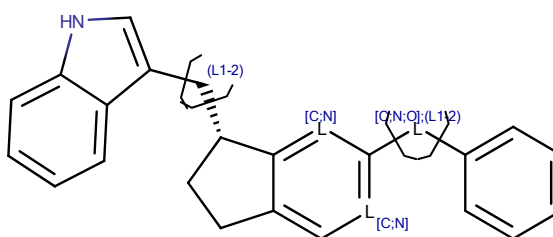
Figure 6.1: A typical Markush structure for enumeration.

Markush Enumeration

The Infochem/Chemaxon 'Markush Enumeration' node was used for enumeration, with the following parameters:

- Calculation: Sequential enumeration.
- Calculation: Generate all enumerations.

- Output: MrvCell

Structure Filters

The Infochem/Chemaxon 'MolSearch' node was used in combination with the 'Marvin Sketch' node for structure filtering, with the following parameters:

- Search type: Substructure

The functionalities specified in 'Marvin Sketch' nodes for removal were the aliphatic heteroatom-heteroatom bonds O-N, N-N and O-O.

6.4.2 Small Molecule Mimetics: Conformer Generation And Alignment To BRCA2

Aim: To generate conformers of BRCA2 mimetic designs, aligning them to the hotspot residues on the BRCA2 helix, and determining the best aligning structures.

Software Package(s): KNIME version 2.9.1, RDKit (Greg Landrum, RDKit: Open-source cheminformatics, http://www.rdkit.org), PyMOL 1.7.6.6 (Schrodinger, LLC, Cambridge, Cambridgeshire, UK, CB1 2JD, https://www.pymol.org).

General Procedure:

Database of structures imported from .sdf file. → Conformers generated using looped 'RDKit Add Conformers' Node. → Conformers aligned to BRCA2 structure using 'RDKit open 3D Alignment' node. ← Structure of BRCA2 edited to only leave hot-spot residues, and imported as .sdf structure.

↓

Results sorted by alignment score, and filtered to only contain the best results. → The best aligning conformers exported to .sdf database for further analysis.

Details and Parameters:

Import of databases

A database of BRCA2 mimetic designs was imported to the KNIME workflow in .sdf format. The database was previously manually curated, or generated using Markush enumeration (6.4.1).

Conformer Generation

Conformers were generated for each structure using the 'RDKit Add Conformers' node. The node was placed into a variable loop, with each iteration of the loop containing one structure for conformer generation. The conformers were concatenated into a single database at the loop end. The 'RDKit Add Conformers' node was configured with the following parameters:

- Number of conformers: 5000
- Maximum number of tries to generate conformers: 200
- Random seed: -1
- RMS threshold for keeping a conformer: -1.0
- Use random co-ordinates as a starting point for distance geometry: No
- Perform a clean-up using universal forcefield after calculation: Yes

Importing BRCA2 Hotspot residues

The necessary BRA2 hotspot residues for the alignment, as well as their backbone connecting atoms, were extracted from the PALB2/BRCA2 crystal structure (3EU7). This was done in PyMOL using the structure editing tools, and saved as an .sdf file for import into the KNIME workflow.
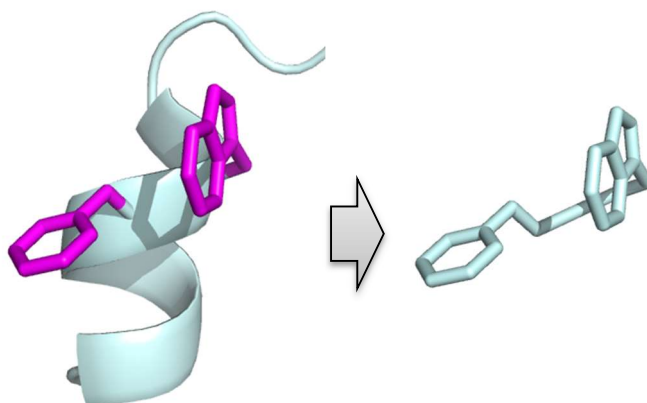
Figure 6.2: Extraction of *i*, *i*+1 residues for mimetic alignment studies.

Aligning Conformers to BRCA2 Hotspots

The conformers generated using the 'RDKit Add Conformers' node were aligned to the modified helix residues using the 'RDKit Open 3D Alignment' node. The parameters set were:

- Allow reflection: No

- Maximal number of iterations: 50

- Accuracy (0 – most accurate, 3 – least accurate): 0

When the alignment was done, the results were sorted by score, and the top 200 results exported to .sdf database format for further analysis.

6.4.3   Small Molecule Mimetics: MOE E_Strain

Aim: To calculate the strain that the best aligning conformers are under using MOE E_Strain tools, to help remove any highly strained conformers.

Software Packages: MOE 2015.1001 (Chemical Computing Group, Montreal, Quebec, Canada, H3A 2R7, https://www.chemcomp.com/).

General Procedure:

Import .sdf file of best aligning conformers into MOE .mdb format. ⇒ Calculate E_Strain for each database entry in MOE.
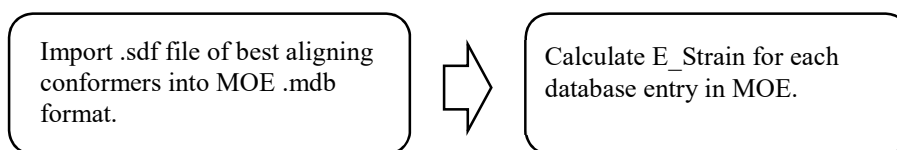
Details and Parameters:

The E_Strain calculations were carried out in the MMFF94x environment in MOE. The MOE database viewer was then used to calculate E_Strain, found in Compute > Descriptors > Calculate > E_Strain. The database could then be manipulated in MOE, or exported an .sdf with E_Strain values for manipulation in KNIME.

6.4.4   Small Molecule Mimetics: Small Molecule Docking in MOE

Aim: To dock the best aligning BRCA2 mimetics into the binding site on PALB2 as a second prediction for binding affinity.

Software Package(s): MOE 2015.1001.

General Procedure:

```
                                              ┌──────────────────┐
                                              │ Best aligning    │
                                              │ conformers with  │
                                              │ low E_Strain     │
                                              │ imported into    │
                                              │ MOE.             │
                                              └──────────────────┘
                                                      ⬇
┌────────────────┐   ┌────────────────┐   ┌────────────────┐   ┌────────────────┐
│ BRCA2/PALB2    │   │ Pharmacophore  │   │ Conformers     │   │ Results of     │
│ 3EU7 .pdb      │   │ parameters     │   │ docked into    │   │ docking study  │
│ structure      │⇨  │ developed based│⇨  │ PALB2 using the│⇨  │ exported as .sdf│
│ imported into  │   │ on the BRCA2   │   │ pharmacophore to│   │ database for   │
│ MOE and        │   │ hot-spot residues│ │ guide the *i* and│  │ analysis.      │
│ prepared.      │   │                │   │ *i*+1 positions.│   │                │
└────────────────┘   └────────────────┘   └────────────────┘   └────────────────┘
```

Details and Parameters:

Import of Protein Crystal Structure into MOE

The 3EU7 crystal structure was downloaded from the protein data bank, and opened in MOE. Before docking was carried out, the structure was prepared using the 'Structure Preparation' and 'Protonate 3D' tools. The BRCA2 helix was redefined as a ligand rather than a protein by isolating the structure and then exporting as a .mol2 file. The .mol2 file was then opened into MOE, causing it to recognise the structure as a small molecule. The session with the modified BRCA2 and the PALB2 structure from 3EU7 was then exported as a .pdb file, so that when it was imported MOE recognised the structure as a protein/ligand rather than a protein/protein interaction.

Preparation of Pharmacophore Parameters for Docking

Pharmacophores were generated from the structure of BRCA2 bound to PALB2. 4 pharmacophores were used to direct the *i* and *i*+1 groups from the designed mimetics into the correct position in the PALB2 binding site. These were two aromatic rings and a nitrogen atom for the *i* indole ring, and an aromatic ring for the *i*+1 phenyl group (Figure 6.3)
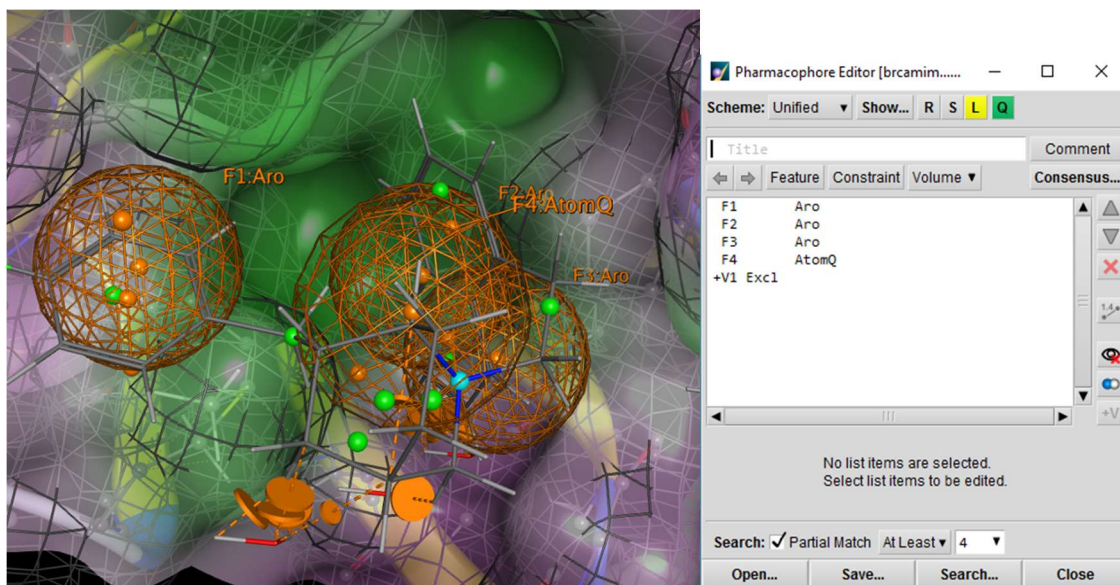
Figure 6.3: Use of pharmacophores to guide the docking of BRCA2 mimetics

Docking Designed BRCA2 Mimetics into PALB2

The best BRCA2 mimetic from alignment studies were imported into MOE as an .sdf database. These were then docked into the PALB2 binding site using the 'Dock' function in MOE with the following parameters:

- Forcefield: Amber10:EHT
- Receptor: Receptor+Solvent
- Site: Ligand atoms
- Placement: α-Triangle
    - Minimum iterations: 300 000
    - Maximum Iterations: 10 000 000
    - Timeout: 600 seconds
- Refinement: Induced fit
    - Termination criterion: Gradient = 0.01, iterations = 500
    - Pharmacophore restraint: Force constant = 100, radius offset = 0.4

6.4.5   Small Molecule Mimetics: 4-Substituted Indole Docking in MOE

Aim: To dock a database of 4-substituted indoles into the PALB2 binding site, to determine whether the Leu binding site can be targeting with these structures.

Software Package(s): MOE 2015.1001.

General Procedure:

```
                              ┌─────────────────┐
                              │ Library of 4-   │
                              │ substituted     │
                              │ indoles         │
                              │ generated by    │
                              │ Markush         │
                              │ enumeration     │
                              │ imported to MOE.│
                              └─────────────────┘
                                       ⇩
┌──────────────┐   ┌──────────────┐   ┌──────────────┐   ┌──────────────┐
│ BRCA2/PALB2  │   │ Pharmacophore│   │ Indoles      │   │ Results of   │
│ 3EU7 .pdb    │ ⇨ │ parameters   │ ⇨ │ docked into  │ ⇨ │ docking study│
│ structure    │   │ developed    │   │ PALB2 using  │   │ exported as  │
│ imported into│   │ based on the │   │ the          │   │ .sdf         │
│ MOE and      │   │ tryptophan   │   │ pharmacophore│   │ database for │
│ prepared.    │   │ indole ring. │   │ to guide the │   │ analysis.    │
│              │   │              │   │ i and i+1    │   │              │
│              │   │              │   │ positions.   │   │              │
└──────────────┘   └──────────────┘   └──────────────┘   └──────────────┘
```

Details and Parameters:

Import of Protein Crystal Structure into MOE

The 3EU7 crystal structure was downloaded from the protein data bank, and opened in MOE. Before docking was carried out, the structure was prepared using the 'Structure Preparation' and 'Protonate 3D' tools. The BRCA2 helix was redefined as a ligand rather than a protein by isolating the structure and then exporting as a .mol2 file. The .mol2 file was then opened into MOE, causing it to recognise the structure as a small molecule. The session with the modified BRCA2 and the PALB2 structure from 3EU7 was then exported as a .pdb file, so that when it was imported MOE recognised the structure as a protein/ligand rather than a protein/protein interaction.

Preparation of Pharmacophore Parameters for Docking

Pharmacophores were generated from the structure of BRCA2 bound to PALB2. A pharmacophore was specified to detail each of the 9 atoms in the indole ring, as well as the two aromatic ring centres. It was found that this allowed MOE to place the indole structures into the correct position, allowing it to optimise the conformation of the 4-position. This resulted in a total of 11 pharmacophores overall.
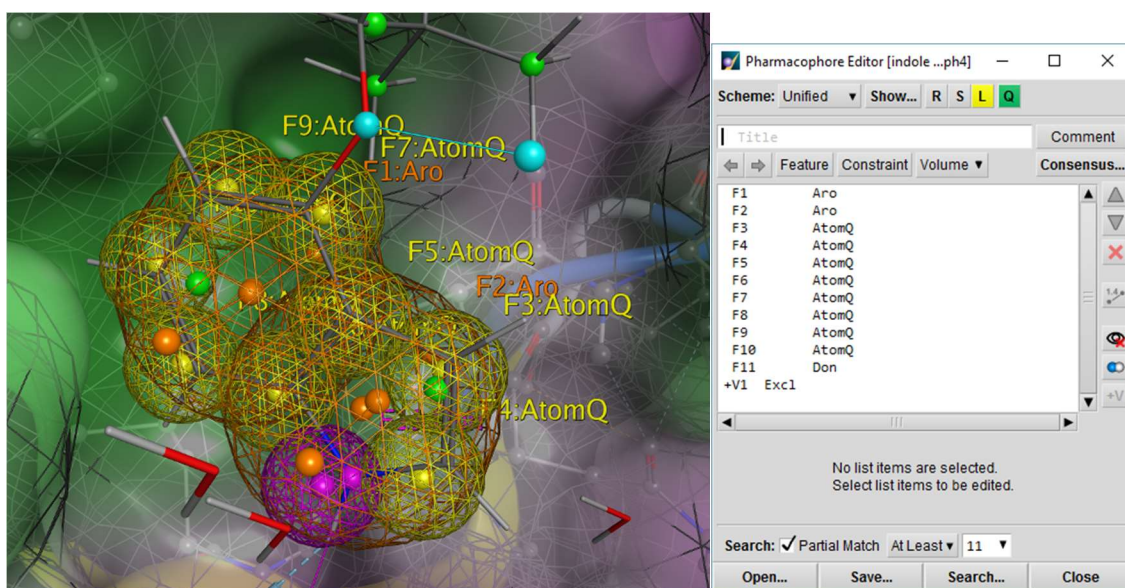
Figure 6.4: Use of pharmacophores to guide the docking of 4-substituted indoles.

Docking Designed BRCA2 Mimetics into PALB2

The 4-substituted indoles generated by Markush enumeration (procedure 6.4.1) were imported into MOE as an .sdf database. These were then docked into the PALB2 binding site using the 'Dock' function in MOE with the following parameters:

- Forcefield: Amber10:EHT

- Receptor: Receptor+Solvent

- Site: Ligand atoms

- Placement: α-Triangle

   o Minimum iterations: 80 000

   o Maximum Iterations: 500 000

   o Timeout: 600 seconds

- Refinement: Induced fit

   o Termination criterion: Gradient = 0.01, iterations = 500

   o Pharmacophore restraint: Force constant = 100, radius offset = 0.4

6.4.6   Miniprotein Design: Homology Modelling in MOE

Aim: To build homology models of modified miniprotein sequences, so that they can be analysed for compatibility with PALB2.

Software Package(s): MOE 2015.1001.
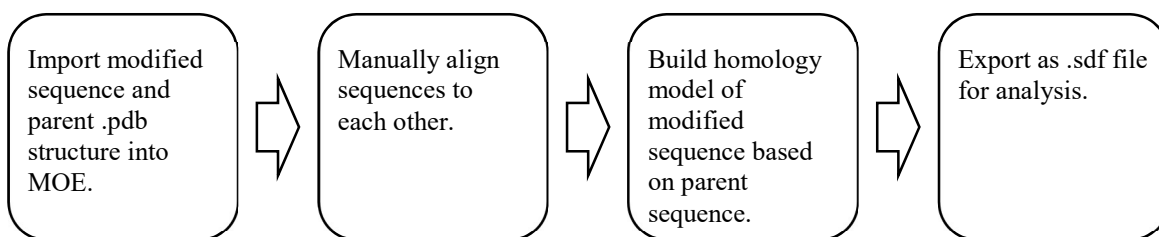
General Procedure:

Import modified sequence and parent .pdb structure into MOE.	⇨	Manually align sequences to each other.	⇨	Build homology model of modified sequence based on parent sequence.	⇨	Export as .sdf file for analysis.

Details and Parameters:

Manually Align Sequences to Each Other

The sequence viewer in MOE was used to manually align the sequences to each other. This was done simply by dragging the modified sequence (middle mouse button) so that it was in line with the parent sequence,

Build Homology Model of Modified Sequence

The 'Homology Model' tool in MOE was used to build a model of the modified sequence. The parameters used were:

- Forcefield: Amber10:EHT (Fixed charges)
- Mainchain models: 10
- Sidechain models: 4, at temperature: 300
- Intermediate refinement: Medium
- Final model: Fine
- Model scoring: GB/VI

Once the model had been built, it was exported in .sdf format for later analysis.

6.4.7 Miniprotein Design: Alignment and Superposition of Miniprotein Homology Models to Test For Steric Incompatibilities

Aim: To superpose the miniprotein homology models with the BRCA2 WFXXL motif, so that they can be analysed for steric overlap.

Software Package(s): MOE 2015.1001.
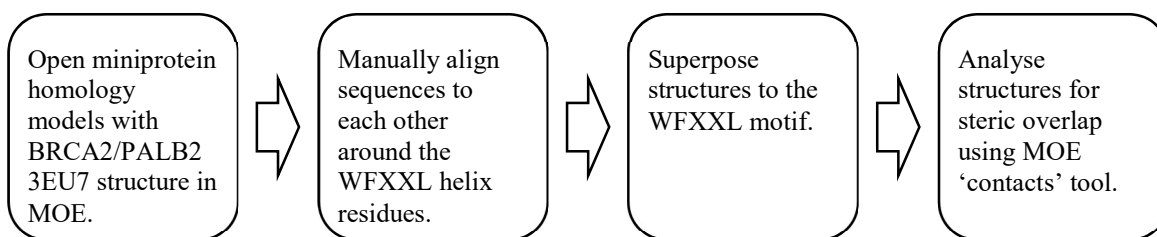
General Procedure:

| Open miniprotein homology models with BRCA2/PALB2 3EU7 structure in MOE. | ⟹ | Manually align sequences to each other around the WFXXL helix residues. | ⟹ | Superpose structures to the WFXXL motif. | ⟹ | Analyse structures for steric overlap using MOE 'contacts' tool. |

Details and Parameters:

Manually Align Sequences to Each Other

The sequence viewer in MOE was used to manually align the sequences to each other, by manually dragging them (middle mouse button) so that the WFXXL motif was in the same position in each case.

Superpose Sequences To The WFXXL Motif

The superpose tool in the MOE sequence editor was used to superpose the miniproteins to the WFXXL motif. The Trp, Phe and Leu residues for each miniprotein were selected, and the align tool was set to superpose to 'Selected Residues', and 'Use Current Alignment'.

Check for Steric Overlap

The MOE 'contacts' tool was used to check for steric overlap. The 'clash' functionality was turned on, showing parts of the miniprotein where steric clash occurred with orange disks (Figure 6.5). The residues where clashing occurred were manually listed, in order to aid the selection of miniproteins for synthesis.
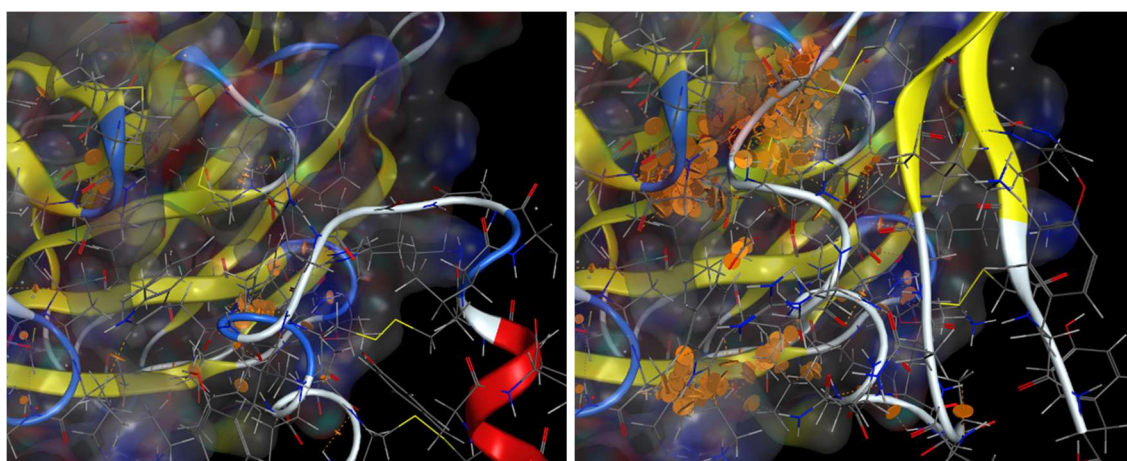
Figure 6.5: Left: Superposed WFXXL miniprotein based on 1WQD[291], with minimal steric clashing. Right: Superposed WFXXL miniprotein based on 2RTZ, showing significant steric clashing denoted by orange disks.

7 Bibliography

1   A. L. Hopkins and C. R. Groom, *Nat. Rev. Drug Discov.*, 2002, **1**, 727–730.

2   G. L. Patrick, *An Introduction to Medicinal Chemistry*, Oxford University Press, Oxford, 2nd edn., 2001.

3   R. D. Smith, L. Hu, J. A. Falkner, M. L. Benson, J. P. Nerothin and H. A. Carlson, *J. Mol. Graph. Model.*, 2006, **24**, 414–425.

4   A. Shulgin and A. Shulgin, *PiHKAL: A Chemical love Story*, Transform Press, 1991.

5   J. A. Wells and C. L. McClendon, *Nature*, 2007, **450**, 1001–1009.

6   L. Lo Conte, C. Chothia and J. Janin, *J. Mol. Biol.*, 1999, **285**, 2177–2198.

7   S. Jones and J. M. Thornton, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 13–20.

8   B. N. Bullock, A. L. Jochim and P. S. Arora, *J. Am. Chem. Soc.*, 2011, **133**, 14220–14223.

9   S. Surade and T. L. Blundell, *Chem. Biol.*, 2012, **19**, 42–50.

10   D. R. Rhodes, S. A. Tomlins, S. A. Tomlins, S. Varambally, V. Mahavisno, T. Barrette, S. Kalyana-sundaram, D. Ghosh, A. Pandey, A. M. Chinnaiyan and A. M. Chinnaiyan, *Nat. Biotechnol.*, 2005, **23**, 951–959.

11   A. Gschwind, O. M. Fischer and A. Ullrich, *Nat. Rev. Cancer*, 2004, **4**, 361–370.

12   A. a Ivanov, F. R. Khuri and H. Fu, *Trends Pharmacol. Sci.*, 2013, 1–8.

13   S. Jones and J. M. Thornton, *J. Mol. Biol.*, 1997, **272**, 121–132.

14   D. Xu, C. J. Tsai and R. Nussinov, *Protein Eng.*, 1997, **10**, 999–1012.

15   T. Clackson and J. Wells, *Science*, 1995, **267**, 383–386.

16   W. Guo, J. A. Wisniewski and H. Ji, *Bioorg. Med. Chem. Lett.*, 2014, **24**, 2546–2554.

17   P. J. Hajduk, J. R. Huth and C. Tse, *Drug Discov. Today*, 2005, **10**, 1675–1682.

18   P. J. Hajduk and J. Greer, *Nat. Rev. Drug Discov.*, 2007, **6**, 211–219.

19   A. Winter, A. P. Higueruelo, M. Marsh, A. Sigurdardottir, W. R. Pitt and T. L. Blundell, *Q. Rev. Biophys.*, 2012, **4**, 1–44.

20   S. F. Altschul, W. Gish, W. Miller, E. W. Myers and D. J. Lipman, *J. Mol. Biol.*, 1990, **215**, 403–410.

21   C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer and T. L. Madden, *BMC Bioinformatics*, 2009, **10**, 1–9.

22   M. Dunkel, S. Günther, J. Ahmed, B. Wittig and R. Preissner, *Nucleic Acids Res.*,

2008, **36**, 55–59.

23   F. Glaser, T. Pupko, I. Paz and R. E. Bell, *Bioinformatics*, 2003, **19**, 163–164.

24   V. Le Guilloux, P. Schmidtke and P. Tuffery, *BMC Bioinformatics*, 2009, **10**, 168.

25   N. J. Burgoyne and R. M. Jackson, *Bioinformatics*, 2006, **22**, 1335–1342.

26   N. Sugaya, K. Ikeda, T. Tashiro, S. Takeda, J. Otomo, Y. Ishida, A. Shiratori, A. Toyoda, H. Noguchi, T. Takeda, S. Kuhara, Y. Sakaki and T. Iwayanagi, *BMC Pharmacol.*, 2007, **7**, 10.

27   N. Sugaya and K. Ikeda, *BMC Bioinformatics*, 2009, **10**, 263.

28   N. Sugaya and T. Furuya, *BMC Bioinformatics*, 2011, **12**, 50.

29   D. R. Koes and C. J. Camacho, *Nucleic Acids Res.*, 2012, **40**, W387–W392.

30   T. Kortemme and D. Baker, *Proc. Natl. Acad. Sci.*, 2002, **99**, 14116–14121.

31   D. M. Kruger and H. Gohlke, *Nucleic Acids Res.*, 2010, **38**, W480–W486.

32   S. Dennis, T. Kortvelyesi and S. Vajda, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 4290–4295.

33   M. R. Landon, D. R. Lancia, J. Yu, S. C. Thiel and S. Vajda, *J. Med. Chem.*, 2007, **50**, 1231–1240.

34   R. Brenke, D. Kozakov, G.-Y. Chuang, D. Beglov, D. Hall, M. R. Landon, C. Mattos and S. Vajda, *Bioinformatics*, 2009, **25**, 621–627.

35   D. Kozakov, D. R. Hall, G.-Y. Chuang, R. Cencic, R. Brenke, L. E. Grove, D. Beglov, J. Pelletier, A. Whitty and S. Vajda, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 13528–13533.

36   P. J. Hajduk, J. R. Huth and S. W. Fesik, *J. Med. Chem.*, 2005, 2518–2525.

37   D. Kozakov, D. R. Hall, R. L. Napoleon, C. Yueh, A. Whitty and S. Vajda, *J. Med. Chem.*, 2015, **58**, 9063–9088.

38   W. H. Elliott and D. C. Elliott, *Biochemistry and Molecular Biology*, Oxford University Press, Third Edit., 2005.

39   R. P. Cheng, S. H. Gellman and W. F. DeGrado, *Chem. Rev.*, 2001, **101**, 3219–3232.

40   D. Seebach, M. Overhand, F. N. M. Kühnle, B. Martinoni, L. Oberer, U. Hommel and H. Widmer, *Helv. Chim. Acta*, 1996, **79**, 913–941.

41   S. H. Gellman, *Acc. Chem. Res.*, 1998, **31**, 173–180.

42   W. S. Horne and S. H. Gellman, *Acc. Chem. Res.*, 2008, **41**, 1399–1408.

43   D. F. Hook, P. Bindschädler, Y. R. Mahajan, R. Sebesta, P. Kast and D. Seebach, *Chem. Biodivers.*, 2005, **2**, 591–632.

44   E. P. English, R. S. Chumanov, S. H. Gellman and T. Compton, *J. Biol. Chem.*, 2006, **281**, 2661–2667.

45   J. A. Kritzer, J. D. Lear, M. E. Hodsdon and A. Schepartz, *J. Am. Chem. Soc.*, 2004, **126**, 9468–9469.

46   E. A. Harker, D. S. Daniels, D. A. Guarracino and A. Schepartz, *Bioorg. Med. Chem.*, 2009, **17**, 2038–2046.

47   R. B. Merrifield, *J. Am. Chem. Soc.*, 1963, **85**, 2149–2154.

48   S. Wang, *J. Am. Chem. Soc.*, 1973, **95**, 1328–1333.

49   F. Regnier, *Science*, 1983, **222**, 245–252.

50   T. W. Von Geldern, T. W. Rockway, S. K. Davidsen, G. P. Budzik, E. N. Bush, M. Y. Chu-Moyer, E. M. Devine, W. H. Holleman and M. C. Johnson, *J. Med. Chem.*, 1992, **35**, 808–816.

51   P. W. Schiller, F. Bellini, G. Dionne, L. A. Maziak, R. Garcia, A. DeLéan and M. Cantin, *Biochem. Biophys. Res. Commun.*, 1986, **138**, 880–886.

52   R. M. Scarborough, A. Arfsten, L. L. Kang, K. Schwartz, G. A. McEnroe, J. G. Porter, M. Suzuki, T. Maack and J. A. Lewicki, *J. Cell. Biochem.*, 1988, **38**, 1–27.

53   P. R. Bovy, J. M. O'Neal, G. M. Olins and D. R. Patton, *J. Med. Chem.*, 1989, **32**, 869–874.

54   R. F. Nutt and D. F. Veber, *Endocrinol. Metab. Clin. North Am.*, 1987, **16**, 19–41.

55   J. D. Baxter, J. A. Lewicki and D. G. Gardner, *Bio/Technology*, 1988, **6**, 529–546.

56   B. Li, J. Y. K. Tom, D. Oare, R. Yen, W. J. Fairbrother, J. A. Wells and B. C. Cunningham, *Science*, 1995, **270**, 1657–1660.

57   G. P. Smith, *Science*, 1985, **228**, 1315–1317.

58   G. P. Smith and V. A. Petrenko, *Chem. Rev.*, 1997, **97**, 391–410.

59   A. C. Braisted and J. A. Wells, *Proc. Natl. Acad. Sci.*, 1996, **93**, 5688–5692.

60   J. Deisenhofer, *Biochemistry*, 1981, **20**, 2361–2370.

61   G. Gish, M. L. McGlone, T. Pawson and J. A. Adams, *"Protein Eng. Des. Sel.*, 1995, **8**, 609–614.

62   T. A. Kunkel, K. Bebenek and J. McClary, *Methods Enzymol.*, 1991, **Volume 204**, 125–139.

63   N. P. Pavletich and C. O. Pabo, *Science*, 1991, **252**, 809–817.

64   K. Wüthrich, *Nat. Struct. Biol.*, 2001, **8**, 923–925.

65   C. Vita, J. Vizzavona, E. Drakopoulou, S. Zinn-Justin, B. Gilquin and A. Menez, *Biopolym. - Pept. Sci. Sect.*, 1998, 47, 93–100.

66   E. Drakopoulou, J. Vizzavona, J. Neyton, V. Aniort, F. Bouet, H. Virelizier, A. Ménez and C. Vita, *Biochemistry*, 1998, **37**, 1292–1301.

67   D. J. Capon and R. H. R. Ward, *Annu. Rev. Immunol.*, 1991, **9**, 649–678.

68   C. Vita, E. Drakopoulou, J. Vizzavona, S. Rochette, L. Martin, a Ménez, C. Roumestand, Y. S. Yang, L. Ylisastigui, a Benjouad and J. C. Gluckman, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 13091–13096.

69   J. C. Martins, F. J. M. Van de Ven and F. A. M. Borremans, *J. Mol. Biol.*, 1995, **253**, 590–603.

70   N. J. Zondlo and A. Schepartz, *J. Am. Chem. Soc.*, 1999, **121**, 6938–6939.

71   A. A. Adzhubei and M. J. E. Sternberg, *J. Mol. Biol.*, 1993, **229**, 472–493.

72   T. E. Ellenberger, C. J. Brandl, K. Struhl and S. C. Harrison, *Cell*, 1992, **71**, 1223–1237.

73   P. König and T. J. Richmond, *J. Mol. Biol.*, 1993, **233**, 139–154.

74   J. W. Chin and A. Schepartz, *J. Am. Chem. Soc.*, 2001, **123**, 2929–2930.

75   J. W. Chin, R. M. Grotzfeld, M. A. Fabian and A. Schepartz, *Bioorganic Med. Chem. Lett.*, 2001, **11**, 1501–1505.

76   J. W. Chin and A. Schepartz, *Angew. Chemie - Int. Ed.*, 2001, **40**, 3806–3809.

77   D. W. Nicholson, *Nature*, 2000, **407**, 810–816.

78   J. K. Buolamwini, *Curr. Opin. Chem. Biol.*, 1999, **3**, 500–509.

79   D. Golemi-Kotra, R. Mahaffy, M. J. Footer, J. H. Holtzman, T. D. Pollard, J. A. Theriot and A. Schepartz, *J. Am. Chem. Soc.*, 2004, **126**, 4–5.

80   F. B. Gertler, K. Niebuhr, M. Reinhard, J. Wehland and P. Soriano, *Cell*, 1996, **87**, 227–239.

81   A. Zarrinpar, R. P. Bhattacharyya and W. A. Lim, *Sci. Signal.*, 2003, **2003**, re8–re8.

82   K. Niebuhr, *EMBO J.*, 1997, **16**, 5433–5444.

83   B. Vogelstein, D. Lane and A. J. Levine, *Nature*, 2000, **408**, 307–310.

84   J. a. Kritzer, R. Zutshi, M. Cheah, F. A. Ran, R. Webman, T. M. Wongjirad and A. Schepartz, *ChemBioChem*, 2006, **7**, 29–31.

85   C. Li, M. Liu, J. Monbo, G. Zou, C. Li, W. Yuan, D. Zella, W. Y. Lu and W. Lu, *J. Am. Chem. Soc.*, 2008, **130**, 13546–13548.

86   A. Szyk, W. Lu, C. Xu and J. Lubkowski, *J. Struct. Biol.*, 2004, **145**, 289–294.

87   C. A. Heinlein and C. Chang, *Endocr. Rev.*, 2004, **25**, 276–308.

88   B. Vaz, S. Möcklinghoff, S. Folkertsma, S. Lusher, J. de Vlieg and L. Brunsveld,

Chem. Commun. , 2009, 5377–5379.

89   M. D. Seoane, K. Petkau-Milroy, B. Vaz, S. Möcklinghoff, S. Folkertsma, L.-G. Milroy and L. Brunsveld, *Med. Chem. Commun.*, 2013, **4**, 187–192.

90   A.-M. Leduc, J. O. Trent, J. L. Wittliff, K. S. Bramlett, S. L. Briggs, N. Y. Chirgadze, Y. Wang, T. P. Burris and A. F. Spatola, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 11273–11278.

91   T. Phan, H. D. Nguyen, H. Göksel, S. Möcklinghoff and L. Brunsveld, *Chem. Commun.*, 2010, **46**, 8207–8209.

92   A. J. Wilson, *Chem. Soc. Rev.*, 2009, **38**, 3289–300.

93   G. L. Verdine and G. J. Hilinski, *Stapled peptides for intracellular drug targets.*, Elsevier Inc., 1st edn., 2012, vol. 503.

94   L. D. Walensky and G. H. Bird, *J. Med. Chem.*, 2014, **57**, 6275–6288.

95   J. A. Miles, D. J. Yeo, P. Rowell, S. Rodriguez-Marin, C. M. Pask, S. L. Warriner, T. A. Edwards and A. J. Wilson, *Chem. Sci.*, 2016, **7**, 3694–3702.

96   M. Pellegrini, M. Royo, M. Chorev and D. F. Mierke, *J. Pept. Res.*, 1997, **49**, 404–414.

97   M. Chorev, E. Roubini, R. L. McKee, S. W. Gibbons, M. E. Goldman, M. P. Caulfield and M. Rosenblatt, *Biochemistry*, 1991, **30**, 5968–5974.

98   H. E. Blackwell and R. H. Grubbs, *Angew. Chemie Int. Ed.*, 1998, **37**, 3281–3284.

99   Y.-W. Kim, T. N. Grossmann and G. L. Verdine, *Nat. Protoc.*, 2011, **6**, 761–771.

100   P. Schwab, R. H. Grubbs and J. W. Ziller, *J. Am. Chem. Soc.*, 1996, **118**, 100–110.

101   M. Scholl, S. Ding, C. W. Lee and R. H. Grubbs, *Org. Lett.*, 1999, **1**, 953–956.

102   T. M. Trnka and R. H. Grubbs, *Acc. Chem. Res.*, 2001, **34**, 18–29.

103   L. D. Walensky, *Science*, 2004, **305**, 1466–1470.

104   L. D. Walensky, K. Pitter, J. Morash, K. J. Oh, S. Barbuto, J. Fisher, E. Smith, G. L. Verdine and S. J. Korsmeyer, *Mol. Cell*, 2006, **24**, 199–210.

105   S. Kneissl, E. J. Loveridge, C. Williams, M. P. Crump and R. K. Allemann, *ChemBioChem*, 2008, **9**, 3046–3054.

106   E. Gavathiotis, M. Suzuki, M. L. Davis, K. Pitter, G. H. Bird, S. G. Katz, H.-C. Tu, H. Kim, E. H.-Y. Cheng, N. Tjandra and L. D. Walensky, *Nature*, 2008, **455**, 1076–1081.

107   C. R. Braun, J. Mintseris, E. Gavathiotis, G. H. Bird, S. P. Gygi and L. D. Walensky, *Chem. Biol.*, 2010, **17**, 1325–1333.

108   E. Gavathiotis, D. E. Reyna, M. L. Davis, G. H. Bird and L. D. Walensky, *Mol.*

Cell, 2010, **40**, 481–492.

109   J. L. LaBelle, S. G. Katz, G. H. Bird, E. Gavathiotis, M. L. Stewart, C. Lawrence, J. K. Fisher, M. Godes, K. Pitter, A. L. Kung and L. D. Walensky, *J. Clin. Invest.*, 2012, **122**, 2018–2031.

110   E. S. Leshchiner, C. R. Braun, G. H. Bird and L. D. Walensky, *Proc. Natl. Acad. Sci.*, 2013, **110**, E986–E995.

111   T. Moldoveanu, C. R. Grace, F. Llambi, A. Nourse, P. Fitzgerald, K. Gehring, R. W. Kriwacki and D. R. Green, *Nat. Struct. Mol. Biol.*, 2013, **20**, 589–597.

112   T. Okamoto, K. Zobel, A. Fedorova, C. Quan, H. Yang, W. J. Fairbrother, D. C. S. Huang, B. J. Smith, K. Deshayes and P. E. Czabotar, *ACS Chem. Biol.*, 2012, 14–19.

113   S. W. Muchmore, M. Sattler, H. Liang, R. P. Meadows, J. E. Harlan, H. S. Yoon, D. Nettesheim, B. S. Chang, C. B. Thompson, S.-L. Wong, S.-C. Ng and S. W. Fesik, *Nature*, 1996, **381**, 335–341.

114   M. Sattler, *Science*, 1997, **275**, 983–986.

115   M. Kvansakul, A. H. Wei, J. I. Fletcher, S. N. Willis, L. Chen, A. W. Roberts, D. C. S. Huang and P. M. Colman, *PLoS Pathog.*, 2010, **6**, e1001236.

116   F. Bernal, A. F. Tyler, S. J. Korsmeyer, L. D. Walensky and G. L. Verdine, *J. Am. Chem. Soc.*, 2007, **129**, 2456–2457.

117   F. Bernal, M. Wade, M. Godes, T. N. Davis, D. G. Whitehead, A. L. Kung, G. M. Wahl and L. D. Walensky, *Cancer Cell*, 2010, **18**, 411–422.

118   A. D. Bautista, J. S. Appelbaum, C. J. Craig, J. Michel and A. Schepartz, *J. Am. Chem. Soc.*, 2010, **132**, 2904–2906.

119   Z. Guo, U. Mohanty, J. Noehre, T. K. Sawyer, W. Sherman and G. Krilov, *Chem. Biol. Drug Des.*, 2010, **75**, 348–359.

120   T. L. Joseph, D. Lane and C. Verma, *Cell Cycle*, 2010, **9**, 4560–4568.

121   A. Gembarska, F. Luciani, C. Fedele, E. A. Russell, M. Dewaele, S. Villar, A. Zwolinska, S. Haupt, J. de Lange, D. Yip, J. Goydos, J. J. Haigh, Y. Haupt, L. Larue, A. Jochemsen, H. Shi, G. Moriceau, R. S. Lo, G. Ghanem, M. Shackleton, F. Bernal and J.-C. Marine, *Nat. Med.*, 2012, **18**, 1239–1247.

122   S. Baek, P. S. Kutchukian, G. L. Verdine, R. Huber, T. A. Holak, K. W. Lee and G. M. Popowicz, *J. Am. Chem. Soc.*, 2012, **134**, 103–106.

123   Y. S. Chang, B. Graves, V. Guerlavais, C. Tovar, K. Packman, K.-H. To, K. A. Olson, K. Kesavan, P. Gangurde, A. Mukherjee, T. Baker, K. Darlak, C. Elkin, Z.

Filipovic, F. Z. Qureshi, H. Cai, P. Berry, E. Feyfant, X. E. Shi, J. Horstick, D. A. Annis, A. M. Manning, N. Fotouhi, H. Nash, L. T. Vassilev and T. K. Sawyer, *Proc. Natl. Acad. Sci.*, 2013, **110**, E3445–E3454.

124 C. J. Brown, S. T. Quah, J. Jong, A. M. Goh, P. C. Chiam, K. H. Khoo, M. L. Choong, M. A. Lee, L. Yurlova, K. Zolghadr, T. L. Joseph, C. S. Verma and D. P. Lane, *ACS Chem. Biol.*, 2013, **8**, 506–512.

125 S. J. Wei, T. Joseph, S. Chee, L. Li, L. Yurlova, K. Zolghadr, C. Brown, D. Lane, C. Verma and F. Ghadessy, *PLoS One*, 2013, **8**, e81068.

126 M. B. Kastan, O. Onyekwere, D. Sidransky, B. Vogelstein and R. W. Craig, *Cancer Res.*, 1991, **51**, 6304–6311.

127 M. Li, C. L. Brooks, F. Wu-Baer, D. Chen, R. Baer and W. Gu, *Science*, 2003, **302**, 1972–5.

128 G. Popowicz, A. Czarna and T. Holak, *Cell Cycle*, 2008, **7**, 2441–2443.

129 H. J.-P. Ryser and R. Hancock, *Science*, 1965, **150**, 501–503.

130 M. Pazgier, M. Liu, G. Zou, W. Yuan, C. Li, C. Li, J. Li, J. Monbo, D. Zella, S. G. Tarasov and W. Lu, *Proc. Natl. Acad. Sci.*, 2009, **106**, 4665–4670.

131 Y. S. Chang, B. Graves, V. Guerlavais, C. Tovar, K. Packman, K.-H. To, K. a. Olson, K. Kesavan, P. Gangurde, A. Mukherjee, T. Baker, K. Darlak, C. Elkin, Z. Filipovic, F. Z. Qureshi, H. Cai, P. Berry, E. Feyfant, X. E. Shi, J. Horstick, D. A. Annis, A. M. Manning, N. Fotouhi, H. Nash, L. T. Vassilev and T. K. Sawyer, *Proc. Natl. Acad. Sci.*, 2013, **110**, E3445–E3454.

132 T. L. Joseph, D. P. Lane and C. S. Verma, *PLoS One*, 2012, **7**, e43985.

133 R. E. Moellering, M. Cornejo, T. N. Davis, C. Del Bianco, J. C. Aster, S. C. Blacklow, A. L. Kung, D. G. Gilliland, G. L. Verdine and J. E. Bradner, *Nature*, 2009, **462**, 182–188.

134 K. Takada, D. Zhu, G. H. Bird, K. Sukhdeo, J.-J. Zhao, M. Mani, M. Lemieux, D. E. Carrasco, J. Ryan, D. Horst, M. Fulciniti, N. C. Munshi, W. Xu, A. L. Kung, R. A. Shivdasani, L. D. Walensky and D. R. Carrasco, *Sci. Transl. Med.*, 2012, **4**, 148ra117–148ra117.

135 T. N. Grossmann, J. T.-H. Yeh, B. R. Bowman, Q. Chu, R. E. Moellering and G. L. Verdine, *Proc. Natl. Acad. Sci.*, 2012, **109**, 17942–17947.

136 H.-K. Cui, B. Zhao, Y. Li, Y. Guo, H. Hu, L. Liu and Y.-G. Chen, *Cell Res.*, 2013, **23**, 581–584.

137 W. Kim, G. H. Bird, T. Neff, G. Guo, M. A. Kerenyi, L. D. Walensky and S. H.

Orkin, *Nat. Chem. Biol.*, 2013, **9**, 643–650.

138   D. Lama, S. T. Quah, C. S. Verma, R. Lakshminarayanan, R. W. Beuerman, D. P. Lane and C. J. Brown, *Sci. Rep.*, 2013, **3**, 3451.

139   H. Zhang, Q. Zhao, S. Bhattacharya, A. A. Waheed, X. Tong, A. Hong, S. Heck, F. Curreli, M. Goger, D. Cowburn, E. O. Freed and A. K. Debnath, *J. Mol. Biol.*, 2008, **378**, 565–580.

140   S. Bhattacharya, H. Zhang, A. K. Debnath and D. Cowburn, *J. Biol. Chem.*, 2008, **283**, 16274–16278.

141   H. Zhang, F. Curreli, X. Zhang, S. Bhattacharya, A. A. Waheed, A. Cooper, D. Cowburn, E. O. Freed and A. K. Debnath, *Retrovirology*, 2011, **8**, 28.

142   Y. Long, S. Huang, Z. Zawahir, Z. Xu, H. Li, T. W. Sanchez, Y. Zhi, S. De Houwer, F. Christ, Z. Debyser and N. Neamati, *J. Med. Chem.*, 2013, **56**, 5601–5612.

143   H.-K. Cui, J. Qing, Y. Guo, Y.-J. Wang, L.-J. Cui, T.-H. He, L. Zhang and L. Liu, *Bioorg. Med. Chem.*, 2013, **21**, 3547–3554.

144   C. Phillips, L. R. Roberts, M. Schade, R. Bazin, A. Bent, N. L. Davies, R. Moore, A. D. Pannifer, A. R. Pickford, S. H. Prior, C. M. Read, A. Scott, D. G. Brown, B. Xu and S. L. Irving, *J. Am. Chem. Soc.*, 2011, **133**, 9696–9699.

145   R. J. Platt, T. S. Han, B. R. Green, M. D. Smith, J. Skalicky, P. Gruszczynski, H. S. White, B. Olivera, G. Bulaj and J. Gajewiak, *J. Biol. Chem.*, 2012, **287**, 20727–20736.

146   C. G. Cummings and A. D. Hamilton, *Curr. Opin. Chem. Biol.*, 2010, **14**, 341–346.

147   M. K. P. Jayatunga, S. Thompson and A. D. Hamilton, *Bioorg. Med. Chem. Lett.*, 2014, **24**, 717–724.

148   D. C. Horwell, W. Howson, G. S. Ratcliffe and H. M. G. Willems, *Bioorg. Med. Chem.*, 1996, **4**, 33–42.

149   B. P. Orner, J. T. Ernst and A. D. Hamilton, *J. Am. Chem. Soc.*, 2001, **123**, 5382–5383.

150   J. Plante, F. Campbell, B. Malkova, C. Kilner, S. L. Warriner and A. J. Wilson, *Org. Biomol. Chem.*, 2008, **6**, 138–146.

151   F. Campbell, J. P. Plante, T. a Edwards, S. L. Warriner and A. J. Wilson, *Org. Biomol. Chem.*, 2010, **8**, 2344–2351.

152   C. G. Cummings, N. T. Ross, W. P. Katt and A. D. Hamilton, *Org. Lett.*, 2009, **11**,

25–28.

153 D. C. Horwell, W. Howson, W. P. Nolan, G. S. Ratcliffe, D. C. Rees and H. M. G. Willems, *Tetrahedron*, 1995, **51**, 203–216.

154 J. T. Ernst, J. Becerril, H. S. Park, H. Yin and A. D. Hamilton, *Angew. Chemie Int. Ed.*, 2003, **42**, 535–539.

155 H. Yin, G.-I. Lee, K. a Sedey, J. M. Rodriguez, H.-G. Wang, S. M. Sebti and A. D. Hamilton, *J. Am. Chem. Soc.*, 2005, **127**, 5463–5468.

156 J. M. Rodriguez and A. D. Hamilton, *Angew. Chemie Int. Ed.*, 2007, **46**, 8614–8617.

157 J. M. Rodriguez, N. T. Ross, W. P. Katt, D. Dhar, G.-I. Lee and A. D. Hamilton, *ChemMedChem*, 2009, **4**, 649–656.

158 S. Thompson and A. D. Hamilton, *Org. Biomol. Chem.*, 2012, **10**, 5780–2.

159 J. P. Plante, T. Burnley, B. Malkova, M. E. Webb, S. L. Warriner, T. A. Edwards and A. J. Wilson, *Chem. Commun.*, 2009, 5091–5093.

160 V. Azzarito, P. Prabhakaran, A. I. Bartlett, N. S. Murphy, M. J. Hardie, C. a. Kilner, T. a. Edwards, S. L. Warriner and A. J. Wilson, *Org. Biomol. Chem.*, 2012, **10**, 6469–6472.

161 N. S. Murphy, P. Prabhakaran, V. Azzarito, J. P. Plante, M. J. Hardie, C. a Kilner, S. L. Warriner and A. J. Wilson, *Chemistry*, 2013, **19**, 5546–50.

162 V. Azzarito, J. A. Miles, J. Fisher, T. A. Edwards, S. L. Warriner and A. J. Wilson, *Chem. Sci.*, 2015, **6**, 2434–2443.

163 I. C. Kim and A. D. Hamilton, *Org. Lett.*, 2006, **8**, 1751–1754.

164 H. Oguri, S. Tanabe, A. Oomura, M. Umetsu and M. Hirama, *Tetrahedron Lett.*, 2006, **47**, 5801–5805.

165 S. M. Biros, L. Moisan, E. Mann, A. Carella, D. Zhai, J. C. Reed and J. Rebek, *Bioorg. Med. Chem. Lett.*, 2007, **17**, 4641–4645.

166 A. Volonterio, L. Moisan and J. Rebek, *Org. Lett.*, 2007, **9**, 3733–3736.

167 B. B. Lao, K. Drew, D. A. Guarracino, T. F. Brewer, D. W. Heindel, R. Bonneau and P. S. Arora, *J. Am. Chem. Soc.*, 2014, **136**, 7877–7888.

168 M. Aeluri, S. Chamakuri, B. Dasari, S. K. R. Guduru, R. Jimmidi, S. Jogula and P. Arya, *Chem. Rev.*, 2014, **114**, 4640–4694.

169 L. T. Vassilev, B. T. Vu, B. Graves, D. Carvajal, F. Podlaski, Z. Filipovic, N. Kong, U. Kammlott, C. Lukacs, C. Klein, N. Fotouhi and E. A. Liu, *Science*, 2004, **303**, 844–848.

170   M. Wade, Y. C. Li and G. M. Wahl, *Nat. Rev. Cancer*, 2013, **13**, 83–96.

171   T. Oltersdorf, S. W. Elmore, A. R. Shoemaker, R. C. Armstrong, D. J. Augeri, B. A. Belli, M. Bruncko, T. L. Deckwerth, J. Dinges, P. J. Hajduk, M. K. Joseph, S. Kitada, S. J. Korsmeyer, A. R. Kunzer, A. Letai, C. Li, M. J. Mitten, D. G. Nettesheim, S. Ng, P. M. Nimmer, J. M. O'Connor, A. Oleksijew, A. M. Petros, J. C. Reed, W. Shen, S. K. Tahir, C. B. Thompson, K. J. Tomaselli, B. Wang, M. D. Wendt, H. Zhang, S. W. Fesik and S. H. Rosenberg, *Nature*, 2005, **435**, 677–681.

172   C. Tse, A. R. Shoemaker, J. Adickes, M. G. Anderson, J. Chen, S. Jin, E. F. Johnson, K. C. Marsh, M. J. Mitten, P. Nimmer, L. Roberts, S. K. Tahir, Y. Xiao, X. Yang, H. Zhang, S. Fesik, S. H. Rosenberg and S. W. Elmore, *Cancer Res.*, 2008, **68**, 3421–3428.

173   C. Ndubaku, E. Varfolomeev, L. Wang, K. Zobel, K. Lau, L. O. Elliott, B. Maurer, A. V Fedorova, J. N. Dynek, M. Koehler, S. G. Hymowitz, V. Tsui, K. Deshayes, W. J. Fairbrother, J. A. Flygare and D. Vucic, *ACS Chem. Biol.*, 2009, **4**, 557–566.

174   L. Li, R. M. Thomas, H. Suzuki, J. K. De Brabander, X. Wang and P. G. Harran, *Science*, 2004, **305**, 1471–1474.

175   C. Krepler, S. K. Chunduru, M. B. Halloran, X. He, M. Xiao, A. Vultur, J. Villanueva, Y. Mitsuuchi, E. M. Neiman, C. Benetatos, K. L. Nathanson, R. K. Amaravadi, H. Pehamberger, M. McKinlay and M. Herlyn, *Clin. Cancer Res.*, 2013, **19**, 1784–1794.

176   F. Falchi, F. Caporuscio and M. Recanatini, *Future Med. Chem.*, 2014, **6**, 343–357.

177   C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug Deliv. Rev.*, 2001, **46**, 3–26.

178   C. A. Lipinski, *Drug Discov. Today Technol.*, 2004, **1**, 337–341.

179   M. Gao, R. E. Nettles, M. Belema, L. B. Snyder, V. N. Nguyen, R. a Fridell, M. H. Serrano-Wu, D. R. Langley, J.-H. Sun, D. R. O'Boyle, J. A. Lemm, C. Wang, J. O. Knipe, C. Chien, R. J. Colonno, D. M. Grasela, N. a Meanwell and L. G. Hamann, *Nature*, 2010, **465**, 96–100.

180   R. E. Nettles, M. Gao, M. Bifano, E. Chung, A. Persson, T. C. Marbury, R. Goldwater, M. P. DeMicco, M. Rodriguez-Torres, A. Vutikullird, E. Fuentes, E. Lawitz, J. C. Lopez-Talavera and D. M. Grasela, *Hepatology*, 2011, **54**, 1956–1965.

181   M. Belema, V. N. Nguyen, C. Bachand, D. H. Deon, J. T. Goodrich, C. a. James,

R. Lavoie, O. D. Lopez, A. Martel, J. L. Romine, E. H. Ruediger, L. B. Snyder, D. R. S. Laurent, F. Yang, J. Zhu, H. S. Wong, D. R. Langley, S. P. Adams, G. H. Cantor, A. Chimalakonda, A. Fura, B. M. Johnson, J. O. Knipe, D. D. Parker, K. S. Santone, R. a. Fridell, J. a. Lemm, D. R. O'Boyle, R. J. Colonno, M. Gao, N. a. Meanwell and L. G. Hamann, *J. Med. Chem.*, 2014, **57**, 2013–2032.

182   A. Alex, D. S. Millan, M. Perez, F. Wakenhut and G. A. Whitlock, *Medchemcomm*, 2011, **2**, 669–674.

183   A. P. Hill and R. J. Young, *Drug Discov. Today*, 2010, **15**, 648–655.

184   R. J. Young, D. V. S. Green, C. N. Luscombe and A. P. Hill, *Drug Discov. Today*, 2011, **16**, 822–830.

185   M. Belema, V. N. Nguyen, J. L. Romine, D. R. St. Laurent, O. D. Lopez, J. T. Goodrich, P. T. Nower, D. R. O'Boyle, J. A. Lemm, R. A. Fridell, M. Gao, H. Fang, R. G. Krause, Y. K. Wang, A. J. Oliver, A. C. Good, J. O. Knipe, N. A. Meanwell and L. B. Snyder, *J. Med. Chem.*, 2014, **57**, 1995–2012.

186   M. Belema and N. A. Meanwell, *J. Med. Chem.*, 2014, **57**, 5057–5071.

187   T. D. Tran, F. Wakenhut, C. Pickford, S. Shaw, M. Westby, C. Smith-Burchnell, L. Watson, M. Paradowski, J. Milbank, R. a. Brimage, R. Halstead, R. Glen, C. P. Wilson, F. Adam, D. Hay, J.-Y. Chiva, C. Nichols, D. C. Blakemore, I. Gardner, S. Dayal, A. Pike, R. Webster and D. C. Pryde, *ChemMedChem*, 2014, **9**, 1378–1386.

188   F. Wakenhut, T. D. Tran, C. Pickford, S. Shaw, M. Westby, C. Smith-Burchnell, L. Watson, M. Paradowski, J. Milbank, D. Stonehouse, K. Cheung, R. Wybrow, F. Daverio, S. Crook, K. Statham, D. Leese, D. Stead, F. Adam, D. Hay, L. R. Roberts, J.-Y. Chiva, C. Nichols, D. C. Blakemore, G. H. Goetz, Y. Che, I. Gardner, S. Dayal, A. Pike, R. Webster and D. C. Pryde, *ChemMedChem*, 2014, **9**, 1387–1396.

189   N. Kuhl, M. N. Hopkinson, J. Wencel-Delord and F. Glorius, *Angew. Chemie Int. Ed.*, 2012, **51**, 10236–10254.

190   D. Alberico, M. E. Scott and M. Lautens, *Chem. Rev.*, 2007, **107**, 174–238.

191   T. W. Lyons and M. S. Sanford, *Chem. Rev.*, 2010, **110**, 1147–1169.

192   D. Kalyani, N. R. Deprez, L. V Desai and M. S. Sanford, *J. Am. Chem. Soc.*, 2005, **127**, 7330–7331.

193   A. R. Dick, K. L. Hull and M. S. Sanford, *J. Am. Chem. Soc.*, 2004, **126**, 2300–2301.

194 S. I. Gorelsky, D. Lapointe and K. Fagnou, *J. Am. Chem. Soc.*, 2008, **130**, 10848–10849.

195 A. Ohta, Y. Akita, A. Inoue, K. Yamamoto, T. Kurihara and M. Shimizu, *Heterocycles*, 1985, **23**, 2327–2333.

196 Y. Akita, Y. Itagaki, S. Takizawa and A. Ohta, *Chem. Pharm. Bull. (Tokyo).*, 1989, **37**, 1477–1480.

197 T. Kuroda and F. Suzuki, *Tetrahedron Lett.*, 1991, **32**, 6915–6918.

198 J. Le Bras, D. K. Mukherjee, S. González, M. Tristany, B. Ganchegui, M. Moreno-Mañas, R. Pleixats, F. Hénin and J. Muzart, *New J. Chem.*, 2004, **28**, 1550–1553.

199 B. B. Touré, B. S. Lane and D. Sames, *Org. Lett.*, 2006, **8**, 1979–1982.

200 B. Liégault, D. Lapointe, L. Caron, A. Vlassova and K. Fagnou, *J. Org. Chem.*, 2009, **74**, 1826–1834.

201 J. M. Joo, B. B. Touré and D. Sames, *J. Org. Chem.*, 2010, **75**, 4911–4920.

202 C. Zhou, M. Garcia-Calvo, S. Pinto, M. Lombardo, Z. Feng, K. Bender, K. D. Pryor, U. R. Bhatt, R. M. Chabin, W. M. Geissler, Z. Shen, X. Tong, Z. Zhang, K. K. Wong, R. S. Roy, K. T. Chapman, L. Yang and Y. Xiong, *J. Med. Chem.*, 2010, **53**, 7251–7263.

203 B. Zhang, Z. Jiang, X. Zhou, S. Lu, J. Li, Y. Liu and C. Li, *Angew. Chemie Int. Ed.*, 2012, **51**, 13159–13162.

204 S. H. Reich, M. Melnick, M. J. Pino, M. A. M. Fuhry, A. J. Trippe, K. Appelt, J. F. Davies, B.-W. Wu and L. Musick, *J. Med. Chem.*, 1996, **39**, 2781–2794.

205 D. Marcovici-Mizrahi, H. E. Gottlieb and A. Nudelman, *J. Org. Chem.*, 1996, **61**, 8402–8406.

206 J. M. Hoover and S. S. Stahl, *J. Am. Chem. Soc.*, 2011, **133**, 16901–16910.

207 J. M. Hoover, J. E. Steves and S. S. Stahl, *Nat. Protoc.*, 2012, **7**, 1161–1166.

208 J. M. Hoover, B. L. Ryland and S. S. Stahl, *J. Am. Chem. Soc.*, 2013, **135**, 2357–2367.

209 M. Uyanik, M. Akakura and K. Ishihara, *J. Am. Chem. Soc.*, 2009, **131**, 251–262.

210 C. A. G. N. Montalbetti and V. Falque, *Tetrahedron*, 2005, **61**, 10827–10852.

211 L. A. Carpino, H. Imazumi, F. J. Ferrer, C. Zhang, Y. Lee, B. M. Foxman, P. Henklein, C. Hanay, M. Clemens, H. Wenschuh, J. Klose, M. Beyermann and M. Bienert, *Angew. Chem. Int. Ed. Engl.*, 2002, **41**, 441–445.

212 J. C. Sauer, *J. Am. Chem. Soc.*, 1947, **69**, 2444–2448.

213 A. C. Laungani, J. M. Slattery, I. Krossing and B. Breit, *Chem. - A Eur. J.*, 2008,

14, 4488–4502.

214   D. R. St Laurent, M. Belema, M. Gao, J. Goodrich, R. Kakarla, J. O. Knipe, J. a Lemm, M. Liu, O. D. Lopez, V. N. Nguyen, P. T. Nower, D. O'Boyle, Y. Qiu, J. L. Romine, M. H. Serrano-Wu, J.-H. Sun, L. Valera, F. Yang, X. Yang, N. a Meanwell and L. B. Snyder, *Bioorg. Med. Chem. Lett.*, 2012, **22**, 6063–6.

215   C. A. Coburn, P. T. Meinke, W. Chang, C. M. Fandozzi, D. J. Graham, B. Hu, Q. Huang, S. Kargman, J. Kozlowski, R. Liu, J. A. McCauley, A. A. Nomeir, R. M. Soll, J. P. Vacca, D. Wang, H. Wu, B. Zhong, D. B. Olsen and S. W. Ludmerer, *ChemMedChem*, 2013, **8**, 1930–1940.

216   F. Wakenhut, T. D. Tran, C. Pickford, S. Shaw, M. Westby, C. Smith-Burchnell, L. Watson, M. Paradowski, J. Milbank, D. Stonehouse, K. Cheung, R. Wybrow, F. Daverio, S. Crook, K. Statham, D. Leese, D. Stead, F. Adam, D. Hay, L. R. Roberts, J.-Y. Chiva, C. Nichols, D. C. Blakemore, G. H. Goetz, Y. Che, I. Gardner, S. Dayal, A. Pike, R. Webster and D. C. Pryde, *ChemMedChem*, 2014, **9**, 1387–1396.

217   B. M. Monks and S. P. Cook, *J. Am. Chem. Soc.*, 2012, **134**, 15297–15300.

218   R. Chinchilla and C. Nájera, *Chem. Rev.*, 2014, **114**, 1783–1826.

219   G. Dyker and A. Kellner, *Tetrahedron Lett.*, 1994, **35**, 7633–7636.

220   T. H. Jonckers, B. U. Maes, G. L. Lemière, G. Rombouts, L. Pieters, A. Haemers and R. A. Dommisse, *Synlett*, 2003, **34**, 615–618.

221   I. C. F. . Ferreira, M.-J. R. . Queiroz and G. Kirsch, *Tetrahedron*, 2002, **58**, 7943–7949.

222   M. J. Mio, L. C. Kopel, J. B. Braun, T. L. Gadzikwa, K. L. Hull, R. G. Brisbois, C. J. Markworth and P. A. Grieco, *Org. Lett.*, 2002, **4**, 3199–3202.

223   M. Von Wantoch Rekowski, A. Pyriochou, N. Papapetropoulos, A. Stößel, A. Papapetropoulos and A. Giannis, *Bioorg. Med. Chem.*, 2010, **18**, 1288–1296.

224   X. Wen, J. El Bakali, R. Deprez-Poulain and B. Deprez, *Tetrahedron Lett.*, 2012, **53**, 2440–2443.

225   R. Wisastra, M. Ghizzoni, A. Boltjes, H. J. Haisma and F. J. Dekker, *Bioorg. Med. Chem.*, 2012, **20**, 5027–5032.

226   K. Kawabata and H. Goto, *Synth. Met.*, 2010, **160**, 2290–2298.

227   A. W. Freeman, M. Urvoy and M. E. Criswell, *J. Org. Chem.*, 2005, **70**, 5014–5019.

228   M. Ceylan, M. B. Gürdere, Y. Budak, C. Kazaz and H. Seçen, *Synthesis*, 2004,

2004, 1750–1754.

229   N. M. Nevar, A. V. Kel'in and O. G. Kulinkovich, *Synthesis*, 2000, **2000**, 1259–1262.

230   A. V. Kel'in and O. G. Kulinkovich, *Synthesis*, 1996, **1996**, 330–332.

231   J. B. Chaires, J. Ren, D. Hamelberg, A. Kumar, V. Pandya, D. W. Boykin and W. D. Wilson, *J. Med. Chem.*, 2004, **47**, 5729–5742.

232   D. Shridhar, M. Jogibhukta, P. Rao and V. Handa, *Synthesis*, 1982, 1061–1062.

233   S. Licciulli, K. Albahily, V. Fomitcheva, I. Korobkov, S. Gambarotta and R. Duchateau, *Angew. Chemie Int. Ed.*, 2011, **50**, 2346–2349.

234   T. O. Moore, M. Paradowski and S. E. Ward, *Org. Biomol. Chem.*, 2016, **14**, 3307–3313.

235   *Nat. Rev. Drug Discov.*, 2014, **13**, 169–169.

236   K. K. Hoe, C. S. Verma and D. P. Lane, *Nat. Rev. Drug Discov.*, 2014, **13**, 217–236.

237   L. Vela and I. Marzo, *Curr. Opin. Pharmacol.*, 2015, **23**, 74–81.

238   P. Wójcik and Ł. Berlicki, *Bioorg. Med. Chem. Lett.*, 2016, **26**, 707–713.

239   M. Pelay-Gimeno, A. Glas, O. Koch and T. N. Grossmann, *Angew. Chemie Int. Ed.*, 2015, **54**, 8896–8927.

240   A. J. Wilson, *Prog. Biophys. Mol. Biol.*, 2015, **119**, 33–40.

241   V. Azzarito, K. Long, N. Murphy and A. Wilson, *Nat. Chem.*, 2013, **5**, 161–173.

242   B. Xia, Q. Sheng, K. Nakanishi, A. Ohashi, J. Wu, N. Christ, X. Liu, M. Jasin, F. J. Couch and D. M. Livingston, *Mol. Cell*, 2006, **22**, 719–729.

243   A. W. Oliver, S. Swift, C. J. Lord, A. Ashworth and L. H. Pearl, *EMBO Rep.*, 2009, **10**, 990–996.

244   S. M. H. Sy, M. S. Y. Huen and J. Chen, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 7155–7160.

245   F. Zhang, J. Ma, J. Wu, L. Ye, H. Cai, B. Xia and X. Yu, *Curr. Biol.*, 2009, **19**, 524–529.

246   F. Zhang, Q. Fan, K. Ren and P. R. Andreassen, *Mol. Cancer Res.*, 2009, **7**, 1110–1118.

247   R. Wooster, S. Neuhausen, J. Mangion, Y. Quirk, D. Ford, N. Collins, K. Nguyen, S. Seal, T. Tran, D. Averill and A. Et, *Science*, 1994, **265**, 2088–2090.

248   K. Yoshida and Y. Miki, *Cancer Sci.*, 2004, **95**, 866–871.

249   S. Jones, R. H. Hruban, M. Kamiyama, M. Borges, X. Zhang, D. W. Parsons, J.

C.-H. Lin, E. Palmisano, K. Brune, E. M. Jaffee, C. A. Iacobuzio-Donahue, A. Maitra, G. Parmigiani, S. E. Kern, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, J. R. Eshleman, M. Goggins and A. P. Klein, *Science*, 2009, **324**, 217–217.

250   A. C. Antoniou, S. Casadei, T. Heikkinen, D. Barrowdale, K. Pylkäs, J. Roberts, A. Lee, D. Subramanian, K. De Leeneer, F. Fostira, E. Tomiak, S. L. Neuhausen, Z. L. Teo, S. Khan, K. Aittomäki, J. S. Moilanen, C. Turnbull, S. Seal, A. Mannermaa, A. Kallioniemi, G. J. Lindeman, S. S. Buys, I. L. Andrulis, P. Radice, C. Tondini, S. Manoukian, A. E. Toland, P. Miron, J. N. Weitzel, S. M. Domchek, B. Poppe, K. B. M. Claes, D. Yannoukakos, P. Concannon, J. L. Bernstein, P. A. James, D. F. Easton, D. E. Goldgar, J. L. Hopper, N. Rahman, P. Peterlongo, H. Nevanlinna, M.-C. King, F. J. Couch, M. C. Southey, R. Winqvist, W. D. Foulkes and M. Tischkowitz, *N. Engl. J. Med.*, 2014, **371**, 497–506.

251   C. L. Tucker and S. Fields, *Nat. Genet.*, 2003, **35**, 204–205.

252   G. Graziani and C. Szabó, *Pharmacol. Res.*, 2005, **52**, 109–118.

253   D. Ford, D. F. Easton, D. T. Bishop, S. A. Narod and D. E. Goldgar, *Lancet*, 1994, **343**, 692–695.

254   H. Farmer, N. McCabe, C. J. Lord, A. N. J. Tutt, D. A. Johnson, T. B. Richardson, M. Santarosa, K. J. Dillon, I. Hickson, C. Knights, N. M. B. Martin, S. P. Jackson, G. C. M. Smith and A. Ashworth, *Nature*, 2005, **434**, 917–921.

255   K. A. Menear, C. Adcock, R. Boulter, X. Cockcroft, L. Copsey, A. Cranston, K. J. Dillon, J. Drzewiecki, S. Garman, S. Gomez, H. Javaid, F. Kerrigan, C. Knights, A. Lau, V. M. Loh, I. T. W. Matthews, S. Moore, M. J. O'Connor, G. C. M. Smith and N. M. B. Martin, *J. Med. Chem.*, 2008, **51**, 6581–6591.

256   P. C. Fong, D. S. Boss, T. A. Yap, A. Tutt, P. Wu, M. Mergui-Roelvink, P. Mortimer, H. Swaisland, A. Lau, M. J. O'Connor, A. Ashworth, J. Carmichael, S. B. Kaye, J. H. M. Schellens and J. S. de Bono, *N. Engl. J. Med.*, 2009, **361**, 123–134.

257   J. F. Liu and U. A. Matulonis, *Curr. Oncol. Rep.*, 2016, **18**, 29.

258   S. R. P. E, T. S. Reddy and G. S. Krushna, *Curr. Res. Drug Target.*, 2011, **1**, 6–17.

259   N. M. Luscombe, D. Greenbaum and M. Gerstein, *Yearb. Med. Inform.*, 2001, 83–100.

260   M. Hann and R. Green, *Curr. Opin. Chem. Biol.*, 1999, **3**, 379–383.

261   W. L. Jorgensen, *Science*, 2004, **303**, 1813–1818.

262　T. Engel, *J. Chem. Inf. Model.*, 2006, **46**, 2267–2277.

263　B. F. Begam and J. S. Kumar, *Procedia Eng.*, 2012, **38**, 1264–1275.

264　M. A. Martí-Renom, A. C. Stuart, A. Fiser, R. Sánchez, F. Melo and A. Šali, *Annu. Rev. Biophys. Biomol. Struct.*, 2000, **29**, 291–325.

265　T. Schmidt, A. Bergner and T. Schwede, *Drug Discov. Today*, 2014, **19**, 890–897.

266　D. T. Jones, W. R. Taylor and J. M. Thornton, *Nature*, 1992, 358, 86–89.

267　J. Peng and J. Xu, *Bioinformatics*, 2010, **26**, i294–i300.

268　D. Xu and Y. Zhang, *Proteins Struct. Funct. Bioinforma.*, 2012, **80**, 1715–1735.

269　K. A. Dill, S. B. Ozkan, T. R. Weikl, J. D. Chodera and V. A. Voelz, *Curr. Opin. Struct. Biol.*, 2007, **17**, 342–346.

270　D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K. P. Tsafou, M. Kuhn, P. Bork, L. J. Jensen and C. Von Mering, *Nucleic Acids Res.*, 2015, **43**, D447–D452.

271　J. C. Fuller, N. J. Burgoyne and R. M. Jackson, *Drug Discov. Today*, 2009, **14**, 155–161.

272　A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani and J. P. Overington, *Nucleic Acids Res.*, 2012, **40**, 1–8.

273　M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kotter, T. Meinl, P. Ohl, K. Thiel and B. Wiswedel, *ACM SIGKDD Explor. Newsl.*, 2009, **11**, 26–31.

274　S. Beisken, T. Meinl, B. Wiswedel, L. F. de Figueiredo, M. Berthold and C. Steinbeck, *BMC Bioinformatics*, 2013, **14**, 257.

275　J. F. Blake, *Curr. Opin. Biotechnol.*, 2000, **11**, 104–107.

276　E. R. Bennett, J. Clausen, E. Linkov and I. Linkov, *Chemosphere*, 2009, **77**, 1412–1418.

277　S. W. Muchmore, J. J. Edmunds, K. D. Stewart and P. J. Hajduk, *J. Med. Chem.*, 2010, **53**, 4830–4841.

278　R. D. Taylor, P. J. Jewsbury and J. W. Essex, *J. Comput. Aided. Mol. Des.*, 2002, **16**, 151–166.

279　E. Yuriev, M. Agostino and P. A. Ramsland, *J. Mol. Recognit.*, 2011, **24**, 149–164.

280　E. Yuriev and P. A. Ramsland, *J. Mol. Recognit.*, 2013, **26**, 215–239.

281　T. a Halgren, *J. Comput. Chem.*, 1996, **17**, 520–552.

282　Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. Kollman, *J. Comput. Chem.*,

2003, **24**, 1999–2012.

283 W. P. Walters, M. T. Stahl and M. a Murcko, *Drug Discov. Today*, 1998, **3**, 160–178.

284 ChemAxon, *https://www.chemaxon.com/*.

285 RDKit: Open-source cheminformatics, *http://www.rdkit.org*.

286 ChemicalComputingGroupInc., *Sci. Comput. Instrum.*, 2004, **22**, 32.

287 *Mol. Oper. Environ. (MOE), 2015.08; Chem. Comput. Gr. Inc., 1010 Sherbooke St. West, Suite #910, Montr. QC, Canada, H3A 2R7*, 2016.

288 A. R. Bayly, A. J. P. White and A. C. Spivey, *European J. Org. Chem.*, 2013, **2013**, 5566–5569.

289 J.-P. Ebejer, G. M. Morris and C. M. Deane, *J. Chem. Inf. Model.*, 2012, **52**, 1146–58.

290 L. Schrödinger, *The PyMOL Molecular Graphics System, Version 1.8*, 2015.

291 B. Chagot, C. Pimentel, L. Dai, J. Pil, J. Tytgat, T. Nakajima, G. Corzo, H. Darbon and G. Ferrat, *Biochem. J.*, 2005, **388**, 263–271.

292 C. Louis-Jeune, M. A. Andrade-Navarro and C. Perez-Iratxeta, *Proteins Struct. Funct. Bioinforma.*, 2012, **80**, 374–381.

293 B. E. Maryanoff and A. B. Reitz, *Chem. Rev.*, 1989, **89**, 863–927.

294 G. Wittig and U. Schöllkopf, *Chem. Ber.*, 1954, **87**, 1318–1330.

295 G. Wittig and W. Haag, *Chem. Ber.*, 1955, **88**, 1654–1666.

296 L. Horner, H. Hoffmann, H. G. Wippel and G. Klahre, *Chem. Ber.*, 1959, **92**, 2499–2505.

297 W. S. Wadsworth and W. D. Emmons, *J. Am. Chem. Soc.*, 1961, **83**, 1733–1738.

298 G. Duret, R. Quinlan, P. Bisseret and N. Blanchard, *Chem. Sci.*, 2015, **6**, 5366–5382.

299 J. C. Tellis, D. N. Primer and G. A. Molander, *Science*, 2014, **345**, 433–436.

300 Y. Yamashita, J. C. Tellis and G. A. Molander, *Proc. Natl. Acad. Sci.*, 2015, **112**, 12026–12029.

301 F. Ullmann and P. Sponagel, *Berichte der Dtsch. Chem. Gesellschaft*, 1905, **38**, 2211–2212.

302 I. Goldberg, *Berichte der Dtsch. Chem. Gesellschaft*, 1906, **39**, 1691–1692.

303 F. Monnier and M. Taillefer, *Angew. Chemie - Int. Ed.*, 2009, **48**, 6954–6971.

304 A. S. Guram and S. L. Buchwald, *J. Am. Chem. Soc.*, 1994, **116**, 7901–7902.

305 J. Louie and J. F. Hartwig, *Tetrahedron Lett.*, 1995, **36**, 3609–3612.

306   A. S. Guram, R. A. Rennels and S. L. Buchwald, *Angew. Chemie Int. Ed. English*, 1995, **34**, 1348–1350.

307   R. F. Heck and J. P. Nolley, *J. Org. Chem.*, 1972, **37**, 2320–2322.

308   K. Sonogashira, Y. Tohda and N. Hagihara, *Tetrahedron Lett.*, 1975, **16**, 4467–4470.

309   R. Chinchilla and C. Nájera, *Chem. Rev.*, 2007, **107**, 874–922.

310   N. Miyaura, K. Yamada and A. Suzuki, *Tetrahedron Lett.*, 1979, **20**, 3437–3440.

311   N. Miyaura and A. Suzuki, *Chem. Rev.*, 1995, **95**, 2457–2483.

312   E. P. Gillis and M. D. Burke, *J. Am. Chem. Soc.*, 2007, **129**, 6716–6717.

313   S. Schröter, C. Stock and T. Bach, *Tetrahedron*, 2005, **61**, 2245–2267.

314   E. Fischer, *Justus Liebig's Ann. der Chemie*, 1886, **236**, 126–151.

315   M. Hilmy Elnagdi, N. A. Al-Awadi and I. Abdelshafy Abdelhamid, in *Advances in Heterocyclic Chemistry*, Academic Press, 2009, vol. Volume 97, pp. 1–43.

316   P. Schmidt and J. Druey, *Helv. Chim. Acta*, 1954, **37**, 1467–1471.

317   S. Evans and E. E. Schweizer, *J. Org. Chem.*, 1977, **42**, 2321–2324.

318   Z. Wang, in *Comprehensive Organic Name Reactions and Reagents*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2010.

319   M. J. Haddadin, S. J. Firsan and B. S. Nader, *J. Org. Chem.*, 1979, **44**, 629–630.

320   M. S. South, T. L. Jakuboski, M. D. Westmeyer and D. R. Dukesherer, *J. Org. Chem.*, 1996, **61**, 8921–8934.

321   H. Bel Abed, O. Mammoliti, O. Bande, G. Van Lommen and P. Herdewijn, *J. Org. Chem.*, 2013, **78**, 7845–7858.

322   H. Bel Abed, O. Mammoliti, O. Bande, G. Van Lommen and P. Herdewijn, *Org. Biomol. Chem.*, 2014, 7159–7166.

323   Q. Gao, Y. Zhu, M. Lian, M. Liu, J. Yuan, G. Yin and A. Wu, *J. Org. Chem.*, 2012, **77**, 9865–9870.

324   M. Terada, K. Soga and N. Momiyama, *Angew. Chemie Int. Ed.*, 2008, **47**, 4122–4125.

325   M. Oliverio, M. Nardi, P. Costanzo, L. Cariati, G. Cravotto, S. Giofrè and A. Procopio, *Molecules*, 2014, **19**, 5599–5610.

326   S. F. Yip, H. Y. Cheung, Z. Zhou and F. Y. Kwong, *Org. Lett.*, 2007, **9**, 3469–3472.

327   D. Maiti and S. L. Buchwald, *J. Org. Chem.*, 2010, **75**, 1791–1794.

328   H. Weingarten, *J. Org. Chem.*, 1964, **29**, 3624–3626.

329   J.-X. Zhang, P. Dubois and R. Jérôme, *Synth. Commun.*, 1996, **26**, 3091–3095.

330   A. Michaelis and R. Kaehne, *Berichte der Dtsch. Chem. Gesellschaft*, 1898, **31**, 1048–1055.

331   R. F. Heck, in *Organic Reactions*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 1982, pp. 345–390.

332   D. S. Daniels and A. Schepartz, *J. Am. Chem. Soc.*, 2007, **129**, 14578–14579.

333   F. Dierschke, A. C. Grimsdale and K. Müllen, *Synthesis*, 2003, 2470–2472.

334   H. S. P. Rao, S. Jothilingam and H. W. Scheeren, *Tetrahedron*, 2004, **60**, 1625–1630.

335   X. F. Wu, J. Schranck, H. Neumann and M. Beller, *Tetrahedron Lett.*, 2011, **52**, 3702–3704.

336   P. Gao, X. Feng, X. Yang, V. Enkelmann, M. Baumgarten and K. Müllen, *J. Org. Chem.*, 2008, **73**, 9207–9213.

337   B. P. Das and D. W. Boykin, *J. Med. Chem.*, 1977, **20**, 1219–1221.

338   Y.-H. Sun, T.-Y. Sun, Y.-D. Wu, X. Zhang and Y. Rao, *Chem. Sci.*, 2016, **7**, 2229–2238.

339   J.-Y. Yeh, M. S. Coumar, H.-Y. Shiao, T.-J. Lin, Y.-C. Lee, H.-C. Hung, S. Ko, F.-M. Kuo, M.-Y. Fang, Y.-L. Huang, J. T. A. Hsu, T.-K. Yeh, S.-R. Shih, Y.-S. Chao, J.-T. Horng and H.-P. Hsieh, *ChemMedChem*, 2012, **7**, 1546–1550.

340   World Patent WO2011/4276 A1, 2011.

341   A. Tlili, F. Monnier and M. Taillefer, *Chem. - A Eur. J.*, 2010, **16**, 12299–12302.

342   B. Liu, M. Yin, H. Gao, W. Wu and H. Jiang, *J. Org. Chem.*, 2013, **78**, 3009–3020.

343   A. K. Macharla, R. Chozhiyath Nappunni, M. R. Marri, S. Peraka and N. Nama, *Tetrahedron Lett.*, 2012, **53**, 191–195.

344   H. S. P. Rao and S. Jothilingam, *J. Org. Chem.*, 2003, **68**, 5392–5394.

345   F. Turksoy, J. D. Wallis, U. Tunca and T. Ozturk, *Tetrahedron*, 2003, **59**, 8107–8116.

346   Y.-N. Duan, L.-Q. Cui, L.-H. Zuo and C. Zhang, *Chem. - A Eur. J.*, 2015, **21**, 13052–13057.

347   H. F. Nour, N. Hourani and N. Kuhnert, *Org. Biomol. Chem.*, 2012, **10**, 4381–4389.

348   J. Yang, D. Wang, W. Liu, X. Zhang, F. Bian and W. Yu, *Green Chem.*, 2013, **15**, 3429–3437.

349   T. M. Konrad, J. T. Durrani, C. J. Cobley and M. L. Clarke, *Chem. Commun.* ,

2013, **49**, 3306–3308.

350   R. Takeuchi and H. Yasue, *J. Org. Chem.*, 1993, **58**, 5386–5392.

351   N. R. Vautravers, D. D. Regent and B. Breit, *Chem. Commun.*, 2011, **47**, 6635–6637.

352   D. R. Boyd, N. D. Sharma, N. I. Bowers, R. Boyle, J. S. Harrison, K. Lee, T. D. H. Bugg and D. T. Gibson, *Org. Biomol. Chem.*, 2003, **1**, 1298–1307.

353   J. Cowell, M. Abualnaja, S. Morton, R. Linder, F. Buckingham, P. G. Waddell, M. R. Probert and M. J. Hall, *RSC Adv.*, 2015, **5**, 16125–16152.

354   K. R. Campos, J. C. S. Woo, S. Lee and R. D. Tillyer, *Org. Lett.*, 2004, **6**, 79–82.

355   S. Tong, Z. Xu, M. Mamboury, Q. Wang and J. Zhu, *Angew. Chemie Int. Ed.*, 2015, **54**, 11809–11812.

ROYAL SOCIETY
OF CHEMISTRY

PAPER

CrossMark
← click for updates

An atom-efficient and convergent approach to the preparation of NS5A inhibitors by C−H activation†

Thomas O. Moore,* Michael Paradowski and Simon E. Ward

A novel approach of the convergent functionalisation of aryl dibromides to form NS5A type inhibitors using C−H activation is reported. The focus of investigation was to reduce the formation of homodimeric side product, as well as to investigate the scope of different aryl dibromides that were tolerated under the reaction conditions. The C−H activation methodology was found to give a viable synthetic route to NS5A inhibitors, with late stage functionalisation of the core portion of the molecule, albeit with some chemical functionalities not tolerated.

1

EC$_{50}$ (pM) G1a=50; G1b=9

Fig. 1 Structure of daclatasvir.

Introduction

The hepatitis C virus (HCV) is estimated to affect 170 million people worldwide and, without treatment, may lead to cirrhosis, hepatocellular carcinoma, liver failure and ultimately death.[1–4] Until recently, the standard of care for HCV patients was a combination of lengthy regimens of pegylated interferon α (pegIFN-α) and ribavirin.[5] Those poorly tolerated and moderately efficacious therapies contributed to the urgent need to identify improved treatments by oral administration. One major focus to achieve this was targeted at the nonstructural protein 5A (NS5A) which is enabled in the HCV genome and plays an essential role for the viral RNA replication.[6–8] Those efforts led to the discovery of daclatasvir (BMS) (Fig. 1), a non 'rule of 5' compliant NS5a inhibitor exhibiting picomolar activity against HCV genotypes 1a and 1b.[9] Subsequently, there have been over 70 patents and publications describing novel NS5A inhibitors related to Daclatasvir.[9–14] Of these, many have maintained its symmetrical biphenylimidazole core with differentiation arising from key structural aryl alterations.[15–18]

The existing strategy for the preparation of dimeric NS5A inhibitors is by chain elongation from an aryl bis acetate core. However, this methodology is linear and does not allow rapid exploration of the central core unit. A number of reports and patents have also described the preparation of symmetrical aryl-imidazoles by a late stage cross coupling of bis-metalated aryl core to the required halogenated imidazole fragments.[17–20] This methodology offers the advantage of a late diversification of the central core units but suffers from the lengthy, expensive and poor atom-efficient processes to install the pre-

requisite organometallic reagents. However, to the best of our knowledge, none of the late stage cross coupling methodologies reported had the key pharmacophores comprising of a pair of (S)-methoxycarbonylvaline capped pyrrolidines already in place and required further synthetic steps to prepare the reported NS5A inhibitors.
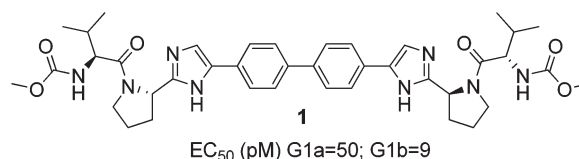
Our search for methodology that allowed us to form this bond as late as possible highlighted the opportunity to consider a double C–H activation strategy. The C–H activation of imidazoles in the desired C-5 position is one that has undergone much development in recent years.[21–25] Arguably the key development was the establishment of broadly applicable conditions for the C–H activation of heterocycles by Fagnou,[24] where the use of pivalic acid as an additive was found to greatly increase the rate and overall yield of reaction due to transition state stabilisation.[26] Given that the SEM group has been found to be well tolerated under C–H activation conditions,[27] a synthetic route based around the reaction between SEM-protected imidazoles and aryl dibromides was developed.

Results and discussion

N-Boc-ʟ-prolinal **3** was obtained from the corresponding prolinol **2** under mild copper/TEMPO mediated aerobic oxidation conditions (Scheme 1). Although the Swern oxidation conditions gave **3** in quantitative yields, we had concerns about

*Sussex Drug Discovery Centre, University of Sussex, Brighton, England BN1 9QJ, UK.
E-mail: T.Moore@sussex.ac.uk*
†Electronic supplementary information (ESI) available. See DOI: 10.1039/c6ob00340k

This journal is © The Royal Society of Chemistry 2016

Org. Biomol. Chem., 2016, **14**, 3307–3313 | **3307**
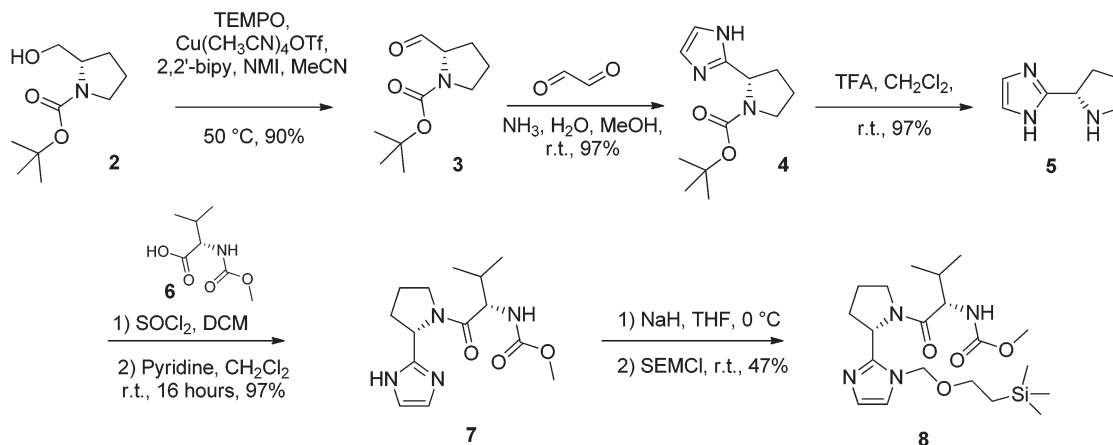
Scheme 1 Preparation of the key (S)-methoxycarbonylvaline capped pyrrolidine **7**.

the potential racemisation of the carbonyl α-carbon due to the presence of triethylamine, and thus opted for milder conditions on scale up. Cyclisation of the prolinal **3** under glyoxal and ammonia conditions followed by removal of the BOC protecting group gave pyrrolidine **5** in 85% yield over three steps. Methyl carbamate-L-valine **6**, obtained *via* reported Schotten–Bauman conditions from L-valine,[28] was then coupled to imidazole pyrrolidine **5** by peptide bond formation, although this proved challenging due to high aqueous solubility of the product formed and its instability when purified by silica chromatography. Mild coupling reagents such as guanidinium salts and carbodiimides were found to be unsuitable mainly due to the difficulty separating the product and byproducts formed. The respective acid chloride was found to react cleanly with pyrrolidine **5**, however we were concerned with the possibility of ketene formation from HCl elimination, leading to racemisation of the chiral valyl carbon as reported by Sauer.[29] Pyridine was used instead of triethylamine to reduce the likelihood of racemisation and gave **7** in 97% yield. D-Valine-carbamate was also coupled with **5** and both NMR and HPLC experiments confirmed the lack of racemisation under the reaction conditions. Subsequent protection of imidazole **7** with 2-(trimethylsilyl)ethoxymethyl (SEM) chloride in THF with sodium hydride as a base gave the key C–H activation precursor **8** in 47% yield.

Pleasingly, **10** was obtained in moderate yields on first attempt by double C–H activation of **8** with 4-4′-dibromobiphenyl under the conditions reported by Fagnou (Pd(dppf)Cl₂, PivOH, DMA, 100 °C, 4 h, Table 1).[24] Unfortunately however a homo-coupled by-product **11** was produced, which proved difficult to separate from the product by standard purification techniques. In proof of concept studies on a simplified imidazole reagent carried out by our group, it was shown that optimisation of conditions focusing on the base, solvent and reaction temperature had the greatest effect on formation of this homo-coupled by-product. We therefore set about modifying these conditions to test whether it was possible to prevent any by-product formation. It was found that the previously established con-

Table 1 Optimisation conditions for the C–H activation

Entry	Base	Solvent	Temperature	Yield 9[a]	10:9[a]
1	K₂CO₃	DMF	100 °C	65%	0.13
2	K₂CO₃	Dioxane	100 °C	10%	0
3	K₂CO₃	NMP	100 °C	38%	0.05
4	K₂CO₃	THF	100 °C	25%	0.11
5	K₂CO₃	DMA	100 °C	78%	0.03
6	K₂CO₃	DMA	60 °C	1%	0
7	K₂CO₃	DMA	80 °C	5%	0.15
8	K₂CO₃	DMA	120 °C	69%	0.07
9	Cs₂CO₃	DMA	100 °C	21%	0.10
10	KOAc	DMA	100 °C	3%	0.23
11	K₃PO₄	DMA	100 °C	72%	0.06

[a] Yield of **9** and ratio of **10:9** determined by LCMS peak integration measured by adsorption at 254 nm. Conditions: (i) Pd(dppf)Cl₂, PivOH, base, 4 hours.

ditions[30] were not only the highest yielding, but also exhibited very low amounts of homo-coupled product by LCMS analysis. When the reaction was carried out in dioxane at 100 °C, or in DMA at 60 °C (Table 1, entries 2 and 6), no homo-coupled product was observed by LCMS, but the drop in overall yield and rate of reaction in both of these cases made them impractical. As such, the Fagnou conditions were deemed the best suited for use in the double C–H activation. The 78% yield observed by LCMS for condition 5 equated to an 84% yield upon scale up and purification by column chromatography.
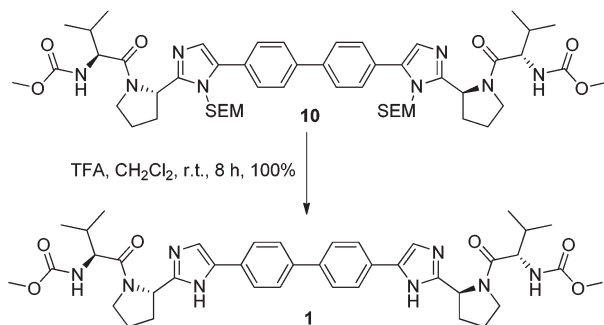
Scheme 2 Deprotection step under protolytic conditions.

Removal of the SEM protecting group post C–H activation using organic solubilised fluoride ions proved unsuccessful even at elevated temperatures and extended reaction times. However, conditions involving trifluoroacetic acid at room temperature cleaved the SEM group slowly but in quantitative yield (Scheme 2). Despite some initial stability concerns, the methyl carbamate functionality proved to be stable under the TFA deprotection conditions.

With the synthesis for **1** optimised, we turned our attention to the scope of this methodology for the rapid synthesis of potential symmetrical NS5A inhibitors (Table 2). Under the reaction conditions developed above, bis-CH activated products were obtained in moderate to good yields over 2 steps. Purification of deprotected product was carried out by prep-scale LCMS mass directed auto purification, giving better separation of the product and homo-dimer by-product. Overall, of 16 different aryl dibromides tested, 9 successfully underwent double C–H activation. The yields of successful couplings consistently ranges from moderate to good on a 0.1 mmol scale. The proportion of homo-dimer to product, however, was found to vary greatly, with a tendency to be highest in electron rich aryl di-bromides, such as entries 4 and 7.

Several functional groups were found to not be compatible with the reaction conditions used. The presence of an alkyne (entry 5) seemed to prevent formation of the anticipated product, and instead produced many unidentifiable side-products. Coordination of alkynes to palladium centres through an η-2 binding mode is well precedented to undergo a wide variety of reactions, and as such this seems likely to be responsible for the lack of desired reactivity observed.[31] Aryl di-bromides containing a 2-phenylpyridine (entries 6 and 15) functionality also did not allow the reaction to proceed as hoped, with the reaction mixture containing starting materials and a mixture of side products after prolonged heating. There is much precedent for the C–H activation directing capability of pyridyl nitrogen's onto 2-phenyl groups,[32] possibly accounting for the loss of desired reactivity, however there was no identifiable side product proving this hypothesis. Furthermore, no reaction was observed in the presence of aryl di-bromides containing an N–H functionality (entries 11, 12, 14 and 16). However, when the amide group in entry 12 was

Table 2 Scope of double C–H activation

Entry	Aryl dibromide	Yield[a] (%)	HD : P[b]
1		78%	0.13
2		80%	0
3		64%	0.37
4		37%	0.72
5		0%	—
6		0%	—
7		42%	0.68
8		67%	0.12
9[c]		45%	0.26
10[c]		58%	0.23
11		0%	—
12		0%	—
13		67%	0
14		0%	—

This journal is © The Royal Society of Chemistry 2016

Org. Biomol. Chem., 2016, **14**, 3307–3313 | **3309**

Table 2 (Contd.)

Entry	Aryl dibromide	Yield^a (%)	HD : P^b
15		0%	—
16		0%	—

^{a} Yield of product determined by LCMS peak integration measured by adsorption at 254 nm. Characterisation was carried out on the fully deprotected product (see Experimental section). ^{b} Ratio of homo-dimer to product. ^{c} Product decomposed under deprotection and characterisation was carried out on the SEM-protected product.

masked with an *N*-methyl functionality the reaction proceeded well (entry 13). This is particularly surprising considering the presence of a carbamate N–H bond within the imidazole starting material, which is tolerated in the reaction conditions.

Subsequent deprotection under TFA conditions led to the preparation of the NS5A-like analogues. To our surprise, the furan and thiophene (entries 9 and 10) were found to decompose upon exposure to TFA during the deprotection. As they were also found to be unreactive towards milder TBAF conditions, they were purified and characterised as the SEM-protected analogue instead.

Conclusions

In summary, we have described a novel, green and efficient approach to the synthesis of symmetrical NS5A-like compounds such as daclatasvir. The overall synthesis was accomplished using a novel double C–H activation as a key step that allows for a rapid synthesis of analogues and late stage structural diversification, which is to our knowledge the first time that palladium catalysed C–H activation has been used to construct symmetrical molecules in such a way. However, we found that intolerance of some co-ordinating functionalities in the reaction conditions gave some limitation to the methodology, and this must be considered when hoping to attempt this strategy.

Experimental section

General procedure for C–H activated cross coupling for compound without deprotection

The appropriate aryl dibromide (0.10 mmol), (*S*)-*tert*-butyl 2-(1-((2-(trimethylsilyl)ethoxy)methyl)-1*H*-imidazol-2-yl)pyrrolidine-1-carboxylate (7, 100 mg, 0.240 mmol), pivalic acid (6.1 mg, 0.060 mmol), K$_2$CO$_3$ (41 mg, 0.30 mmol) and Pd(ddpf)Cl$_2$ (7.3 mg, 0.010 mmol) were placed in a sealed reaction vial. The reaction vial was placed under reduced pressure then purged with nitrogen, repeating the process 5 times. The contents of the sealed vial were dissolved in dimethylacetamide (1 mL), and the solvent was deoxygenated by bubbling with nitrogen for 20 minutes. The solution was then irradiated in a microwave at 100 °C for 4 hours. The resulting solution was quenched with water (15 mL), and extracted with EtOAc (2 × 10 mL). The organic phase was washed with concentrated sodium bicarbonate solution (2 × 15 mL) water (2 × 15 mL), and brine (2 × 15 mL), before being concentrated to a solid under reduced pressure. This solid was then purified to the desired product by mass-directed auto purification.

Dimethyl(2*S*,2′*S*)-1,1′-((2*S*,2′*S*)-2,2′-(5,5′-(biphenyl-4,4′-diyl)bis(1-((2-(trimethylsilyl)ethoxy)methyl)-1*H*-imidazole-5,2-diyl))bis(pyrrolidine-2,1-diyl))bis(3-methyl-1-oxobutane-2,1-diyl)dicarbamate (10). ^{1}H NMR (500 MHz, CD$_3$OD) 7.79 (4H, d, *J* 7.8 Hz, ArH), 7.59 (4H, d, *J* 7.9 Hz, ArH), 7.04 (2H, s, ArH), 5.79 (2H, d, *J* 10.9 Hz, CH$_2$), 5.32 (2H, d, *J* 10.9 Hz, CH$_2$), 5.24 (2H, s, CH), 4.22 (2H, d, *J* 7.5 Hz, CH), 4.11–4.03 (2H, m, CH$_2$), 3.95–3.86 (2H, m, CH$_2$), 3.66 (6H, s, CH$_3$), 3.61–3.44 (4H, m, CH$_2$), 2.37 (4H, m, CH$_{AB}$), 2.21–1.96 (6H, m, CH$_{AB}$), 0.90 (16H, m, CH$_2$, CH$_3$), −0.03 (18H, s, Si(CH$_3$)$_3$); ^{13}C NMR (126 MHz, CD$_3$OD) 171.7 (CO), 158.0, O*C*(O)N), 150.9 (N*C*N), 140.0 (Ar*C*), 133.1 (Ar*C*), 129.4 (Ar*C*H), 128.8 (Ar*C*), 127.0 (Ar*C*H), 125.4 (Ar*C*H), 72.1 (*C*H$_2$), 65.6 (*C*H$_2$), 58.3 (*C*H), 53.2 (*C*H), 51.3 (*C*H$_3$), 47.6 (*C*H$_2$), 31.6 (*C*H$_2$), 30.1 (*C*H), 24.7 (*C*H$_2$), 18.4 (*C*H$_3$), 17.30 (*C*H$_2$), −2.84 (Si(*C*H$_3$)$_3$); **IR** (neat, ν_{max}, cm^{-1}) 3276 (bw), 2959 (m), 1719 (s), 1631 (s), 1507 (m), 1441 (s), 1248 (s), 1189 (m), 1178 (s), 1033 (m), 820.21 (s), 693 (m); **HRMS (ESI):** calcd for C$_{52}$H$_{78}$N$_8$Na$_1$O$_8$Si$_2$ [M + H]$^+$ 1021.5368, found 1012.5373; [α]$_D$ −39.4 (*c* 0.6 in CHCl$_3$, *T* 21.0 °C).

Dimethyl((2*S*,2′*S*)-((2*S*,2′*S*)-2,2′-(5,5′-(furan-2,5-diylbis(4,1-phenylene))bis(1-((2-(trimethylsilyl)ethoxy)methyl)-1*H*-imidazole-5,2-diyl))bis(pyrrolidine-2,1-diyl))bis(3-methyl-1-oxobutane-2,1-diyl))dicarbamate (12). ^{1}H NMR (500 MHz, CDCl$_3$) δ 7.80 (4H, d, *J* 8.0 Hz, Ar*H*), 7.49 (4H, d, *J* 8.1 Hz, Ar*H*), 7.05 (2H, s, Furyl*H*), 6.82 (2H, s, Ar*H*), 5.79 (2H, d, *J* 10.6 Hz, CH$_2$), 5.29 (2H, d, *J* 9.1 Hz, N*H*), 5.24–5.16 (4H, m, CH$_{AB}$, CH), 4.35 (2H, dd, *J* 9.2, 6.4 Hz, CH), 3.98–3.87 (4H, m, CH$_2$), 3.67 (6H, s, CH$_3$), 3.59–3.46 (4H, m, CH$_2$), 2.55–2.42 (2H, m, CH$_{AB}$), 2.39–2.26 (4H, m, CH$_2$), 2.15–1.91 (4H, m, CH$_{AB}$, CH), 1.01–0.87 (10H, m, CH$_3$, CH$_2$), 0.80 (6H, d, *J* 6.7 Hz, CH$_3$), 0.00 (18H, s, Si(CH$_3$)$_3$); ^{13}C NMR (126 MHz, CDCl$_3$) δ 170.9 (*C*O), 157.1 (O*C*(O)N), 153.3 (Ar*C*), 151.6 (Ar*C*), 150.7 (Ar*C*), 133.8 (Ar*C*), 132.8 (Ar*C*), 129.3 (Ar*C*H), 124.0 (Ar*C*H), 108.0 (Ar*C*H), 88.3 (Furyl*C*H), 65.9 (*C*H$_2$), 57.6 (*C*H), 52.9 (*C*H), 52.2 (*C*H$_3$), 47.6 (*C*H$_2$), 31.8 (*C*H$_2$), 31.1 (*C*H), 25.2 (*C*H$_2$), 19.3 (*C*H$_3$), 18.0

(CH_2), 17.3 (CH_3), −1.4 (Si(CH_3)$_3$); **IR** (neat, ν_{max}, cm^{-1}) 2955 (m), 1719 (s), 1640 (s), 1503 (m), 1440 (m), 1249 (s), 1188 (m), 1082 (s), 1030 (w), 837 (s), 697 (w); **HRMS (ESI):** calcd for $C_{56}H_{81}N_8O_7Si_2$ [M + H]$^+$ 1065.5660, found 1065.5642; $[\alpha]_D$ −102.1 (c 0.3 in CHCl$_3$, T 22.5 °C).

Dimethyl((2S,2′S)-((2S,2′S)-2,2′-(5,5′-(thiophene-2,5-diylbis-(4,1-phenylene))bis(1-((2-(trimethylsilyl)ethoxy)methyl)-1H-imidazole-5,2-diyl))bis(pyrrolidine-2,1-diyl))bis(3-methyl-1-oxobutane-2,1-diyl))dicarbamate (13). ^{1}H NMR (500 MHz, CDCl$_3$) δ 7.71 (4H, d, J 7.8 Hz, ArH), 7.47 (4H, d, J 8.0 Hz, ArH), 7.38 (2H, s, thiopheneH), 7.12 (2H, s, ArH), 5.81 (2H, d, J 10.6 Hz, CH$_{AB}$), 5.29–5.14 (6H, m, CH$_{AB}$, CH, NH), 4.39–4.31 (2H, m, CH), 4.21–4.05 (2H, m, CH$_{AB}$), 3.99–3.84 (2H, m, CH$_{AB}$), 3.66 (6H, s, CH$_3$), 3.59–3.43 (4H, m, CH$_2$), 2.58–2.28 (6H, m, CH$_{AB}$, CH$_2$), 2.17–1.93 (4H, m, CH$_{AB}$, CH), 1.01–0.85 (10H, m, CH$_3$, CH$_2$), 0.79 (6H, d, J 6.7 Hz, CH$_3$), 0.00 (18H, s, Si(CH$_3$)$_3$); ^{13}C NMR (126 MHz, CDCl$_3$) δ 171.4 (CO), 157.1 (OC(O)N), 150.3 (NCN), 143.2 (ArC), 134.5 (ArC), 134.1 (ArC), 132.9 (ArC), 129.7 (ArCH), 125.9 (ArCH), 124.8 (thiopheneCH), 118.4 (ArCH), 72.9 (CH$_2$), 66.4 (CH$_2$), 57.7 (CH), 52.8 (CH), 52.2 (CH$_3$), 47.7 (CH$_2$), 31.7 (CH$_2$), 30.6 (CH), 25.4 (CH$_2$), 19.4 (CH$_3$), 18.0 (CH$_2$), 17.2 (CH3), −1.4 (Si(CH$_3$)$_3$); **IR** (neat, ν_{max}, cm^{-1}) 3681 (bw), 2967 (m), 2844 (m), 1718 (m), 1637 (m), 1508 (w), 1455 (m), 1429 (m), 1055 (s), 1033 (s), 1015 (s), 838 (m); **HRMS (ESI):** calcd for $C_{56}H_{81}N_8O_6Si_2S$ [M + H]$^+$ 1081.5431, found 1081.5449; $[\alpha]_D$ −69.1 (c 1.2 in CHCl$_3$, T 22.7 °C).

General procedure for C–H activated cross coupling and subsequent SEM-removal

The appropriate aryl dibromide (0.10 mmol), (S)-tert-butyl 2-(1-((2-(trimethylsilyl)ethoxy)methyl-1H-imidazol-2-yl)pyrrolidine-1-carboxylate (7, 100 mg, 0.240 mmol), pivalic acid (6.1 mg, 0.060 mmol), K$_2$CO$_3$ (41 mg, 0.30 mmol) and Pd(ddpf)Cl$_2$ (7.3 mg, 0.010 mmol) were placed in a sealed reaction. The reaction vial was placed under reduced pressure then purged with nitrogen, repeating the process 5 times. The contents of the sealed vial were dissolved in dimethylacetamide (1 mL), and the solvent was deoxygenated by bubbling with nitrogen for 20 minutes. The solution was then irradiated in a microwave at 100 °C for 4 hours. The resulting solution was quenched with water (15 mL), and extracted with EtOAc (2 × 10 mL). The organic phase was washed with concentrated sodium bicarbonate solution (2 × 15 mL) water (2 × 15 mL), and brine (2 × 15 mL), before being concentrated to a solid under reduced pressure. This solid material was then re-dissolved in dichloromethane (0.5 mL) and trifluoroacetic acid (0.50 mL, 6.5 mmol) was added. The solution was then stirred in inert atmosphere for 4 hours, before being diluted with dichloromethane (15 mL). The solution was then washed with concentrated sodium bicarbonate solution (2 × 30 mL), taking care due to the production of gas as the solution is neutralised. The organic phase was then concentrated to a solid under reduced pressure. This solid was then purified to the desired product by mass-directed auto purification.

Dimethyl((2S,2′S)-((2S,2′S)-2,2′-(5,5′-([1,1′-biphenyl]-4,4′-diyl)bis(1H-imidazole-5,2-diyl))bis(pyrrolidine-2,1-diyl))bis(3-methyl-1-

oxobutane-2,1-diyl))dicarbamate (1). ^{1}H **NMR** (500 MHz, CD$_3$OD) 7.73 (4H, d, J 8.2 Hz, ArH), 7.66 (4H, d, J 8.0, Ar), 7.37 (2H, s, ArH), 5.19 (2H, dd, J 8.0 Hz, 6.3 Hz, CH), 4.24 (2H, d, J 7.4 Hz, CH), 4.05–3.95 (2H, m, CH$_{AB}$), 3.92–3.84 (2H, m, CH$_{AB}$), 3.66 (6H, s, CH$_3$), 2.41–2.17 (6H, m, CH$_2$, CH$_{AB}$), 2.11–2.04 (4H, m, CH$_{AB}$, CH), 0.93 (12H, dd, J 22.0 Hz, 6.8 Hz, i-PrCH$_3$); ^{13}C **NMR** (126 MHz, CD$_3$OD) 172.0 (CO), 158.0 (NCN), 149.6 (OC(O)N), 139.0 (ArC), 136.1 (ArC), 130.9 (ArC), 126.6 (ArCH), 125.0 (ArCH), 116.2 (ArCH), 58.4 (CH), 54.8 (CH), 51.3 (CH$_3$), 47.4 (CH$_2$), 30.9 (CH$_2$), 30.1 (CH), 24.7 (CH$_2$), 18.4 (i-PrCH$_3$), 17.0 (i-PrCH$_3$); **IR** (neat, ν_{max}, cm^{-1}) 2964 (w), 2875 (w), 1705 (s), 1632 (s), 1446 (s), 1393 (s), 1258 (w), 1035 (m), 829 (m), 773 (m); **HRMS (ESI):** calcd for $C_{40}H_{51}N_8O_6$ [M + H]$^+$ 739.3926, found 739.3937; $[\alpha]_D$ −170.0 (c 0.6 in CHCl$_3$, T 22.2 °C).

Dimethyl((2S,2′S)-((2S,2′S)-2,2′-(5,5′-(1,4-phenylene)bis(1H-imidazole-5,2-diyl))bis(pyrrolidine-2,1-diyl))bis(3-methyl-1-oxo-butane-2,1-diyl))dicarbamate (14). ^{1}H **NMR** (500 MHz, CD$_3$OD) 7.88 (4H, s, Ar), 7.48 (2H, s, ArH), 5.29–5.19 (2H, m, CH), 4.29–4.17 (2H, m, CH), 4.14–3.97 (2H, m, CH$_{AB}$), 3.93–3.77 (2H, m, CH$_{AB}$), 3.66 (6H, s, CH$_3$), 2.60–2.46 (CH$_{AB}$, m, 2H), 2.32–1.95 (8H, m, CH$_{AB}$, CH$_2$, CH), 0.96–0.85 (12H, m, i-PrCH$_3$); ^{13}C **NMR** (126 MHz, CD$_3$OD) 172.7 (CO), 158.1 (NCN), 149.4 (OC(O)N), 147.5 (ArC), 126.0 (ArC), 118.7 (ArCH), 115.3 (ArCH), 58.5 (CH), 53.7 (CH), 53.2 (CH$_2$), 51.4 (CH$_3$), 29.8 (CH), 24.9 (CH$_2$), 18.4 (CH$_2$), 16.7 (i-PrCH$_3$); **IR** (neat, ν_{max}, cm^{-1}) 3284 (bw), 2962 (w), 1711 (s), 1625 (s), 1521 (s), 1444 (s), 1370 (w), 1264 (s), 1192 (m), 1098 (m), 1037 (m), 846 (m), 734 (s); **HRMS (ESI):** calcd for $C_{34}H_{47}N_8O_6$ [M + H]$^+$ 663.3613, found 663.3608; $[\alpha]_D$ −288.6 (c 0.3 in CHCl$_3$, T 22.6 °C).

Dimethyl((2S,2′S)-((2S,2′S)-2,2′-(5,5′-((E)-ethene-1,2-diylbis(4,1-phenylene))bis(1H-imidazole-5,2-diyl))bis(pyrrolidine-2,1-diyl))bis(3-methyl-1-oxobutane-2,1-diyl))dicarbamate (15). ^{1}H **NMR** (500 MHz, CD$_3$OD) 8.16 (2H, s, alkeneH), 7.92 (2H, d, J 8.6 Hz, ArCH), 7.82 (2H, d, J 8.3 Hz, ArCH), 7.66 (2H, s, ArCH), 5.24 (2H, dd, J 8.1 Hz, 6.2 Hz, CH), 4.26 (2H, d, J 7.3 Hz, CH), 4.09–4.02 (2H, m, CH$_{AB}$), 3.97–3.81 (2H, m, CH$_{AB}$), 3.67 (6H, s, CH$_3$), 2.52–2.41 (2H, m, CH$_{AB}$), 2.32–2.02 (8H, m, CH$_{AB}$, CH$_2$, CH), 0.94 (12H, dd, J 24.3, 6.7, i-PrCH$_3$); ^{13}C **NMR** (126 MHz, CD$_3$OD) 172.4 (CO), 163.2 (NCN), 158.1 (ArC), 149.6 (OC(O)N), 133.0 (ArC), 128.5 (ArC), 125.0 (alkeneCH), 123.7 (ArCH), 123.0 (ArCH), 115.8 (ArCH), 58.5 (CH), 54.4 (CH), 51.4 (CH$_3$), 47.5 (CH$_2$), 30.9 (CH$_2$), 30.0 (CH), 24.8 (CH$_2$), 18.4 (CH$_3$), 16.9 (i-PrCH$_3$); **IR** (neat, ν_{max}, cm^{-1}) 2966 (w), 2876 (w), 1701 (s), 1631 (s), 1521 (m), 1443 (m), 1264 (w), 1201 (s), 1132 (m), 1033 (w), 799 (w), 719 (m); **HRMS (ESI):** calcd for $C_{42}H_{53}N_8O_6$ [M + H]$^+$ 765.4083, found 765.4118; $[\alpha]_D$ −538.3 (c 0.3 in CHCl$_3$, T 22.6 °C).

Dimethyl((2S,2′S)-((2S,2′S)-2,2′-(5,5′-(naphthalene-2,6-diyl)bis(1H-imidazole-5,2-diyl))bis(pyrrolidine-2,1-diyl))bis(3-methyl-1-oxobutane-2,1-diyl))dicarbamate (16). ^{1}H **NMR** (500 MHz, CD$_3$OD) 7.64–7.54 (4H, m, ArH), 7.41 (2H, s, ArH), 7.20–7.13 (2H, m, ArCH), 5.19 (2H, t, J 7.0 Hz, CH), 4.24 (2H, d, J 7.3 Hz, CH), 4.06–3.97 (2H, m, CH$_{AB}$), 3.92–3.83 (2H, m, 22/26-CH$_{AB}$), 3.66 (6H, s, CH$_3$), 2.43–2.32 (2H, m, CH$_{AB}$), 2.31–1.97 (8H, m, CH$_{AB}$, CH$_2$, CH), 0.93 (12H, dd, J 22.8 Hz, 6.9 Hz, i-PrCH$_3$);

This journal is © The Royal Society of Chemistry 2016

Org. Biomol. Chem., 2016, 14, 3307–3313 | 3311

^{13}C NMR (126 MHz, CD$_3$OD) 172.2 (CO), 158.0 (NCN), 149.4 (OC(O)N), 136.7 (ArC), 135.6 (ArC), 130.0 (ArC), 127.8 (ArC), 126.7 (ArCH), 124.8 (ArCH), 115.9 (ArCH), 58.4 (CH), 54.6 (CH), 51.4 (CH$_3$), 47.5 (CH$_2$), 30.9 (CH$_2$), 30.1 (CH), 24.7 (CH$_2$), 18.4 (i-PrCH$_3$), 16.9 (i-PrCH$_3$); IR (neat, ν_{max}, cm^{-1}) 3287 (bw), 2962 (w), 2875 (w), 1711 (s), 1628 (w), 1517 (m) 1443 (m), 1370 (w), 1264 (m), 1181 (m), 1131 (m), 1035 (m), 947 (w), 841 (m), 775 (m), 732 (s); HRMS (ESI): calcd for C$_{38}$H$_{49}$N$_8$O$_6$ [M + H]$^+$ 713.3770, found 713.3779; $[\alpha]_D$ −222.1 (c 0.6 in CHCl$_3$, T 22.1 °C).

Dimethyl((2S,2′S)-((2S,2′S)-2,2′-(5,5′-(oxybis(4,1-phenylene)) bis(1H-imidazole-5,2-diyl))bis(pyrrolidine-2,1-diyl))bis(3-methyl-1-oxobutane-2,1-diyl))dicarbamate (17). ^{1}H NMR (500 MHz, CD$_3$OD) 8.13 (2H, s, NH), 7.68 (4H, d, J 8.1 Hz, ArH), 7.48 (2H, s, ArH), 7.08 (4H, d, J 8.0 Hz, ArH), 5.20 (2H, t, J 7.1 Hz, CH), 4.24 (2H, d, J 7.2 Hz, CH), 4.09–3.99 (2H, m, CH$_{AB}$), 3.92–3.82 (2H, m, CH$_{AB}$), 3.66 (6H, s, CH$_3$), 2.50–2.35 (2H, m, CH$_{AB}$), 2.31–1.97 (8H, m, CH$_{AB}$, CH$_2$, CH), 0.91 (12H, dd, J 20.4 Hz, 6.8 Hz, i-PrCH$_3$); ^{13}C NMR (126 MHz, CD$_3$OD) 172.3 (CO), 163.5 (NCN), 158.1 (OC(O)N), 156.9 (ArC), 149.1 (ArC), 134.9 (ArC), 126.6 (ArCH), 118.9 (ArCH), 114.8 (ArCH), 58.4 (CH), 54.3 (CH), 51.4 (CH$_3$), 47.5 (CH$_2$), 30.9 (CH$_2$), 30.0 (CH), 24.8 (CH$_2$), 18.4 (i-PrCH$_3$), 16.8 (i-PrCH$_3$); IR (neat, ν_{max}, cm^{-1}) 3294 (bw), 2962 (w), 2875 (w), 1712 (m), 1627 (s), 1497 (m), 1424 (m), 1237 (s), 1199 (m), 1198 9m), 1130 (m), 1033 (m), 798 (m), 733 (s), 701 (m); HRMS (ESI): calcd for C$_{40}$H$_{51}$N$_8$O$_7$ [M + H]$^+$ 755.3875, found 755.3882; $[\alpha]_D$ −114.3 (c 0.6 in CHCl$_3$, T 22.0 °C).

Methyl N-[(2S)-1-[(2S)-2-{5-[4-(5-{2-[(2S)-1-[(2S)-2-[(methoxy-carbonyl)amino]-3-methylbutanoyl]pyrrolidin-2-yl]-1H-imid-azol-5-yl}-1,3-benzoxazol-2-yl)phenyl]-1H-imidazol-2-yl}pyrrolidin-1-yl]-3-methyl-1-oxobutan-2-yl]carbamate (18). ^{1}H NMR (500 MHz, CD$_3$OD) δ 8.33–8.20 (2H, m, ArH), 8.02 (1H, d, J 8.3 Hz, ArH), 7.94–7.84 (2H, m, ArH), 7.83–7.70 (2H, m, ArH), 7.72–7.59 (2H, m, ArH), 5.27–5.18 (2H, m, CH), 4.25 (2H, t, J = 7.3 Hz, CH), 4.13–3.98 (2H, m, CH$_{AB}$), 3.95–3.84 (2H, m, CH$_{AB}$), 3.66 (6H, s, CH$_3$), 2.57–2.35 (2H, m, CH$_{AB}$), 2.32–2.02 (8H, m, CH$_{AB}$, CH$_2$, CH), 0.99–0.89 (12H, m, iPrCH$_3$); ^{13}C NMR (126 MHz, CD$_3$OD $_3$) δ 173.7 (CO), 165.0 (ArC), 159.5 (OC(O)N), 152.9 (ArC), 152.4 (Ar–C), 152.0 (Ar–C), 151.5 (ArC), 150.3 (NCN), 136.6 (ArC), 136.0 (ArC), 129.3 (ArCH), 126.5 (ArCH), 124.5 (ArCH), 118.0 (ArCH), 117.5 (ArCH), 116.4 (ArCH), 112.5 (ArCH), 59.9 (CH), 55.4 (CH$_3$), 52.8 (CH$_2$), 49.0 (CH$_2$), 32.4 (CH$_2$), 31.5 (CH), 26.3 (CH$_2$), 19.8 (iPrCH$_3$), 18.3 (iPrCH$_3$); IR (neat, ν_{max}, cm^{-1}) 3284 (bw), 2966 (w), 2875 (w), 1699 (m), 1628 (s), 1517 (m), 1424 (m), 1368 (w), 1265 (m), 1180 (s), 1132 (s), 1039 (m), 825 (m), 798 (m), 732 (s); HRMS (ESI): calcd for C$_{41}$H$_{50}$N$_9$O$_7$ [M + H]$^+$ 780.3828, found 780.3804; $[\alpha]_D$ −446.6 (c 0.3 in CHCl$_3$, T 22.4 °C).

Methyl N-[(2S)-1-[(2S)-2-(5-{4-[(4-{2-[(2S)-1-[(2S)-2-[(methoxy-carbonyl)amino]-3-methylbutanoyl]pyrrolidin-2-yl]-1H-imidazol-5-yl}phenyl)(methyl)carbamoyl]phenyl}-1H-imidazol-2-yl)pyrro-lidin-1-yl]-3-methyl-1-oxobutan-2-yl]carbamate (19). ^{1}H NMR (500 MHz, CDCl$_3$) δ 7.51–7.39 (2H, m, ArH), 7.35–7.26 (2H, m, ArH), 7.26–7.20 (2H, m, ArH), 7.15–7.04 (2H, m, ArH), 6.98 (2H, d, J 8.2 Hz, ArH), 5.58 (2H, s, NH), 5.18 (2H, s, CH), 4.30 (2H, t,

J 8.1 Hz, CH), 3.90–3.77 (2H, m, CH$_{AB}$), 3.74–3.63 (8H, m, CH$_{AB}$, CH$_3$), 3.46 (3H, s, CH$_3$), 2.35–2.11 (4H, m, CH$_{AB}$, CH$_{AB}$), 2.09–1.85 (6H, m, CH$_{AB}$, CH$_{AB}$, CH), 0.88–0.77 (12H, m, i-PrCH$_3$); ^{13}C NMR (100 MHz, CDCl$_3$) δ 172.6 (CO), 170.2 (CO), 157.1 (OC(O)N), 148.5 (NCN), 143.6 (ArC), 138.5 (ArC), 134.3 (ArC), 132.3 (ArC), 129.3 (ArCH), 127.1 (ArCH), 125.5 (ArCH), 124.6 (ArC), 123.9 (ArCH), 115.3 (ArCH), 57.7 (CH), 54.2 (CH), 52.3 (CH$_3$), 47.8 (CH$_2$), 38.4 (CH$_3$), 30.8 (CH), 28.4 (CH$_2$), 25.2 (CH$_2$), 19.2 (i-PrCH$_3$), 17.43 (i-PrCH$_3$); IR (neat, ν_{max}, cm^{-1}) 3227 (bw), 2958 (w), 1698 (s), 1630 (s), 1532 (m), 1445 (m), 1396 (m), 1203 (m), 1137 (m), 1038 (w), 847 (w); HRMS (ESI): calcd for C$_{41}$H$_{50}$N$_9$O$_7$ [M + H]$^+$ 796.4141, found 796.4240; $[\alpha]_D$ −243.3 (c 0.6 in CHCl$_3$, T 22.4 °C).

Acknowledgements

References

1 R. S. Brown, Nature, 2005, 436, 973–978.

2 D. Lavanchy, Liver Int., 2009, 29, 74–81.

3 T. K. H. Scheel and C. M. Rice, Nat. Med., 2013, 19, 837–849.

4 D. L. Thomas, Nat. Med., 2013, 19, 850–858.

5 F. J. Torriani, M. Rodriguez-Torres, J. K. Rockstroh, E. Lissen, J. Gonzalez-García, A. Lazzarin, G. Carosi, J. Sasadeusz, C. Katlama, J. Montaner, H. Sette, S. Passe, J. De Pamphilis, F. Duff, U. M. Schrenk and D. T. Dieterich, N. Engl. J. Med., 2004, 351, 438–450.

6 R. Bartenschlager and V. Lohmann, J. Gen. Virol., 2000, 81, 1631–1648.

7 K. L. LeMay, J. Treadaway, I. Angulo and T. L. Tellinghuisen, J. Virol., 2013, 87, 1255–1260.

8 M. Hughes, S. Griffin and M. Harris, J. Gen. Virol., 2009, 90, 1329–1334.

9 M. Belema and N. A. Meanwell, J. Med. Chem., 2014, 57, 5057–5071.

10 M. Belema, O. D. Lopez, J. A. Bender, J. L. Romine, D. R. St. Laurent, D. R. Langley, J. A. Lemm, D. R. O'Boyle, J. H. Sun, C. Wang, R. A. Fridell and N. A. Meanwell, J. Med. Chem., 2014, 57, 1643–1672.

11 M. Belema, V. N. Nguyen, C. Bachand, D. H. Deon, J. T. Goodrich, C. A. James, R. Lavoie, O. D. Lopez, A. Martel, J. L. Romine, E. H. Ruediger, L. B. Snyder,

3312 | Org. Biomol. Chem., 2016, 14, 3307–3313

This journal is © The Royal Society of Chemistry 2016

D. R. S. Laurent, F. Yang, J. Zhu, H. S. Wong, D. R. Langley, S. P. Adams, G. H. Cantor, A. Chimalakonda, A. Fura, B. M. Johnson, J. O. Knipe, D. D. Parker, K. S. Santone, R. A. Fridell, J. A. Lemm, D. R. O'Boyle, R. J. Colonno, M. Gao, N. A. Meanwell and L. G. Hamann, *J. Med. Chem.*, 2014, **57**, 2013–2032.

12 D. R. St. Laurent, M. H. Serrano-Wu, M. Belema, M. Ding, H. Fang, M. Gao, J. T. Goodrich, R. G. Krause, J. A. Lemm, M. Liu, O. D. Lopez, V. N. Nguyen, P. T. Nower, D. R. O'Boyle, B. C. Pearce, J. L. Romine, L. Valera, J. H. Sun, Y. K. Wang, F. Yang, X. Yang, N. A. Meanwell and L. B. Snyder, *J. Med. Chem.*, 2014, **57**, 1976–1994.

13 M. Belema, V. N. Nguyen, D. R. St. Laurent, O. D. Lopez, Y. Qiu, A. C. Good, P. T. Nower, L. Valera, D. R. O'Boyle, J. H. Sun, M. Liu, R. A. Fridell, J. A. Lemm, M. Gao, J. O. Knipe, N. A. Meanwell and L. B. Snyder, *Bioorg. Med. Chem. Lett.*, 2013, **23**, 4428–4435.

14 M. Belema, V. N. Nguyen, J. L. Romine, D. R. St. Laurent, O. D. Lopez, J. T. Goodrich, P. T. Nower, D. R. O'Boyle, J. A. Lemm, R. A. Fridell, M. Gao, H. Fang, R. G. Krause, Y. K. Wang, A. J. Oliver, A. C. Good, J. O. Knipe, N. A. Meanwell and L. B. Snyder, *J. Med. Chem.*, 2014, **57**, 1995–2012.

15 D. A. DeGoey, J. T. Randolph, D. Liu, J. Pratt, C. Hutchins, P. Donner, A. C. Krueger, M. Matulenko, S. Patel, C. E. Motter, L. Nelson, R. Keddy, M. Tufano, D. D. Caspi, P. Krishnan, N. Mistry, G. Koev, T. J. Reisch, R. Mondal, T. Pilot-Matias, Y. Gao, D. W. A. Beno, C. J. Maring, A. Molla, E. Dumas, A. Campbell, L. Williams, C. Collins, R. Wagner and W. M. Kati, *J. Med. Chem.*, 2014, **57**, 2047–2057.

16 O. D. Lopez, V. N. Nguyen, D. R. St. Laurent, M. Belema, M. H. Serrano-Wu, J. T. Goodrich, F. Yang, Y. Qiu, A. S. Ripka, P. T. Nower, L. Valera, M. Liu, D. R. O'Boyle, J.-H. Sun, R. A. Fridell, J. A. Lemm, M. Gao, A. C. Good, N. A. Meanwell and L. B. Snyder, *Bioorg. Med. Chem. Lett.*, 2013, **23**, 779–784.

17 I. K. Mangion, C. Chen, H. Li, P. Maligres, Y. Chen, M. Christensen, R. Cohen, I. Jeon, A. Klapars, S. Krska, H. Nguyen, R. A. Reamer, B. D. Sherry and I. Zavialov, *Org. Lett.*, 2014, **16**, 2310–2313.

18 F. Wakenhut, T. D. Tran, C. Pickford, S. Shaw, M. Westby, C. Smith-Burchnell, L. Watson, M. Paradowski, J. Milbank, D. Stonehouse, K. Cheung, R. Wybrow, F. Daverio, S. Crook, K. Statham, D. Leese, D. Stead, F. Adam, D. Hay, L. R. Roberts, J.-Y. Chiva, C. Nichols, D. C. Blakemore, G. H. Goetz, Y. Che, I. Gardner, S. Dayal, A. Pike, R. Webster and D. C. Pryde, *ChemMedChem*, 2014, **9**, 1387–1396.

19 M. Zhong, E. Peng, N. Huang, Q. Huang, A. Huq, M. Lau, R. Colonno and L. Li, *Bioorg. Med. Chem. Lett.*, 2014, **24**, 5738–5742.

20 T. D. Tran, F. Wakenhut, C. Pickford, S. Shaw, M. Westby, C. Smith-Burchnell, L. Watson, M. Paradowski, J. Milbank, R. A. Brimage, R. Halstead, R. Glen, C. P. Wilson, F. Adam, D. Hay, J.-Y. Chiva, C. Nichols, D. C. Blakemore, I. Gardner, S. Dayal, A. Pike, R. Webster and D. C. Pryde, *ChemMedChem*, 2014, **9**, 1378–1386.

21 F. Bellina, S. Cauteruccio, L. Mannina, R. Rossi and S. Viel, *J. Org. Chem.*, 2005, **70**, 3997–4005.

22 F. Bellina, S. Cauteruccio, L. Mannina, R. Rossi and S. Viel, *Eur. J. Org. Chem.*, 2006, 693–703.

23 F. Bellina, S. Cauteruccio, A. Di Fiore and R. Rossi, *Eur. J. Org. Chem.*, 2008, 5436–5445.

24 B. Liégault, D. Lapointe, L. Caron, A. Vlassova and K. Fagnou, *J. Org. Chem.*, 2009, **74**, 1826–1834.

25 R. Goikhman, T. L. Jacques and D. Sames, *J. Am. Chem. Soc.*, 2009, **131**, 3042–3048.

26 S. I. Gorelsky, D. Lapointe and K. Fagnou, *J. Am. Chem. Soc.*, 2008, **130**, 10848–10849.

27 J. M. Joo, B. B. Touré and D. Sames, *J. Org. Chem.*, 2010, **75**, 4911–4920.

28 I. H. Bae, J. K. Choi, C. Chough, S. J. Keum, H. Kim, S. K. Jang and B. M. Kim, *ACS Med. Chem. Lett.*, 2014, **5**, 255–258.

29 J. C. Sauer, *J. Am. Chem. Soc.*, 1947, **69**, 2444–2448.

30 B. Liégault, D. Lapointe, L. Caron, A. Vlassova and K. Fagnou, *J. Org. Chem.*, 2009, **74**, 1826–1834.

31 R. Chinchilla and C. Nájera, *Chem. Rev.*, 2014, **114**, 1783–1826.

32 D. Alberico, M. E. Scott and M. Lautens, *Chem. Rev.*, 2007, **107**, 174–238.

This journal is © The Royal Society of Chemistry 2016

Org. Biomol. Chem., 2016, **14**, 3307–3313 | **3313**