

Activity-Directed Discovery of Inhibitors of the p53/Human-MDM2 Protein-Protein Interaction

Adam Ieuan Green

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds

School of Chemistry

and

Astbury Centre for Structural Molecular Biology

September 2020

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

A section of the work presented in Chapter 3 has appeared in the following publication:

“Efficient Approaches for the Synthesis of Diverse α -Diazo Amides”, S. Chow, **A. I. Green**, C. Arter, S. Liver, A. Leggott, L. Trask, G. Karageorgis, S. Warriner, and A. Nelson, *Synthesis*, 2020, **52**(11), 1695–1706.

From the work contained in this publication, the candidate developed methods for the synthesis of compounds containing α -diazo amide functional groups (methods A and C), and synthesised nine compounds. Synthetic methods were also developed by SC (methods A and B). Synthesis was also performed by SC, CA, SL, AL, LT, and GK. The research project was supervised by AN, and conceived by AN and SW. The manuscript was written by AN, and the supporting information was written by CA and SL.

The section of the work presented in Chapter 3 and Chapter 4 has appeared in the following publication:

“Activity-Directed Synthesis of Inhibitors of the p53/*hDM2* Protein–Protein Interaction”, **A. I. Green**, F. Hobor, C. P. Tinworth, S. Warriner, A. J. Wilson, and A. Nelson, *Chemistry – A European Journal*, 2020, **26**, 10682–10689.

From the work contained in this publication, the candidate performed the described chemical synthesis in its entirety, and the binding studies using the fluorescence polarisation assay. The candidate also performed the data analysis for the NMR binding studies, and the molecular similarity analyses. The candidate and FH acquired $^1\text{H}/^{15}\text{N}$ -HSQC spectra jointly, and FH and Pallavi Ramsahye expressed *hDM2*₁₇₋₁₂₅ protein. CPT performed a computational docking study and generated a scaffold-hopping analysis. CPT, SW, and AN conceived the research project, and AN provided supervision. AJW provided additional supervision for FH. The candidate and AN wrote an initial draft of the manuscript, and AN wrote the final version for submission.

Other contributions:

Dr Natalie Fey (Senior Lecturer, University of Bristol) provided additional supervision of the candidate for the work described in Chapter 2. Dr Natalie Fey also performed conformational searches for two rhodium(II) complexes described in this thesis (**3d** and **4g**), which were used for further DFT calculations by the candidate. Dr Christopher P. Tinworth (GlaxoSmithKline) calculated chromLogP and chromLogD values for the library of known MDM2 ligands that was used to benchmark the molecular properties of ADS products in Chapter 4. Dr Mark Howard (University of Leeds) assisted in the acquisition of the ¹⁹F-CPMG NMR spectra described in Chapter 4.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The right of Adam Ieuan Green to be identified as Author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

© September 2020 The University of Leeds and Adam Ieuan Green

Acknowledgements

I've been incredibly lucky to have had fantastic scientific mentors guide me through my PhD studies. I'm indebted to Prof. Adam Nelson, who I'd like to thank for pushing me as a scientist and for always supporting my work. I've grown a great deal professionally over the past four years, and that's because Adam consistently allowed me to explore ideas independently and with autonomy. He invested a lot of time in my development and I hope that I can repay the favour one day. I'd also like to thank Dr Stuart Warriner for teaching me a great deal about chemical biology, and for his supervision. I'm also very thankful to Dr Natalie Fey for all her work as a collaborator, and as a mentor – it takes a lot of effort to teach a complete beginner!

I was lucky to have been funded by the EPSRC and GSK, who I thank for their investment in my training. Alongside this, I'd like to thank my industrial supervisor Dr Chris Tinworth who stepped in and helped guide my research to its conclusion. Chris arranged a fantastic placement at GSK Stevenage, where I was exposed to great science and also a great team. It was a wonderful time for me personally and professionally, thank you!

I will miss working in the Nelson lab and would like to thank all the past and present group members who made my time so enjoyable. Working with you all was a rewarding experience and I hope you all achieve your goals. To Sam, Jacob and Chloe I'd like to express my heartfelt gratitude for your friendship. I couldn't have expected to make three of my closest friends over the last four years and I'll always remember the time we shared at Leeds fondly.

I also want to thank my family who have supported me throughout my life-long studies. To my parents who taught me to chase my ambitions and pursue my dreams, your love and support has made me feel as though I can do anything. Finally, I'd love to thank my partner Beth who has supported me unwaveringly and has always been by my side. You gave me the courage to move to Leeds and I couldn't have done any of it without you. Thank you.

Abstract

Methods for the discovery of bioactive small molecules are a constant source of innovation in drug discovery. Typically, medicinal chemists discover molecules in design-make-test cycles and invest equal resources into each molecule regardless of biological function. New and emerging workflows for the rapid discovery of bioactive small molecules aim to redistribute resources to high value, active small molecules. This thesis focuses on developing a discovery workflow that invests resources exclusively in high-value, active molecules, and circumvents limitations of traditional discovery workflows.

Chapter 1 gives an overview of modern drug discovery practices and focuses on emerging methods for the integrated and high-throughput discovery of bioactive small molecules. Chapter 1 also outlines the biological and medicinal chemistry of the p53/MDM2 protein–protein interaction, and proposes the PPI as a target for activity-directed small molecule discovery.

Chapter 2 describes the development of a computational map for the selection of catalysts in high-throughput reaction arrays. Principal component analysis of a library of 48 DFT-optimised rhodium(II) catalysts was used to build the map from a collection of bespoke computational descriptors. The map was compared to a variety of experimental data sources and found to be a useful tool for interpreting reaction outcomes.

Chapter 3 describes the design and implementation of two high-throughput reaction arrays for the activity-directed discovery of inhibitors of the p53/MDM2 protein–protein interaction. 346 microscale reactions were performed and seven products were isolated from the scale-up of hit reaction mixtures.

Chapter 4 describes the characterisation of hit molecules identified from the activity-directed synthesis workflow. The products were tested in orthogonal biological assays and four distinct series were found to have low micromolar inhibitory activity of the p53/MDM2 protein–protein interaction. Similarity analysis also demonstrated that the products have promising ligand efficiencies and that the products can provide new starting points for drug discovery.

Table of Contents

Acknowledgements.....	iv
Abstract.....	v
Table of Contents	vi
List of Abbreviations.....	ix
Chapter 1 Chemical Approaches for the Discovery of Small Molecules in Drug Discovery	1
1.1 An Overview of Early-Stage Drug Discovery.....	2
1.2 Chemical Transformations Used to Synthesise Small Molecules in Drug Discovery.....	5
1.3 Integrated Approaches for High-Throughput Chemistry and Molecular Discovery.....	7
1.3.1 DNA-Encoded Libraries.....	8
1.3.2 Integration of Nanoscale Chemical Synthesis and Screening for Small Molecule Discovery.....	12
1.3.3 Activity-Directed Synthesis.....	15
1.4 The p53/MDM2 Protein-Protein Interaction.....	20
1.4.1 Inhibition of the p53/MDM2 Protein-Protein Interaction.....	21
1.4.2 Dissection of the Nutlin RG7112	23
1.5 Project Outline.....	24
1.5.1 Development of a Catalyst Map for Rhodium(II) Catalyst Selection in Activity-Directed Synthesis	25
1.5.2 Activity-Directed Discovery of Inhibitors of the p53/MDM2 Protein-Protein Interaction	26
Chapter 2 Construction of a Catalyst Map for Rhodium(II) Catalysts	27
2.1 Computational Parameterisation for Catalysis	27
2.2 An Overview of Rhodium(II)-Carbene Chemistry	33
2.3 Design of a Knowledge Base for Rhodium(II) Catalysts.....	36
2.4 Principal Component Analysis for the Construction of a Catalyst Map	44
2.5 Correlation of the Catalyst Map with Experimental Outcomes	55
2.6 Conclusions.....	61
Chapter 3 Design and Implementation of Activity-Directed Synthesis Reaction Arrays 1 and 2	63
3.1 Design of an Initial Reaction Array	64

3.1.1 Synthesis of Diazo Substrates for an Initial Reaction Array	68
3.2 Establishment of a Fluorescence Anisotropy Assay for the p53/MDM2 Protein-Protein Interaction	72
3.3 Implementation and Analysis of Reaction Array 1	76
3.4 Design and Implementation of Reaction Array 2	81
3.5 Identification of Products from Hit Reaction Mixtures	85
3.6 Conclusions.....	87
Chapter 4 Validation and Characterisation of Hit Compounds from Activity-Directed Synthesis.....	88
4.1 Characterisation of Putative Inhibitors Using a Fluorescence Anisotropy Assay	88
4.2 Characterisation of the Binding of Putative Inhibitors to Human-MDM2 by NMR Spectroscopy	92
4.2.1 Titration Experiments for the Characterisation of Putative Inhibitors.....	92
4.2.2 Structural Insights into the Binding of Putative Inhibitors to Human-MDM2.....	97
4.3 Development of Structure-Activity Relationships for Selected Human-MDM2 Inhibitor Classes	102
4.4 Molecular Similarity and Comparisons with Known MDM2 Ligands.....	105
4.5 Conclusions.....	111
Chapter 5 Experimental	115
5.1 General Information and Instrumentation	115
5.2 Materials and Methods for the Catalyst Knowledge Base	117
5.2.1 Synthesis of Compounds for the High-Throughput Screening of Model Reactions	117
5.2.2 Protocols for the Screening of Model Reactions	121
5.2.3 Computational Details	129
5.2.4 Design of a Descriptor Database	132
5.2.5 Exemplar Submission Script for Gaussian09	137
5.2.6 Exemplar Script for the Automated Extraction of Descriptors.....	138
5.2.7 Data Extracted from DFT-Converged Rhodium(II) Complexes	139
5.2.8 Exemplar Python Code for Principal Component Analysis.....	144
5.3.9 Overlay of Alternative Model Reactions onto the PCA Map	145

5.3 Materials and Methods for Activity-Directed Synthesis.....	147
5.3.1 Synthesis of Diazo Compounds	147
5.3.2 Synthesis of Co-Substrates.....	157
5.3.3 Synthesis of MDM2 Ligands.....	157
5.3.4 Implementation of High-Throughput Chemistry for Activity-Directed Synthesis Reaction Arrays	168
5.3.4.1 LC/MS Analysis of Round One Reaction Mixtures	169
5.3.5 Fluorescence Anisotropy Assay for the Inhibition of the p53/MDM2 Protein-Protein Interaction	173
5.3.5.1 Binding of p53 ¹⁵⁻³¹ Flu to hDM2 ¹⁷⁻¹²⁵	173
5.3.5.2 Inhibition of the p53 ¹⁵⁻³¹ Flu/hDM2 ¹⁷⁻¹²⁵ protein- protein interaction with Nutlin-3a	175
5.3.6 Procedure for the Screening of Reaction Mixtures.....	176
5.3.8 Determining EC ₅₀ Values for Isolated Compounds	180
5.3.9 NMR Measurements for K _d Estimation.....	185
5.3.9.1 K _d Estimation Using ¹ H/ ¹⁵ N-HSQC NMR.....	185
5.3.9.1 ¹ H/ ¹⁵ N-HSQC Spectra and Fitting	186
5.3.9.2 ¹⁹ F-CPMG Ligand-Observed NMR.....	193
5.4 Computational Methods for the Design and Evaluation of Reaction Arrays.....	193
5.4.1 Exemplar Reaction SMARTS for the Enumeration of Virtual Products from Reaction Arrays	195
5.4.2 A Python Script for the Enumeration of Virtual Reaction Arrays using RDKit.....	196
5.4.3 An Exemplar Virtual Reaction Array	198
5.5 Similarity Analysis of Hit Molecules	198
5.6 Molecular Docking.....	199
List of References	201
Appendix A Establishing the p53/hDM2 Fluorescence Anisotropy Assay for Photoredox Reaction Conditions	215

List of Abbreviations

δ	chemical shift
ADS	activity-directed synthesis
ASMS	affinity-selection mass-spectrometry
ASO	average steric occupancy
BRAF	gene encoding B-Raf proto-oncogene, serine/threonine kinase
CH ₂ Cl ₂	Dichloromethane
CHK1	Checkpoint kinase 1
CN	nitrile
CPMG	Carr-Purcell-Meiboom-Gill pulse sequence
d	doublet
DBU	Diazabicyclo[5.4.0]undec-7-ene
dd	doublet of doublets
DEL	DNA-encoded library
DFT	density-functional theory
DMSO	dimethylsulfoxide
DNA	Deoxyribonucleic acid
DOS	diversity-oriented synthesis
DTS	DNA-templated synthesis
DZP	double-zeta polarisation
EC ₅₀	half maximal effective concentration
EDG	electron donating group
<i>e.g.</i>	<i>example gratia</i> ; for example
ERK2	Extracellular Signal-Regulated kinase 2
ES	electrospray ionisation
ESAC	encoded self-assembling libraries
<i>etc.</i>	<i>et cetera</i> ; and so forth
Et	Ethyl

EWG	electron withdrawing group
GSK	GlaxoSmithKline
FA	fluorescence anisotropy
FBDD	fragment-based drug discovery
FDA	United States Food and Drug Administration
HA	heavy atoms, or number of non-hydrogen atoms
<i>hDM2</i>	Human murine double minute 2
HPLC	high performance liquid chromatography
HRMS	high resolution mass spectrometry
HTE	high-throughput experimentation
HTS	high-throughput screening
<i>i.e.</i>	<i>id est</i> , that is
IC ₅₀	half-maximal inhibitory concentration
<i>in silico</i>	via computer simulation
<i>in vivo</i>	within the living
IR	infrared
ITC	isothermal titration calorimetry
<i>J</i>	spin-spin coupling
K _d	dissociation constant
LC-MS	liquid chromatography mass spectrometry
LE	ligand efficiency
LKB	ligand knowledge base
LLE	lipophilic ligand efficiency
logP	octanol-water partition coefficient
m	multiplet
Me	methyl
MDM2	Murine double minute 2
MK2	Mitogen-Activated Protein kinase 2
ML	machine learning

MM	molecular mechanics
MS	mass spectrometry
NanoSAR	nanoscale synthesis with affinity ranking
NMR	nuclear magnetic resonance
OAc	Acetate
<i>p</i> -ABSA	<i>para</i> -acetamidobenzenesulfonyl azide
PAINS	pan-assay interference motifs
PC	principal component
PCA	principal component analysis
PDB	protein data bank
PPI	protein-protein interaction
ppm	parts per million
PTFE	Polytetrafluoroethylene
q	quartet
R_f	retention factor
RIP1	Receptor Interacting Protein 1 kinase
rt	room temperature
UPLC	ultra-high-performance liquid chromatography
UPLC-MS	ultra-high-performance liquid chromatography - mass spectrometry
s	singlet
SAR	structure-activity relationships
SPR	surface plasmon resonance
t	triplet
THF	Tetrahydrofuran
TLC	thin layer chromatography
TR-FRET	Time-Resolved Fluorescence Energy Transfer

Chapter 1

Chemical Approaches for the Discovery of Small Molecules in Drug Discovery

The identification of bioactive small molecules underpins drug discovery and is essential for the development of new therapeutics for diseases with an unmet clinical need. The drug discovery process has a high attrition rate with over 90% of candidates failing in clinical trials.^{1,2} While small molecule discovery is not the rate limiting step for the development of new medicines, a highly productive drug discovery pipeline is required to counteract the high failure rates of clinical trials.³ This is highlighted by the number of FDA drug approvals between 2010 and 2019, where new molecular entities make up over 75% of approvals.⁴ Efforts to reduce clinical failure rates include extensive target validation; ensuring a robust link between target biology and disease mechanisms; and optimisation of the chemical structure of a candidate for parameters such as selectivity, toxicity, metabolic liabilities, pharmacokinetics and pharmacodynamics.⁵⁻⁷

Hit molecules suitable for progression through the discovery process can usually be identified a large majority of the time, but many campaigns target protein classes that are known to be compatible with small molecules.^{8,9} Failure to identify a hit for a challenging class of protein targets can limit the development of first in class treatments.^{10,11} The inability to identify a hit can be due to a lack of structural information that can be used to direct screening efforts;¹²⁻¹⁴ poor protein stability *in vitro*;¹⁵ poor assay performance;¹⁶ or low availability of the isolated protein meaning high throughput screening cannot be conducted.¹⁷ In some cases, the chemical space covered by a screening library is not relevant to the targeted activity and prevents the identification of synthetically tractable hits.¹⁸⁻²⁰ Deficiencies that impede the drug discovery process can lead to the discontinuation of projects, meaning alternative methods for molecular discovery could help bring first in class medicines to the clinic.²¹⁻²³

The aim of the research described in this thesis was to develop and demonstrate the effectiveness of activity-directed synthesis (ADS) as a method for the discovery and synthesis of small molecules against a challenging class of protein targets, protein-protein interactions. The key objective was to further develop ADS as an approach that can enable hit identification without recourse to elucidation of the structural basis for a

protein-ligand interaction. It was envisaged that the ADS will expand the toolkit of discovery techniques available for drug discovery and lead to the identification of new active small molecules that could not have been designed *a priori*.

1.1 An Overview of Early-Stage Drug Discovery

The drug discovery pipeline, for small molecules, is a linear process which begins with target identification and validation and is followed by hit discovery and hit-to-lead optimisation. The pipeline ends after successful progress through clinical trials (Figure 1.1).⁷ Biological targets are selected based on either an unmet clinical need to treat a disease, or to improve the standard of care for a disease that already has approved drugs available.^{24,25} Thorough identification and validation of the biological target ensures that, after successful hit-to-lead optimisation, a clinical candidate molecule can modulate the targeted biological mechanisms that underpin disease progression and act as a drug.^{7,26} In order to support the low success rate of clinical trials, where fewer than one in ten candidates are successfully marketed as drugs, the discovery pipeline must have an ample supply of new chemical matter.^{27,28} Therefore, technologies that enable fast and efficient hit discovery can improve downstream productivity and help bring drugs to market more quickly.²⁹⁻³¹

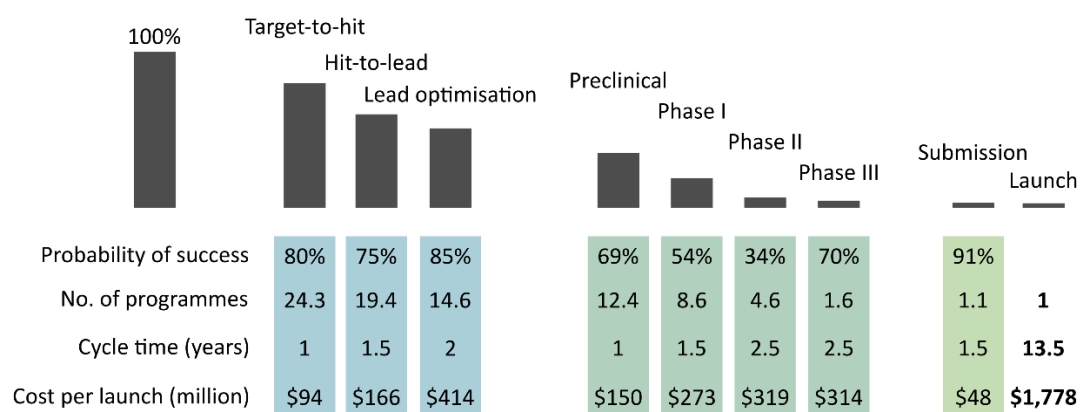


Figure 1.1. An outline of the number of active drug discovery programmes required, at each stage of the pipeline, to support the launch of one new small molecule drug.¹ Each stage has a breakdown of likelihood of success, cycle time and cost.

During the early 2000s high-throughput screening (HTS) became a staple technique for the discovery of hit compounds, and pharmaceutical companies amassed libraries containing typically $>10^6$ molecules for screening.^{29,32} HTS relies upon screening an entire compound library against a drug target in a highly automated array process. Compounds found in HTS screening libraries often have molecular properties that obey the Lipinski “rule of five”, or stricter criteria such as lead-like space,³³ and have a molecular weight below 500 Da, a logP lower than five, no more than five hydrogen bond donors, and no more than 10 hydrogen bond acceptors.³⁴ Identified hits are then characterised by structure-activity-relationship (SAR) studies and subjected to a rigorous optimisation programme to improve various properties including potency, ligand efficiency, lipophilicity and solubility.^{33,35–37} Lead compounds are then optimised further to improve pharmacokinetic and pharmacodynamic properties, and to elucidate structure-toxicity relationships to avoid failure in preclinical studies.³⁸ HTS techniques can have significant shortcomings, including an inability to produce hits against new or challenging targets as compound libraries often contain large and complex molecules,³⁹ and false positive results that are often caused by aggregators.^{40,41}

Fragment-based drug discovery (FBDD) is another approach for the discovery of small molecules that utilises carefully curated compound libraries to screen several thousand low-molecular weight, low-affinity molecules with tightly controlled molecular properties.^{39,42–44} A large number of early fragment libraries were based on the “rule of three”, where molecular weight was less than 300 Daltons (approximately 20 heavy atoms), and where each fragment contains fewer than three hydrogen bond donors and acceptors, less than three rotatable bonds and clogP less than three.^{45,46} Modern fragment libraries are curated with a great deal of care and employ much stricter criteria on molecular properties including; molecular weight (fragments usually fall within 140 to 230 Da or 10 to 15 heavy atoms); logP (-1 to 3); fragment complexity (3D-shape and fraction of sp^3 character); synthetic tractability; pan-assay interference (PAINS) motifs; and solubility.^{47–52}

Fragments are screened against a biological target using sensitive high-throughput biophysical or biochemical assays, allowing for the identification of weak binders (typically between 1 and 10 mM K_d).⁵³ Examples of screening methods include X-ray crystallography; Surface Plasmon Resonance; protein- and ligand-observed NMR; and fluorescence-based

assays using labelled proteins to detect binding.^{54–56} In practice, fragments can form highly productive interactions with protein binding sites, and when elaborated, the subsequent lead compounds tend to conserve the original fragment binding modes (Figure 1.2).⁵⁷ As fragments bind with low-affinity, the concept of ligand efficiency has been used to aid prioritising high quality hits and is defined below in equation 1.1.⁵⁸ Generally, hits with ligand efficiency of 0.3 kcal mol⁻¹ HA⁻¹ or greater are prioritised and considered alongside SAR, binding model and potential for synthetic elaboration.^{59,60}

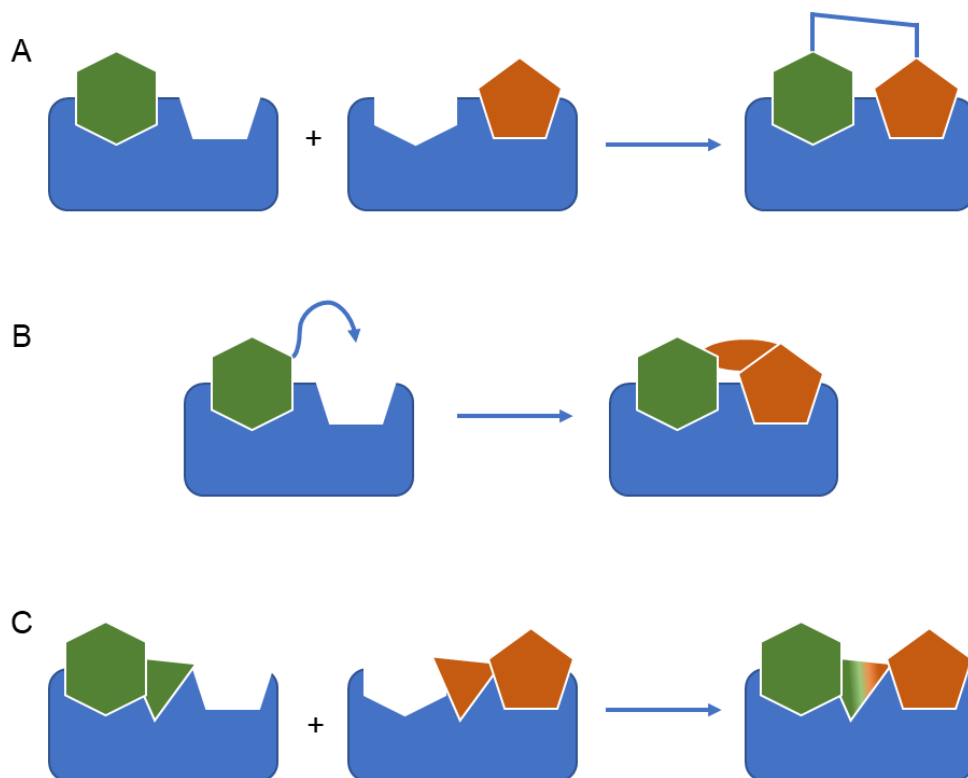
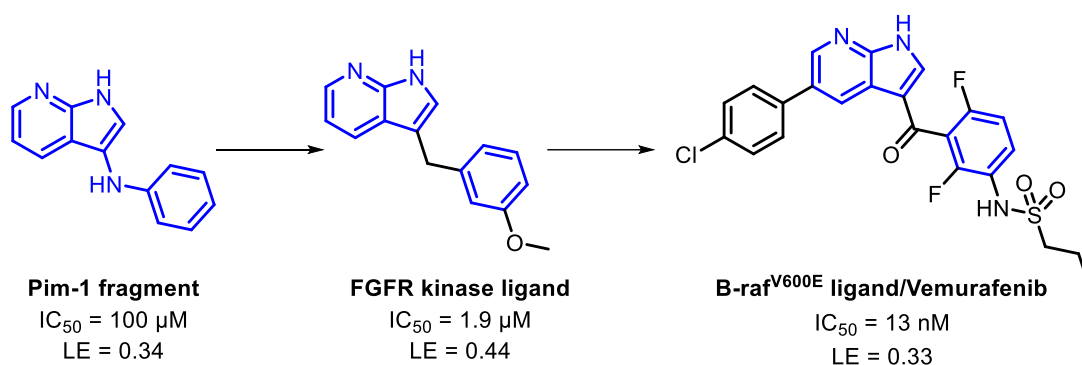


Figure 1.2. Fragment-based hit to lead strategies. A: Fragment linking – two fragments are linked together to conserve the combined activity of each individual fragment. B: Fragment growing – a single fragment is grown into adjacent binding sites. C: Fragment merging – two compounds that share structural similarities are combined into one molecule.⁵⁷

Equation 1.1:

$$\text{Ligand Efficiency (kcal mol}^{-1} \text{ HA}^{-1}) = \frac{\Delta G^{\ominus}}{\text{Heavy Atom Count}}$$
$$\approx \frac{1.37}{\text{HA}} \times \text{pIC}_{50}$$

The first successful example of FBDD was the discovery of Vemurafenib for the treatment of *BRAF*^{V600E} mutant metastatic melanoma.^{61–63} The discovery campaign was initiated by screening a 20,000-fragment library against a panel of structurally diverse serine/threonine kinases to identify new scaffolds for FBDD.⁶¹ Initially, 238 hit fragments were identified using X-ray crystallography and evaluated based on their synthetic tractability for further development as new fragment scaffolds. A 3-substituted 7-azaindole scaffold was selected and quickly optimised into a potent kinase inhibitor. Over 100 solved co-crystal structures of compounds bound to B-raf^{V600E} were produced as part of the optimisation (Scheme 1.1).



Scheme 1.1. The development of Vemurafenib. The portions of the primary fragment are highlighted in blue.

1.2 Chemical Transformations Used to Synthesise Small Molecules in Drug Discovery

Medicinal chemists performing organic synthesis must produce active small molecules quickly, and iteratively, to advance from an initial hit to a lead compound that has the potential to be developed into a clinical candidate.⁷ The chemical transformations commonly used by medicinal chemists tend to focus on reactions with well-established precedent and substrate scope, so that analogues can be generated without the need for reaction optimisation.^{64–69} An analysis of submissions to the *Journal of Medicinal Chemistry*, totalling 40,000 compounds, between 1959 and 2009 revealed that the molecular properties of drug-like molecules have significantly changed during this period.⁶⁵ Modern molecules are on average heavier, more complex and flatter.^{64,67} Modern drug-like molecules also have larger polar surface areas, are more lipophilic and feature more hydrogen bond donors than acceptors.⁶⁴ Overall, only three of the top twenty reactions featured in the *Journal of Medicinal Chemistry* in 2014 were developed after

1974, and two of the top six reactions were protecting group manipulations (Figure 1.3).

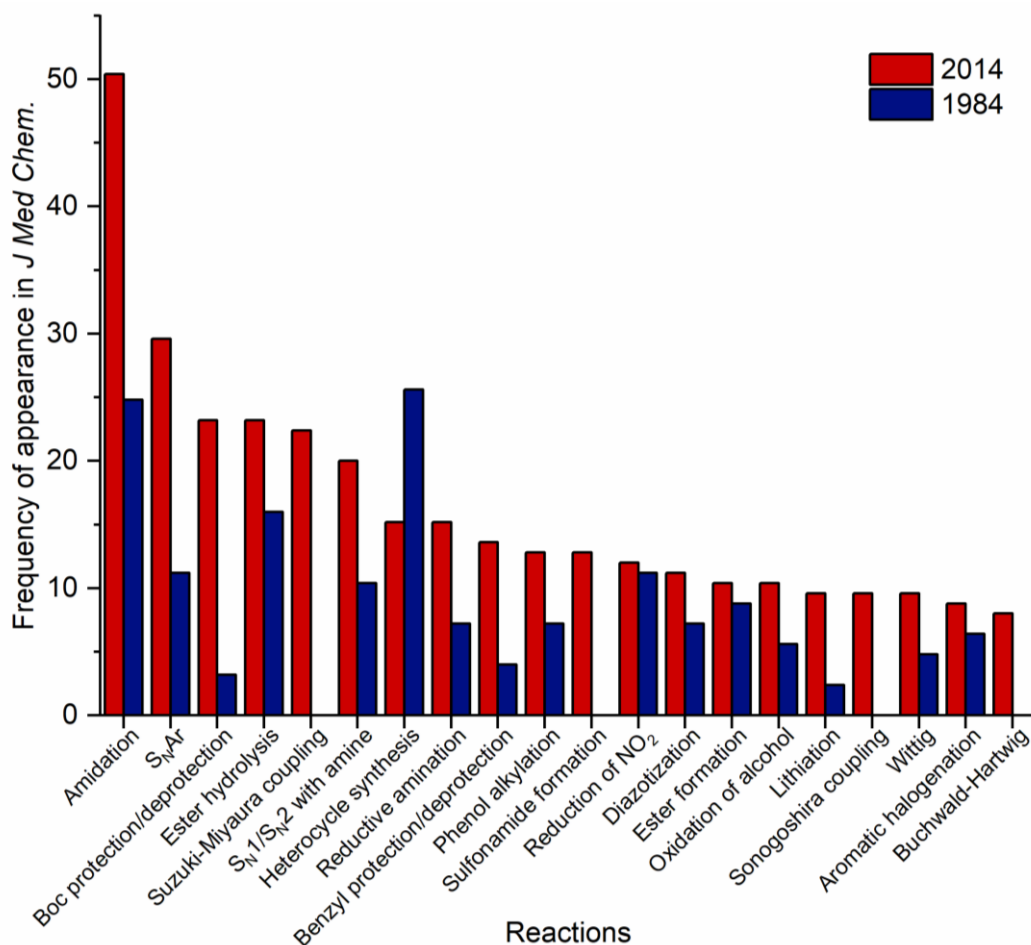


Figure 1.3. The frequency of appearance for each transformation in papers published in *J Med Chem* in 1984 and 2014 (n = 125). Suzuki-Miyaura coupling was first described in 1981; Sonogoshira coupling in 1975; and Buchwald-Hartwig in 1994.⁶⁵

Metal-catalysed cross-coupling reactions are frequently used in the early stages of drug discovery as a method of forming key carbon-carbon and carbon-heteroatom bonds in drug-like molecules.⁶⁵ Such cross-couplings dominate the reactions reported in pharmaceutical patents,⁶⁸ and newer carbon-carbon bond forming reactions have not been readily embraced. Medicinal chemists are often slow to incorporate new synthetic methodology into discovery workflows, instead relying on methods with a higher likelihood of success, which can limit the diversity of compounds synthesised and tested.⁷⁰ However, recent developments in photoredox chemistry and direct C–H bond functionalisation have been well received and implemented to support discovery efforts.^{70–72} Ultimately the value of a given molecule, or

class of molecules, drives medicinal chemists to try new methodologies. Molecules that fit robustly into a biological design hypothesis are more likely to have time invested into their synthesis due to their higher perceived value.⁷⁰ Therefore, the application of new synthetic methodologies by medicinal chemists is linked to the rational design of small molecules. Advancements in high-throughput chemistry and micro-scale synthesis have enabled the design of predictive models for reactions between pairs of densely functionalised reagents, by facilitating the generation of large databases consisting of experimental data.⁷³⁻⁷⁵ These models could have potential applications in drug discovery by allowing medicinal chemists to select optimal reaction conditions without prior testing and bridge the gap between emerging synthetic methodologies and medicinal chemistry.

1.3 Integrated Approaches for High-Throughput Chemistry and Molecular Discovery

Integrated approaches for molecular discovery have several advantages over FBDD and HTS technologies. Firstly, there is no need to purchase or maintain an expensive library or screening collection which can result in substantial cost savings for research and development. HTS libraries typically cost between \$400 million and \$2 billion USD to acquire, maintain and screen, whereas integrated technologies can be implemented for a tiny fraction of the cost (less than 0.1%).⁷⁶ Secondly, the chemical diversity of an HTS library cannot be readily adjusted to accommodate a specific protein meaning tractable chemical matter may not be found for challenging drug targets.²⁹ Integrated technologies are frequently used in drug discovery and have even produced clinical candidates for a range of diseases where HTS or FBDD approaches failed.⁷⁶⁻⁷⁸ Integrated approaches typically conduct chemical synthesis directly before biological screening using highly efficient and optimised processes to create libraries of compounds that can be screened on-demand.⁷⁶ These libraries can consist of pure compounds, crude reaction mixtures, or cocktails of compounds that can be individually identified by a tag containing information for each molecule. Reactions are usually conducted on the micro-scale and require only milligrams, or even micrograms, of each building block meaning expensive reagents can be routinely used.⁷⁹⁻⁸³

1.3.1 DNA-Encoded Libraries

DNA-encoded library (DEL) technology enables the synthesis and screening of millions (or billions) of small molecules simultaneously by encoding the synthetic history of each molecule with a strand of covalently linked DNA.^{76,84,85} The process begins by mixing a pool of building blocks that are individually identified by a unique strand of DNA. The building blocks are then split into several pools that are each reacted with a new building block and a new section of DNA is ligated to the original DNA tag conferring the synthetic history of the new compound.^{76,86–88} The pools of compounds are then recombined and then subjected to successive rounds of the split and mix strategy to generate a DEL.^{89,90} Once a DEL has reached a desired size, or achieved a large enough level of chemical diversity, it will be screened for active molecules using a single affinity-selection assay with purified protein immobilised on a solid support. Often, DELs are stored in an aqueous buffer, in a single Eppendorf tube, and can be conveniently screened on-demand after synthesis.⁷⁶ The DNA tags of compounds that bind to the target protein are then amplified and sequenced to obtain the identity of hits (Figure 1.4).⁹¹ DEL's can be constructed using various technologies including DNA-recording,^{89,90} DNA-templated synthesis (DTS),⁹² yoctoReactor,^{93,94} DNA routing,⁹⁵ and encoded self-assembling libraries (ESAC).⁹⁶

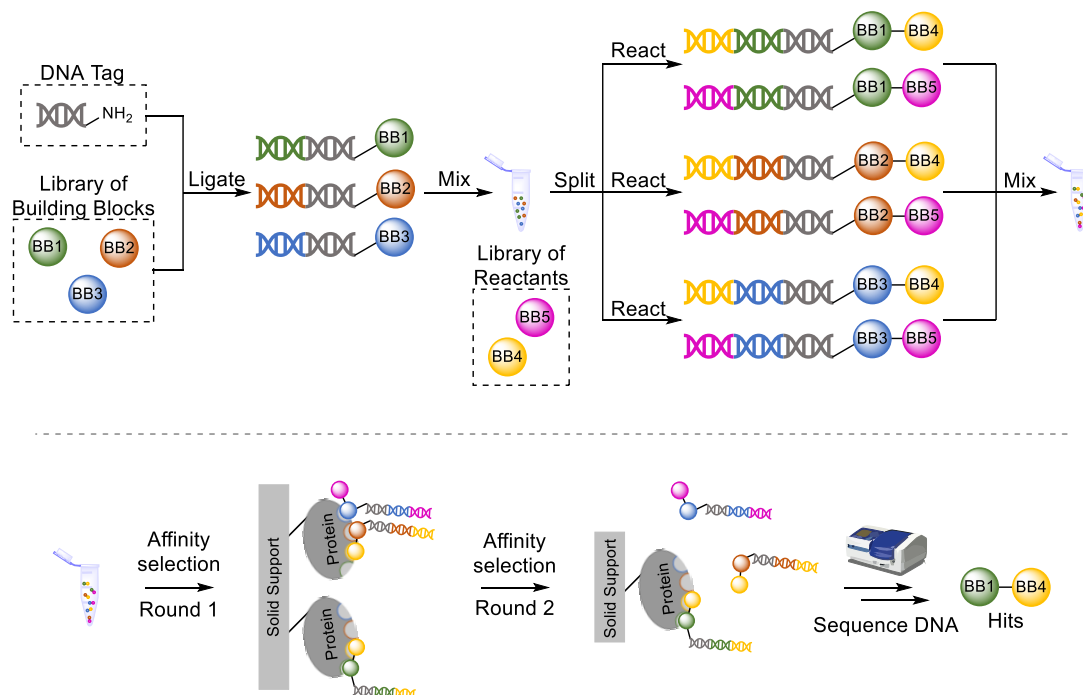


Figure 1.4. The split and mix strategy used in the synthesis of DNA encoded libraries. The DEL workflow illustrated above is an example of a DNA recording strategy.

DEL technology has delivered clinical candidates across different classes of protein targets and for a range of diseases with unmet clinical need.⁷⁶ An example of successful hit identification using DEL technology through to a proposed clinical candidate is the discovery of GSK2982772, a small molecule targeting Receptor Interacting Protein 1 (RIP1) kinase.^{78,97–101} Initially, the GlaxoSmithKline (GSK) kinase inhibitor library was screened to identify an inhibitor of RIP1 kinase, but no selective inhibitors were found, and the hits had poor synthetic tractability.⁹⁸ The GSK HTS compound collection, comprised of approximately 2 million compounds, was then screened and a lead series was identified but the best inhibitor had poor oral bioavailability, and poor synthetic tractability, and was discontinued.⁹⁹ The GSK collection of DNA-encoded small molecule libraries, containing 7.7 billion compounds, was then screened and an inhibitor of RIP1 kinase with a unique chemotype was identified (Figure 1.5).¹⁰¹ The discovery programme was then successfully advanced to give GSK2982772, a clinical candidate for the treatment of immune-mediated inflammatory diseases.^{78,97,100}

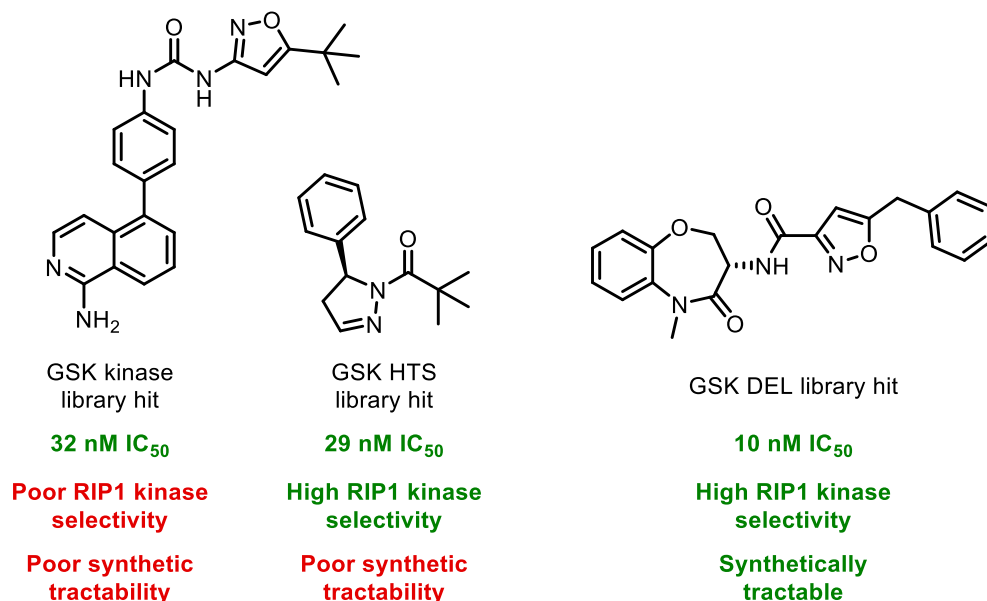


Figure 1.5. Hits from GSK compound libraries screened against RIP1 kinase.

DEL technology often cannot fully exploit the full range of modern synthetic chemistry due to limitations in reaction conditions.^{76,102–104} This is commonly due to incompatibility of new transformations with the aqueous, buffered, conditions required for maintaining the integrity of DNA.^{105,106} Advances in DEL technology have demonstrated that libraries can be synthesised in organic solvents and under inert atmospheres^{107,108} but have yet to be

translated to molecular discovery. Examples of reactions commonly used in DEL synthesis include amidation, metal-catalysed cross-couplings, Diels-Alder reactions, click chemistry, and nucleophilic substitution.¹⁰⁵ Designing DELs that cover a broad area of chemical space can also prove challenging due to the limited toolkit of reactions available for consideration, meaning libraries tend to contain compounds with many aromatic rings and rotatable bonds.^{76,102–104} DELs that incorporate the principles of diversity-oriented synthesis (DOS) in building block design have been proposed to address the lack of structural and stereochemical diversity.^{102,109} One example of a DOS-DEL, focussed on incorporating stereoisomeric small molecules, used a palladium-catalysed C–H arylation of azetidines and pyrrolidines to create a library of DOS-inspired initial building blocks.¹¹⁰ The building blocks were then attached to DNA and subjected to two rounds of split and mix synthesis (Figure 1.6). One set of building blocks were installed using a DNA-compatible Suzuki reaction and the other set using a range of nitrogen-capping reactions. Cheminformatic analysis of the DOS-DEL demonstrated that most compounds had drug-like physicochemical properties, and similarity analysis showed that the library contained extensive chemical diversity. The DOS-DEL was also screened against carbonic anhydrase and several potent inhibitors were identified, showing that specific DNA-compatible complexity generating reactions are not required for increasing the diversity of DELs.

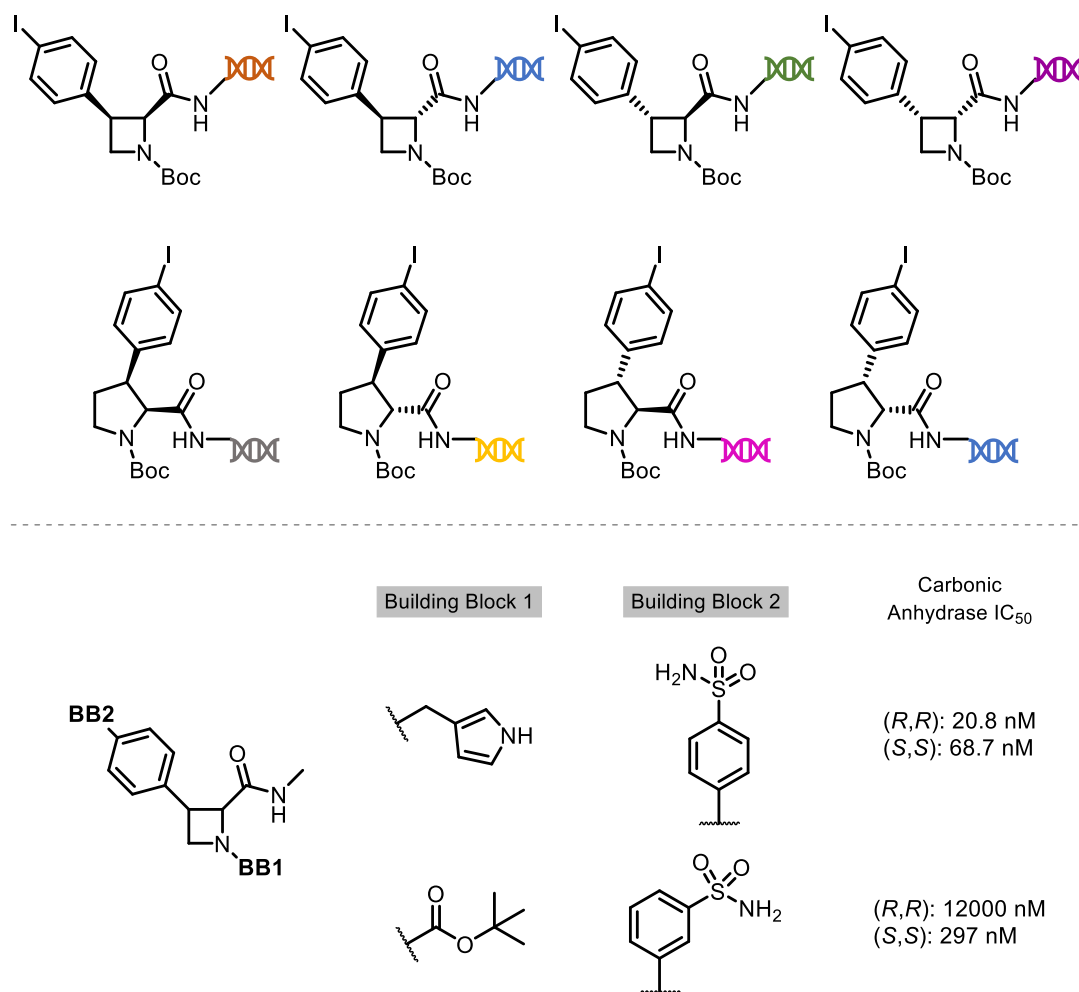


Figure 1.6. Library design for a DOS inspired DEL. Top panel: the DOS inspired scaffolds used to synthesise the DEL. Bottom panel: active products identified and synthesised off-DNA for characterisation.

DEL technology is largely limited to affinity-selection assays for hit identification, rather than activity-based assays, meaning that inhibitors of targets such as transcription factors, ion channels, receptors and protein-protein interactions cannot be directly identified.^{76,111,112} Instead individual hit molecules must be retested after affinity selection, and re-synthesised off-DNA, using an activity-based assay to confirm the desired biological mechanism of action. Advancements in microfluidics have demonstrated that it is possible to directly screen a DEL using an activity-based assay, but targets such as protein-protein interactions are still yet to be directly interrogated.^{111–113}

In summary, DEL technology offers several advantages over HTS by providing a flexible method for screening huge libraries of compounds without the overhead of synthesising and maintaining the corresponding

screening collection. The ability to tailor the properties of a DEL towards a specific target also allows for the progression of discovery campaigns where more traditional methods have not produced desired results. DELs also allow for hit identification against targets where structural information has not been elucidated, meaning that discovery for targets that are challenging for FBDD can be progressed.

1.3.2 Integration of Nanoscale Chemical Synthesis and Screening for Small Molecule Discovery

The miniaturisation of chemistry allows hundreds or thousands of reactions to be conducted in parallel on the microscale.^{75,79,80,82,83,114–116} Each reaction uses only a small amount of material, usually less than a milligram per reaction (approximately 1 μmol), meaning that large libraries of building blocks can be investigated for both reactivity and bioactivity.¹¹⁴ This has led to the optimisation of challenging reactions in drug discovery and development^{73,115} more quickly than when using traditional reaction optimisation strategies. The coupling of microscale chemical synthesis and biological screening has allowed for a substantial increase in the chemical space tested against a target.^{75,80,81} The leveraging of accessible reactive space for one or more chemical transformations allows for the discovery of unexpected reaction products with interesting biological properties.⁸³ Integrated approaches for microscale chemical synthesis and biological testing typically synthesise compounds in batch, plate-based formats, and screen crude reaction mixtures, but some workflows include in-line purification, or use flow chemistry.^{79,81–83,117–123}

Reaction arrays have been conducted on the nanoscale, using micrograms of material per reaction (typically between 50 and 5 μg) and total volumes between 1 and 2 μL .^{75,80,81} Nanoscale synthesis with affinity ranking (NanoSAR) is an example of this approach where both reaction space and biological activity space were investigated to yield potent inhibitors of kinases.⁸¹ Active products were identified using affinity-selection mass-spectrometry (ASMS) assays to identify hit compounds from crude reaction mixtures. NanoSAR was validated by synthesising screening libraries to identify new inhibitors of Extracellular Signal-Regulated kinase 2 (ERK2), Mitogen-Activated Protein Kinase 2 (MK2), and Checkpoint Kinase 1 (CHK1). The reactions used to synthesise the library were based on the most popular transformations used in small molecule discovery and included amidation, Suzuki-Miyaura cross-couplings, and Buchwald-Hartwig cross-

couplings.^{64,65,67,68} Each reaction was analysed by ultra-high-performance liquid chromatography – mass spectrometry (UPLC-MS) to identify reactions giving the desired product and to estimate the yield.⁷⁵ Initially, productive conditions were identified for each class of reactions before moving onto library synthesis. In total over 1,500 reactions were conducted to scout ideal reaction conditions (Figure 1.7).

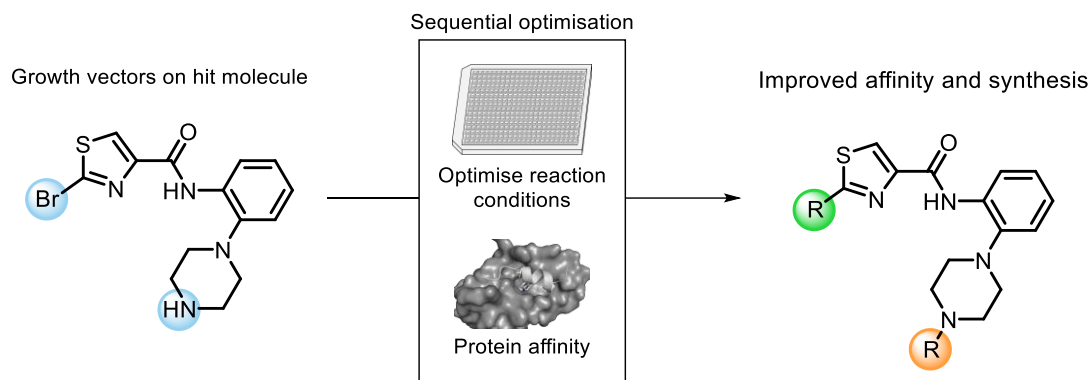


Figure 1.7. NanoSAR workflow for the synthesis and optimisation of small molecules. Reactions are performed in a 384-well plate, at 1.2 μ L total volume, in a glovebox and dispensed using microfluidic dispensing robotics. Active products are identified using affinity-selection mass-spectrometry for a protein target.

Each protein target was assigned a different scaffold building block that was elaborated with one or more of the coupling reactions. For each type of coupling reaction used to decorate the scaffolds, the reaction conditions and capping groups were widely varied. In total, 435 building blocks were used across all three targets and UPLC-MS analysis showed that 396 of the combinations gave a product (over 1,700 reactions conducted). The libraries were then screened for affinity towards ERK2, MK2 or CHK1 depending on the starting scaffold and a range of hits were identified (Figure 1.8). Each well was screened at multiple concentrations to estimate the potency of each hit by decreasing the concentration of protein in the ASMS assay. The products that function as ligands were identified from the protein-bound fraction and selectivity was tested by the presence of a fixed concentration of a competitor protein. Each active product was also resynthesized on a 20 mg scale and re-tested in the ASMS assay, and tested in a functional biochemical assay, to confirm the nanoscale results correlated with the activity of the pure compounds.

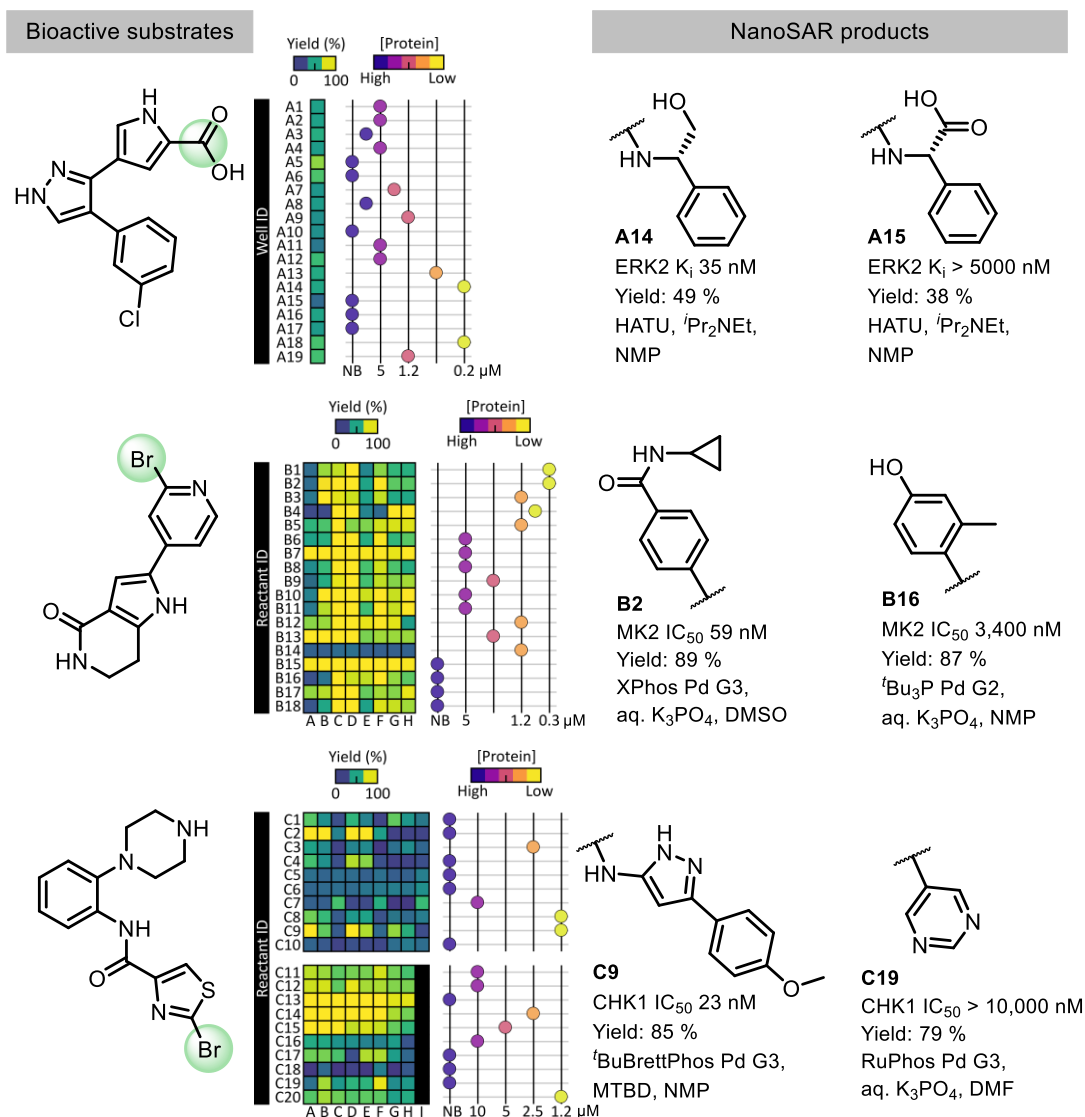


Figure 1.8. Example results from three NanoSAR screening libraries for ERK2, MK2 and CHK1 using amidation, Suzuki-Miyaura cross-coupling, and Buchwald-Hartwig C–N cross-coupling reactions respectively. Other reactions employed included the Sonogashira coupling, and S–H and O–H coupling. The most potent compounds are those with detectable binding via ASMS at the lowest protein concentrations.

Interestingly, if the libraries were synthesised using a single set of reaction conditions only a fraction of the available chemical space would have been sampled. For example, the C–N cross-couplings required extensive optimisation of reaction conditions before the nanoscale screening library could be synthesised. This is because the optimised reaction conditions varied as a function of substrate combinations. Twenty-four different reaction conditions had to be tested with five different amine building blocks to identify four reaction conditions that were robust enough to be used with the

96 amines selected for the nanoscale library. Overall, 92 of the 96 possible C–N cross-coupling products were successfully synthesised as part of the library. Therefore, NanoSAR may enable a diversification of the toolkit of reactions used to synthesise small molecules as methodologies with narrow scopes, or that are challenging to optimise, can be explored as part of a drug discovery workflow.

1.3.3 Activity-Directed Synthesis

Biosynthetic pathways evolve in nature when a natural product confers a competitive advantage to the host organism – these pathways are structure-blind and driven purely by the function of bioactive products. Activity-directed synthesis (ADS) takes some inspiration directly from the evolution of biosynthetic pathways and applies it to the discovery of active small molecules. Active products emerge in tandem with their synthetic routes through iterative rounds of microscale reaction arrays and integrated biological screening of crude reaction mixtures (Figure 1.9). Selection pressure can be increased by raising the threshold required for a crude reaction mixture to be identified as active, and efforts can be focussed solely on the characterisation of active products. Due to the low purification and characterisation overhead, a less established toolkit of chemistry can be employed, with minimal risk, to aid the discovery of new bioactive molecules.

The utility of ADS has been shown for the discovery of bioactive small molecules.^{82,83,123} Two successive publications described the use of α -diazo compounds with metal catalysts for the discovery previously unknown classes of ligands for the androgen receptor.^{82,83} Metal-catalysed carbene chemistry was chosen to intentionally leverage the possibility that multiple products could be formed in each reaction mixture. Reaction arrays were carried out in 100 μ L 96-well PTFE plates, under ambient conditions, and the product mixtures were scavenged for metals and assayed using a time-resolved fluorescence resonance energy transfer (TR-FRET) assay for agonism against the androgen receptor. The activity data was then analysed to identify active product mixtures, which subsequently informed the design of a second-generation array. Selection pressure was increased after each round, by screening at successively lower concentrations, optimising for both activity and yield. After the final round was complete, active product mixtures were scaled-up, purified and characterised to determine the identity of the active compounds. ADS has also been shown effective in the discovery of

anti-microbials using phenotypic assays and a different toolkit of chemistry.¹²³

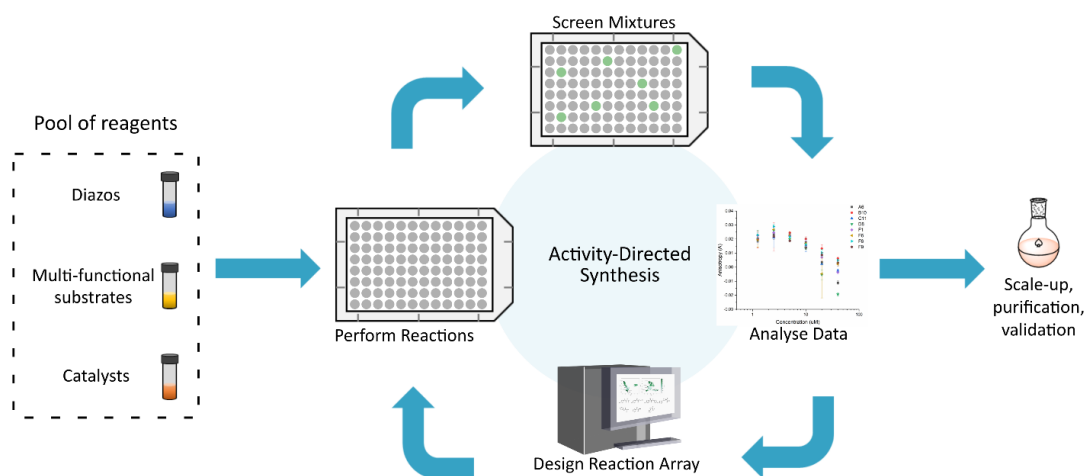


Figure 1.9. Activity-directed synthesis workflow for the discovery of agonists of the androgen receptor.

Initially, three rounds of ADS were conducted with 16 α -diazo amides (**1.1** – **1.16**) bearing a 4-cyano-3-trifluoromethylphenyl moiety and a range of metal catalysts to harness intramolecular transformations (Figure 1.10).⁸² In round one 36 reactions were conducted using 12 α -diazo amides **1.1** – **1.12**, three catalysts and one solvent, and the products were assayed at a 10 μ M total product concentration. Four diazo substrates **1.1**, **1.3**, **1.6** and **1.7** were identified as yielding active products which informed the design of the next reaction array. In round two, 192 reactions were conducted using six substrates **1.1**, **1.3**, **1.6**, **1.7**, **1.11** and **1.12** in combination with eight catalysts and four solvents, and the products were assayed at 1 μ M total product concentration. Diazo substrates **1.11** and **1.12** did not produce any reaction mixtures with detectable activity and were discounted. Two substrates **1.1** and **1.3** were identified as responsible for the most active products and four analogues **1.13** – **1.16** were synthesised to exploit this. In round three 108 reactions were conducted using six substrates **1.1**, **1.3**, **1.13** – **1.16**, six catalysts and three solvents, and the products were assayed at 100 nM total product concentration. Eight reactions were identified as producing the most promising bioactive products and scaled-up to obtain the pure products for analysis. Three active compounds **1.17** – **1.19** were identified with sub-micromolar activity demonstrating that ADS can be used to discover potent and novel scaffolds. ADS also enables the optimisation of the most active product between rounds of reaction arrays meaning activity and yield can be optimised simultaneously.

chemistry was utilised, but this time for activity-directed fragment growth. A range of co-substrates with diverse structure and reactivity was selected to enable the growth of an *N*-[4-cyano-3-(trifluoromethyl)phenyl]-*N*-methylacetamide fragment (IC_{50} : 92 μ M). In round one 192 reactions, out of a possible 480, were conducted featuring a random combination of four substrates, nine co-substrates (plus one without co-substrate), six catalysts and two solvents. The products were assayed at 10 μ M total product concentration and two active product mixtures were identified, both of which originated from diazo substrate **1.20** with either cyclohexene or indole as co-substrate, $Rh_2(S-DOSP)_4$ and DCM. In round two 86 reactions, out of a possible 360, were conducted using two diazo substrates, 18 co-substrates based on cyclohexene and indole, five catalysts and two solvents. The products were assayed at 5 μ M total product concentration and five active product mixtures were identified originating from diazo substrate **1.20** with either dihydronaphthalene, dihydropyran or indene. In round three all 48 possible reactions were conducted with diazo substrate **1.20**, 12 co-substrates and four catalysts, and the products were assayed at 1 μ M total product concentration. Co-substrate 6-benzopyran-carbonitrile yielded active product mixtures with all four catalysts, however, the dihydropyran co-substrate only gave an active mixture with $Rh_2(R-DOSP)_4$. Consequently, the most active product mixtures and a selection of inactive mixtures were scaled up and characterised to study the emergence of the bioactive products.

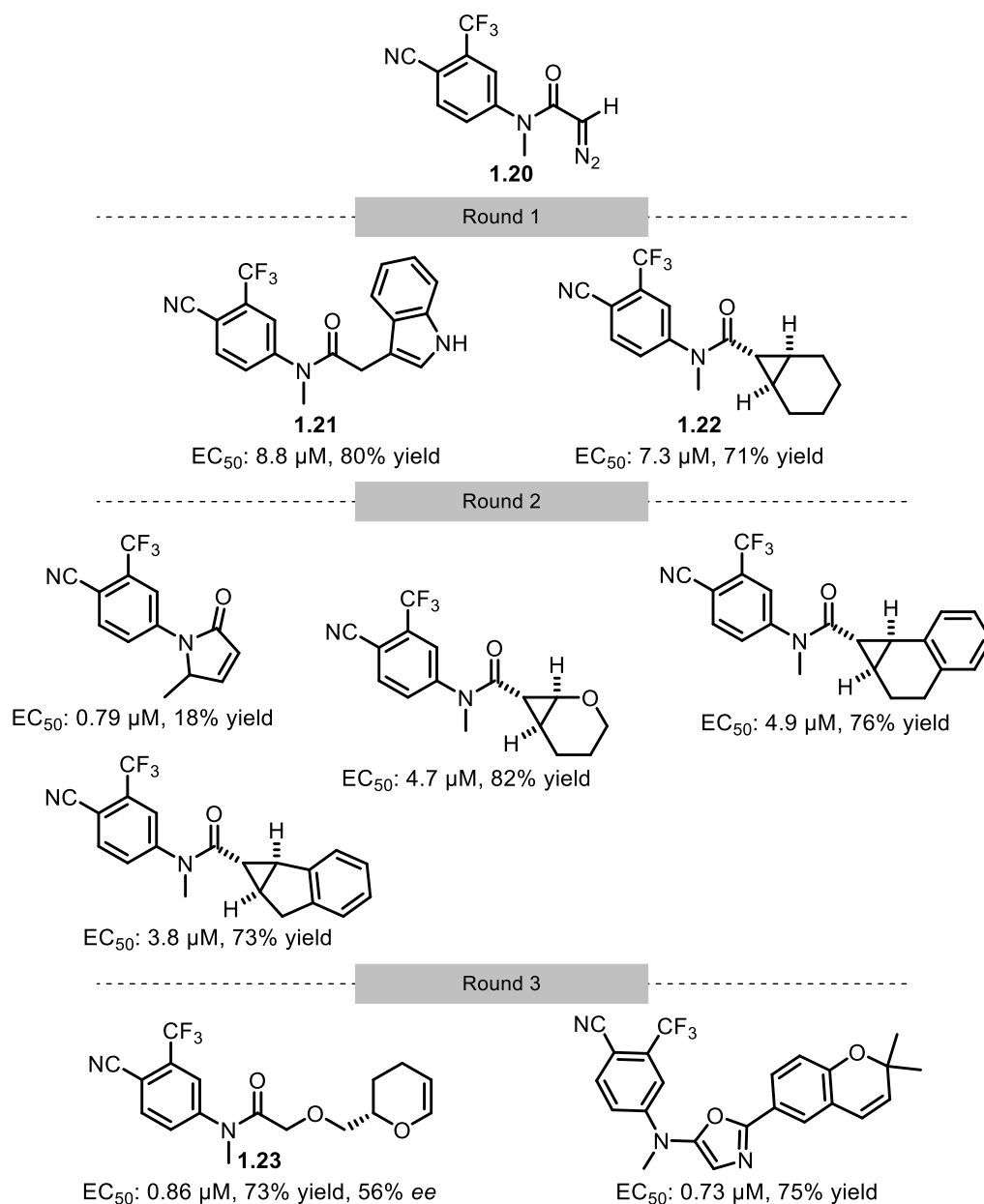


Figure 1.11. ADS using intermolecular reactions for the discovery of agonists of the androgen receptor.

Compounds **1.21** and **1.22** were the only active compounds identified during round one and both the product of intermolecular transformations, a C–H insertion into the C3 position of indole and a cyclopropanation of cyclohexene respectively. Characterisation of a selection of inactive wells from round one revealed that a diverse range of products had been formed and a large area of chemical space explored but ultimately discarded in favour of the active products. Round two capitalised on the introduction of a varied set of alternative substrates to expand the range of cyclopropanation products and identified alternative compounds. Finally, co-substrates

containing new functionality (alcohol and nitrile groups) were introduced in round three and yielded active products from unexpected reactions. The reaction forming **1.23** was unlikely to have been rationally predicted as enantioselective O–H insertions using racemic starting material were not previously known. ADS has, therefore, been shown to be capable of identifying novel transformations through the pursuit of bioactivity, mimicking the action of evolutionary biosynthesis.

ADS is a complementary method to other medicinal chemistry strategies for molecular discovery and could be integrated into existing drug discovery workflows. ADS has been demonstrated as an effective strategy for scaffold discovery,⁸² fragment growth,⁸³ and expansion of SAR.¹²³ Scaffold hopping is a key challenge in medicinal chemistry,¹²⁴ and ADS has not yet been used in this context. ADS has also been used for molecular discovery against a limited range of targets with evolved small molecules binding sites (androgen receptor^{82,83} and the penicillin binding protein of *Staphylococcus aureus*¹²³). Employing ADS against non-traditional drug targets would widen the number of ways ADS can be applied in medicinal chemistry settings.

1.4 The p53/MDM2 Protein-Protein Interaction

Murine double minute 2 (MDM2) is an oncoprotein that acts as a negative regulator of the p53 tumour suppressor, downregulating p53 and limiting the transcription of genes associated with tumour suppression.^{125–127} Under normal cellular conditions p53 is controlled through an autoregulatory feedback loop in which p53 activates MDM2 expression, and then MDM2 suppresses p53 action.^{128–133} MDM2 acts through binding of p53 at its transactivation domain, activation of cellular signalling pathways that lead to p53 nuclear export, and activation of a ubiquitin ligase that promotes p53 degradation.^{134,135}

The first crystal structure of the p53/MDM2 complex characterised the 109-residue *N*-terminal domain of MDM2 bound to the 15-residue transactivation domain of p53.¹³⁶ It revealed that MDM2 has a deep apolar cleft, to which p53 binds as an amphipathic α -helix. The interface of this interaction relies on steric compatibility between the MDM2 peptide binding cleft and the apolar face of the p53 α -helix, referring to F19, W23 and L26 residues of p53 (Figure 1.12). Alanine scanning studies, which systematically replaced each residue in the p53 transactivation domain with an alanine residue,

demonstrated that p53 analogues containing F19A or W23A substitutions lost all affinity towards MDM2.^{137,138} An L26A substitution also significantly reduced binding affinity towards MDM2 demonstrating that F19, W23 and L26 are key hot-spot residues for the p53/MDM2 protein-protein interaction.¹³⁷

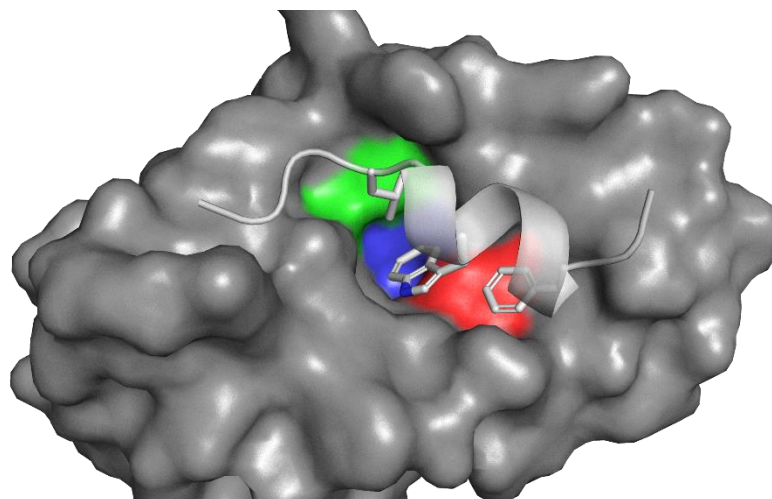


Figure 1.12. Co-crystal structure of the p53 transactivation domain bound to Human-MDM2 (PDB: 1YCR).¹³⁶ The sub-pockets targeted by p53 hotspot residues F19 (red), W23 (blue) and L26 (green) are shown.

1.4.1 Inhibition of the p53/MDM2 Protein-Protein Interaction

MDM2 is overexpressed in some cancers, therefore, targeting of the p53/MDM2 PPI and restoring the tumour suppressor functions of p53 has potential to be a mechanism for the development of new cancer therapeutics.¹²⁷ Protein-protein interactions typically occur over a large surface area and binding interactions are usually driven by the interaction of hotspot residues of one protein with another.¹³⁹⁻¹⁴¹ Inhibitor design for PPIs has traditionally been guided by structural and experimental analysis of an interaction and designs tend to mimic the hotspot residues that bind to a target protein.¹⁴²⁻¹⁴⁷ Inhibitors of the p53/MDM2 PPI usually target all three of the MDM2 sub-pockets for the p53 hotspot residues F19, W23 and L26. This allows a small molecule to mimic the p53 transactivation domain and act as an antagonist of MDM2 (Figure 1.13). Several small molecule antagonists of MDM2 have been progressed to clinical trials for solid tumours, lymphoma, neuroblastoma and other cancers (Figure 1.14).¹⁴⁸

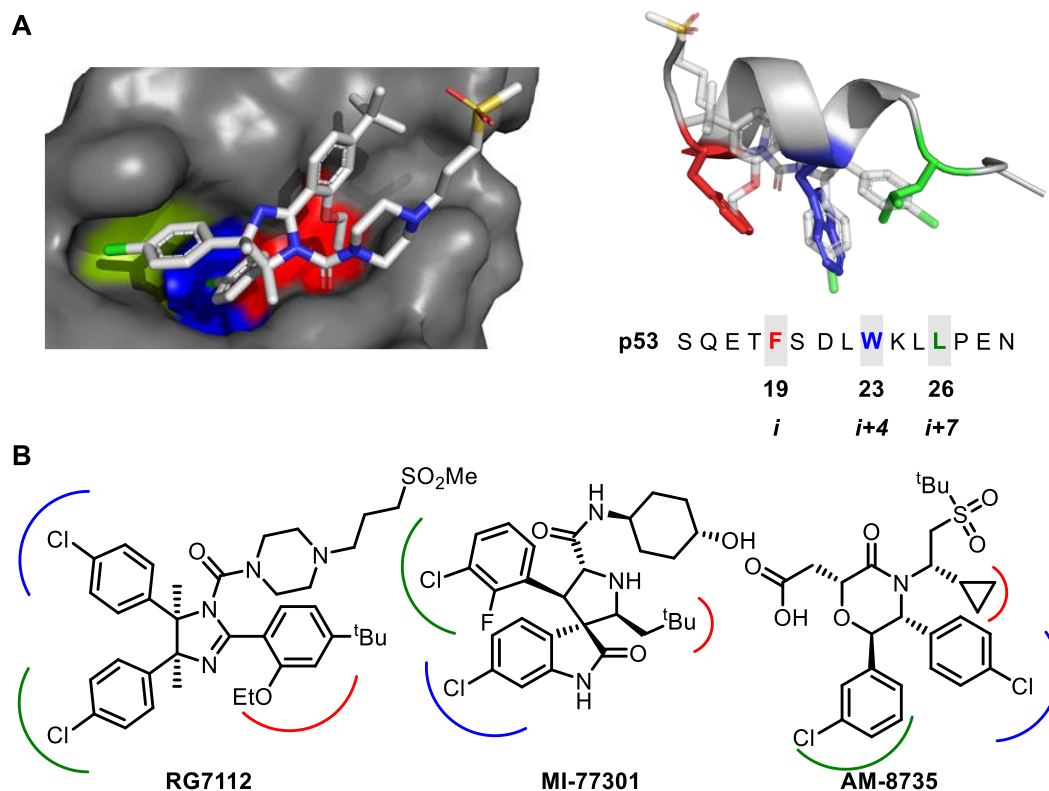


Figure 1.13. p53/Human-MDM2 PPI inhibitors, the sub-pockets targeted by p53 hotspot residues F19 (red), W23 (blue) and L26 (green) are shown. Panel A: Structure of MDM2 in complex with RG7112 (PDB: 4IPF) and overlay of RG7112 with p53 transactivation domain (PDB: 1YCR). Panel B: Examples of clinical candidates that target the p53/MDM2 PPI.^{149–152}

Clinical candidates that target the p53/MDM2 PPI employ many different approaches for orienting functional groups that can act as mimics of p53 hotspot residues.¹⁴⁸ Some scaffolds, for example RG7388¹⁵³ and AMG232,¹⁵⁴ use a small saturated ring system to project aromatic moieties into the W23 and L26 sub-pockets. Alternatively, CGM097¹⁵⁵ uses a tetrahydroisoquinoline core to interact with the L26 subpocket, and place an aromatic ring in the W23 subpocket, and then project an extended series of rings along the peptide binding cleft, similarly to the p53 peptide. RG7112¹⁴⁹ also uses this strategy with a methanesulfonylpropyl group to mimic the p53 peptide (Figure 1.13, panel A). Other scaffolds, such as MI-77301^{150,151} and ATX-265,¹⁵⁶ use a fused ring system to place an oxindole or isoindole ring in the W23 subpocket and orient substituents to mimic the alpha helix of p53. The diversity in approaches for building scaffolds mirrors the techniques that have been used to initiate drug discovery campaigns against the p53/MDM2

PPI and include virtual screening, structure-based drug discovery and HTS.¹⁴⁸

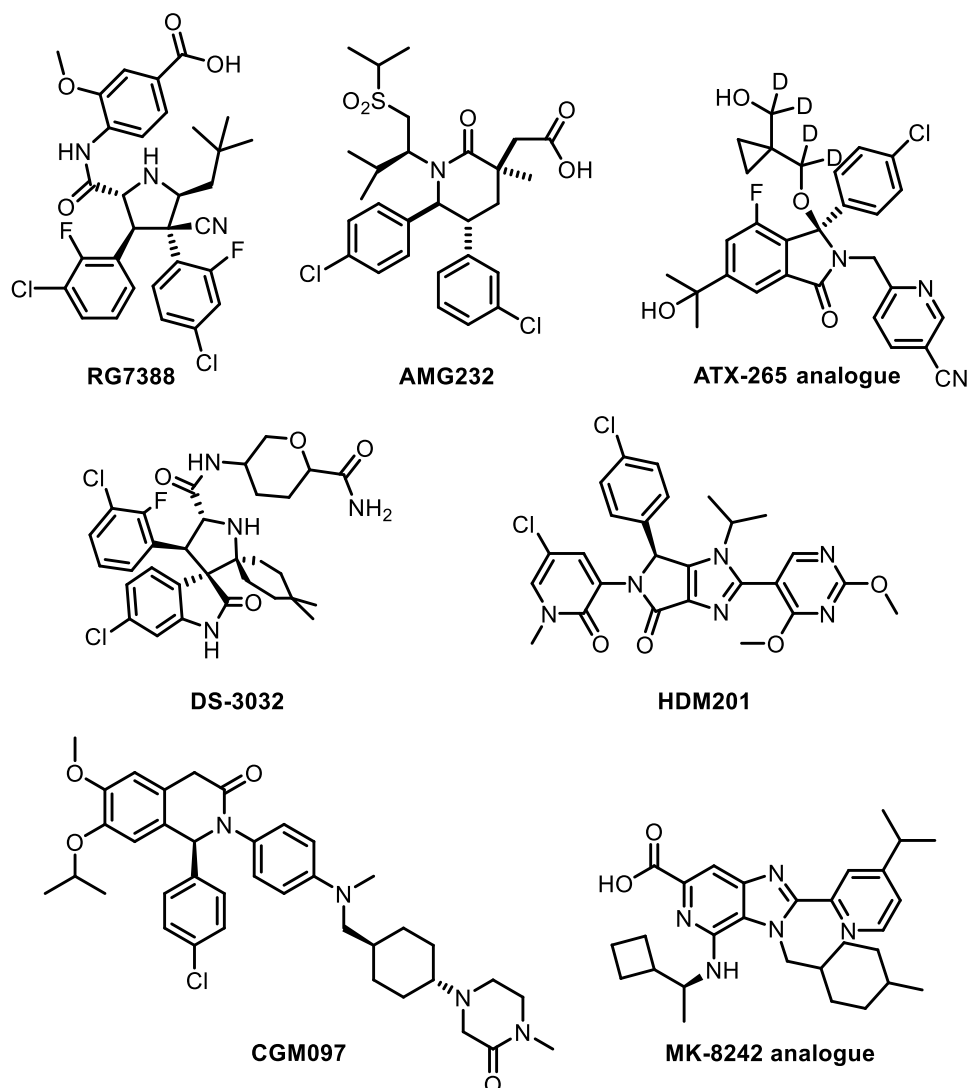


Figure 1.14. A selection of clinical molecules with known structure that have recruited, conducted, or reported results from clinical trials. RG7112 was discontinued after Phase 1b clinical trials in favour of RG7388, and the development of MI-77301 was also discontinued. Data obtained using Pharmaprojects, accessed 08/06/2020. Each highlighted molecule targets all three of the MDM2 sub-pockets.

1.4.2 Dissection of the Nutlin RG7112

The nutlin class of MDM2 antagonists are based on a *cis*-imidazolidine core scaffold that allows the projection of the para-chlorophenyl moieties into the MDM2 sub-pockets targeted by p53 hotspot residues W23 and L26.^{149,157} The anisole ring engages with the sub-pocket usually targeted by the p53 residue F19 and is important to overall affinity. The sulfone cap installed

onto the piperidine ring of RG7112 extends outwards into the solvent face, engaging MDM2 residue G58 and the peptide binding cleft.

The clinical candidate RG7112 was retrospectively deconstructed to assess whether a drug discovery campaign, in principle, could have been initiated by FBDD instead of HTS.¹⁵⁸ Functional groups were systematically removed to give a range of fragment-sized molecules containing substructures of RG7112 (Figure 1.15). The fragments were then tested for MDM2 binding using NMR and SPR assays to identify the smallest molecules with detectable binding. The fragments containing mimics for two of the three hotspot residues, where one of the mimics targets the W23 sub-pocket, were found to bind to MDM2 (using biophysical methods) demonstrating that a hit molecule could have been discovered from a fragment-based approach. A limitation of this approach is that none of the fragments with detectable binding to MDM2 were commercially available, meaning a custom fragment library would have to be synthesised to target the p53/MDM2 PPI.

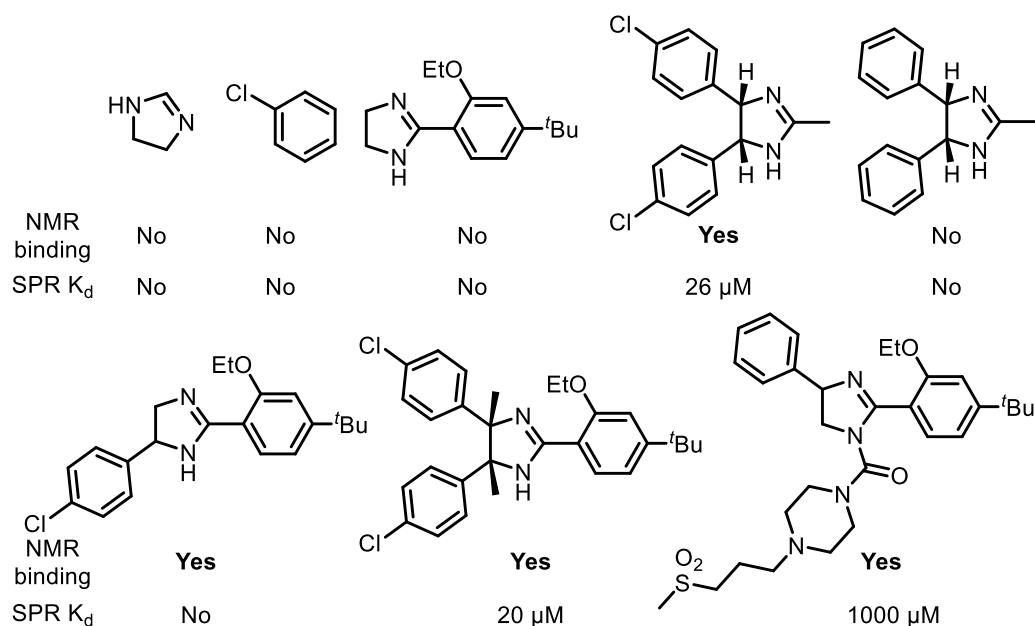


Figure 1.15. Deconstruction of RG7112 (SPR K_d 0.22 μM) to determine key groups for binding to MDM2.

1.5 Project Outline

Activity-directed synthesis is a proven and powerful method for the discovery of ligands of protein targets with an evolved small molecule binding site. A key aim is to build upon the existing activity-directed framework and enable the discovery of ligands for more challenging biological targets, specifically a

target without an evolved small molecule binding site. Protein-protein interactions, specifically the p53/MDM2 PPI, offer an ideal platform for the validation of ADS as a discovery tool for more challenging protein targets. The medicinal chemistry of the p53/MDM2 PPI has been well explored but has not been targeted by integrated methods of synthesis and biological screening for small molecule discovery. Therefore, if ADS can facilitate the discovery of inhibitors containing novel chemotypes for inhibitors of the p53/MDM2 PPI then the approach will be showcased as an efficient tool for identifying new bioactive chemical space.

1.5.1 Development of a Catalyst Map for Rhodium(II) Catalyst Selection in Activity-Directed Synthesis

Intermolecular reactions between rhodium(II)-carbenes and multi-functional substrates, as part of an ADS experiment, offer opportunities for the identification of new small molecule scaffolds that can modulate protein function. This is because transformations catalysed by rhodium(II) complexes can be controlled through their corresponding ligand properties (Chapter 2, section 2.2). It was proposed that a knowledge base for interpreting the reactive space of rhodium(II)-carbenes would facilitate the selection of catalysts with complementary ligand properties and potentially facilitate unexpected reaction outcomes. This could allow for a better exploration of the reactive space within an ADS reaction array as each catalyst could direct reactions towards products in different areas of chemical space.

To validate and exemplify a knowledge base approach for catalyst selection a database of computational parameters describing rhodium(II) complexes must be built and analysed. This would create a map for navigating catalyst properties and would aid in the selection of catalysts for ADS array design. The chemical relevance of the map must also be demonstrated to ensure correlation between the knowledge base and the observed reactivity in rhodium(II)-catalysed reactions. Ideally, the knowledge base will enable the optimisation of ADS reaction arrays by reducing the number of catalysts required to identify structurally distinct classes of ligands. Further detailed objectives for this work are described in Chapter 2.

1.5.2 Activity-Directed Discovery of Inhibitors of the p53/MDM2 Protein-Protein Interaction

MDM2 does not have an evolved small molecule binding site meaning the identification of fragment-sized ligands is extremely challenging. Fragment libraries often do not contain molecules suitable for screening against protein-protein interactions, and have not been used for drug discovery targeting the p53/MDM2 PPI. It was proposed that ADS could be used for experimental scaffold hopping to target pairs of MDM2 sub-pockets and discover structurally distinct inhibitors of this PPI. Reaction arrays would be designed to leverage mimics of p53 hot spot residues, and rhodium(II)-carbene chemistry, to identify new scaffolds that target the p53/MDM2 PPI. Successful application of ADS in this context would demonstrate that the method can fill gaps in current discovery methods without extensive investment in chemical synthesis. This work, including detailed objectives, is described in Chapters 3 and 4. Chapter 3 focussed on the design and implementation of two ADS reaction arrays, and the isolation of products from the scale-up of hit crude reaction mixtures. Chapter 4 focussed on the characterisation of the isolated products as inhibitors of the p53/MDM2 protein-protein interaction.

Chapter 2

Construction of a Catalyst Map for Rhodium(II) Catalysts

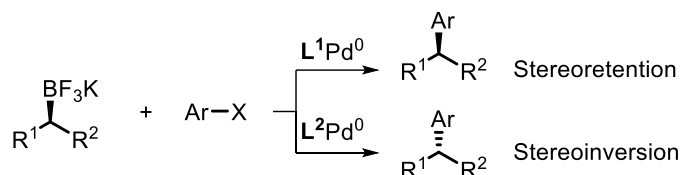
Metal-catalysed cross-coupling reactions, such as the Suzuki-Miyaura and Buchwald-Hartwig reactions, are frequently used to aid in the discovery of bioactive small molecules.^{65–69,159–161} More recent advances in metal-catalysed carbon-carbon and carbon-heteroatom bond forming reactions have not yet been widely adopted by the medicinal chemistry community.¹⁵⁹ The slow uptake of these methods may stem from a poor understanding of substrate scope, a perceived low success rate or poor availability of newly reported catalysts.^{65,69,159}

Recent advances in high-throughput experimentation (HTE) have allowed hundreds, or thousands, of metal-catalysed reactions between densely functionalised pairs of substrates to be optimised and employed in molecular discovery workflows.^{73,75,80,81,114,115,162} The increased availability of these information rich datasets has also spurred along work deploying machine learning (ML) or statistical learning approaches to predict catalyst performance and gain mechanistic insights.^{73,74,163–166} These approaches rely on a combination of experimental data and computational parameterisation data.^{167,168} ML techniques often lead to the development of virtual screening platforms where computational parameters describe features of the substrate pairs,^{73,74,169,170} whereas statistical learning approaches can reveal mechanistic details by describing features of catalytic intermediates or transition states.^{164,171,172} Both approaches rely on experimental and computational data and can reveal trends that were not obvious from the standard processing of experimental data. Models that allow medicinal chemists to easily interpret the most important factors for successfully implementing new reactions could stimulate their uptake into molecular discovery workflows.^{168,173–176}

2.1 Computational Parameterisation for Catalysis

The computational parameterisation of catalysts, or ligands, has been used in combination with experimental data to rationalise outcomes when developing new synthetic methodologies.¹⁷⁶ An example of this approach is the development of a ligand-controlled stereospecific Pd-catalysed C–C

bond forming reaction.¹⁶⁴ A suite of second-generation molecular descriptors for phosphine ligands was developed so that statistical learning methods could be applied to the optimisation and mechanistic rationalisation of the enantio-divergent Pd-catalysed coupling of alkylboron nucleophiles with aryl halides (Scheme 2.1).



Scheme 2.1. Ligand-controlled enantio-divergent cross-coupling reactions using alkyl trifluoroborate nucleophiles.

The authors noted that electron-withdrawing substituents on the aryl halide substrate lead to diminished stereo-retention, whereas electron-donating substituents maintained stereospecificity. This trend suggested that manipulating the electronics of the system could influence the mechanism of transmetallation and the product stereochemistry. A high-throughput screen of phosphine ligands was then initiated, and correlation analysis performed to identify computational parameters that captured factors important to the stereospecificity of transmetallation.

The modelling workflow began by performing molecular mechanics (MM) conformational searches for each ligand to identify low-energy conformers, followed by geometry optimisation using DFT. Parameters were then extracted, and seven computational descriptors (four electronic and three steric) were responsive to the observed stereo-retention. Four additional descriptor subsets were also defined to capture the conformational dynamics of the ligands, including Boltzmann weighted average parameters. Two ligand classes were eventually identified from a virtual screen of phosphine ligands (Figure 2.1): biaryl phosphines bearing electron-withdrawing substituents that resulted in stereo-retention (**L1**); and bulky tri-alkyl phosphines that resulted in stereo-inversion (**L2**).

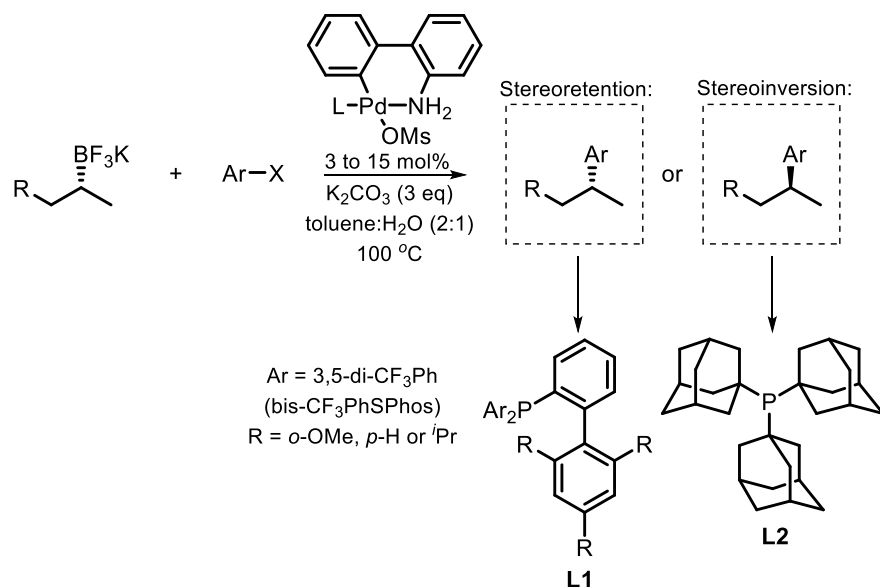


Figure 2.1. Experimental investigation for the ligand controlled stereospecific Pd-catalysed arylation of alkyl boron nucleophiles with a range of aryl chlorides. **L1** gives stereo-retention and **L2** gives stereo-inversion.

Machine learning techniques have also been used for the prediction of highly selective chiral phosphoric acid catalysts for the acid-catalysed thiol addition to *N*-acylimines.¹⁶⁶ The workflow began with an *in silico* library of over 800 chiral phosphoric acid catalysts with two different scaffolds, one with fully aromatic binaphthyl backbones and another with a saturated ring in place of the fused binaphthyl aromatic ring. Approximately 400 synthetically accessible substituents were included at the 3,3' positions of the two scaffolds to create the library of catalysts (Figure 2.2). The authors then invented a new descriptor called average steric occupancy (ASO) designed to capture steric information about an entire ensemble of conformers for a single molecule. ASO provides this information as a matrix of values that represents the weighted spatial occupancy of substituents for each conformer of the ligand, as a function of the energetic profile for the ensemble of conformers.

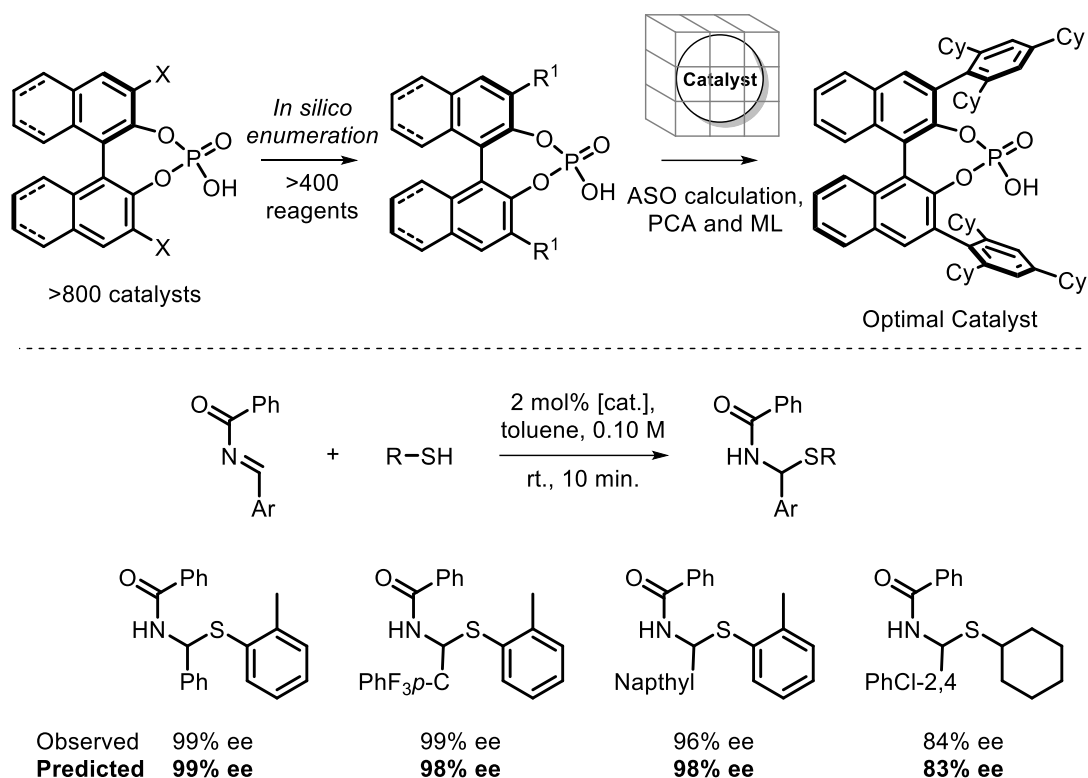


Figure 2.2. Design of computational workflow for the prediction of enantioselectivity for the phosphoric acid-catalysed formation *N,S*-acetals.

A subset of ligands covering the broadest area of available chemical space within the *in silico* library was then selected using principal component analysis (PCA) combined with the Kennard-Stone algorithm. The authors attribute this sampling method to the ultimate success of the workflow as it guarantees uniform selection of catalysts across all regions of chemical space (including boundary cases and outliers). The ASO descriptor and catalyst training set (selected using PCA and Kennard-Stone) were then validated by creating a predictive model for the enantioselective formation of *N,S*-acetals. The authors were able to build high performance statistical models for: the prediction of catalyst performance against reactions forming new products; the prediction of new catalysts external to the model against known products; and prediction of new catalysts forming new products, both external to the model. Finally, a deep feed-forward neural network was able to accurately recapitulate the experimental selectivity data, successfully predicting the most selective reactions when half of the training data was omitted. These results demonstrate the power of chemically relevant molecular descriptors when applied to the virtual screening of catalysts.

Computational parameterisation of rhodium(II)-carbenes has also been undertaken to investigate the Rh_2esp_2 -catalysed intermolecular C–H insertion of β -carbonyl ester carbenes.¹⁷⁷ The model is based upon parameterisation of a β -carbonyl diazo ester substrate and the resulting carbene formed with Rh_2esp_2 . The authors quantified steric effects related to the diazo substituents by calculating the 1,5-interaction between the diazo substituent and the ester carbonyl (captured in terms of an r value) in the Rh_2esp_2 carbene complex. Electronic effects were also quantified by calculating the proton affinity (PA) of the β -carbonyl. Analysis of the calculations formed a potential energy surface allowing the relative enthalpy (ΔH kcal mol⁻¹) to be mapped onto the r and PA values for each carbene complex. The authors noted that regions of large ΔH values correlated with the formation of side-products through the Wolff rearrangement. This inspired the design of a series of β -carbonyl diazo ester substrates that could undergo alkyl C–H insertion reactions in moderate yields (Figure 2.3). The experimental results showed that increasing the size of the β -carbonyl substituent leads to a decreased yield of the C–H insertion product.

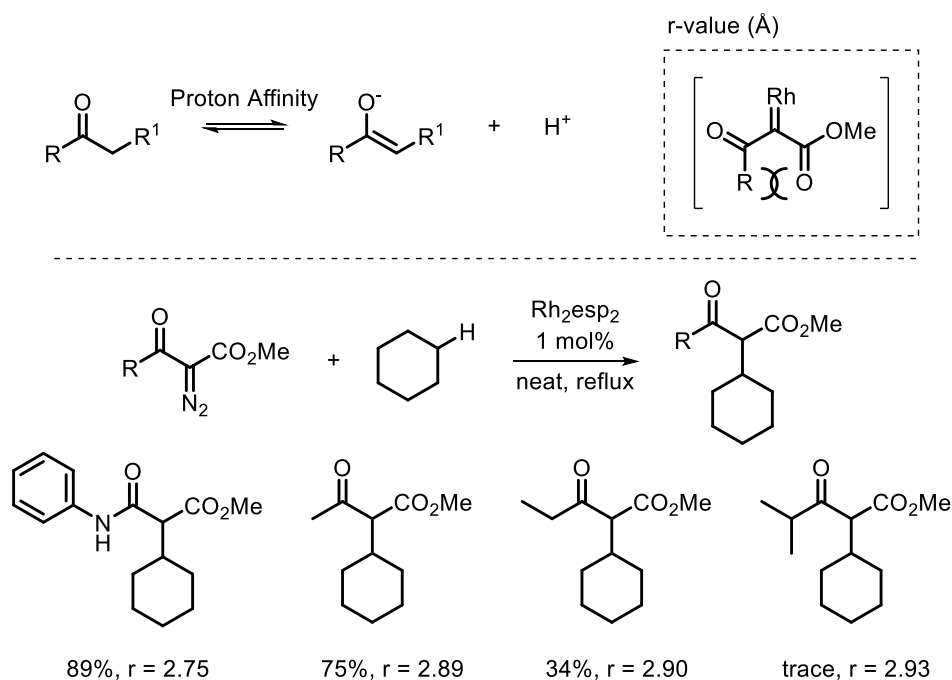


Figure 2.3. Parameters for β -carbonyl diazo esters and observed trends with experimental data.

The Bristol Ligand Knowledge Base (LKB)^{178–182} was based on a wide range of computational descriptors for monodentate P-donor ligands (348 ligands) and has been expanded to include bidentate ligands. The monodentate LKB

uses a set of 28 electronic and steric descriptors that were derived from DFT optimised geometries of gold, platinum and palladium complexes. The descriptors included frontier molecular orbital energies, proton affinities, adduct formation, and metal coordination.¹⁷⁹ The knowledge base was then used to build a PCA model, and thereby create a map of chemical space from which underlying chemical features could be elucidated for rational phosphorus(III) ligand design (Figure 2.4).^{179,180}

PCA is a method for revealing a simplified structure from a highly dimensional dataset by extracting relevant information that is often hidden in higher dimensional space.¹⁸³ PCA achieves dimension reduction by setting the goal to compute the most meaningful basis to represent a dataset. PCA is a linear transformation, and it simplifies the number of possible basis sets by assuming that the data is continuous and is, therefore, restricted to re-expressing the data as a linear combination of its basis vectors (*i.e.* as principal components).^{183,184} PCA retains the integrity of the data, in this case the chemical descriptors, as the underlying data is not manipulated. In the LKB catalysts with similar properties will be near neighbours in the PCA map, while dissimilar catalysts will be further apart. The benefit of PCA is that the resulting model is representative of the original chemical observations as PCA represents the variance between data points, not the data points themselves, so the analysis does not project absolute values.

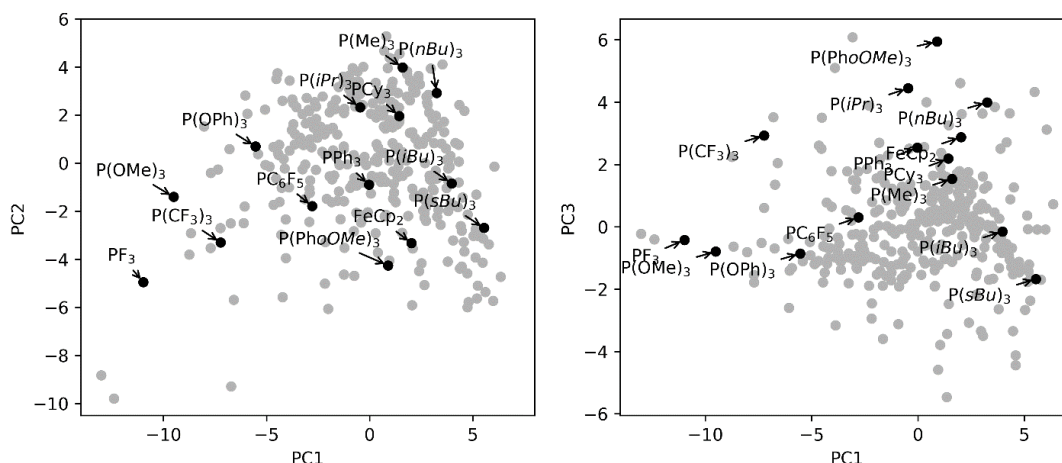


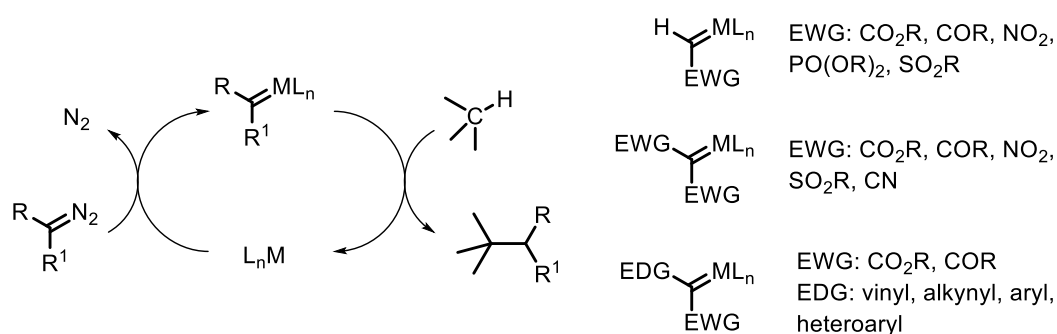
Figure 2.4. Bristol LKB PCA map for monodentate P-donor ligands.¹⁸⁰

The Bristol LKB effectively captured ligand similarities and differences by sampling a wide range of chemical space. The LKB included 117 commercially available ligands meaning the model could be used for the design of experiments and reaction optimisation.¹⁸⁰ The PCA model also

captured some trends in the experimental data for a palladium-catalysed amination reaction and some of the descriptors were used to build a multi-variate predictive model for product yield with good accuracy. A model to predict bond dissociation energies was also built using descriptors from the knowledge base which would allow for optimisation of a catalytic process through rational ligand design. The LKB, and catalyst map, has proved to be a useful tool for the optimisation of palladium-catalysed cross-coupling reactions and allows for the rational exploration of catalyst space.^{167,184}

2.2 An Overview of Rhodium(II)-Carbene Chemistry

Rhodium(II)-catalysed carbenoid chemistry provides access to a wide range of transformations including selective insertion into C–H bonds, and insertion into O–H, N–H and S–H bonds, cyclopropanation and ylide formation.^{185–190} The selectivity of many rhodium(II)-catalysed transformations can be tuned by considered selection of both the functional groups that decorate a diazo substrate and the catalyst.^{191–195} Diazo precursors, for the formation of rhodium(II)-carbenes, can be separated into two categories: those with only electron withdrawing substituents (acceptor and acceptor/acceptor); and those with both electron donating and withdrawing substituents (donor/acceptor) (Scheme 2.2).¹⁸⁷



Scheme 2.2. An overview of a catalytic cycle for C–H insertion¹⁸⁵ and common substituents incorporated into acceptor, acceptor/acceptor, and donor/acceptor diazo compounds.

Acceptor and acceptor/acceptor carbenes are highly reactive and often undergo intramolecular transformations, preferentially forming five-membered rings due to their electron withdrawing nature and inability to stabilise the electrophilic carbene centre – although the formation of four and six membered rings are also common.^{190–192} Donor/acceptor substituted

carbenes are less reactive than acceptor and acceptor/acceptor carbenes due to their greater ability to stabilise the electrophilic carbene centre.^{185,187,196} This allows donor/acceptor carbenes to undergo the site-selective intermolecular C–H functionalisation of complex molecules at primary, secondary, or tertiary C–H bonds.^{185,194,195,197,198} Selective C–H insertions are achieved through steric and electronic control of the carbene and substrate, not through site-directing functional groups which is common for other C–H activation chemistries.¹⁹⁹ Substrates that feature sites able to stabilise a build-up of positive charge react significantly faster than those unable to do so, and sites adjacent to a heteroatom are also more reactive.¹⁸⁵ However, the reaction rate can be negatively impacted if the site is sterically hindered. For example, tertiary C–H bonds have slower reaction rates than secondary C–H bonds despite their ability to stabilise positive charge more effectively.^{185,197,198} As such, electronic effects are not the sole contributions to consider and a much more complicated analysis of electronic and steric effects for both substrate and catalyst are required to determine reactivity (Figure 2.5).

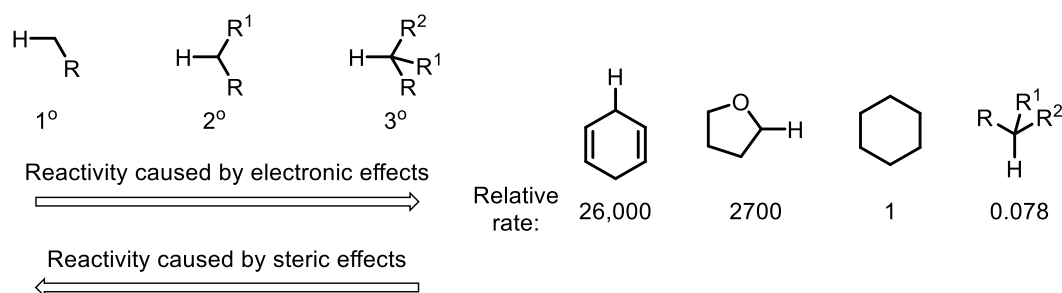
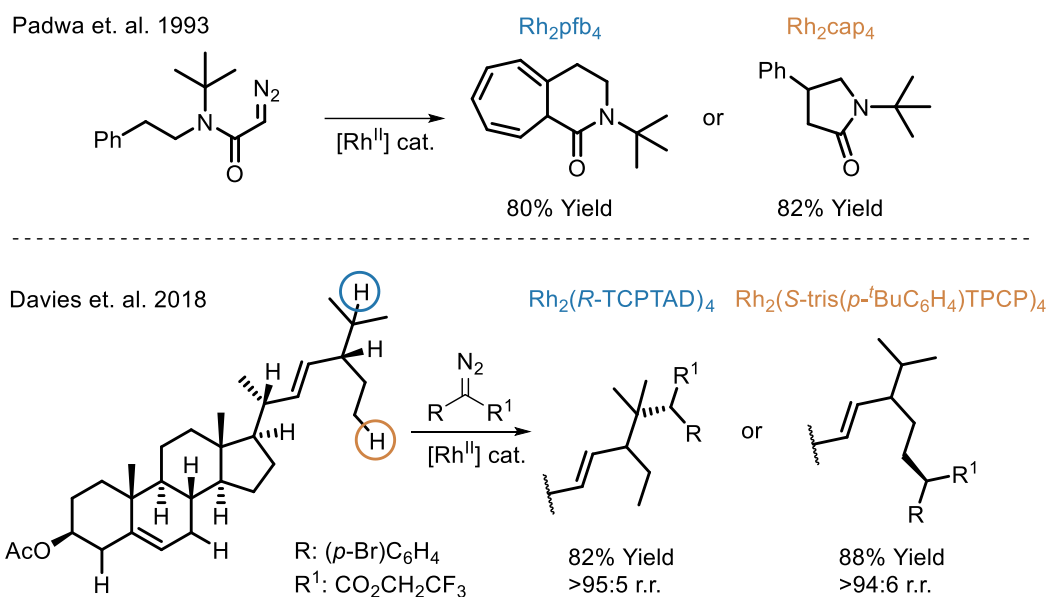


Figure 2.5. Controlling factors for increasing reactivity controlled by electronic and steric effects in reactions with donor/acceptor diazo compounds and $\text{Rh}_2(\text{S-DOSP})_4$.^{185,197} Relative rates refer to the rate of C–H insertion into a cyclohexyl C–H bond, normalised to one.

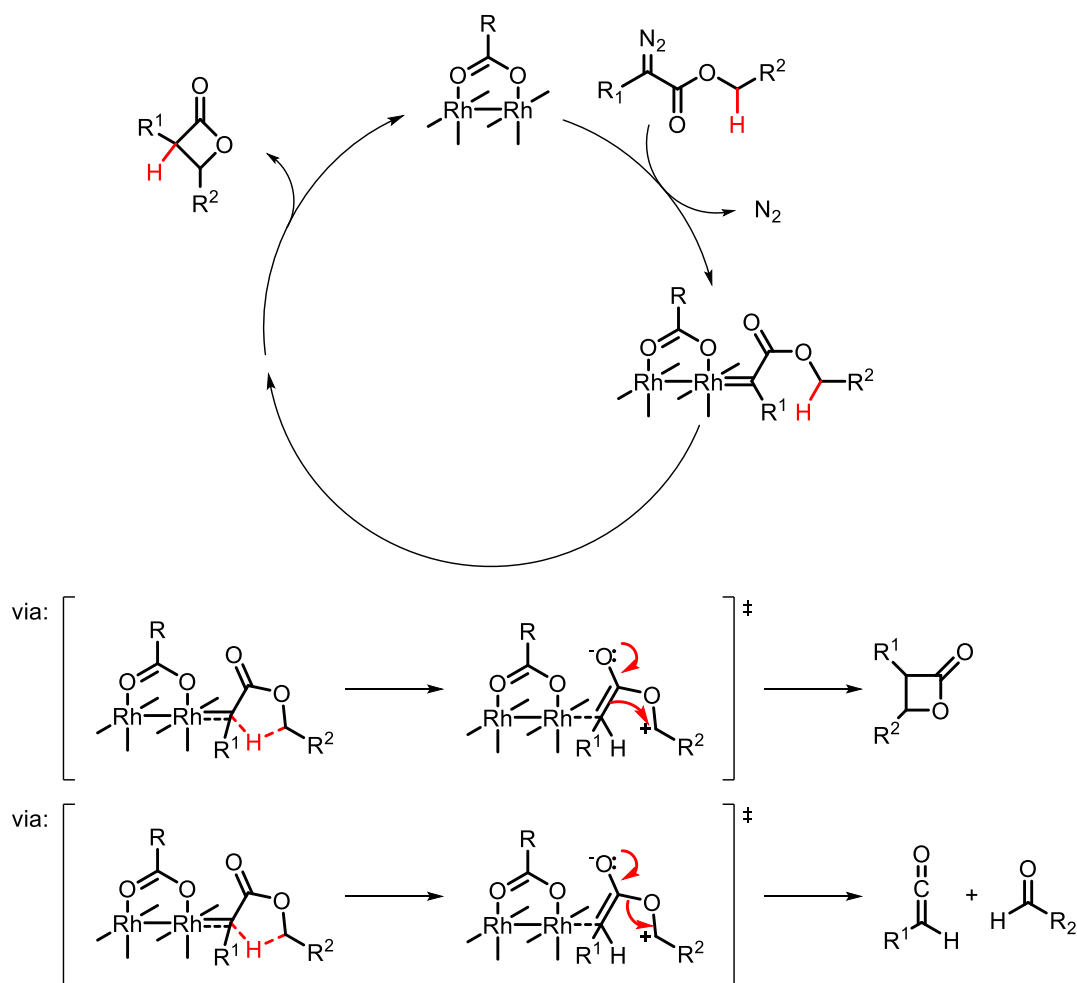
The ligand controlled nature of rhodium(II) catalysts allows for high levels of chemo-, regio-, and stereo-selectivity in a variety of contexts.²⁰⁰ For example, donor atom type and fluorination of the bidentate ligand backbone leads to switchable chemoselectivity in a substrate with multiple reactive sites (Scheme 2.3).^{191,192} Intermolecular C–H insertion reactions can also be tuned through ligand effects to give highly regioselective insertion into primary or tertiary C–H centres (Scheme 2.3).¹⁹⁸ These catalysts employ large aliphatic ligands that intrude into the co-ordination sphere of the rhodium(II) core and give control over unactivated and unfunctionalized

substrates. This reactivity is challenging to analyse and predictive modelling of this reaction^{170,201} has resorted to substrate observed parameters due to the size of the catalysts and the complexity of the ligand effects.



Scheme 2.3. Control over two alternative reaction pathways by Rh_2pfb_4 and Rh_2cap_4 ,¹⁹¹ and ligand controlled regio- and stereoselective C–H insertion by $\text{Rh}_2(\text{R-TCPTAD})_4$ and $\text{Rh}_2(\text{S-tris}(p\text{-}^t\text{BuC}_6\text{H}_4)\text{TPCP})_4$.¹⁹⁸ (r.r. regioisomer ratio).

The variety of reactivity of rhodium(II)-carbenes makes accurate mechanistic studies challenging. In the above examples, modification of the ligand backbone affects the selectivity of the proceeding reaction when multiple competitive pathways are accessible. A further complication is that rhodium(II)-carbenes can undergo transition state bifurcations (Scheme 2.4).²⁰² If the products of the transition state bifurcation can then react to form the desired product then any modelling effort is only relevant to the specific modelled case. Catalyst parameterisation approaches have used calculated steric and electronic parameters for substrates,²⁰¹ or a surrogate,¹⁷⁰ or structural analysis of the relationship between the catalyst and substrate.¹⁷⁷ For these approaches, careful parameter design enabled the development of predictive models for specific reactions, but are not generalisable for other rhodium(II)-catalysed reactions.



Scheme 2.4. Transition state bifurcations for the intramolecular processes of rhodium(II) carbenes. Each rhodium complex is bound to four bidentate ligands, however three are omitted for clarity.

2.3 Design of a Knowledge Base for Rhodium(II) Catalysts

Knowledge bases, such as the Bristol LKB, are databases of parameters, or descriptors, that describe the steric and electronic properties of interesting organometallic complexes.^{167,184} Parameters are usually captured from DFT-optimised geometries which allows for the investigation of many more complexes than would be experimentally plausible. Development of a knowledge base for rhodium(II) catalysts could enable the inclusion of underutilised carbon-carbon and carbon-heteroatom bond forming reactions into molecular discovery workflows. The rhodium(II) knowledge base was built in collaboration with Dr Natalie Fey at the University of Bristol and employs similar techniques as the Bristol LKB for catalyst parameterisation. The parameters used to construct the catalyst knowledge base were

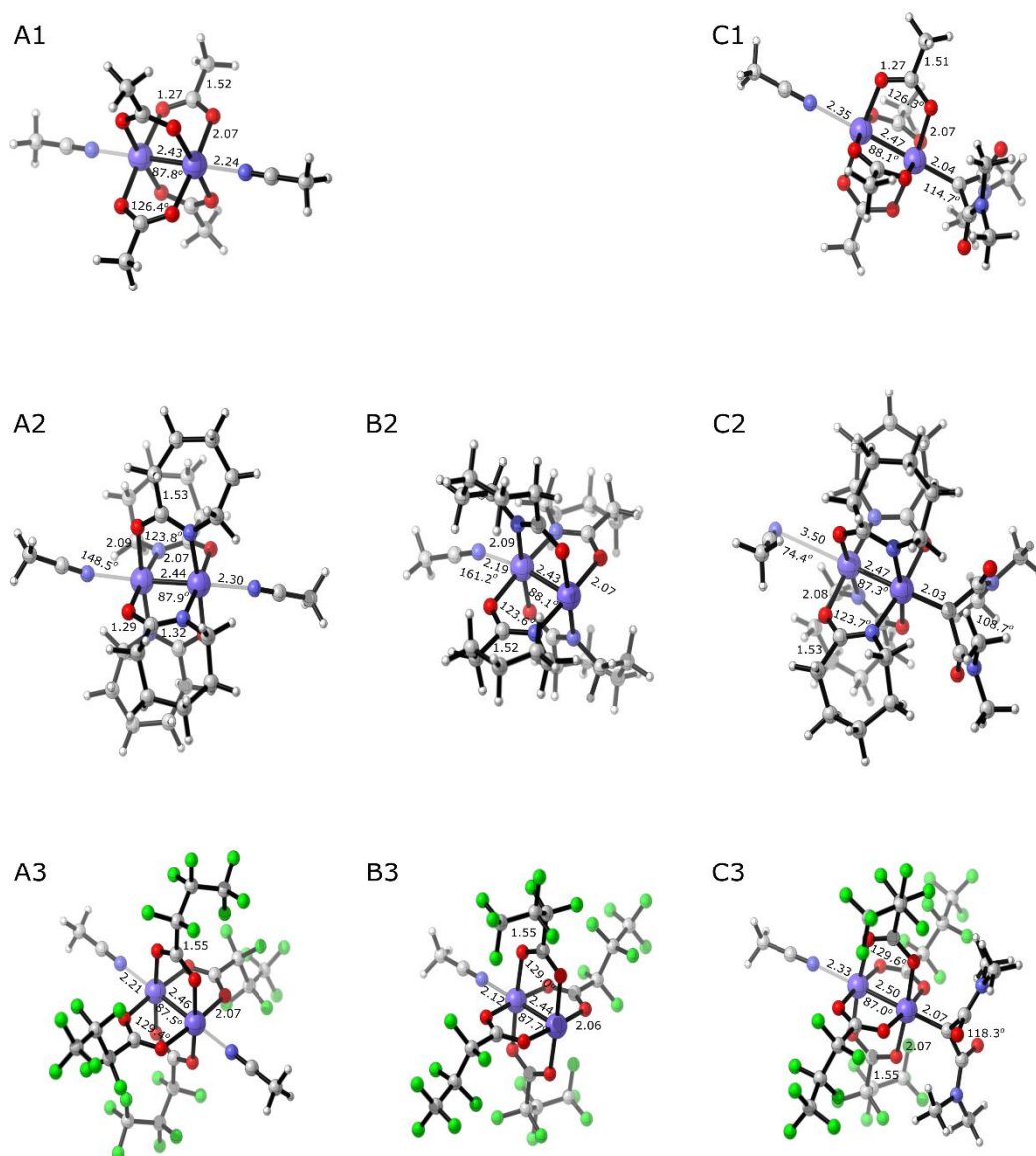


Figure **2.6**. Example acetonitrile adducts calculated at B3LYP/6-31G(d)/MWB28 level of theory. A1: Complex **3a** and 2 acetonitrile ligands. C1: Complex **3a** with a carbene ligand and one acetonitrile ligand. A2: Complex **4a** with one acetonitrile ligands. B2: Complex **4a** with one acetonitrile ligand. C2: Complex **4a** with a carbene ligand and one acetonitrile ligand. A3: Complex **3h** with two acetonitrile ligands. B3: Complex **3h** with one acetonitrile ligand. C3: Complex **3h** with a carbene ligand and one acetonitrile ligand. Complex **3a** with one acetonitrile ligand could not be converged and was omitted.

Moving to the BP86^{213,214} density functional using the 6-31G(d) DZP basis set and MWB28 core potential allowed for the rapid generation of 96 converged rhodium(II) complexes. The BP86 function may produce slight

over-binding effects but is computationally robust and allowed for expansion of the scope from the three initial complexes **3a**, **4a** and **3h** to 48 diverse complexes (Figure 2.7). A conformer search, using molecular mechanics, conducted by Dr Natalie Fey at the University of Bristol, found that the DFT optimised geometries accurately represented each complex.

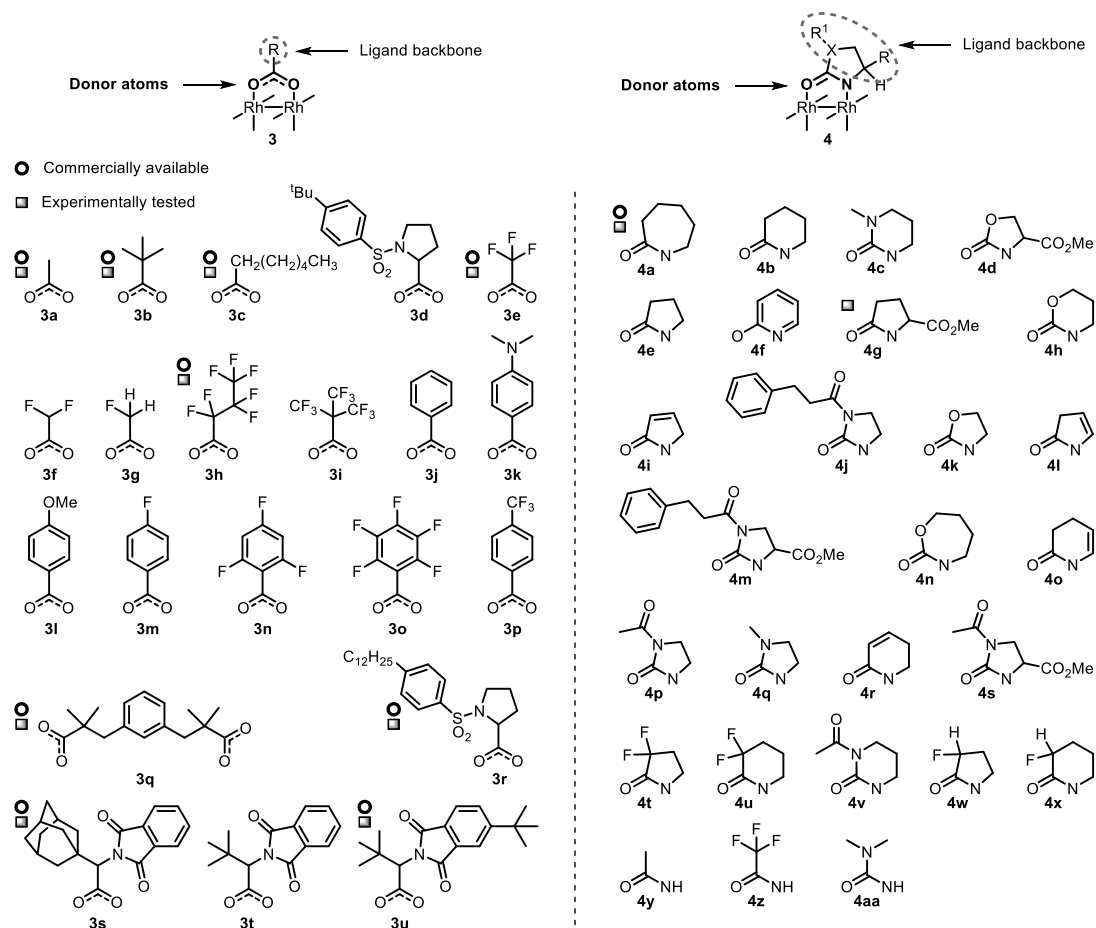


Figure 2.7. All 48 ligands used to generate the knowledge base. Ligands are drawn with the coordinating atoms at the bottom and the labels describe the rhodium(II) complex bound to these ligands. Catalysts that were purchasable commercially are highlighted (circle) or that were later experimentally tested (square).

The knowledgebase contains a wide range of carboxylate (*O,O*) and carboxamidate (*N,O*) ligands that were selected, or designed, to give an even distribution of empirically predictable electronic and steric features such as donor atom type, substitution patterns, steric bulk and synthetic feasibility. The ability to select and synthesise any given catalyst is an important feature for this knowledgebase to have wider applications in the activity-directed discovery of molecules because many rhodium(II) catalysts are not commercially available and available sets offer poor catalytic

diversity. Complexes bearing ligands with features including heteroatom substitution (e.g. **3k**, **3l**, **4c**, **4h**, **4k**, **4n** and **4q**), fluorination (**3e-3i**, **4t-4u** and **4w-4x**) and unsaturation (**4i**, **4l**, **4o** and **4r**) were included to probe electronic effects propagated through the ligand backbone. Carboxamidate ligands with different cyclic ring sizes were also designed to investigate structural changes to the rhodium(II) core (**4a-4x**). Finally, several ligands bearing large ligand substituents that can control the nearby environment of the metal coordination sphere (e.g. **3r-3u** and **4e**, **4m**, **4p** and **4s**) were included to complete the reported scope of ligand effects.¹⁹⁸

A wide range of descriptors were then captured, from DFT-optimised geometries, to investigate the range of possible electronic and steric effects arising from different ligand types (Figure 2.8 and Table 2.1). The features captured by each descriptor were classified into three categories: electronic (HOMO, σ , LUMO, σ , Q Rh, Q(L, mean), Q(Donor Atoms, mean) and $\Delta E(\text{FMO})$); electronic and steric ($r(\text{Rh-Rh})$, $\chi(\text{Rh-Rh-L})$, $\chi(\text{O-C-X})$, $\chi(\text{C-C-C})$, $r(\alpha\text{C-R}^1)$, $r(\text{Rh-C})$ and $\Delta E(\text{coord})$); and steric (He_8 and $|\text{wV}|$). The descriptor He_8 ¹⁸⁰ approximates the steric influence of a ligands substituents by modelling the approach of a reactant to the complex with a ring of eight Helium atoms. $|\text{wV}|$ ^{215,216} measures the steric bulk of a ligand while accounting for the proximity of the ligand to an important atom, in this case the rhodium atom forming the carbon bond. Both descriptors were calculated using the optimised geometry of the corresponding carbene complex **2** with the carbene ligand removed and aligned with the rhodium-rhodium bond.

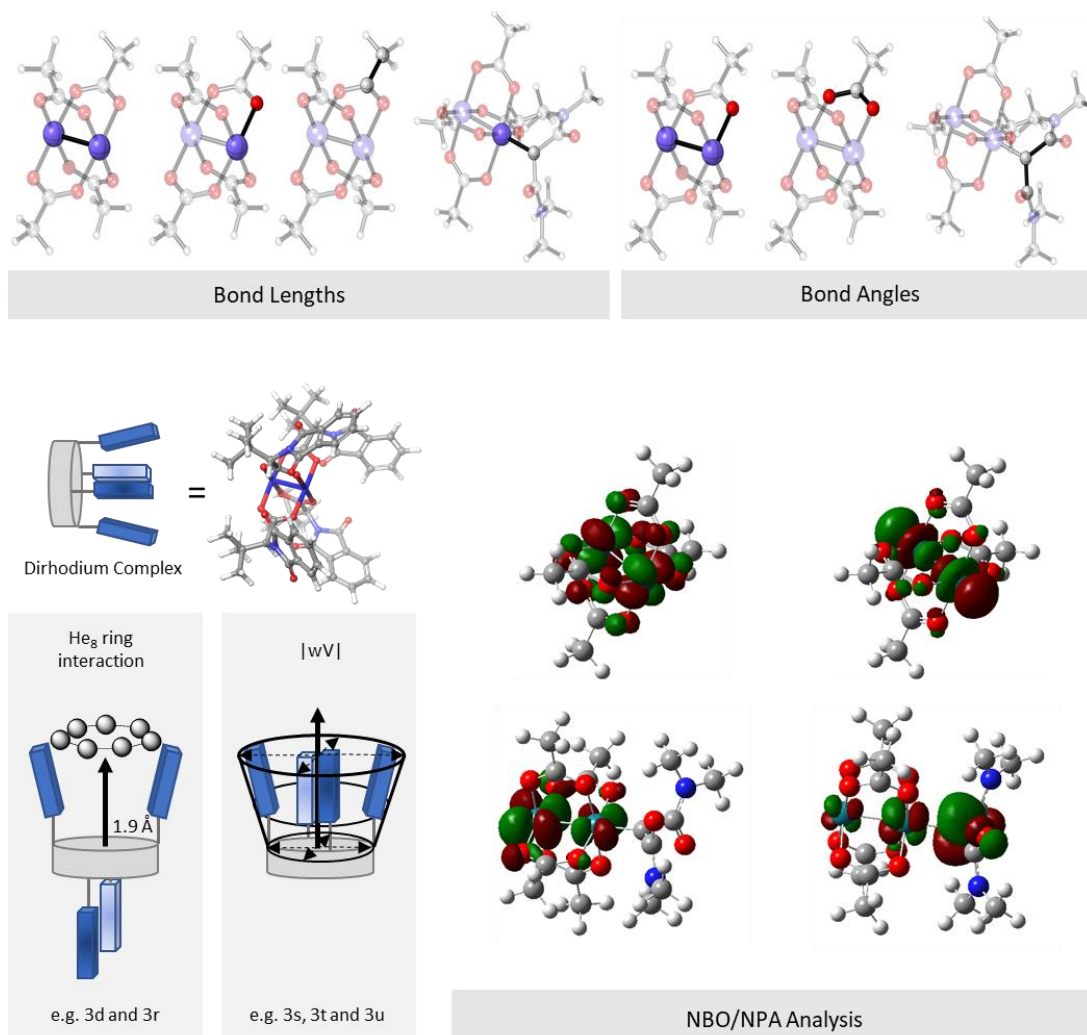









Figure 2.8. Examples of calculated structural, steric and electronic parameters used in construction of the knowledgebase. For He₈ the ring of Helium atoms was positioned 1.9 Å from the Rh atom forming the carbene bond, and |wV| was aligned along the vector of the Rh-Rh bond.

Table 2.1. Computational parameters considered for inclusion in the knowledge base.

Descriptor	Derivation	Diagram	Median Value	Range
r(Rh-Rh)	Rh-Rh bond length (Å), 1 and 2		2.41 Å	0.09 Å
r(Rh-L)	Average Rh-Ligand bond length (Å), 1 and 2		2.06 Å	0.02 Å
∠(Rh-Rh-L)	Average Rh-Ligand bond angle, 1 and 2		88.57	1.29
∠(O-C-X)	Average ligand bite angle, 1 , 2		125.99	7.34
∠(C-C-C)	Carbene bite angle, 2		-	-
r(Rh-C)	Rhodium-Carbene bond length (Å), 2		-	-
r(αC-R1)	Average αC-Ligand Backbone (R1) bond length (Å), 1		1.52 Å	0.19 Å
HOMO, 1	Energy of HOMO for complex 1 (a.u.)	Figure 2.8	-0.15 a.u.	0.12 a.u.
LUMO, 1	Energy of LUMO for complex 1 (a.u.)	Figure 2.8	-0.13 a.u.	0.10 a.u.
Q Rh, 1	Charge on Rhodium atoms	-	0.66	0.30
Q(L, mean), 1	Mean charge on ligands	-	-0.85	0.55
Q(Donor Atoms, mean), 1	Mean charge on ligand donor atoms	-	-4.27	1.30
ΔE(FMO), 1	ΔE between HOMO and LUMO (a.u.)	-	0.03 a.u.	0.03 a.u.
wV	Distance-Weighted Volume, ¹⁹ 2	$V_{W,k,l} = \sum_{i=1}^n \frac{r_i^k}{d_i^l}$	9.35	40.20
He ₈	Interaction energy for 2 and ring of 8 Helium atoms ¹⁶ (kcal mol ⁻¹)	He ₈ = E(He ₈ .[Rh-Rh]) - E(He ₈) - E([Rh-Rh])	35.33 kcal mol ⁻¹	47.75 kcal mol ⁻¹
ΔE(coord)	Energy for diazo precursor to form the carbene complex (kcal mol ⁻¹), 2	ΔE(coord) = (E1 + EDiazo) - (E2 + EN ₂)	19.75 kcal mol ⁻¹	24.09 kcal mol ⁻¹

Correlation analysis was then performed on the knowledge base to evaluate the performance of the chosen descriptors and to contextualise the information captured (Figure 2.9). Strong correlations between $r(\text{Rh-Rh})$ and $\angle(\text{Rh-Rh-L})$, with R values between 0.8 and 1.0, reflect the rigidity of the paddlewheel Rh_2L_4 scaffold. Changes to the Rh-donor atom distances, either through electronic effects or steric clashes, affect the Rh-Rh bonding along with the geometry of the ligand coordination. Trends such as increasing $r(\text{Rh-Rh})$ between the axially unbound complexes and carbene complexes were universal across donor atom type (0.079 Å increase on average for O,O ligands and 0.072 Å for N,O ligands). However, N,O ligands with cyclic backbones have shorter Rh-Rh bonds on average (2.404 Å and 2.387 Å for unbound $O,O/N,O$ ligands respectively), in part, due to the constrained geometry of the amide functional group when part of a cyclic ring (e.g. $\angle(\text{O-C-X})$ 126.7 ° for O,O ligands and 125.5 ° for N,O ligands).

A different form of steric effect, described as the extent to which the ligand substituents intrude into the site of reaction, was captured by $|wV|$ and was important for controlling the geometry of the rhodium(II)-carbene and the subsequent angle of attack of a reactant. This steric effect was likely responsible for control of the regio- and stereo-selectivity of each catalyst.^{198,211,212} Catalysts **3r-3u** and **4d**, **4g**, **4m** and **4s** have the largest $|wV|$ values due to ligand projection over the axial face. Weak correlations between $|wV|$ and all other parameters confirms that this was a unique and purely steric descriptor, with R values ranging between 0 and 0.4. Electronic effects are also important for regio- and chemo-selectivity.¹⁹⁰⁻¹⁹³ Modulating the Lewis acidity of a catalyst can bias the outcome of a transformation where multiple competing pathways occur and was an important feature to capture.¹⁹¹ Moderate correlations between purely electronic parameters and those that capture a mixture of electronic and steric effects, such as $Q(L, \text{mean})$ and $r(\text{Rh-Rh})$ (R 0.5 – 0.6), demonstrates that steric effects do not dominate the knowledge base.

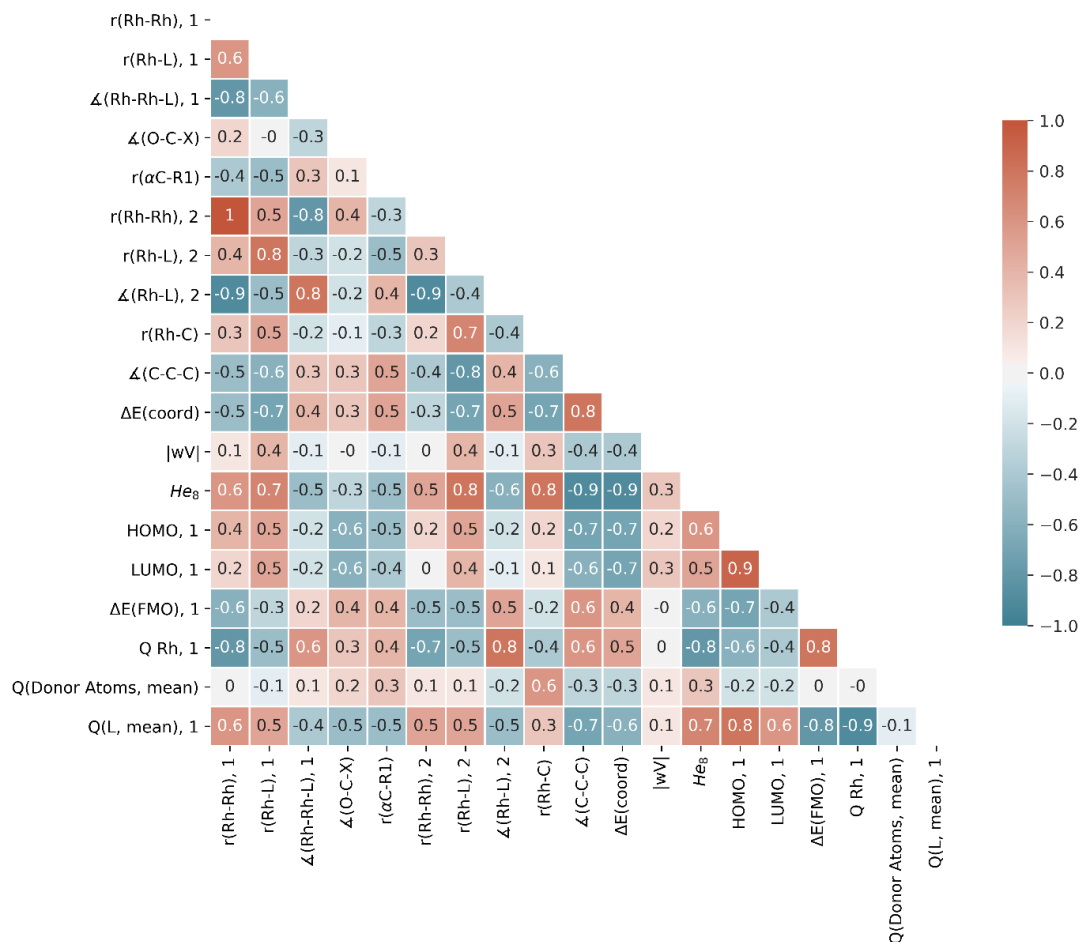


Figure 2.9. Calculated Pearson R correlation coefficient map

2.4 Principal Component Analysis for the Construction of a Catalyst Map

A model for interpreting the knowledge base using principal component analysis (PCA) was then implemented. A relevant subset of descriptors must be selected for use in principal component analysis to ensure the subsequent model can be reliably interpreted and deconvoluted. This means that the model must be generated from descriptors that can be related to chemical features important to the properties of the catalyst. Inclusion of redundant descriptors will make the PCA model harder to interpret as each principal component will not contribute as clearly to clustering.

To simplify the number of PCA plots that need to be analysed manually the best solution for each discrete number of descriptors (n) was calculated (Figure 2.10). PCA results can be ranked by the total variance captured and by the mean squared error of projection, which is the amount of information

lost through PCA. The top solution for n descriptors was selected by the highest percentage variance captured (for solutions, see Table 2.2). For qualitative and unsupervised analyses, such as PCA, there is often no correct number of parameters for a model, instead models must also be considered in terms of their descriptor loading and interpretability. Models that give even descriptor loadings are favoured over those that bias results towards a fewer number of descriptors, as they are often better at highlighting complex relationships in the original data.

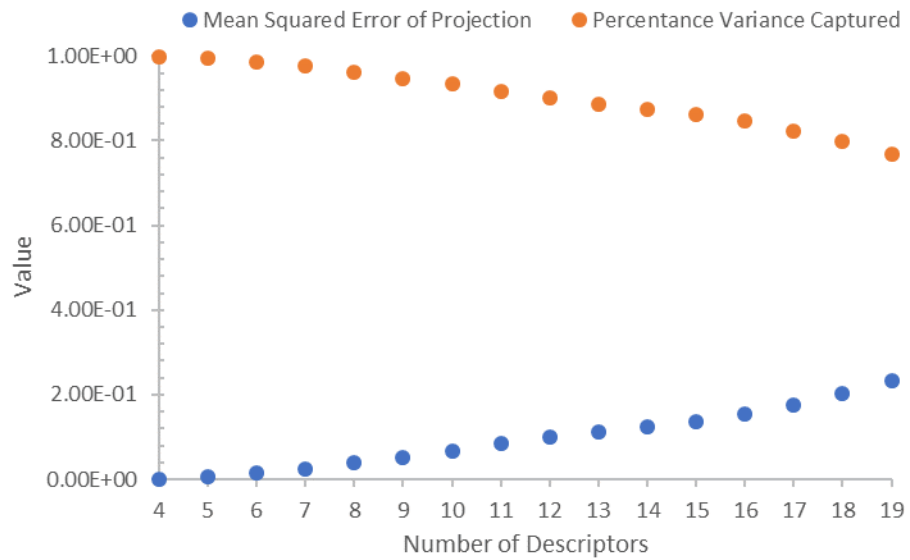


Figure 2.10. Percentage variance and mean squared error of projection for PCA solutions of n descriptors.

Table 2.2. Descriptors used in each highest performing PCA solution for the knowledge base.

<i>n</i>	Descriptors	Mean Squared Error of Projection	% Variance Captured
4	r(Rh-Rh) 2, HOMO, 1, LUMO, 1, ΔE(FMO)	1.82x10 ⁻³¹	99.9
5	r(Rh-Rh) 1, r(Rh-Rh) 2, HOMO, 1, LUMO, 1, ΔE(FMO)	0.0049	99.5
6	r(Rh-Rh) 1, r(Rh-Rh) 2, HOMO, 1, LUMO, 1, Q Rh, 1, Q(L, mean), 1	0.014	98.6
7	r(Rh-Rh) 1, ∠(O-C-X), r(Rh-Rh) 2, HOMO, 1, LUMO, 1, Q Rh, Q(L, mean), 1	0.0234	97.7
8	r(Rh-Rh) 1, ∠(O-C-X), r(Rh-Rh) 2, Rh-L Angle Carbene, HOMO, 1, LUMO, 1, Q Rh, 1, Q(L, mean), 1	0.038	96.2
9	r(Rh-Rh) 1, ∠(Rh-Rh-L) 1, r(Rh-Rh) 2, ∠(Rh-Rh-L) 2, HOMO, 1, LUMO, 1, ΔE(FMO), Q Rh, 1, Q(L, mean), 1	0.0516	94.8
10	r(Rh-Rh) 1, ∠(Rh-Rh-L) 1, ∠(O-C-X), r(Rh-Rh) 2, ∠(Rh-Rh-L) 2, HOMO, 1, LUMO, 1, ΔE(FMO), Q Rh, 1, Q(L, mean), 1	0.0662	93.4
11	r(Rh-Rh) 1, r(Rh-Rh) 2, ∠(Rh-Rh-L) 2, r(Rh-C), ∠(C-C-C), ΔE(coord), He ₈ Ring Interaction, HOMO, 1, LUMO, 1, Q Rh, 1, Q(L, mean), 1	0.0846	91.5
12	r(Rh-Rh) 1, ∠(O-C-X), r(Rh-Rh) 2, ∠(Rh-Rh-L) 2, ∠(C-C-C), ΔE(coord), He ₈ Ring Interaction, HOMO, 1, LUMO, 1, ΔE(FMO), Q Rh, 1, Q(L, mean), 1	0.0993	90.1
13	r(Rh-Rh) 1, ∠(Rh-Rh-L) 1, ∠(O-C-X), r(Rh-Rh) 2, ∠(Rh-Rh-L) 2, ∠(C-C-C), ΔE(coord), He ₈ Ring Interaction, HOMO, 1, LUMO, 1, ΔE(FMO), Q Rh, 1, Q(L, mean), 1	0.1132	88.7
14	r(Rh-Rh) 1, ∠(Rh-Rh-L) 1, ∠(Rh-Rh-L) 1, ∠(O-C-X), r(Rh-Rh) 2, r(Rh-L) 2, ∠(Rh-Rh-L) 2, ∠(C-C-C), ΔE(coord), He ₈ Ring Interaction, HOMO, 1, ΔE(FMO), Q Rh, 1, Q(L, mean), 1	0.1253	87.4
15	r(Rh-Rh) 1, r(Rh-L) 1, ∠(Rh-Rh-L) 1, ∠(O-C-X), r(Rh-Rh) 2, r(Rh-L) 2, ∠(Rh-Rh-L) 2, ∠(C-C-C), ΔE(coord), He ₈ Ring Interaction, HOMO, 1, LUMO, 1, ΔE(FMO), Q Rh, 1, Q(L, mean), 1	0.1373	86.3
16	r(Rh-Rh) 1, r(Rh-L) 1, ∠(Rh-Rh-L) 1, ∠(O-C-X), r(Rh-Rh) 2, r(Rh-L) 2, ∠(Rh-Rh-L) 2, r(Rh-C), ∠(C-C-C), ΔE(coord), He ₈ Ring Interaction, HOMO, 1, LUMO, 1, ΔE(FMO), Q Rh, 1, Q(L, mean), 1	0.1537	84.6
17	r(Rh-Rh) 1, ∠(Rh-L) 1, ∠(Rh-Rh-L) 1, ∠(O-C-X), r(Rh-Rh) 2, r(Rh-L) 2, ∠(Rh-Rh-L) 2, r(Rh-C), ∠(C-C-C), ΔE(coord), He ₈ Ring Interaction, HOMO, 1, LUMO, 1, ΔE(FMO), Q Rh, 1, Q Ligand Donor Atoms, Q(L, mean), 1	0.1764	82.4
18	r(Rh-Rh) 1, r(Rh-L) 1, ∠(Rh-Rh-L) 1, ∠(O-C-X), r(αC-R ¹), r(Rh-Rh) 2, r(Rh-L) 2, ∠(Rh-Rh-L) 2, r(Rh-C), ∠(C-C-C), ΔE(coord), He ₈ Ring Interaction, HOMO, 1, LUMO, 1, ΔE(FMO), Q Rh, 1, Q Ligand Donor Atoms, Q(L, mean), 1	0.203	79.7
19	All available descriptors	0.2328	76.7

As more descriptors were added to the PCA model, the percentage of variance captured decreased, and the mean squared error of projection increased. This was because the solutions with fewer descriptors have less information to capture and can effectively be represented with only three principal components. While a PCA model that captures 100% of the available information appears useful from a metric point of view, it fails to consider the quality of the clustering in the model or the relevance of the captured information. Therefore, the PCA plots were evaluated for their individual interpretability in the context of rhodium(II) chemistry and an optimal model was selected. Analysis of the first two principal components (PC1 and PC2) (Figure 2.11) shows that for all combinations of n descriptors catalysts were separated by their ligand donor atoms along PC1. Further separation of catalyst sub-types begins along PC1 once 11 descriptors were utilised, highlighting complexes **4y**, **4z** and **4aa** (acyclic *N,O* ligands). PC2 clusters the most electron-withdrawing carboxylate ligands **3e**, **3f**, **3h** and **3i** immediately, but does not delineate complexes **3g**, **3o** or **3p** from the bulk carboxylate cluster again until 11 descriptors were used.

Analysis of the first and third principal components (PC1 and PC3) (Figure 2.12) shows that full separation of complexes by ligand donor atoms did not fully occur until 6 descriptors were used, allowing the formation of discrete clusters with three principal components. Complexes **4y**, **4z** and **4aa** were fully separated from the bulk carboxamidate cluster with 11 descriptors, and the 13 descriptor model highlighted carboxamidate complexes with sterically demanding substituents (**4d**, **4g**, **4m** and **4s**) in a new cluster with PC3. For both series of plots (PC1/PC2 and PC1/PC3), the clustering boundaries become less clear when all the available descriptors were added to the PCA model, which was consistent with lower amounts of captured variance and lower descriptor loadings. While the PCA models with greater numbers of descriptors are not bad, models using between 11 and 15 descriptors have the most well-defined clusters and are, therefore, easiest to interpret.

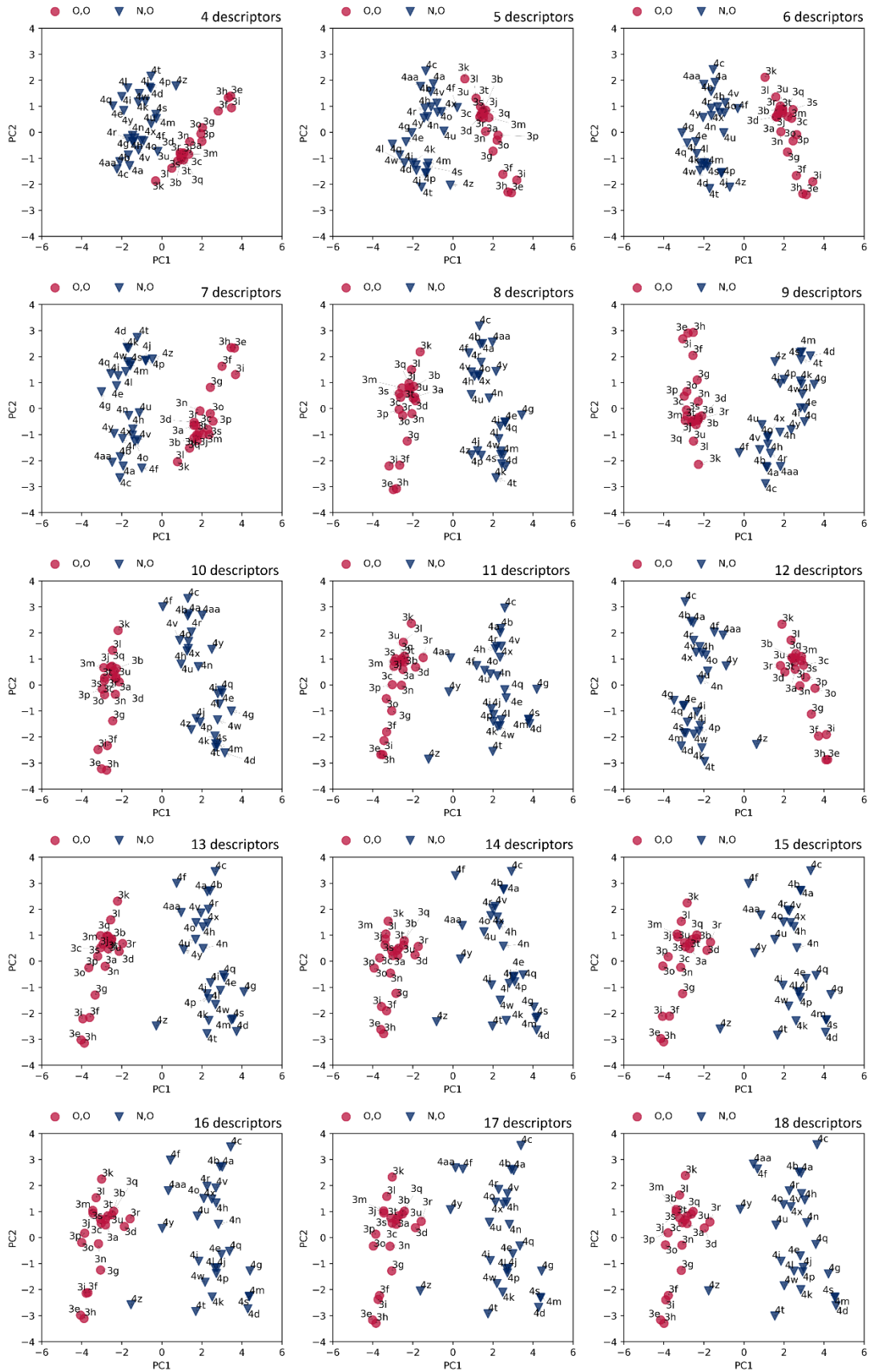


Figure 2.11. PC1/PC2 plots for the highest performing solutions with n descriptors.

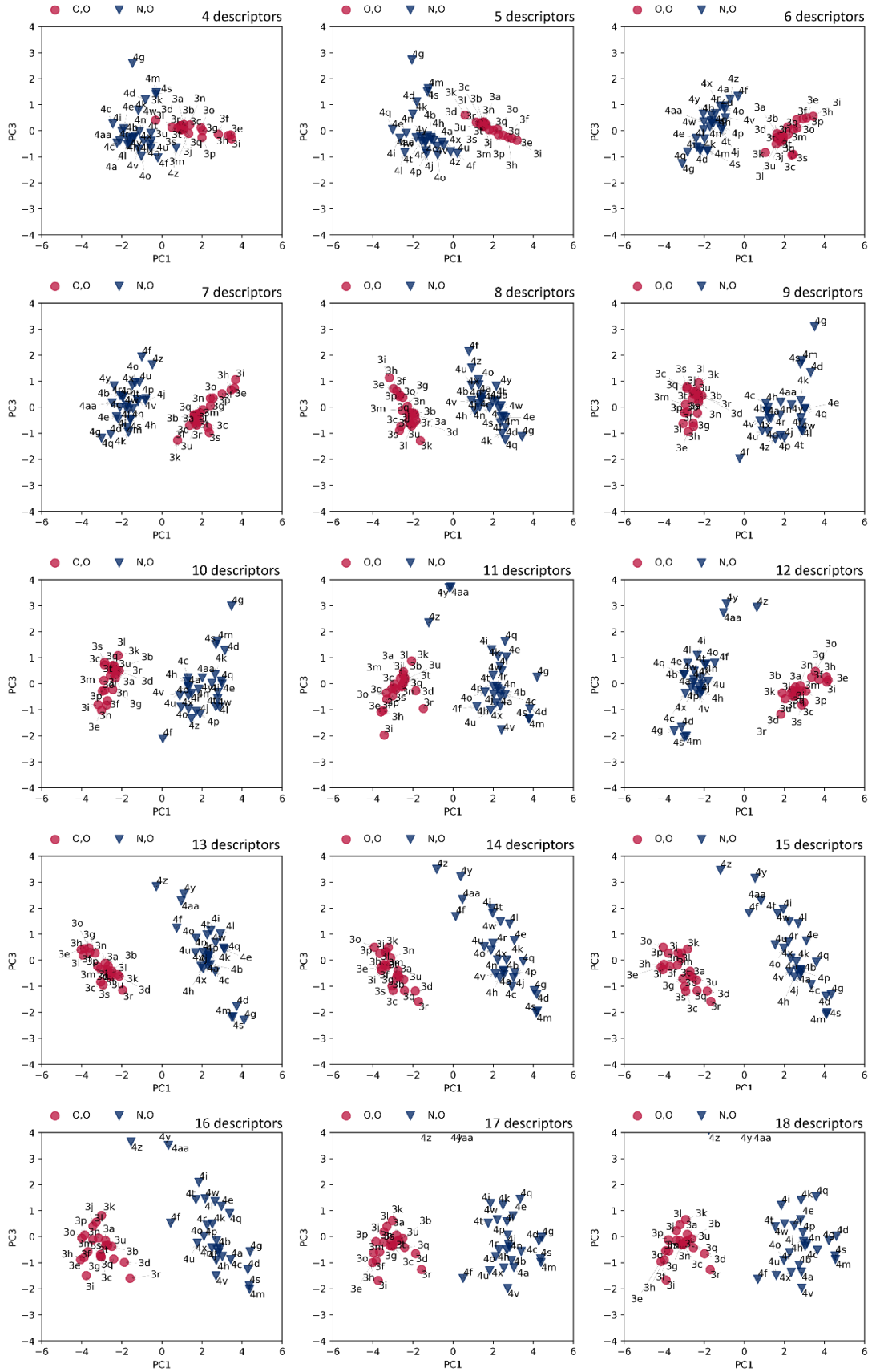


Figure 2.12. PC1/PC2 plots for the highest performing solutions with n descriptors.

For solutions using 11 to 15 descriptors most of the main catalyst groups were captured by discrete clusters (O,O vs. N,O, EDG-substituents, EWG-substituents, acyclic N,O ligands and sterically demanding N,O ligands). However, the sterically demanding carboxylate complexes were not represented by a distinct cluster in any of these models. An extra descriptor, $|wV|$, was added to the 13 descriptor PCA model to create a new 14 descriptor solution (Figure 2.13) that formed a new cluster for the bulky O,O ligands (3d, 3r, 3s, 3t and 3u). Comparative models with 11, 12, 13, 14 and 15 descriptors were also generated, but the 14-descriptor model gave the best clustering performance and was chosen as the final model.

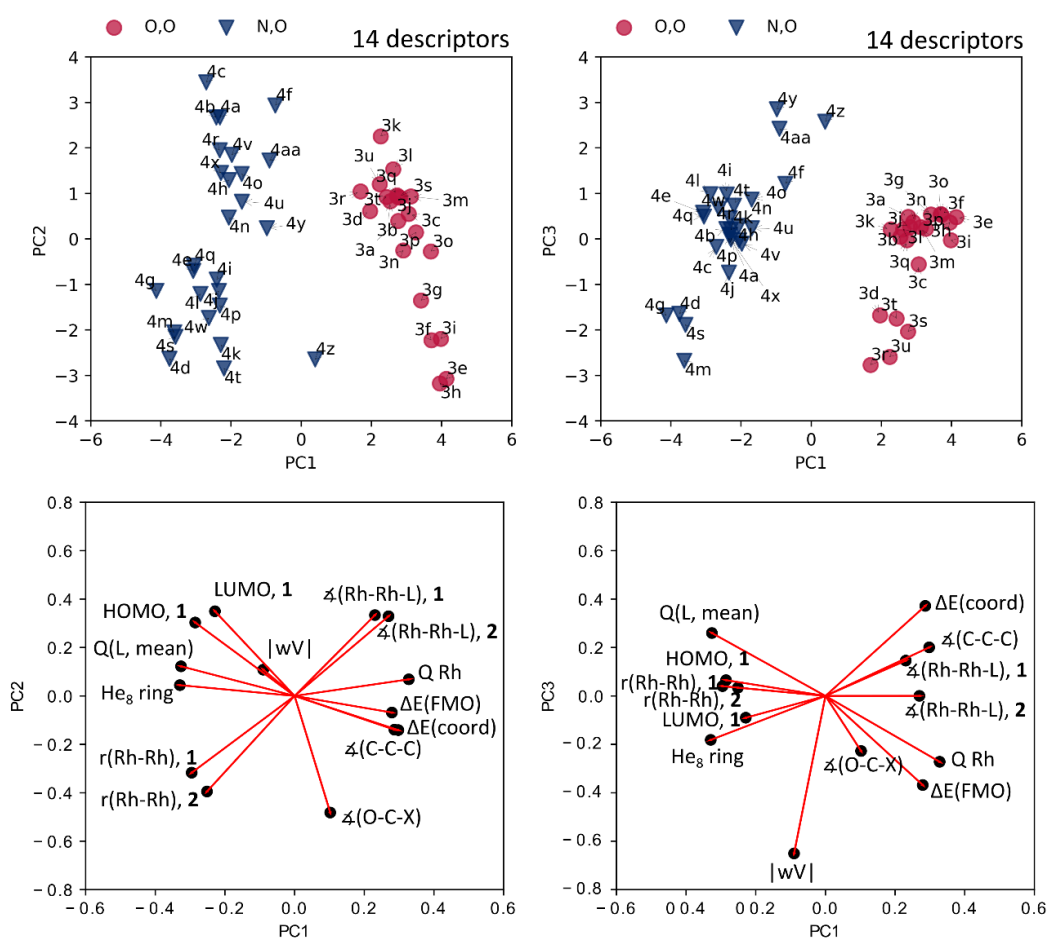


Figure 2.13. Optimal solution for the PCA of Rhodium(II) catalysts capturing 85.5% total variance. PC1/PC2/PC3 explained variance: 53.6, 21.7 and 10.5% respectively. Mean squared error loss from projection: 0.142. The corresponding descriptor loading plots for the PC1/PC2 and PC1/PC3 PCA plots are also shown.

The final PCA model gives high descriptor loading performance implying that there was minimal bias towards a specific subset of ligand features (Table 2.3). Analysis of the eigenvectors for the 14 descriptors demonstrates that each has unique contributions to the observed clustering. The separation of complexes along PC1 by ligand donor atom type was driven by the electronic descriptors Q Rh and $\Delta E(\text{FMO})$, and the combined electronic and steric descriptor $\Delta E(\text{coord})$ for carboxylate complexes. Carboxamidate clustering along PC1 was driven by Q(L, mean) and the steric descriptor He_8 highlighting the larger ligand backbones. The increased fluorination of the carboxylate ligands, leading to decreased PC2 values, was represented by a mixture of electronic and steric descriptors indicating that PC2 captures the Lewis acidity of the complexes. PC2 also captures the ring size of the carboxamidate ligands with 5- and 6/7-membered ligands separated into distinct clusters. PC3 captures predominantly steric information, creating new clusters for carboxylate and carboxamidate complexes with sterically demanding ligand substituents (**3d** and **3r-3u**, and **4d**, **4g**, **4m** and **4s**) and a separate cluster for acyclic carboxamidates (**4y-4aa**).

Table 2.3. Eigenvectors for PC1/2/3 capturing 53.6, 21.7 and 10.5% variance respectively.

Descriptor	Eigenvector PC1	Eigenvector PC2	Eigenvector PC3
% contribution to variance	53.6	21.7	10.5
r(Rh-Rh), 1	-0.2955	-0.3171	0.0404
χ (Rh-L), 1	0.2306	0.3339	0.1460
χ (C=O- α C-X), 1	0.1023	-0.4817	-0.2274
r(Rh-Rh) 2	-0.2516	-0.3947	0.0341
χ (Rh-L) 2	0.2697	0.3292	0.0003
χ (N-C-N)	0.2978	-0.1411	0.2007
$\Delta E(\text{coord})$	0.2866	-0.1394	0.3720
He_8 Ring Interaction	-0.3294	0.0445	-0.1815
HOMO, 1	-0.2854	0.3033	0.2599
LUMO, 1	-0.2287	0.3498	0.0652
$\Delta E(\text{FMO})$	0.2793	-0.0685	-0.0900
Q Rh, 1	0.3278	0.0686	-0.3680
Q(L, mean), 1	-0.3260	0.1223	-0.2720
wV	-0.0901	0.1084	-0.6506

Analysis of the PCA scores and loadings plots demonstrates that many of the descriptors capture information on only a few features of each catalyst subset, creating an underlying trend that drives the distinct clustering observed. For example, the large, aliphatic carboxylate ligands were separated from the largest cluster of carboxylate ligands by greater |wV| and

He₈ loadings values on PC3 (Figure **2.13**). Complexes bearing carboxylate ligands were separated from carboxamidate ligands by the descriptor $r(\text{Rh-Rh})$, and 5- and 6/7- membered cyclic *N,O* ligands were separated by $\chi(\text{Rh-Rh-L})$ and He₈. These contributions illustrate the usefulness of PCA to highlight complex chemical relationships, from appropriate chemical descriptors, where many complexes have remarkably similar values for many other parameters.

A separate PCA decomposition was also conducted for the *O,O* and *N,O* sub-classes of rhodium(II) complexes to investigate the variation in catalyst properties for each ligand donor atom type. PCA of the carboxylate subclass (Figure **2.14**) separates complexes by electronic properties along PC1, with complexes bound to electron-withdrawing ligands (**3e** – **3i**) between PC1: 4-6, and by steric properties along PC2 with the complexes bound to sterically demanding ligands (**3d**, **3r** – **3u**) between PC2: 1-4. These trends were similar to the full PCA model but with greater distance between each complex highlighting the large range of ligand properties within the carboxylate cluster.

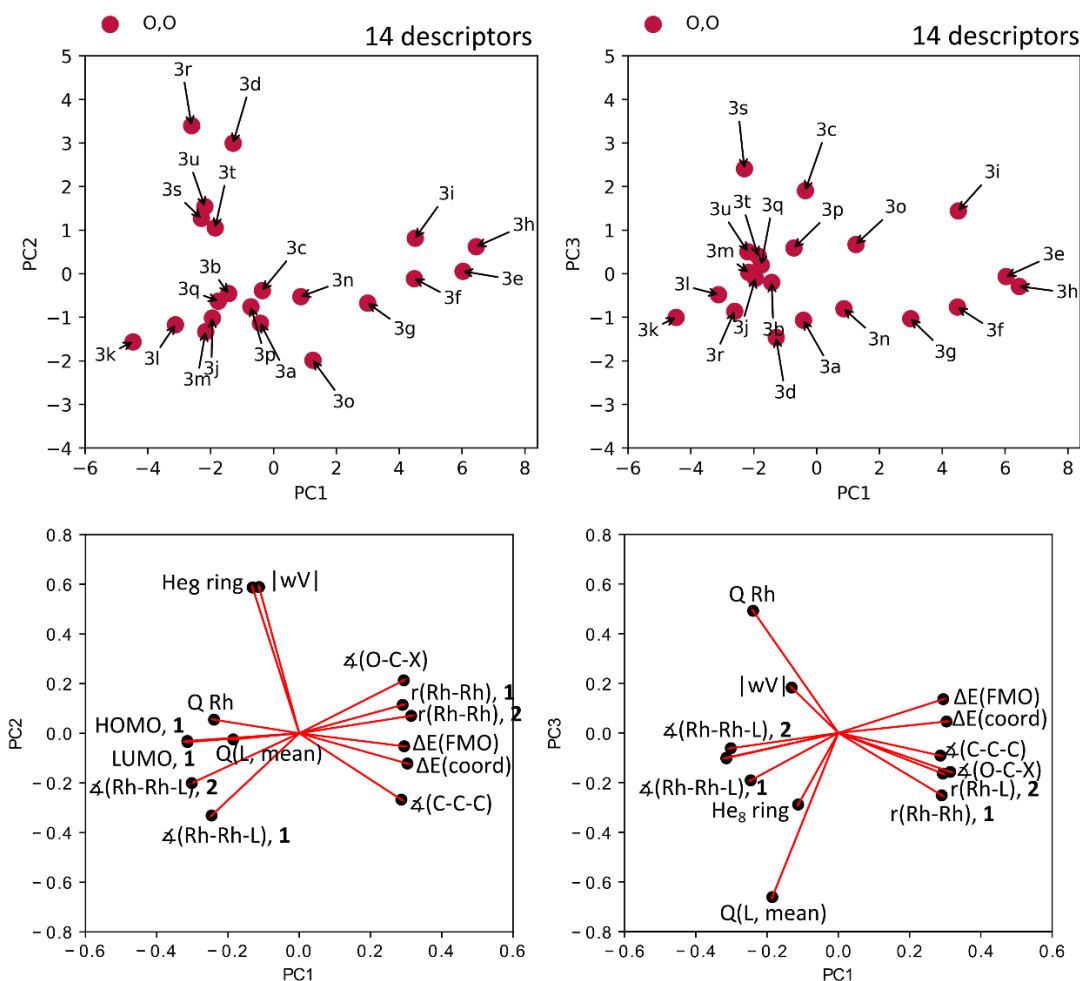


Figure 2.14. PCA plot of O,O dirhodium(II) catalysts capturing 87.4% total variance. PC1/PC2/PC3 explained variance: 66.6, 13.8 and 7.0% respectively. Mean squared error loss from projection: 0.126.

PCA of the carboxamidate sub-class (Figure 2.15) separates complexes based on the cyclic backbone ring size (5 or 6/7-membered rings) and electronic properties along PC1. For example, **4t**, a 5-membered lactam ring substituted with fluorine atoms, was at the peak of the right-hand cluster (PC1: 3) and the urea **4q**, an unsubstituted 5-membered lactam ring, was on the opposing edge (PC1: 0). This trend was similar in the left-hand cluster with **4u**, a 6-membered lactam ring substituted with fluorine atoms, located at PC1: -1 and the catalyst on the opposing edge was the unsubstituted 6-membered lactam **4c**. PC2 separates the *N,O* complexes by steric factors with the acyclic ligands located at the top of the plot (PC2: 2-6) and the sterically demanding complexes at the bottom (PC2: -2 - -4). Analysis of the carboxamidate cluster individually gives greater resolution of electronic ligand effects, when compared to the full model, due to the lower density of

clustering, and shows that the Lewis acidity of carboxamidate complexes was also captured by the model.

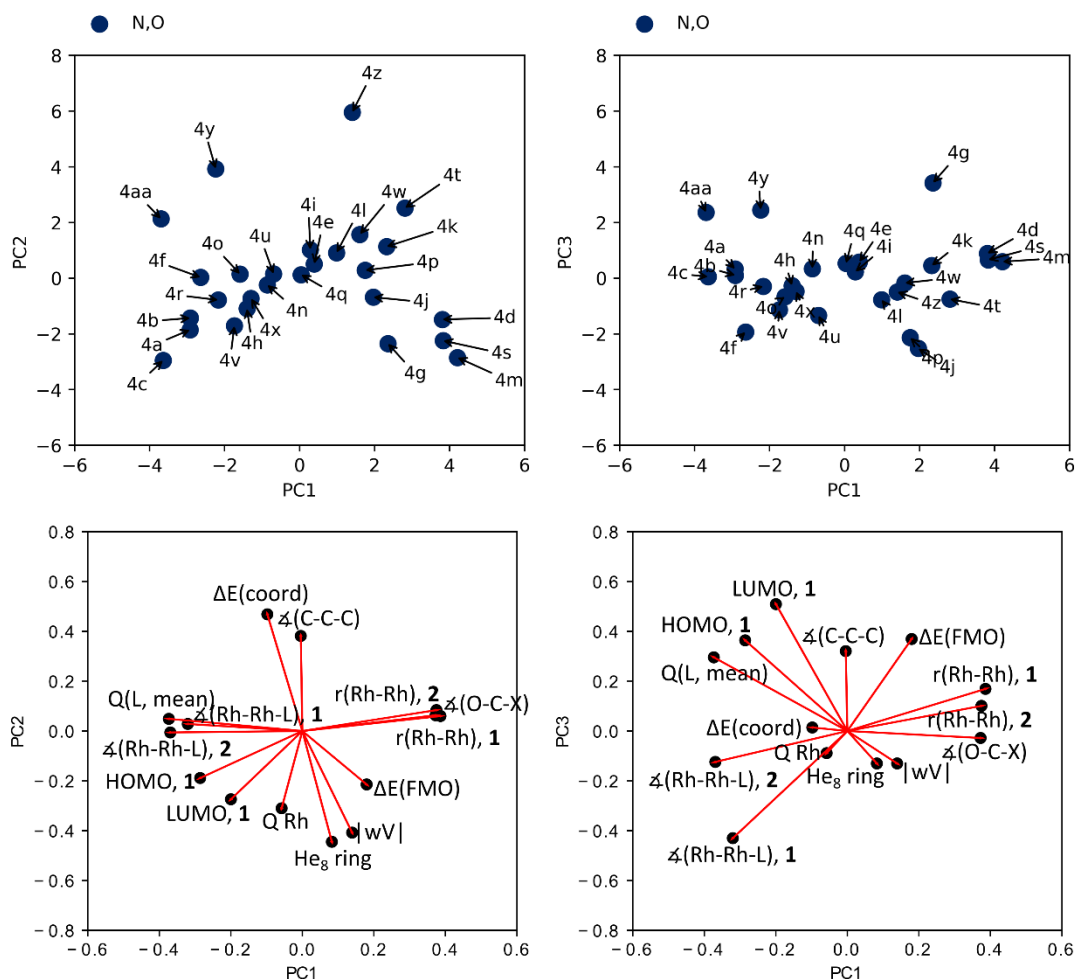
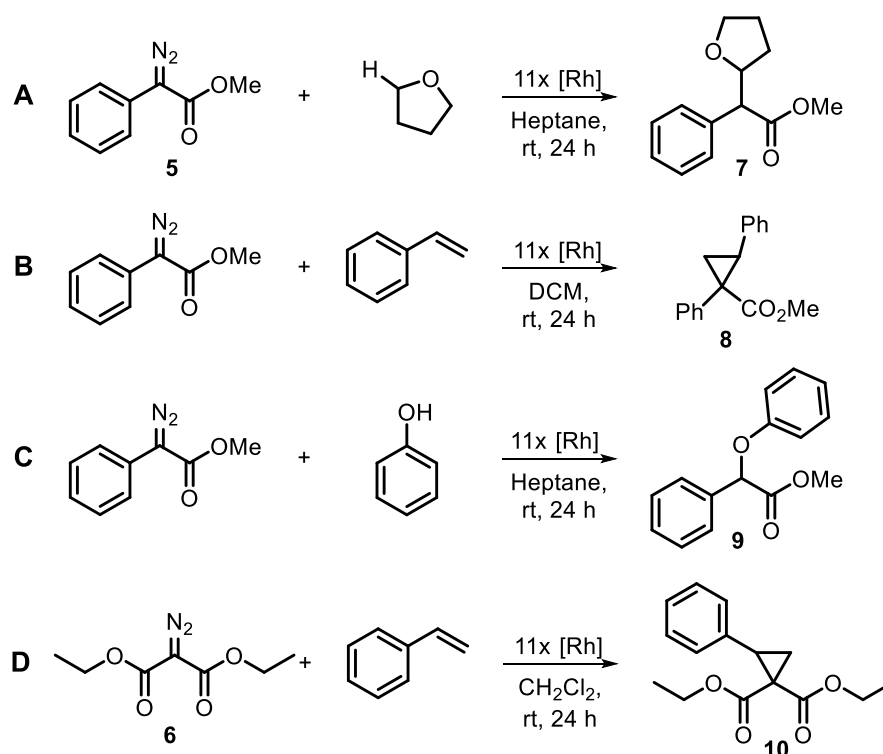


Figure 2.15. PCA plot of *N,O* dirhodium(II) catalysts capturing 80.3% total variance. PC1/PC2/PC3 explained variance: 39.5, 28.7 and 12.1% respectively. Mean squared error loss from projection: 0.196.

In summary, the PCA model generated from the catalyst knowledge base offers an interpretable representation of the chemical space covered by a broad range of rhodium(II) complexes. Key ligand features drive clustering and give a map that delineates catalyst properties into a continuum of effects, such as steric hinderance and Lewis acidity. Principal component analysis of the carboxylate and carboxamidate complexes separately showed that there is considerable diversity within each class and that the final model captures a huge range of catalyst diversity.

2.5 Correlation of the Catalyst Map with Experimental Outcomes

Creating computational models that interpret experimental data can provide useful information for the design of experiments and could promote a broader adoption of metal-catalysed reactions into medicinal chemistry workflows.^{73,114,159,167,184} Four different model reactions, representative of rhodium(II)-catalysed transformations, were investigated experimentally to test whether the principal components derived from the knowledgebase correlate with observed reactivity. A C–H insertion involving tetrahydrofuran (**7**),¹⁹⁷ a cyclopropanation of styrene (**8**),^{217,218} and an O–H insertion involving phenol (**9**)²¹⁹ using a donor/acceptor diazo ester (**5**), and a cyclopropanation of styrene (**10**) using an acceptor/acceptor diazo ester (**6**) (Scheme 2.6). Catalysts were selected according to their actual availability, and structural and electronic diversity where possible.



Scheme 2.6. The four reactions selected for comparison with the PCA model.

The reactions were performed in a high-throughput, micro-scale, format in 96-position plates equipped with 700 μL glass vials and stirring fleas. Each vial contained 20 μmol of a diazo compound (**5** or **6**), 200 μmol substrate

(THF, styrene or phenol) and 1 mol% catalyst. After 24 hours, the product mixtures were analysed by quantitative UPLC (Table 2.4).

Table 2.4. Observed yields and standard deviations for the screened catalysts in all four transformations.

Catalyst	C-H Insertion, 7	Cyclopropanation, 8		O-H Insertion, 9	Cyclopropanation, 10
	Yield %	Yield %	dr ratio ^a	Yield %	Yield %
Rh ₂ OAc ₄ , 3a	44 ± 5	38 ± 4	98:2	30 ± 5 ^d	22 ± 2
Rh ₂ PIV ₄ , 3b	51 ± 11	38 ± 5	96:4	38 ± 1 ^d	41 ± 2
Rh ₂ Oct ₄ , 3c	51 ± 1 (52 ^c)	49 ± 1	98:2	36 ± 9 ^d	35 ± 2
Rh ₂ tfa ₄ , 3e	45 ± 3	80 ± 2 (70 ^c)	95:5	42 ± 10 ^d	36 ± 7
Rh ₂ pf ₄ , 3h	44 ± 2	75 ± 4	95:5	49 ± 3 ^d (57 ^c)	38 ± 1
Rh ₂ esp ₂ , 3q	50 ± 4	61 ± 6	96:4	39 ± 1 ^d	39 ± 4
Rh ₂ S-DOSP ₄ , 3r	53 ± 7	56 ± 8	97:3	27 ± 12 ^d	32 ± 2
Rh ₂ S-tertP ₄ , 3u	56 ± 12	53 ± 7	97:3	40 ± 11 ^d	42 ± 4
Rh ₂ S-PTAD ₄ , 3s	43 ± 13	61 ± 4 (48 ^c)	99:1	40 ± 13 ^d	45 ± 2
Rh ₂ cap ₄ , 4a	19 ± 5	20 ± 14	96:4	10 ± 4 ^d	11 ± 7
Rh ₂ 5R-MEPY ₄ , 4g	19 ± 5 ^d	22 ± 8	96:4	34 ± 10 ^d	15 ± 2

Outcome of the four model reactions, including standard deviations for replicates ($n = 3$) completed on different days from fresh stock solutions. ^a Yield determined by UPLC with an external standard. ^b Diastereomer ratio determined by UPLC. ^c Isolated yield of purified product using 100 mg of the diazo **5** or **6** as substrate. ^d Yield obtained in duplicate.

The high-throughput UPLC screening gave a range of yields for **7**, **8**, **9**, and **10** with carboxylate catalysts generally giving higher yields than carboxamidate catalysts. The C–H insertion gave a moderate range of yields from 18-56% for product **7**. Intermolecular C–H insertion is known to be dependent on the rate of diazo addition to the reaction mixture, which cannot be controlled while using multi-channel liquid dispensing and appears to be limiting the observed yield for the best performing catalysts.¹⁹⁷ O–H insertion gave a similar, but narrower, range of yields (10-49%) for product **9**. Phenols are also documented to be poor substrates for rhodium(II) catalysts, and the narrow range of yields was expected.²²⁰ Cyclopropanation of the acceptor/acceptor diazo ester **10** also had a narrower range of yields (10-45%) due to the effect of using dichloromethane as a solvent (for full details, see Chapter 5, section 5.2.1). The cyclopropanation of styrene gave the largest range of yields for product **8**, from 20-80%, and **3e** and **3h** stood out as the best catalysts for this reaction. Several selected reactions were also repeated on a 50-fold larger scale in conventional labware to assess

reproducibility, demonstrating that reaction optimisation on the microscale was translatable to reactions on preparative scale.

Many of the descriptors gave moderate positive or negative correlations with all three reactions (Figure 2.16). Q Rh, an electronic descriptor, had the strongest positive correlation with the yield of **7** ($R = 0.89$) indicating electronic properties of the catalyst were important. He₈ also had a strong negative correlation ($R = -0.89$) for the yield of **7**, further highlighting the importance of the steric properties of the ligand backbone. $\Delta E(\text{coord})$ gave the best correlation with the observed yield for products **8** ($R = 0.85$) and **9** ($R = 0.83$) demonstrating that the calculated descriptors captured the Lewis acidity of the catalyst. Q Rh also gave a strong positive correlation with the observed yield of **10** ($R = 0.9$), along with a strong negative correlation with He₈ ($R = -0.84$), indicating that electronic and steric descriptors were relevant for interpreting reaction outcomes with an acceptor/acceptor diazo.

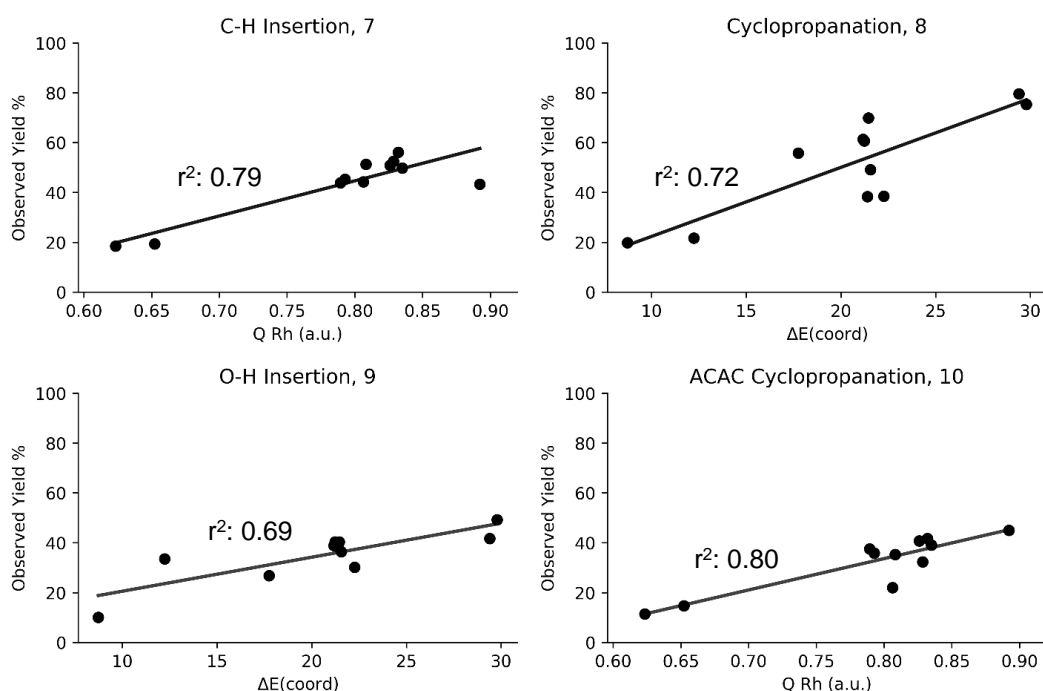


Figure 2.16. Best descriptors for univariate correlation with experimental screening data. Carbene binding energy is given in kcal mol⁻¹.

The yield of the cyclopropanation reaction between **5** and styrene varied widely as a function of the catalyst, indicating there was a catalyst effect. Mapping the observed yields for the cyclopropanation onto the PCA model showed two clusters of higher yield: one cluster populated by catalysts with fluorinated ligands (**3e** and **3h**); and the other by catalysts with sterically demanding ligands (**3r**, **3s** and **3u**) (Figure 2.17). Based on this clustering

PC2, representing the Lewis acidity of the catalyst, and PC3, representing steric intrusion into the reactive site, can represent catalyst performance for this reaction. For the catalysts with carboxylate ligands, the yield of the cyclopropanation product **8** and the principal components were correlated (PC1: $R = 0.81$ and $r^2 = 0.66$, and PC2: $R = -0.48$ and $r^2 = 0.23$). Catalysts with higher PC1 values and lower PC2 values gave higher yields in this cyclopropanation reaction indicating that the PCA model captures relevant information for further model building for the prediction of reaction performance. The remaining reactions, C–H and O–H insertion (products **7** and **9**), and acceptor/acceptor diazo cyclopropanation (product **10**), are also responsive to the PCA model demonstrating that the map is useful for interpreting the results for different types of reactions (for overlay plots, see Chapter 5 section 5.3.9)

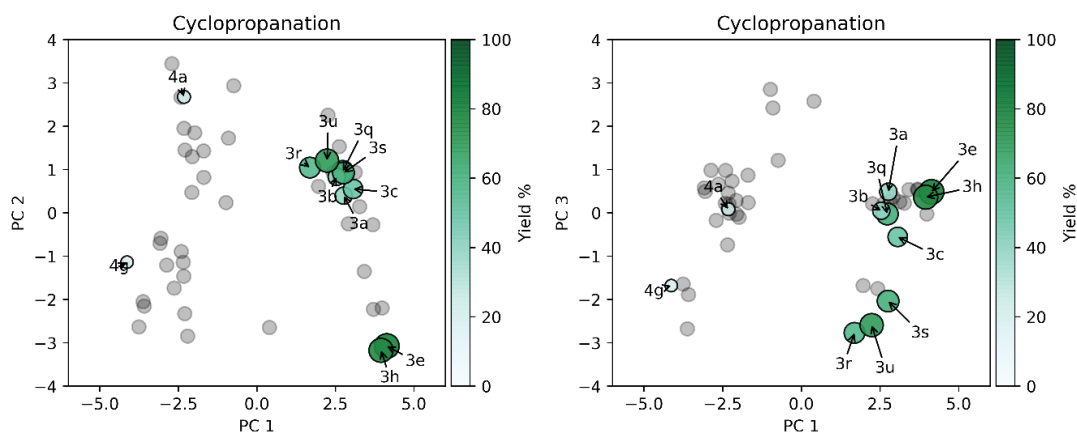


Figure 2.17. An overlay of the outcome of a cyclopropanation reaction with the PCA map.

Rhodium(II) catalysts can also give selectivity between two competitive reaction pathways under identical conditions.¹⁹¹ The reported selectivity for a rhodium(II)-catalysed intramolecular cyclisation was shown to be highly dependent on the specific catalyst used (Figure 2.18). Catalysts with carboxylate ligands were selective for cycloheptatriene formation, and catalysts with carboxamidate ligands were selective for γ -lactam formation. Principal component 1 captures the features responsible for this selectivity as catalysts with positive PC1 values were selective for cycloheptatriene formation and those with negative PC1 values were selective for the γ -lactam. The descriptor loadings plot (Figure 2.13) indicates catalysts with positive PC1 values had been clustered due to the covariance of PC1 with the descriptors Q Rh, $\Delta E(\text{FMO})$, $\Delta E(\text{coord})$ and $\angle(\text{C-C-C})$, whereas catalysts

with negative PC1 values had been clustered due a more complicated subset of ligand features. Complexes **4a** and **4y** have negative PC1 and positive PC2 values due to the covariance of both PC1 and PC2 with the descriptors Q(L, mean), HOMO, **1** and LUMO, **1** suggesting the electronic effects are predominantly responsible for the differences in selectivity between catalyst classes.

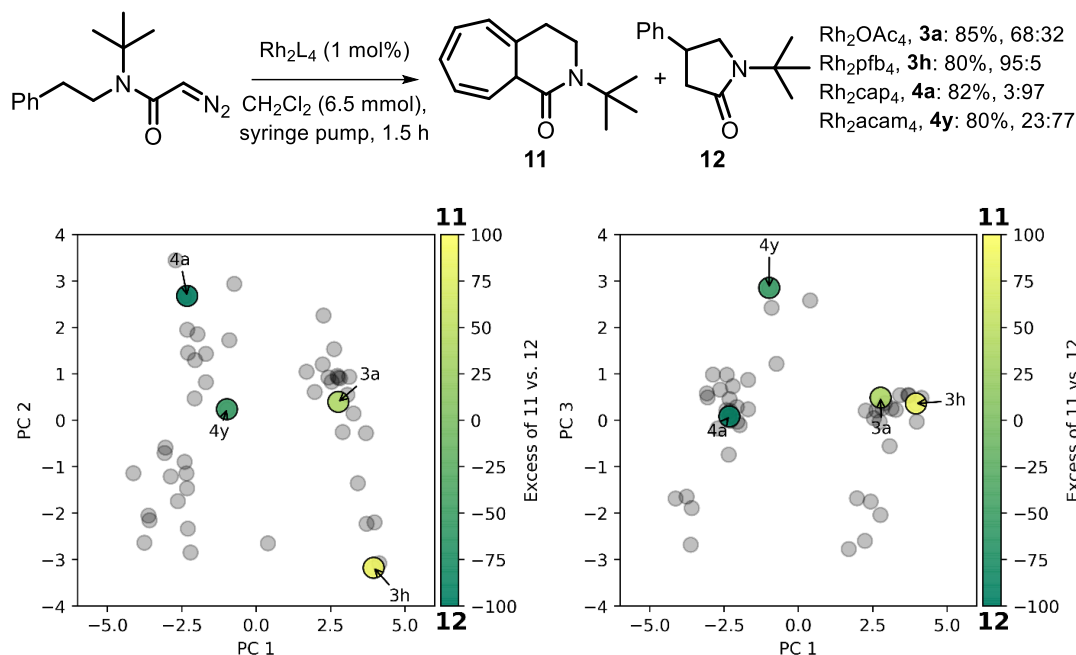


Figure 2.18. The effect of catalyst on reaction outcome for a reaction with two precedented pathways.

Rhodium(II) catalysts had previously been used in the activity-directed discovery of androgen receptor agonists by harnessing intramolecular cyclisation reactions of α -diazo amides.⁸² Over three rounds of ADS, four reaction solvents (CH_2Cl_2 , THF, Toluene and THF) and ten different rhodium(II) catalysts were used: **3a**, **3r**, and **4g** in round one; **3a**, **3e**, **4a**, **4d**, and **4m** in round two; and **3a**, **3c**, **3e**, **3h**, **3q**, and Rh_2tpa_4 in round three. In round one catalyst **3a** afforded the most active reaction mixture with diazo **1.3** and was used to guide the design of the round two reaction array. Round two expanded the coverage of catalyst space by including more catalysts with carboxamidate ligands (**4a**, **4d** and **4m**) and adding an electron-deficient carboxylate ligand (**3e**). Catalysts **3a** and **4a** gave the most active reaction mixtures with diazo **1.3**, but only **3a** and **3e** were carried forward into round three. An overlay of percentage activity onto the PCA map shows that the observed activity of crude reaction mixtures was clustered based on the location of the catalyst (Figure 2.19).

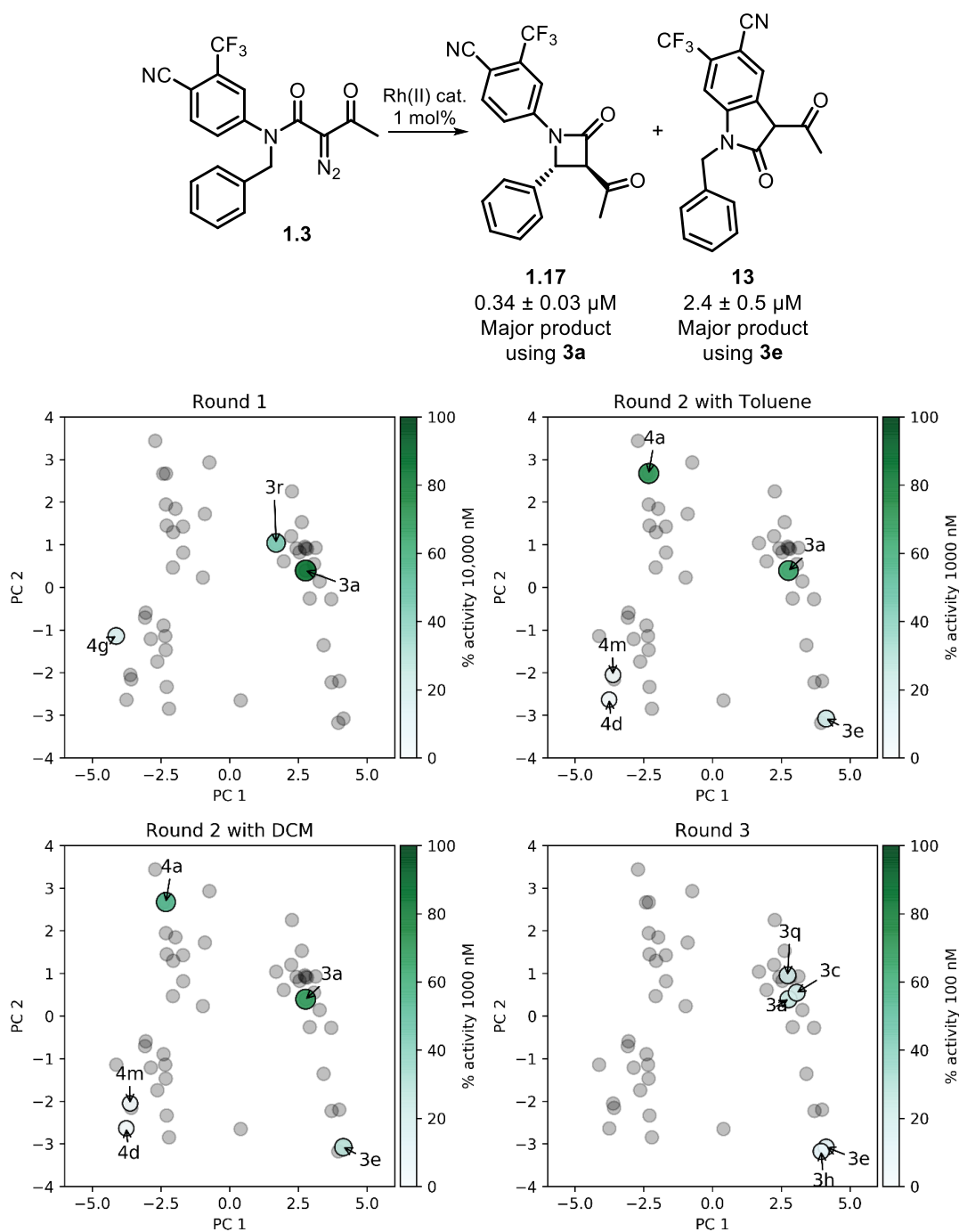


Figure 2.19. The relationship between the catalyst space captured by the PCA map and the activity of the crude reaction mixtures for rounds one, two, and three of ADS using intramolecular reactions.⁸² Screening concentrations decreased with each round (10, 1, and 0.1 μM respectively).

Catalysts **3a**, **3c** and **3q** produced the best agonist **1.17** in similar yields (68-75%) but with different reaction solvents (EtOAc for **3a**, and CH_2Cl_2 for **3c** and **3q**) in round three. These catalysts were clustered by both PC1 and PC2 on the PCA map demonstrating that selective formation of the most

active product was controlled by catalyst properties. Catalyst **3e** gave a different major product **13** when reacted with the same diazo providing further evidence that the map was responsive to catalyst properties. Unfortunately, none of the active reaction mixtures using a catalyst with a carboxamidate ligand were scaled up, but were highlighted as potentially interesting by the PCA map. The round two crude reaction mixture containing carboxamidate catalyst **4a** showed particularly promising activity, in both toluene and CH_2Cl_2 , but was discarded in favour of catalysts with carboxylate ligands. The PCA map could, therefore, serve as a tool for the selection of diverse catalysts in an initial reaction array, and as a tool for the selection of related catalysts when peaks of activity are found.

Overall, the PCA map built from the rhodium(II) knowledge base was responsive to different types of reactions, demonstrating that the parameterisation of the catalysts was independent of mechanism. Catalysts that have been clustered together have similar reaction performance and the map could be used to design experiments using rhodium(II) catalysts. The map shows that the diversity of readily available catalysts with carboxylate ligands was relatively poor (Figure **2.13**). This was further highlighted by the modest variation in the yields for the products **7**, **9**, and **10**. However, the available chemical space was large and could be better utilised (Figure **2.14**), and the catalyst map acts a useful tool for selecting new catalysts that expand this space.

2.6 Conclusions

The development of tools for reaction design has shaped the classes of transformations routinely used by process chemists in pharmaceutical companies.^{115,161,221} New tools could also enable a wider incorporation of new synthetic methodologies by medicinal chemists into molecular discovery workflows.^{159,168,174,175} The catalyst map could serve as a useful tool for the selection of rhodium(II) catalysts for specific transformations, such as late-stage C–H functionalisation. The catalyst map gives a reliable, and general, interpretation of reactive space for rhodium(II)-carbenes and could enable the identification of fertile regions of largely unexplored catalyst space. The range of commercially available rhodium(II) complexes was relatively poor and the catalyst map could serve as a guide for creating screening libraries comparable to those for phosphorous ligands, such as the Bristol LKB. Application of the catalyst map in drug discovery could also increase the

diversity of the molecules synthesised by medicinal chemists, because many of the bonds formed by rhodium(II)-catalysed reactions are underrepresented in the scaffolds of small molecule ligands.

Further computational development of the catalyst map could include implementing more complicated steric descriptors, such as average steric occupancy,^{166,222} to better evaluate the influence of very large ligand substituents. This may allow for greater resolution of the interplay between electronic and steric effects for site selective C–H insertion reactions, and aid in the development of more predictive models. Synthesis of a diverse set of rhodium(II) complexes, chosen using the catalyst map, could also allow for the creation of a broad catalyst screening library for the optimisation of rhodium(II)-catalysed reactions. This screening library could also have considerable utility for future activity-directed synthesis experiments using rhodium(II) carbene chemistry.

Activity-directed synthesis experiments rely on the diverse reactivity and selectivity of rhodium(II)-carbenes to identify novel bioactive compounds. An ideal ADS experiment would use many functionally dissimilar catalysts to cover the largest range of possible reaction outcomes. The catalyst map could enable the selection of dissimilar catalysts that complement the range of potential transformations in a reaction array, and aid in the selection of similar catalyst in subsequent reaction arrays. Greater synergy between the catalysts, diazo substrates, and multi-functional co-substrates could also promote a larger number of successful reactions across an array, meaning a greater likelihood of identifying interesting bioactive molecules. As such, the catalyst map can form part of the workflow for array design and inform catalyst selection for optimising the compatibility between the designed diazo substrates, co-substrates and catalysts.

Chapter 3

Design and Implementation of Activity-Directed Synthesis Reaction Arrays 1 and 2

It was proposed that ADS could be used for the discovery of inhibitors of protein targets without an evolved small molecule binding site. The p53/MDM2 PPI represents an appropriate target for demonstrating the utility of ADS for small molecule discovery. Fragment-sized molecules have not previously been identified as inhibitors of the PPI during drug discovery campaigns, and FBDD approaches have also not been used. However, molecules that can interact with a pair of MDM2 sub-pockets, where one of the interactions is with the tryptophan MDM2 sub-pocket, can have detectable binding to MDM2.¹⁵⁸ It was proposed that ADS could be used to identify new scaffolds that link pairs of fragment-sized reactants together to give a range of structurally-diverse small-molecule antagonists of MDM2. A series of activity-directed experiments for the discovery of new inhibitors of the p53/MDM2 PPI was subsequently designed and executed to test this hypothesis.

An initial exhaustive microscale reaction array of diazo substrates and co-substrates was designed to consider a number of factors including: similarity to known MDM2 ligands; reactivity with rhodium(II) catalysts; and coverage of the available chemical space. The set of diazo substrates proposed for synthesis, and co-substrates selected for purchase, were chosen to consider the number of potential array products that would be capable of making interactions with two or three of the MDM2 sub-pockets (Figure 3.1). Combinations that do not obviously form potential products capable of interacting with two sub-pockets were also included to avoid biasing the outcome of the array towards products with higher similarity to known MDM2 ligands.

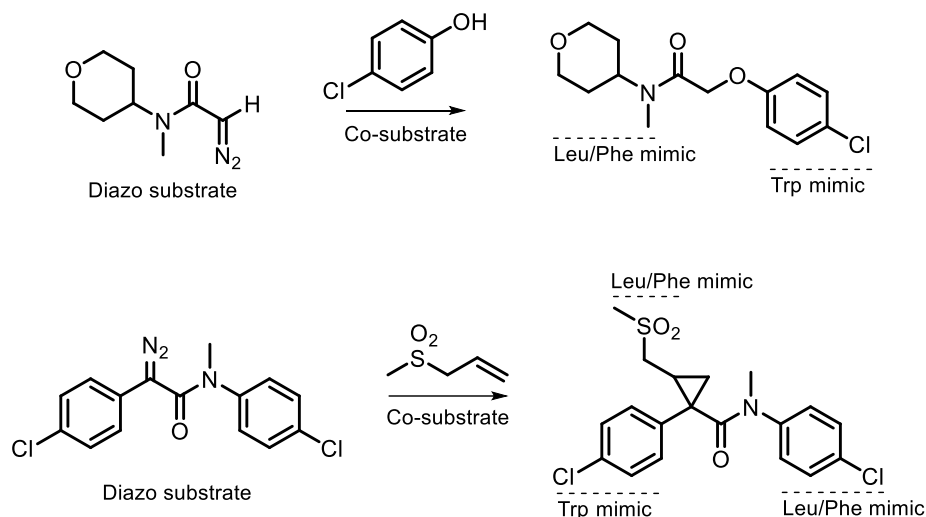


Figure 3.1. Combinations of diazo substrate and co-substrate from the initial activity-directed synthesis array showing hypothetical products containing two or three MDM2 hot-spot mimics.

3.1 Design of an Initial Reaction Array

Design of the first reaction array was aided using cheminformatic software (Pipeline Pilot and RDKit²²³) to filter chemical supplier databases and find relevant molecules that could then be selected manually to complete libraries of diazo substrates and co-substrates. The finalised array used all combinations involving seven diazo substrates (**D1** – **D7**), ten co-substrates (**S1** – **S10**) and two rhodium(II) catalysts, giving 154 microscale reactions (including in-line controls, described in section 3.3).

A database containing commercially available compounds from Sigma Aldrich, Fluorochem, Enamine and Alfa Aesar was filtered through two separate workflows in Pipeline Pilot to select amines to be used in diazo synthesis, and molecules to be used as co-substrates (Figure 3.2). The aim was to identify potential substrates and co-substrates with favourable molecular properties. Compounds with molecular weight lower than 300 Da, fewer than three rotatable bonds, and cLogP values between -2 and 5 were prioritised for inclusion. Both workflows filtered the database of compounds to remove undesirable molecules, such as compounds containing functional groups found in Pan-Assay Interference Compounds (PAINS).^{224,225} The first workflow was designed to identify primary and secondary amines that could be used as reactants in the diazotization methodologies outlined in section 3.1.1 to yield diazo substrates. The second workflow was implemented to

select co-substrates that contain functional groups known to react with rhodium(II)-carbenes. This was achieved by matching the database of commercially available compounds with a set of substructures to select potentially reactive molecules.

Potential amines to be used for diazo synthesis and co-substrate molecules identified by the Pipeline Pilot or RDKit workflows were then evaluated to select a library of compounds to synthesize or purchase (Figure **3.3** and **3.4** respectively). Amines were selected manually from the enumerated diazo library to create a library of diazo reactants with a wide range of functional groups capable of making interactions with the sub-pockets in the peptide binding cleft of MDM2, for example: the alkyl groups on **3.3**, **3.4**, **3.5** and **3.6** could interact with the leucine sub-pocket; the aromatic groups on **3.1**, **3.12** and **3.13** could interact with the phenylalanine sub-pocket; and the functional groups on **3.7** and **3.15** could interact with the tryptophan sub-pocket. Several of the designed substrates could also interact with two sub-pockets, for example **3.5**, **3.6** and **3.9** could interact with the leucine and tryptophan or the phenylalanine and tryptophan sub-pockets.

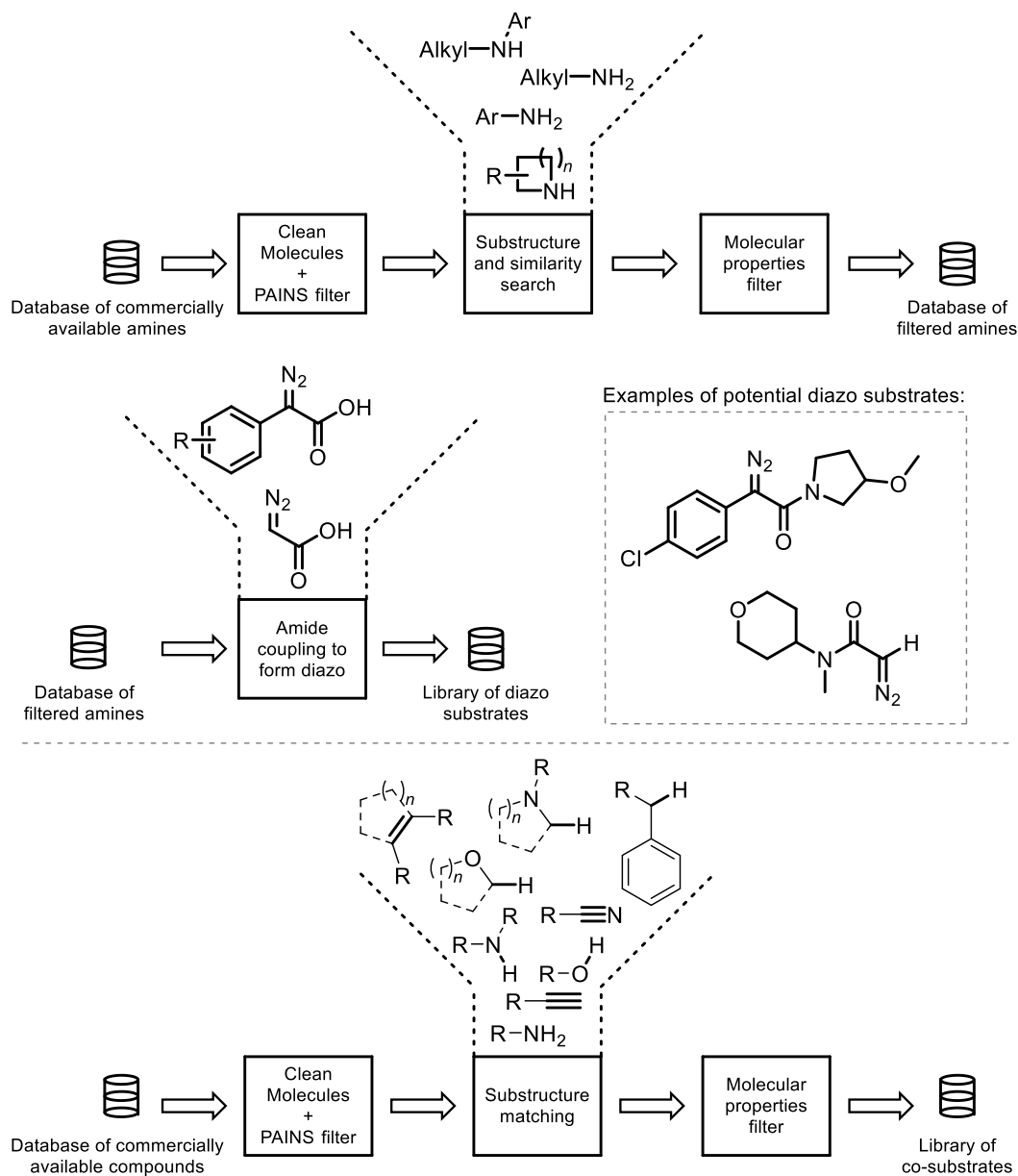


Figure 3.2. Workflows for the selection of amines for diazo synthesis and co-substrates. Molecular properties were calculated within Pipeline Pilot or RDKit, and molecular fingerprints were generated in pipeline pilot using the ECFP4 fingerprint algorithm. Top panel: Workflow for the enumeration of diazo substrates. Bottom panel: Workflow for the selection of co-substrates.

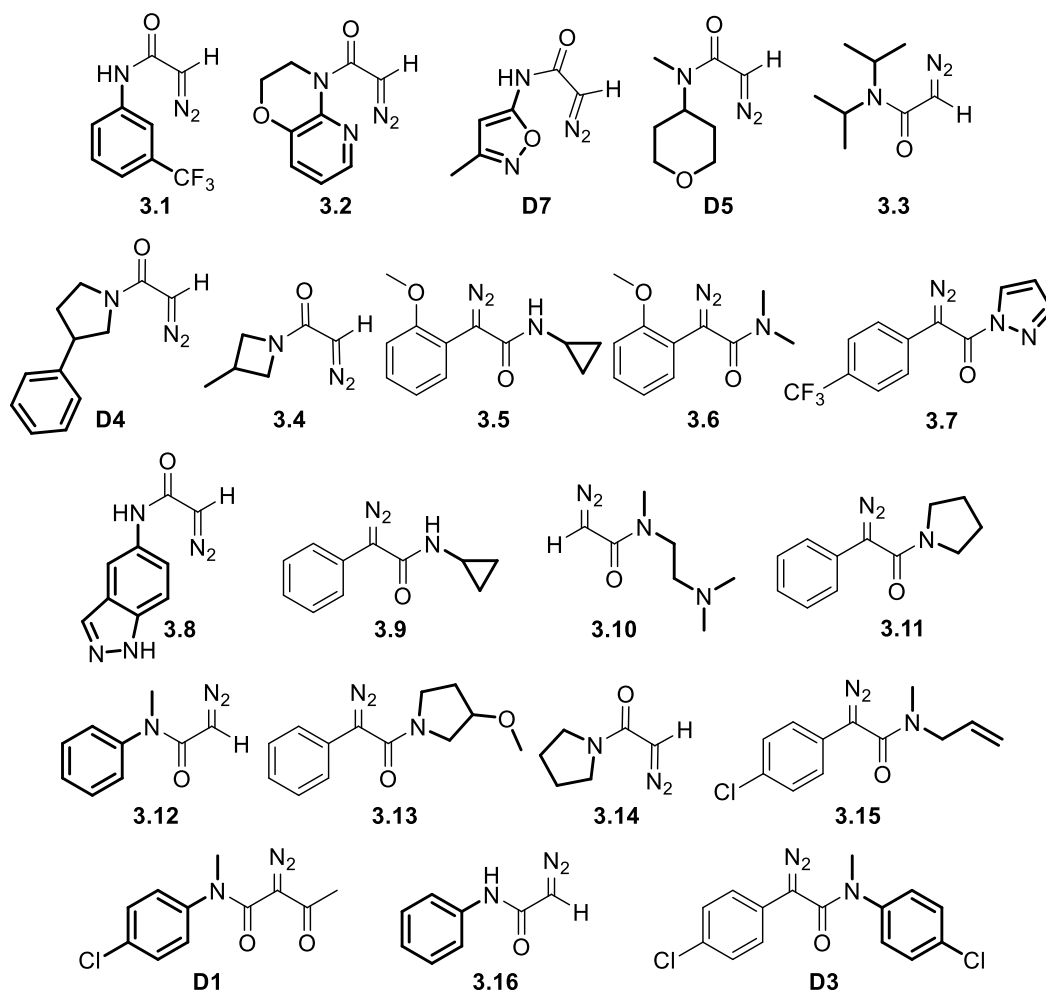


Figure 3.3. Diazo substrates selected for synthesis from the enumerated diazo library.

Co-substrates containing multiple potentially reactive sites with rhodium(II) carbenes were prioritised for selection in the library. For example, alkenes for cyclopropanation (**S4**, **S5**, **S9**, and **S10**), alcohols for O–H insertion (**S2** and **S8**), and C–H bonds next to α -nitrogen atoms or in benzylic positions (**S3**, **S4** and **S5**). Co-substrates that were used in the initial reaction array reaction array (**S1** – **S10**) contain compatible functional groups in contexts that are preceded by rhodium(II)-catalysed reactions, for example indole-C2 and -C3 C–H insertion (**S1**), and C–H insertion at positions adjacent to nitrogen occur most commonly when the nitrogen is part of an amide or carbamate (**S4** and **S5**).^{185,226,227} C–H insertion reactions α -to oxygen atoms or at allylic positions are significantly more dependent on the electronic and steric properties of the rhodium(II)-carbene species, and also the extent the desired reaction site is activated by the oxygen atom, so co-substrates were

not prioritised based on containing these centres despite several co-substrates matching these fingerprints (**S6**, **S8**, **S9** and **S10**).¹⁸⁵

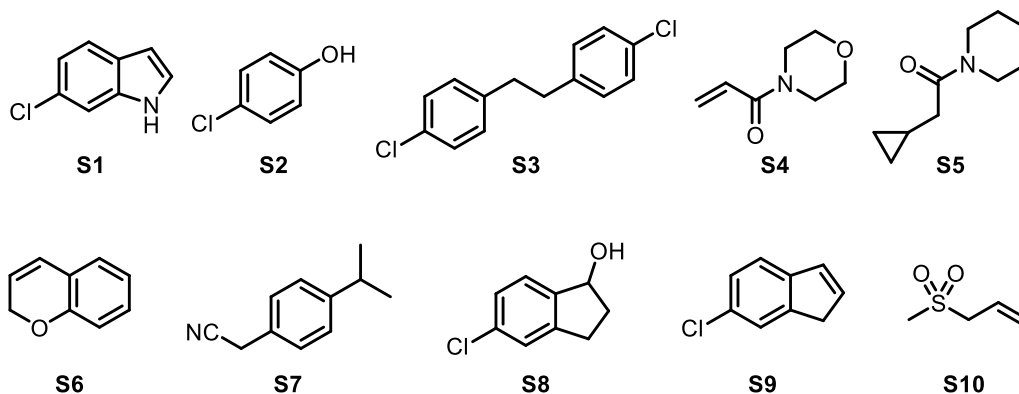


Figure 3.4. Selected co-substrates for synthesis or purchase.

Two rhodium(II) catalysts, Rh_2piv_4 and Rh_2pfb_4 , were chosen on the basis of the principal component analysis in Chapter 2 for use in the initial reaction array, completing the array design. There are five main classes of rhodium(II) catalyst, as identified in Chapter 2, and catalysts from the carboxylate cluster of the analysis were prioritised due to significantly greater literature precedence for intermolecular transformations when compared to rhodium(II) catalysts with carboxamidate ligands.²¹¹ Rh_2piv_4 was selected due to favourable solubility in organic solvents when compared to close analogues, such as Rh_2OAc_4 . Rh_2pfb_4 was chosen because of the low predicted similarity with Rh_2piv_4 and contrasting contributions from the strongly electron-withdrawing perfluorobutyrate ligand. In summary, a library of potential diazo substrates (Figure 3.3) was prioritised for synthesis and 10 co-substrates (**S1** – **S10**) (Figure 3.4) were selected for purchase in the first reaction array.

3.1.1 Synthesis of Diazo Substrates for an Initial Reaction Array

The next step towards completing an ADS reaction array is to synthesise the required reagents. α -Diazo amides are commonly accessed through five general synthetic routes (Figure 3.5): decomposition of a hydrazone; acylation; diazo transfer; α -substitution of α -diazo amides by cross-coupling; and direct diazotization of primary amines.^{187,228} Reaction conditions from synthetic routes using decomposition and diazo transfer strategies were prioritised for the synthesis of the proposed diazo compounds due to

superior tolerance of substituent functionality compared to direct diazotization and acylation strategies.¹⁸⁷

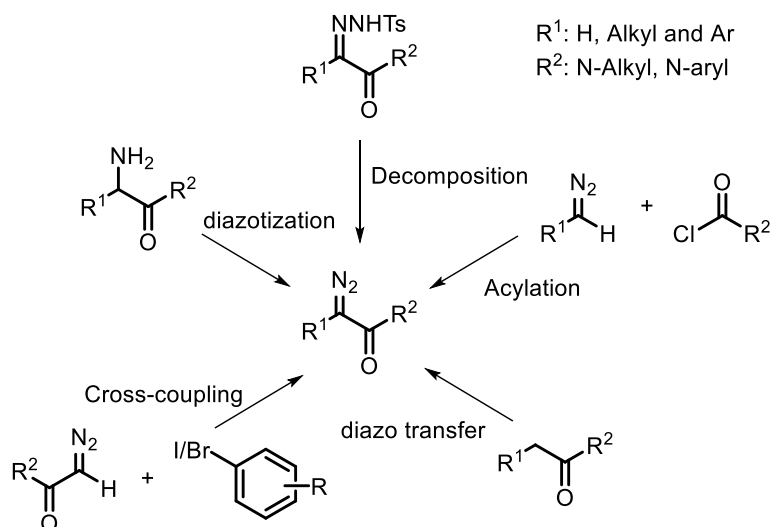
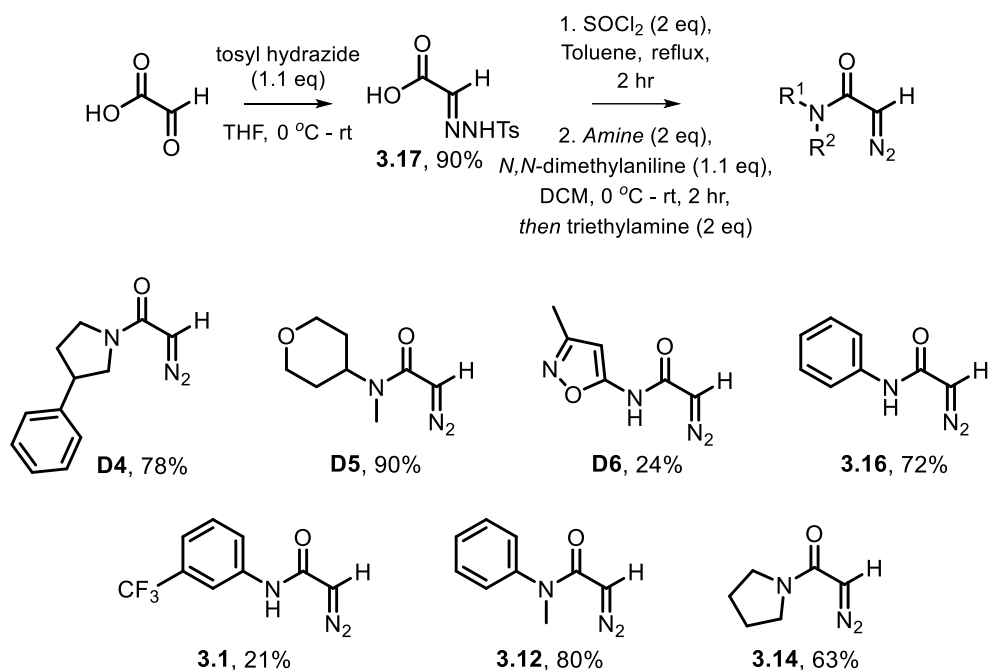


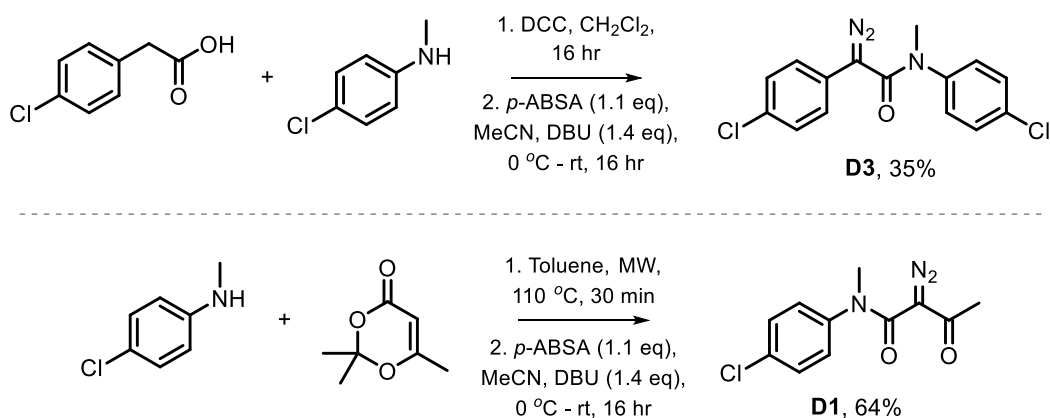
Figure 3.5. Strategies for synthesising α -diazo amides.

Synthesis of diazo substrates **D4** – **D6**, and **3.1**, **3.12**, **3.14** and **3.16** through decomposition of the corresponding α -hydrazone amide was achieved using reported conditions (Scheme 3.1).²²⁹ Accordingly, glyoxylic acid was condensed with tosyl-hydrazide to afford the intermediate hydrazone **3.17**, which was then subjected to an acid chloride amide coupling with the parent amines, and finally decomposed using a non-nucleophilic base, triethylamine, to give the α -diazo amide products in moderate to good yields (24-90%). The diazo substrates **3.2**, **3.4**, **3.8** and **3.10** proved to be synthetically inaccessible via decomposition of the corresponding α -hydrazone amides due to degradation of either the product α -diazo amide or α -hydrazone amide intermediate into a complex mixture.^{230,231} The diazo **3.11** could be synthesised through decomposition of the corresponding α -phenylhydrazone amide in trace yields and was not considered for inclusion as a substrate. In the case of diazo **3.3** the α -hydrazone intermediate could be readily formed from diisopropylamine, but subsequent decomposition to the corresponding α -diazo amide did not proceed using triethylamine or 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) as a base. Using considerably more forcing conditions (12 M aqueous sodium hydroxide, aliquat-336 (a phase transfer catalyst), heating between 30-60 °C and toluene as solvent) were also screened but the hydrazone remained stable and could not be decomposed to the diazo. Ultimately seven α -diazo amides were synthesised through the decomposition of hydrazone **3.17** and considered for use as substrates in the initial reaction array.



Scheme 3.1. Synthesis of diazo compounds via decomposition of a hydrazone.

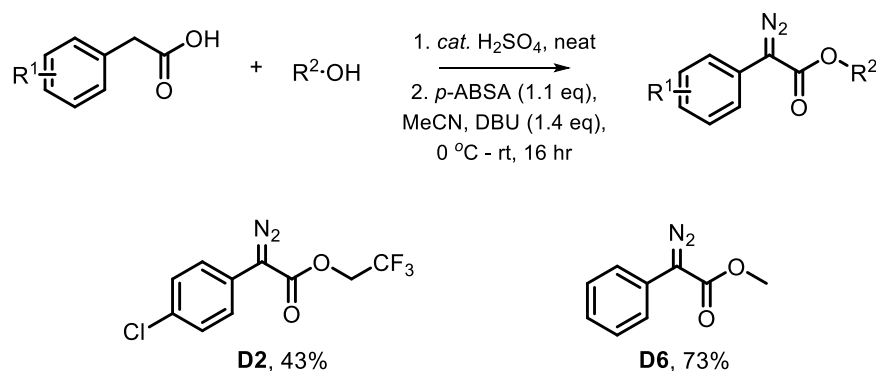
Diazo substrate **D3** was synthesised by diazo transfer to the parent diphenylacetamide in one pot. An amide coupling between 4-chlorophenylacetic acid, 4-chloro-*N*-methylaniline and DCC formed the intermediate diphenylacetamide which was then treated with *p*-ABSA and DBU to yield the product diazo (Scheme 3.2). Synthesis of **D1** proceeded with acetylation of 4-chloro-*N*-methylaniline, forming the β -ketoamide which then underwent diazo transfer using identical conditions to **D3** (Scheme 3.2).



Scheme 3.2. Synthesis of **D3** and **D1** by diazo transfer using *p*-ABSA.

Attempts to synthesise diazo substrates **3.5** – **3.7**, **3.9**, **3.13** and **3.15** by diazo transfer were unsuccessful despite consumption of the parent

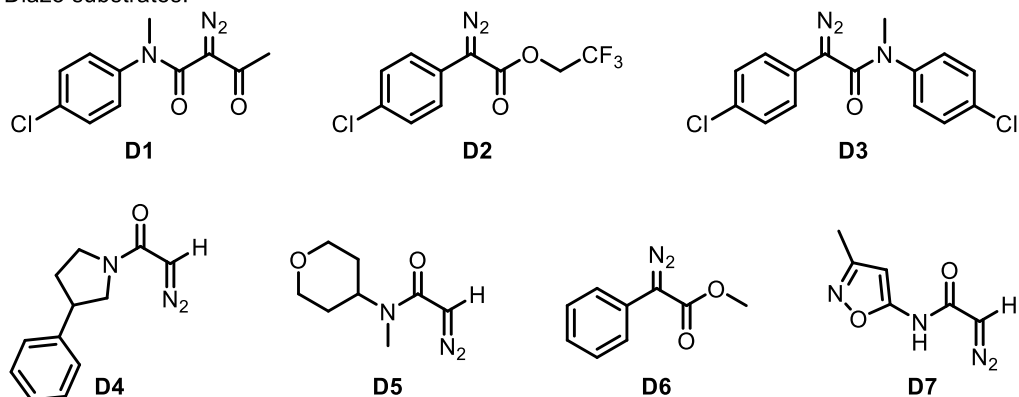
phenylacetamide and *p*-ABSA diazo transfer reagent.²³² Due to the limited success in synthesising α -diazo acetamides using diazo transfer and hydrazone decomposition methodologies two diazo esters, **D2** and **D6**, were synthesised to try and exploit more established diazo synthesis that were not originally considered (Figure 3.2). The corresponding phenyl acetates were synthesised by Fischer esterification, using catalytic sulfuric acid, and then diazotized by diazo transfer using *p*-ABSA and DBU to yield the final two diazo compounds **D2** and **D6** (Scheme 3.3).



Scheme 3.3. Synthesis of **D2** and **D6** via diazo transfer using *p*-ABSA.

In summary, nine α -diazo acetamides (**D1**, **D3** – **D5**, **D7**, **3.1**, **3.12**, **3.14** and **3.16**), from the initially proposed set of 21 substrates, were synthesised and available for selection in the first reaction array along with two additional diazo esters (**D2** and **D6**). All ten selected co-substrates were available for purchase and included in the final array design. The final array design includes diazo substrates **D1** – **D7** and co-substrates **S1** – **S10**, as shown in Figure 3.6. Diazo compounds **D1**, **D2** and **D3** were selected as they contain a *para*-chlorophenyl moiety that mimics the hotspot tryptophan residue of the p53 transactivation domain and has been extensively reported in known MDM2 ligands. Diazo compounds **D4** and **D6** were chosen as they feature phenyl rings that can mimic the conserved phenylalanine or leucine residues on the p53 transactivation domain. Finally, diazo compounds **D5** and **D7** were chosen to increase the diversity of functionality featured on each diazo, **D5** is substituted with a homologated analogue of morpholine with the nitrogen atom extruded from the six membered ring, and **D7** contains a small isoxazole heterocycle which has not been included in any of the other substrates.

Diazo substrates:



Co-substrates:

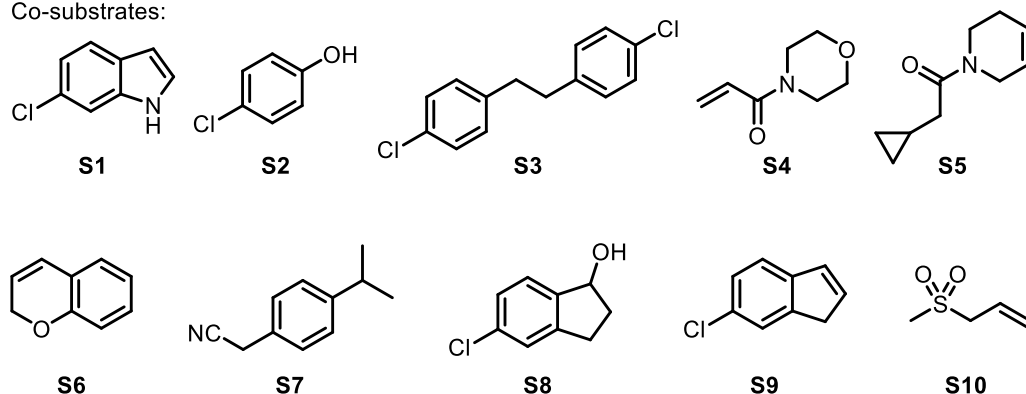


Figure 3.6. Summary of reaction components for the initial reaction array. Rh_2OAc_4 and Rh_2pfb_4 were selected as catalysts.

3.2 Establishment of a Fluorescence Anisotropy Assay for the p53/MDM2 Protein-Protein Interaction

In order to effectively identify active crude reaction mixtures from the initial reaction array a primary, high-throughput assay was required. The assay needed to be capable of screening hundreds of reactions quickly and have adequate controls for evaluating positive and negative results. A fluorescence anisotropy assay, using the p53-transactivation domain and Human-MDM2 (*hDM2*), was chosen as the main assay for screening the crude reaction mixtures.

Fluorescence anisotropy is a sensitive technique for studying protein-protein and protein-ligand interactions.²³³ Typically a protein or peptide is labelled with a fluorescent tag, such as fluorescein or rhodamine, to create a molecular probe, commonly called a tracer, that can then be used to investigate the binding interaction between a tracer and an unlabelled protein.²³⁴ At a molecular level, when a fluorophore is excited with polarised

light the emitted light is also polarised, and the degree of polarisation is related to the macro-molecular properties of the fluorophore meaning that it can be used as a sensor.^{235,236} The extent of polarisation is a function of the Brownian molecular rotation of the peptide-tracer or protein/peptide-tracer complex, so a protein-protein interaction can be studied by measuring changes in polarisation caused by changes in the rate of the Brownian molecular rotation.²³⁷ An unbound tracer will have a lower anisotropy value than the protein-bound tracer as the unbound tracer depolarises the plane-polarised light by rotational diffusion during the lifetime of the fluorescence (Figure 3.7).

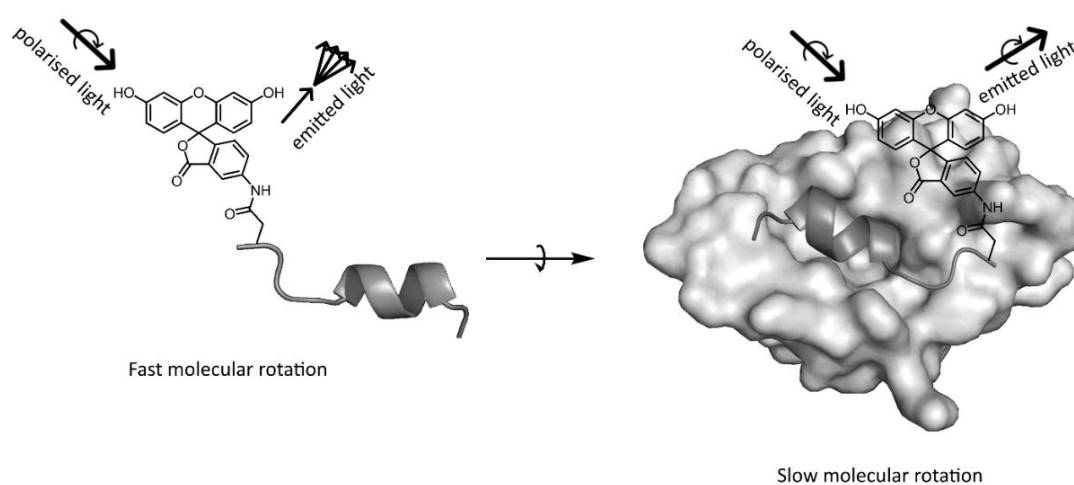


Figure 3.7. Binding of a fluorescently labelled peptide to an unlabelled protein changing the extent of emitted polarised light (PDB: 1YCR).

A fluorescence anisotropy assay using a fluorescein-tagged p53 peptide (p53₁₅₋₃₁ Flu), spanning the p53 transactivation domain, and *hDM2* (serine-17 to asparagine-125, *hDM2*₁₇₋₁₂₅) was chosen as the primary high-throughput screening technique for assessing the outcome of the activity-directed reaction array outlined in section 3.1. The assay described here is based on a reported assay and adapted to improve the performance in a high-throughput single point screen.^{238,239} Initially, a serial dilution of *hDM2*₁₇₋₁₂₅ (20 – 0.006 μM) was added to a fixed concentration of p53₁₅₋₃₁ Flu (54.5 nM) in an aqueous phosphate buffer pH 7.5 (40 mM phosphate, 200 mM NaCl and 0.02 mg/mL Bovine Serum Albumin, PBSA buffer) to give a 60 μL total volume per assay well. Each dilution was performed in triplicate and K_d was estimated from the fraction of bound p53₁₅₋₃₁ Flu tracer (Figure 3.8 and Figure 3.9).

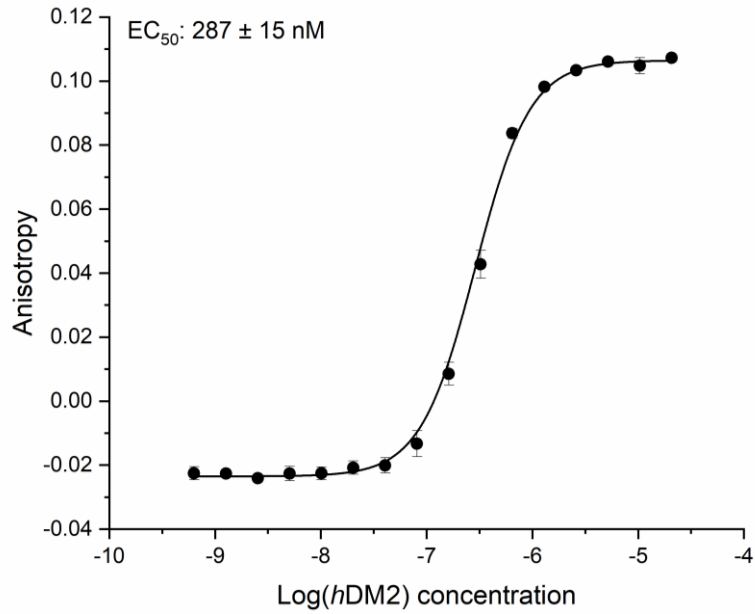


Figure 3.8. Fluorescence anisotropy titration of *hDM2*₁₇₋₁₂₅ (20.75 to 0.0006 μ M) into a fixed concentration of p53₁₅₋₃₁ Flu (54.5 nM) in PBSA buffer.

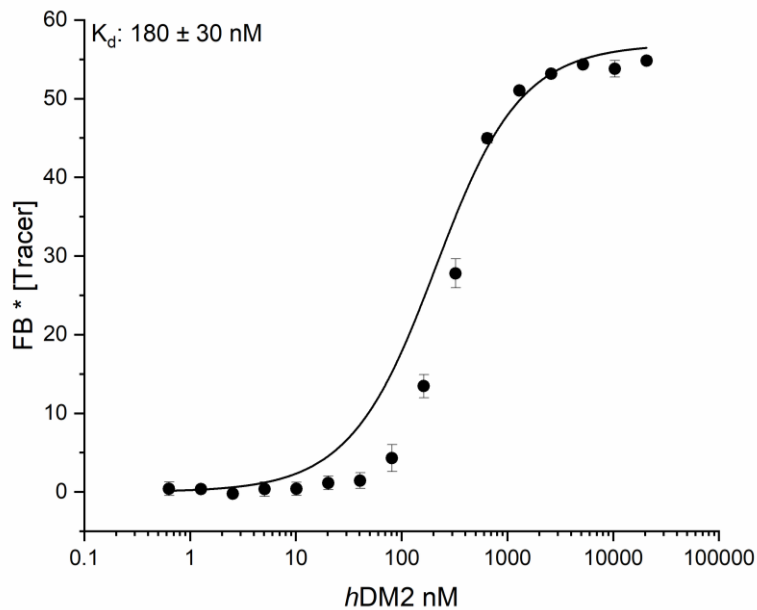


Figure 3.9. Fluorescence anisotropy titration of *hDM2*₁₇₋₁₂₅ (20.75 to 0.0006 μ M) into a fixed concentration of p53₁₅₋₃₁ Flu (54.5 nM) in PBSA buffer. FB = fraction of bound tracer. FB * [Tracer] = FB * Tracer concentration.

To test the performance of the assay for measuring the inhibition of the *hDM2*₁₇₋₁₂₅/p53₁₅₋₃₁ Flu protein-protein interaction with a small-molecule ligand a serial dilution of Nutlin-3a, a potent and extensively studied inhibitor, was added to fixed concentrations of *hDM2*₁₇₋₁₂₅ (150 nM) and p53₁₅₋₃₁ Flu (25 nM)

in PBSA buffer (Figure 3.10). The measured IC_{50} for Nutlin-3a in this assay is 95.2 ± 1.6 nM and the literature reported value is 90 nM,¹⁵⁷ demonstrating that the assay performs well in competition experiments for determining the inhibition of the p53₁₅₋₃₁ Flu/*hDM2*₁₇₋₁₂₅ protein-protein interaction. The serial dilution of Nutlin-3a also gives the expected r_{min} and r_{max} values (-0.025 and 0.04 respectively, estimated from Figure 3.8) for the assay conditions described above, so Nutlin-3a can be used as a positive control in the upcoming single concentration high-throughput screening of crude reaction mixtures.

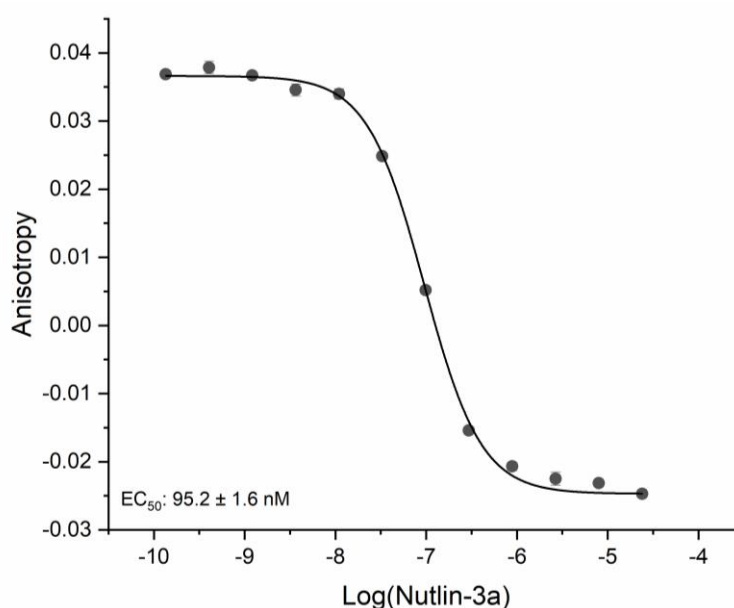


Figure 3.10. Dose response of Nutlin-3a in the *hDM2*₁₇₋₁₂₅/*p53*₁₅₋₃₁ Flu (150 nM and 25 nM concentrations respectively) fluorescence anisotropy assay using PBSA buffer. Observed EC_{50} : 95.2 ± 1.6 nM, reported EC_{50} : 90 nM.¹⁵⁷

The components for the first reaction array were then pre-screened in the fluorescence anisotropy assay to detect any compounds that could interfere with the results by inhibiting the p53₁₅₋₃₁ Flu/*hDM2*₁₇₋₁₂₅ protein-protein interaction before a reaction had taken place. The substrates and co-substrates, outlined in Figure 3.6, were screened at single concentrations in the fluorescence anisotropy assay to investigate the compatibility of the assay with high concentrations of reactants (Figure 3.11). Diazo substrates were screened at 50 μ M, and co-substrates at 100 and 250 μ M concentrations to determine the activity of a crude reaction mixture up to 50 μ M total product concentration. Diazo substrate **D3** showed a small amount of inhibition of the PPI, but not enough to warrant exclusion from the initial

reaction array, and all other substrates and co-substrates showed no residual inhibition.

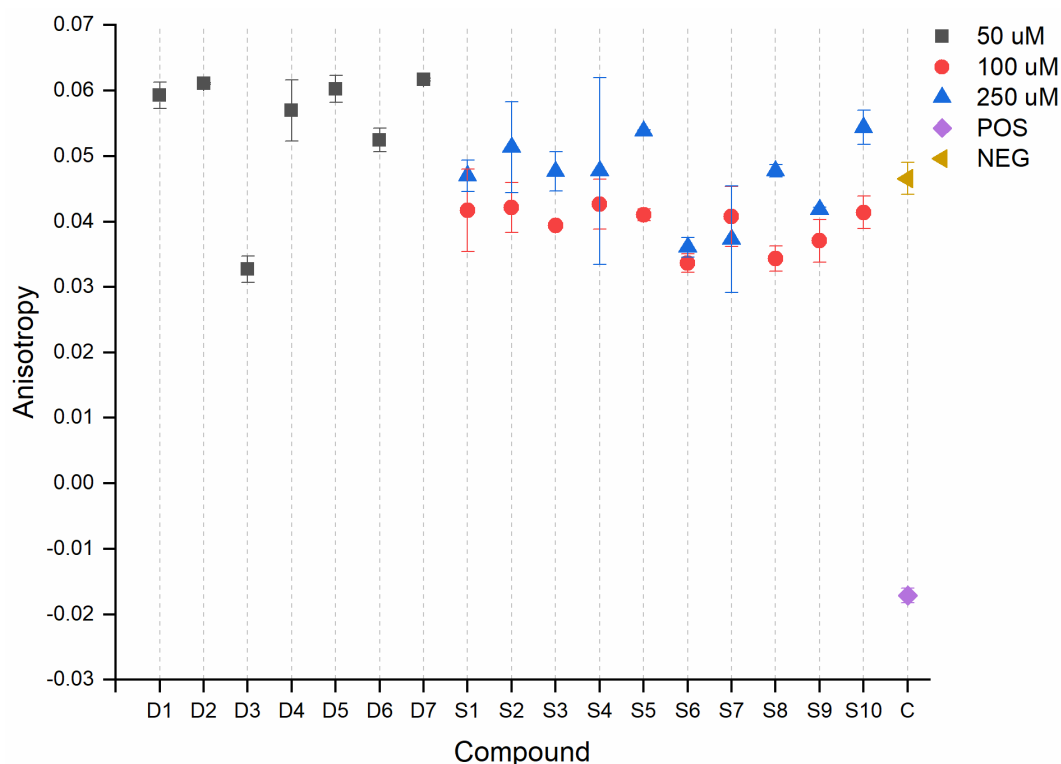


Figure 3.11. Substrate and co-substrate controls for the initial reaction array. Diazo substrates were screened at 50 μM , and co-substrates at 100 and 250 μM , in duplicate with *hDM2*₁₇₋₁₂₅ (150 nM) and p53₁₅₋₃₁ Flu (25 nM), with Nutlin-3a as a positive control (10 μM) and 1% DMSO as a negative control.

Overall, the fluorescence anisotropy assay using p53₁₅₋₃₁ Flu and *hDM2*₁₇₋₁₂₅ was a robust assay and compatible with the screening conditions required for investigating the activity of crude reaction mixtures. Nutlin-3a acts as a suitable positive control for detecting activity in the assay and will be used as a reference compound for determining the activity of crude reaction mixtures. It was decided that the initial reaction array would be screened at 20 μM total product concentration (referenced to the diazo substrate starting concentration), with five equivalents of co-substrate, as there will be no background activity associated with the diazo substrates or co-substrates.

3.3 Implementation and Analysis of Reaction Array 1

The initial reaction array was performed using the design outlined in Figure 3.6. The reaction array consisted of 154 reactions, using diazo substrates

D1 – D7 and co-substrates **S1 – S10**, and was conducted at 100 mM diazo concentration, with a co-substrate (500 mM) and a rhodium(II) catalyst (1 mM, Rh₂piv₄ or Rh₂pf₆) in dichloromethane (100 μL total volume, 100 mM final diazo concentration). In-line controls to detect activity from reactions between a diazo substrate and a rhodium(II) catalyst, or a co-substrate and a rhodium(II) catalyst were also included in the reaction array.

The reactions were performed in a 96-well reaction block with micro-scale borosilicate vials equipped with a stirring flea and sealed using a Teflon sheet. The diazo substrates and co-substrates were dissolved in dichloromethane and the rhodium(II) catalysts were dissolved in tetrahydrofuran (1.25 M, 6.25 M, and 12.5 mM respectively) to create individual stock solutions. Catalysts were added to the reaction vials and the solvent was allowed to evaporate under a stream of nitrogen gas. The reaction solvent (80 μL) was then added and the wells stirred for 5 minutes. The reagents were then dispensed serially across the reaction block using a 8-channel pipette, sealed and allowed to react for 24 hours (for plate maps and experimental details: see Chapter 5). The crude reaction mixtures were then scavenged to remove the catalysts, using a Quadrapure™ TU resin (30 mg per well), and the bulk solvent evaporated under a stream of nitrogen gas. The crude material was then dissolved in DMSO and filtered to remove the resin, giving the master stocks (50 mM) from reaction array 1 which were used for all analysis.

All 154 reactions from the initial reaction array were analysed by LC-MS to investigate how many combinations had produced a mass corresponding to an expected product (Figure 3.12). All samples were diluted to 1 mg/mL concentrations (with respect to the starting diazo concentration) from the original 50 mM DMSO master stock. Reaction wells containing diazo and substrate were analysed for intermolecular products by searching for a combined mass of the substrate and co-substrate minus N₂, and blank control wells were analysed for intramolecular products.

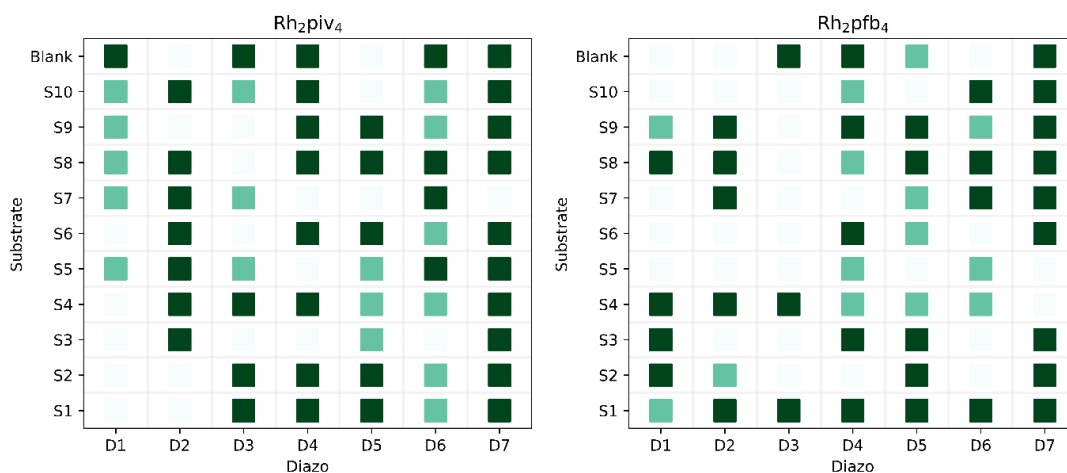


Figure 3.12. LC-MS heatmap for reaction array one. Dark green squares indicate clear m/z for the desired product(s) and a clear corresponding UV peak(s). Light green squares indicate m/z for the desired product(s) and either weak or no corresponding UV peak(s). Blank squares indicate that no m/z was observed for the desired product(s).

Overall, 112 out of 154 reactions (73%) showed the presence of an expected product mass, by LC-MS, of which 78 reactions (51%) also showed distinct UV peaks. The number of successful reactions varied per diazo substrate and per catalyst, presumably as a function of the substrate scope of each rhodium(II)-carbene. For example, the crude reaction mixture containing **D2**, **S10** and Rh₂piv₄ shows a clear new peak corresponding to a product whereas analysis for the same diazo substrate and co-substrate with Rh₂pfb₄ indicates no reaction occurred (Figure 3.13, panel A). Similarly, the crude mixture containing **D6** and **S3** shows no new products with either Rh₂piv₄ or Rh₂pfb₄ as catalyst, but the reaction mixture with **D7** and **S3** shows product formation with both catalysts (Figure 3.13, panels B and C respectively). For all 70 combinations of diazo and co-substrate, excluding intramolecular controls, only 8 combinations out of 70 (11%) failed to give detectable product mass when considering reactions across both catalysts. The functionality of the co-substrates, therefore, contributes less to reaction outcomes than the reactivity of the rhodium(II)-carbenes across the first reaction array. This highlights the importance of including multiple catalysts with different reactivity to improve the number of productive substrate and co-substrate pairs.

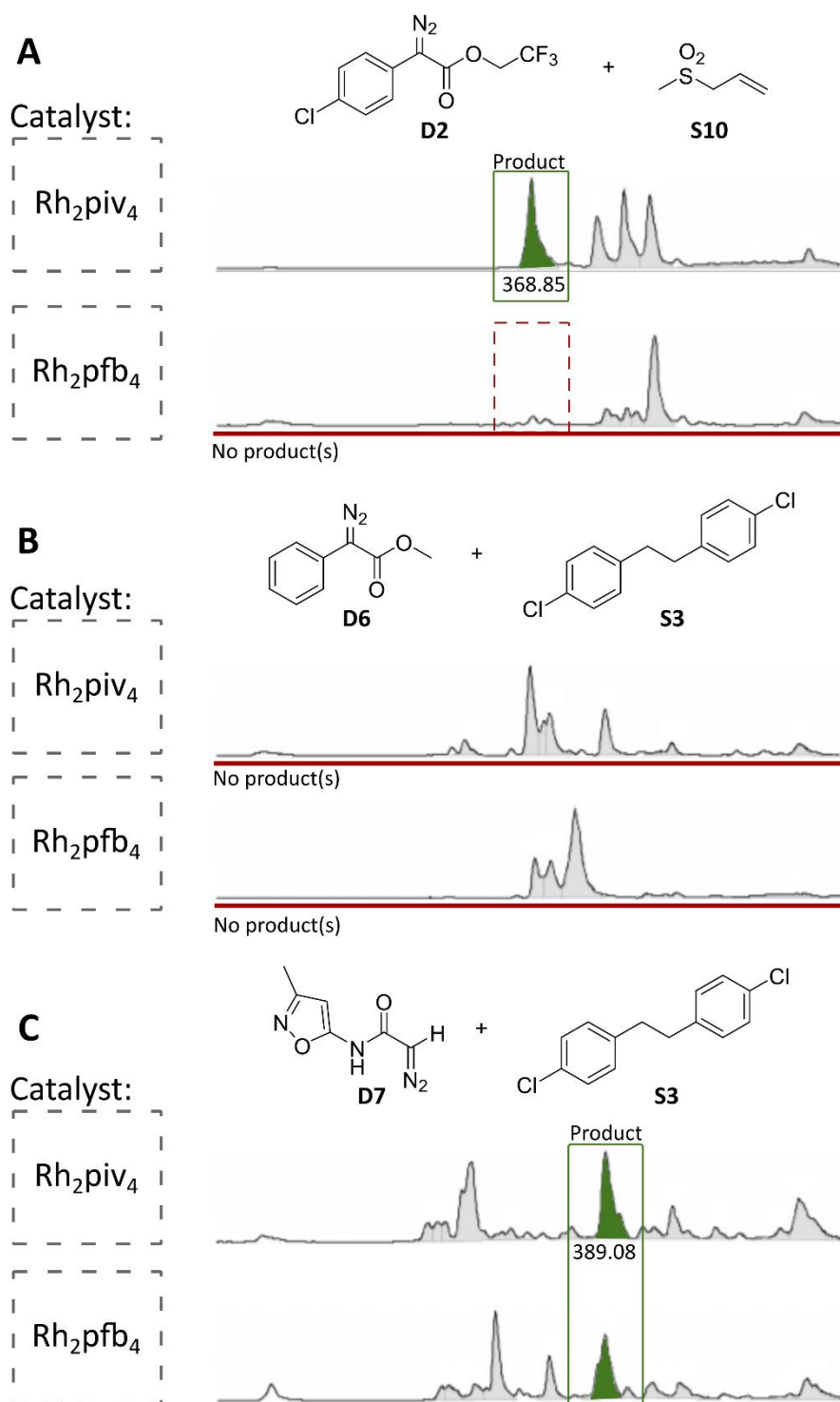


Figure 3.13. LC-MS extracted ion chromatograms for crude reaction mixtures. **A:** **D2** and **S10**. **B:** **D6** and **S3**. **C:** **D7** and **S3**.

All 154 crude reaction mixtures were then assessed for inhibition of the p53₁₅₋₃₁ Flu/hDM2₁₇₋₁₂₅ protein-protein interaction using the fluorescence polarisation assay (Chapter 3.2) at a single concentration of 20 μM total product concentration (Figure 3.14). All reaction wells containing **D3** showed

a decrease in anisotropy, including wells without a co-substrate, indicating that a product formed from only **D3** was likely responsible for the activity. Four other reaction mixtures, containing both a diazo and co-substrate, showed greater than 35% reduction in anisotropy when compared to the activity of Nutlin-3a at 10 μ M. Three of the reactions were catalysed by Rh₂piv₄ (**D1** and **S6**, **D2** and **S10**, and **D4** and **S8**) and one by Rh₂pfb₄ (**D2** and **S1**), of which three reaction mixtures also showed the formation of a desired product by LC-MS (**D2** and **S10**, **D4** and **S8**, and **D2** and **S1**).

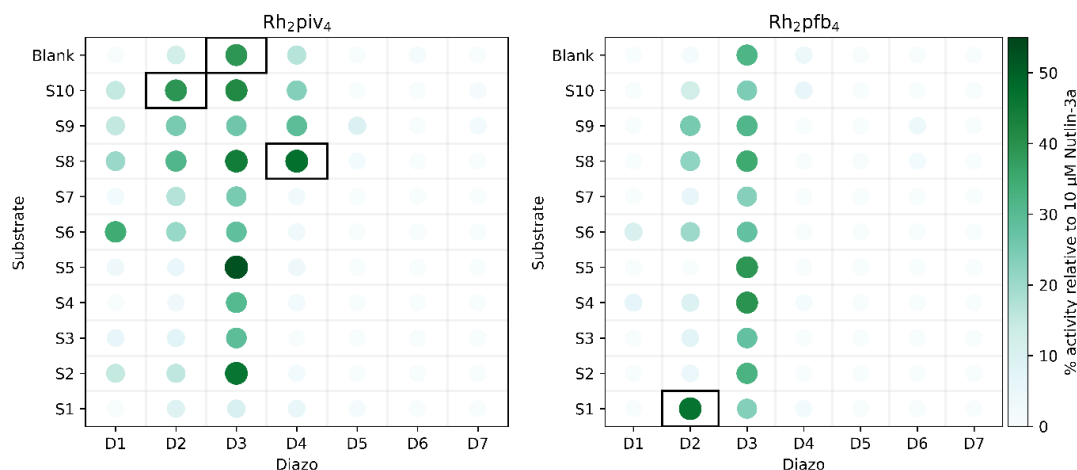


Figure 3.14. Single concentration screening data for the first activity-directed reaction array at 20 μ M total product concentration (referenced to the initial concentration of each diazo substrate).

The four hit crude reaction mixtures were then assayed in dose-response mode to enable the validation of the activity observed in the single concentration high-throughput screen (Figure 3.15). Crude reaction mixtures containing **D2** and **S1** (Rh₂pfb₄, panel A), **D3** (panel B), and **D4** and **S8** (Rh₂piv₄, panel C) show clear dose-dependent decreases in anisotropy consistent with inhibition of the p53₁₅₋₃₁ Flu/*hDM2*₁₇₋₁₂₅ interaction. The crude reaction mixture containing **D2** and **S10** (Rh₂piv₄, panel A) results in a much smaller decrease in anisotropy which could indicate a product with lower activity or a product that was formed in a lower yield. As a result, all four active combinations were taken forward for scale-up (see section 3.5).

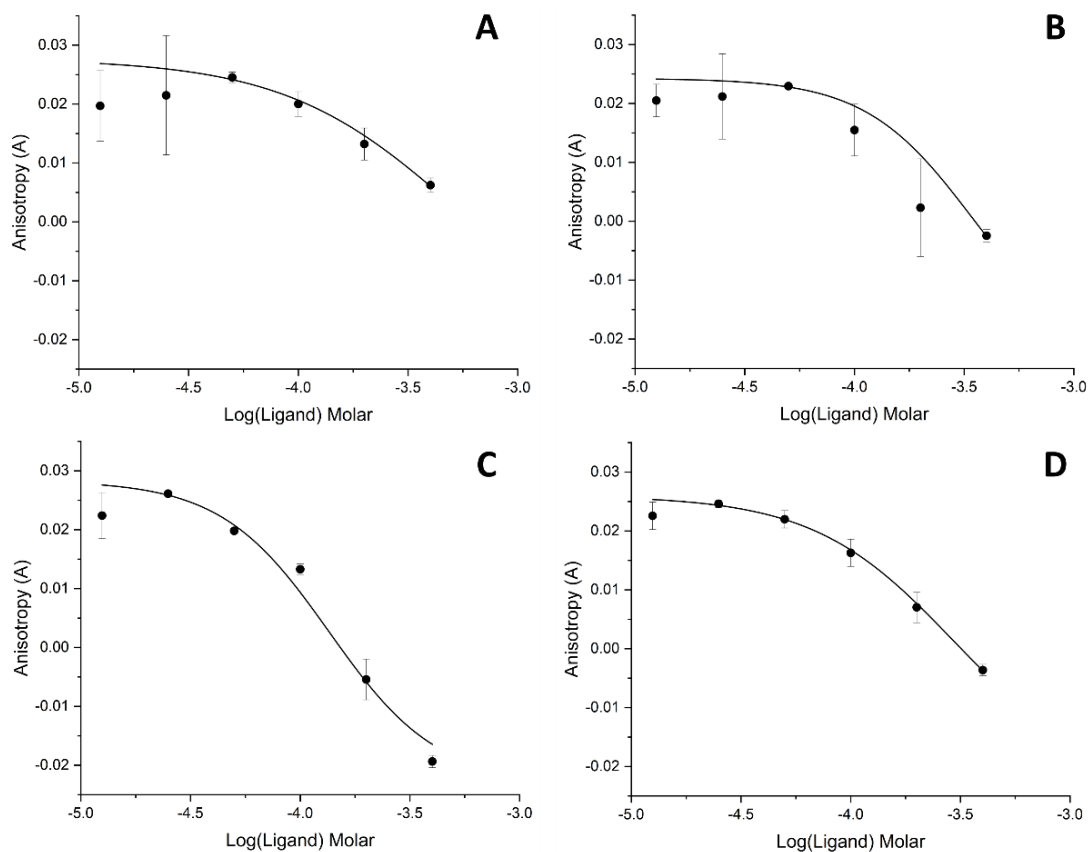
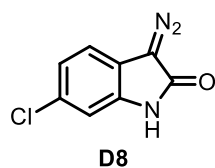


Figure 3.15. Dose-dependent inhibition of the p53¹⁵⁻³¹ FlU/hDM2¹⁷⁻¹²⁵ interaction by crude reaction mixtures. **A:** **D2**, **S10** and Rh₂piv₄. **B:** **D3** and Rh₂piv₄. **C:** **D4**, **S8** and Rh₂piv₄. **D:** **D2**, **S1** and Rh₂pfb₄.

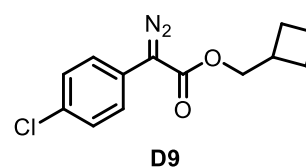
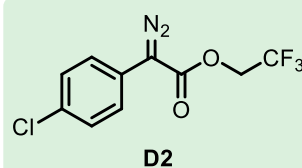
3.4 Design and Implementation of Reaction Array 2

A second reaction array containing a range of new diazo compounds and co-substrates, from active combinations in the initial reaction array, was designed (Figure 3.16). New diazo substrates (**D8** and **D9**) and co-substrates (**S11** – **S21**) were inspired by hit reactions identified in the previous reaction array. It was expected that **D10** could mirror the observed reactivity of **D7**, where all but three reactions had been observed by LC-MS to give intermolecular products, but none which inhibited the p53¹⁵⁻³¹ FlU/hDM2¹⁷⁻¹²⁵ interaction. Co-substrate **S9** was also included in the second reaction array as it had produced several reaction mixtures with lower levels of activity, specifically with **D2** and **D4**, so could prove productive with new diazo substrates.

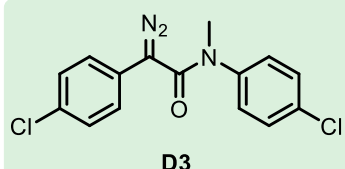
Round 2 Diazos:



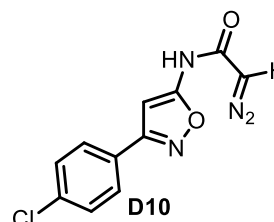
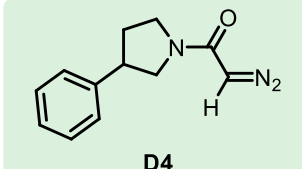
Round 1 hit:



Round 1 hit:

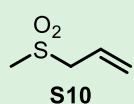
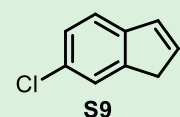
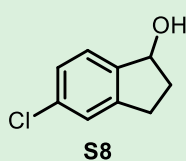
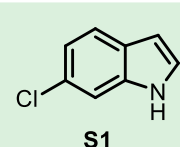


Round 1 hit:



Round 2 Co-substrates:

Round 1 hits:



Round 2 analogues:

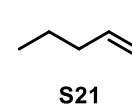
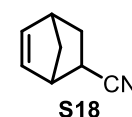
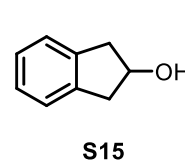
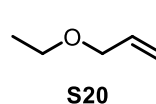
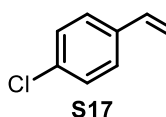
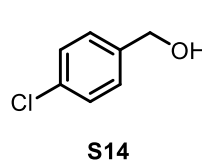
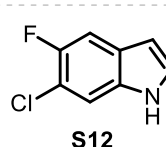
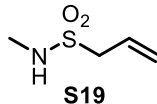
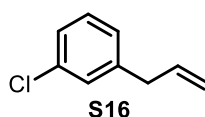
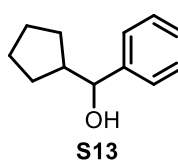
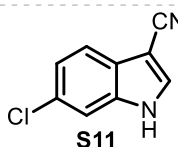
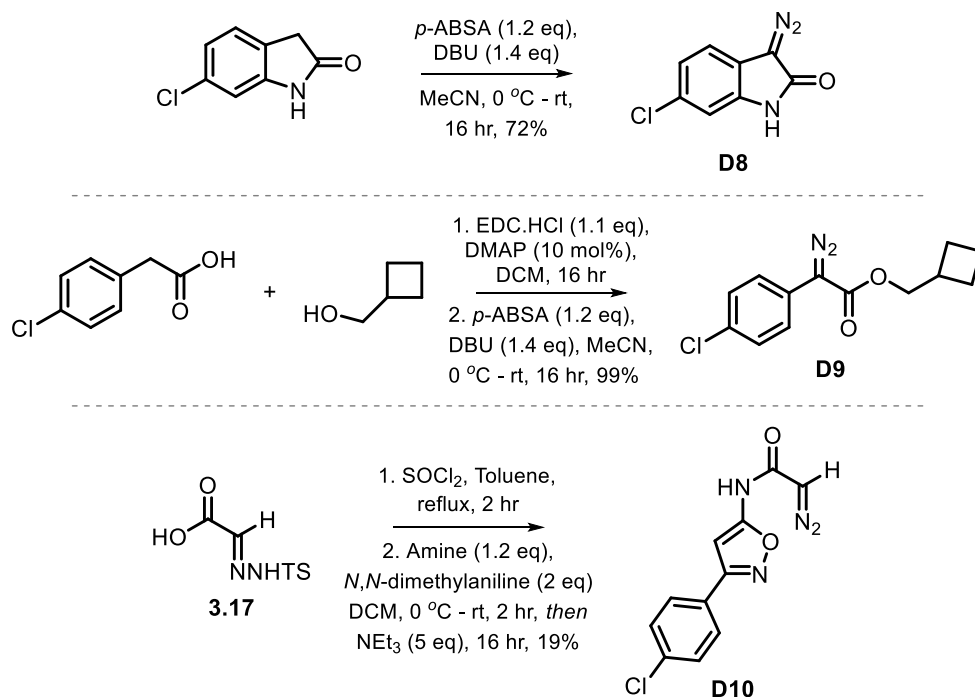


Figure 3.16. Diazo substrates and co-substrates for the second reaction array.

Diazo substrates **D8** and **D9** were synthesised by diazo transfer, using identical conditions to **D1** – **D3** and **D6**, and **D10** was synthesised by decomposition of a hydrazone. Diazo **D8** was synthesised directly from 6-chloro-1,3-dihydroindol-2-one by diazo transfer in a moderate yield (Scheme 3.4, panel A), likely due an azo-coupling side reaction depleting the product diazo.²³² The precursor ester for **D9** was first synthesised from cyclobutanemethanol and 4-chlorophenylacetic acid using Steglich esterification conditions,²⁴⁰ then diazotized by diazo transfer (Scheme 3.4,

panel B) to give **D9**. Diazo **D10** was synthesised by coupling 3-(4-chlorophenyl)-1,2-oxazol-5-amine to **3.17** to give the intermediate α -hydrazone amide which was then decomposed to the product diazo in good yield (Scheme 3.4, panel C).



Scheme 3.4. Synthesis of new diazo substrates, for the second reaction array.

The second reaction array totalled 192 reactions, comprised of six diazo substrates (**D2** – **D4** and **D8** – **D10**), 16 co-substrates (**S1**, **S8**, **S9**, **S10** and **S11** – **S21**) and two catalysts (Rh₂piv₄ and Rh₂pfb₄). The reaction components were pre-screened at two concentrations (5 and 25 μ M for diazo substrates, and 25 and 100 μ M for co-substrates) to detect residual activity that could interfere with the interpretation of the reaction array assay results. The pre-screening assay gave control data for effective concentrations between 5 μ M and 20 μ M for screening crude reaction mixtures and showed that none of the reagents have significant residual activity (Figure 3.17).

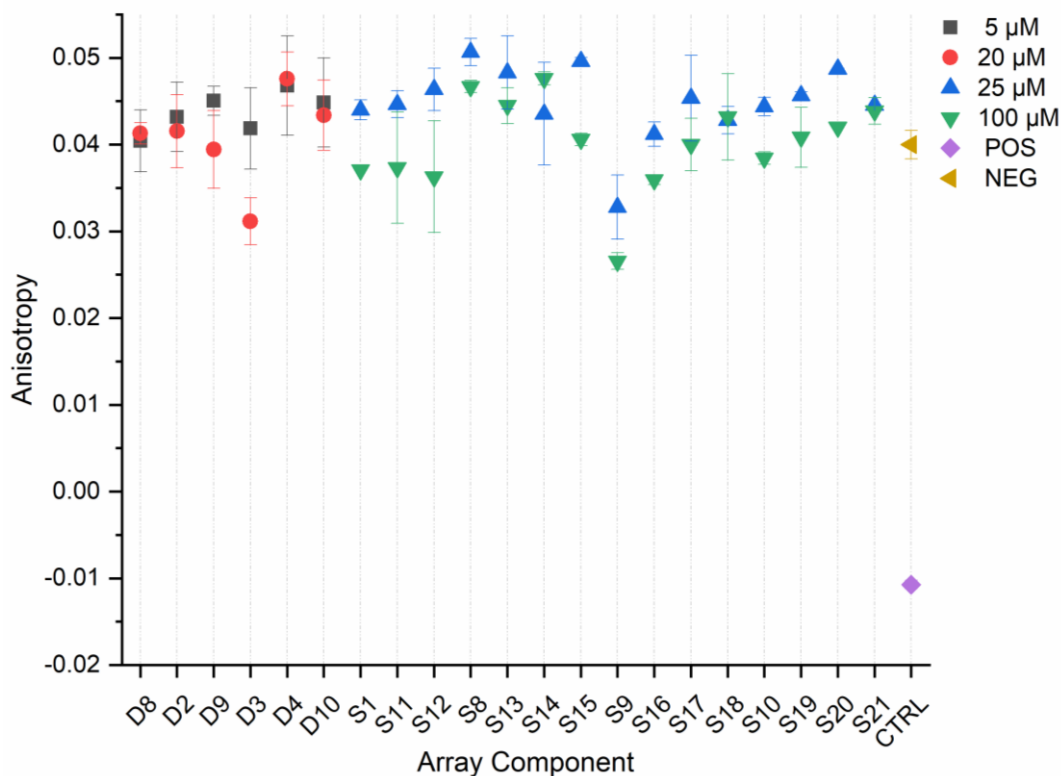


Figure 3.17. Diazo substrate and co-substrate controls for the initial reaction array. Diazo substrates were screened at 5 and 20 μM , and co-substrates at 25 and 100 μM , in duplicate with *hDM2*₁₇₋₁₂₅ (150 nM) and p53_{15-31 Flu} (25 nM), with Nutlin-3a as a positive control (10 μM) and 1% DMSO as a negative control.

An identical workflow to the initial reaction array was followed, and after 24 hours the reactions were scavenged and evaporated, and the crude reaction mixtures were screened initially at 5 μM total product concentration in the fluorescence anisotropy assay (for data see: Chapter 5). However, no hits were identified from the single concentration high-throughput screen of the crude reaction mixtures at 5 μM . The reaction mixtures were re-screened at 20 μM total product concentration and five new hit combinations were identified (Figure 3.18). New reaction mixtures containing **D8** and **S12**, **D8** and **S17**, **D8** and **S20** and **D2** and **S12**, catalysed by Rh₂piv₄, all showed at least a 30% reduction of anisotropy compared to 10 μM Nutlin-3a. One new reaction mixture containing **D8** and **S1**, catalysed by Rh₂pfb₄, also decreased anisotropy by greater than 30%, and the **D2** and **S1** active combination from the initial reaction array reappeared in the second array. Five new reaction mixtures from the second reaction array, along with the four active reaction mixtures from the initial reaction array, were then

prioritised for scale-up in conventional lab-ware to identify products responsible for the observed activity.

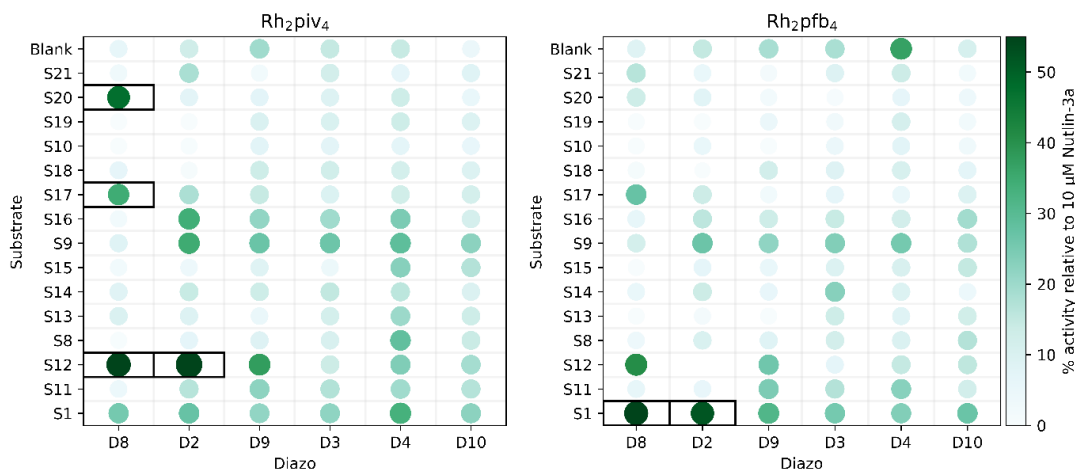
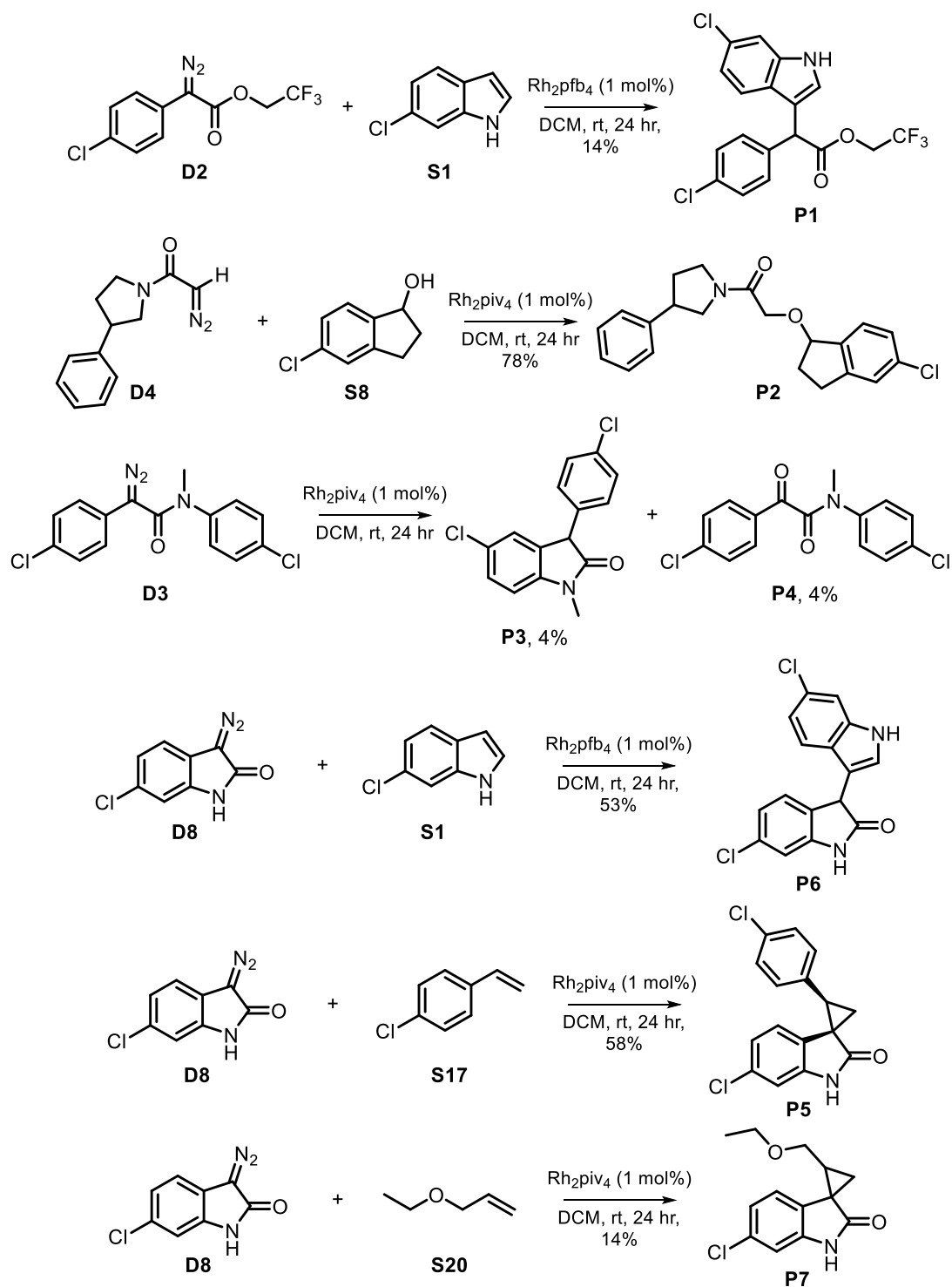


Figure 3.18. Single concentration screening data for the second reaction array at 20 μM total product concentration.

3.5 Identification of Products from Hit Reaction Mixtures

After completing both high-throughput reaction arrays active reaction mixtures were scaled-up to a 50-fold larger scale. Products were isolated from the reaction mixtures by flash column chromatography with the aim of identifying the compounds responsible for the observed activity. The Rh_2piv_4 -catalysed intramolecular reaction of diazo **D3** was initially scaled up and found to yield two products, the oxindole **P3** and the α -keto amide **P4** (Scheme 3.5). The other Rh_2piv_4 -catalysed reactions with **D4** and **S8**, **D8** and **S17** or **S20** were found to yield products **P2**, **P5** and **P7** respectively. **P5** was isolated as a single diastereomer, consistent with literature reports of rhodium(II)-carbene catalysed cyclopropanation reactions with styrene.²¹⁷ **P5** also produced a diagnostic ^1H NMR shift for the indol-2-one-4H hydrogen at 5.84 ppm indicating an interaction between the indol-2-one-4H hydrogen and the 4-chlorophenyl aromatic ring and confirming that the cyclopropane ring has a *cis*-configuration.²⁴¹



Scheme 3.5. Scale-up of hit reaction mixtures from the first and second reaction arrays.

The Rh_2piv_4 -catalysed reaction between **D2** and **S10** was not found to give an intermolecular product when repeated on a larger scale. The Rh_2pfb_4 -catalysed reactions with **D2** and **S1**, and **D8** and **S1** were found to give

products **P1** and **P6** respectively. Reactions between **D2** and **S12**, and **D8** and **S12** did not proceed when repeated at a larger scale, instead immediate precipitation of the Rh₂pfb₄ catalyst occurred when **S12** was added to the reaction mixture.

Three types of rhodium(II)-catalysed reactions yielded products from the scale-up of the selected reaction mixtures, including O-H insertion (**P2** and **P4**), C-H insertion (**P1**, **P3** and **P6**) and cyclopropanation (**P5** and **P7**). The diversity of productive reactions was enabled by using rhodium(II) catalysts with significant differences in their reactivity and substrate scope, as highlighted by the C-H insertion between **D2** and **S1**, where only Rh₂pfb₄ gives an active product, and the O-H insertion between **D4** and **S8**, where only Rh₂piv₄ gives an active product. In summary, seven products were isolated from the nine reactions prioritised for scale-up after testing the crude reaction mixtures in the fluorescence anisotropy assay (Scheme 3.5).

3.6 Conclusions

Two high-throughput reaction arrays were designed and executed in series to identify potential inhibitors of the p53/MDM2 protein-protein interaction, leading to the scale-up of nine hit reaction mixtures and isolation of seven new compounds (**P1** – **P7**). The first reaction array was designed to consider the reactivity of diazo substrates and co-substrates with rhodium(II) catalysts, and similarity with features of known MDM2 ligands. The second reaction array was inspired by hit combinations of diazo substrates and co-substrates from the first reaction array. Catalyst choice for both arrays was aided by the catalyst map described in Chapter 2, and several products were isolated from reactions catalysed by only one of the catalysts demonstrating the utility of exploring catalyst reactivity.

A total of 346 crude reaction mixtures was screened in the fluorescence polarisation assay, at 20 µM total product concentration, and all 154 crude reactions were analysed by LC-MS from the initial reaction array indicating up to 73% had yielded an intermolecular product. Nine potential hit reaction mixtures were identified and scaled up to give seven products (**P1** – **P7**) from both the first and second reaction array. The characterisation and validation of the purified products (**P1** – **P7**) as *hDM2*₁₇₋₁₂₅ ligands and inhibitors of the p53_{15-31 Flu}/*hDM2*₁₇₋₁₂₅ protein-protein interaction is described in Chapter 4.

Chapter 4

Validation and Characterisation of Hit Compounds from Activity-Directed Synthesis

Typically hit molecules that emerge from discovery campaigns are tested in orthogonal assays to corroborate the observed activity and gain information on the binding interaction with the protein target.^{7,16,242} Hit molecules that emerge from the activity-directed synthesis workflow need to be tested as pure compounds to validate their activity and characterise their function. For the seven products identified from the activity-directed synthesis workflow (**P1** – **P7**, Figure 4.1) a series of experiments using a competition assay and a binding assay were devised to identify inhibition of the p53/*hDM2* protein-protein interaction.

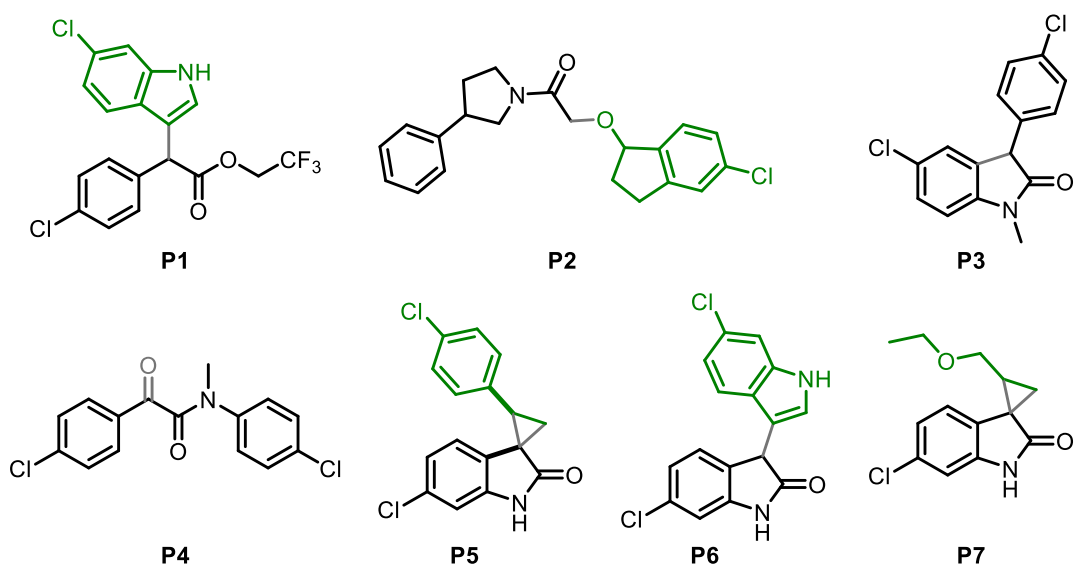


Figure 4.1. Products isolated from the scale-up of hit reactions from reaction arrays 1 and 2 (described in Chapter 3). The origin of each structure is highlighted in black for contributions from the diazo substrate and green for contributions from the co-substrate, and new bonds are highlighted in grey.

4.1 Characterisation of Putative Inhibitors Using a Fluorescence Anisotropy Assay

The purified products were tested in the fluorescence anisotropy assay as an initial step towards validation as inhibitors of the p53₁₅₋₃₁ Flu/*hDM2*₁₇₋₁₂₅ protein-protein interaction. The products (**P1** – **P7**) were serially diluted in

DMSO, using a two-fold or three-fold dilution series, and then diluted further using an aqueous phosphate buffer pH 7.5 to achieve effective concentrations between 0.24 and 480 μM and 3% DMSO. The serially diluted products were then diluted three-fold and tested in the assay containing 150 nM *hDM2*₁₇₋₁₂₅, 25 nM *p53*₁₅₋₃₁ Flu, and 1% DMSO (final product concentrations: 0.08-160 μM). Products **P2** and **P6** gave full dose-dependent decreases in anisotropy, consistent with inhibition of the protein-protein interaction, and IC_{50} values of $15.0 \pm 0.1 \mu\text{M}$ and $0.94 \pm 0.03 \mu\text{M}$ respectively (Figure 4.2).

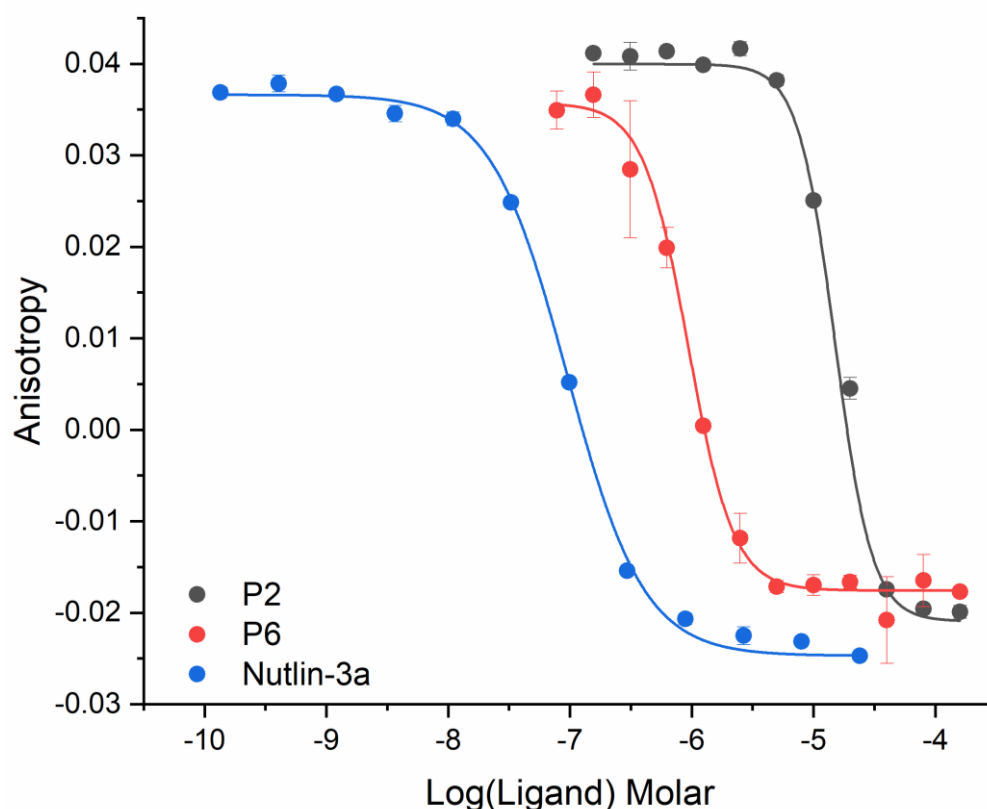


Figure 4.2. Evaluation of the dose-dependent activity of products **P2** (black, IC_{50} : $15.0 \pm 0.1 \mu\text{M}$) and **P6** (red, $0.94 \pm 0.03 \mu\text{M}$) in the *p53*₁₅₋₃₁ Flu/*hDM2*₁₇₋₁₂₅ fluorescence polarisation assay using aqueous phosphate buffer pH 7.5 containing 40 mM phosphate, 200 mM NaCl and 0.02 mg/mL Bovine Serum Albumin. Nutlin-3a (blue, IC_{50} : $95 \pm 1.6 \text{ nM}$) is shown for comparison.

Products **P3** – **P5** also gave dose-dependent decreases in anisotropy that were consistent with inhibition of the interaction. However, the products were not sufficiently soluble in aqueous phosphate buffer pH 7.5, with final DMSO concentrations of between 1 and 3%, to reach product concentrations capable of producing the minimum anisotropy value for the free tracer

(Figure 4.3). IC₅₀ values for products **P3** – **P5** were, therefore, estimated by fitting the bottom asymptote of the dose-response curve to the minimum anisotropy value of the Nutlin-3a control, giving 102 ± 20 µM, 36 ± 11 µM and 11.4 ± 1.2 µM values respectively. Product **P7** was not active in the fluorescence anisotropy assay and did not give a dose-dependent decrease in anisotropy at higher concentrations (up to 320 µM) suggesting that the active components in the **D8**, **S20**, and Rh₂piv₄ reaction mixture had not been isolated.

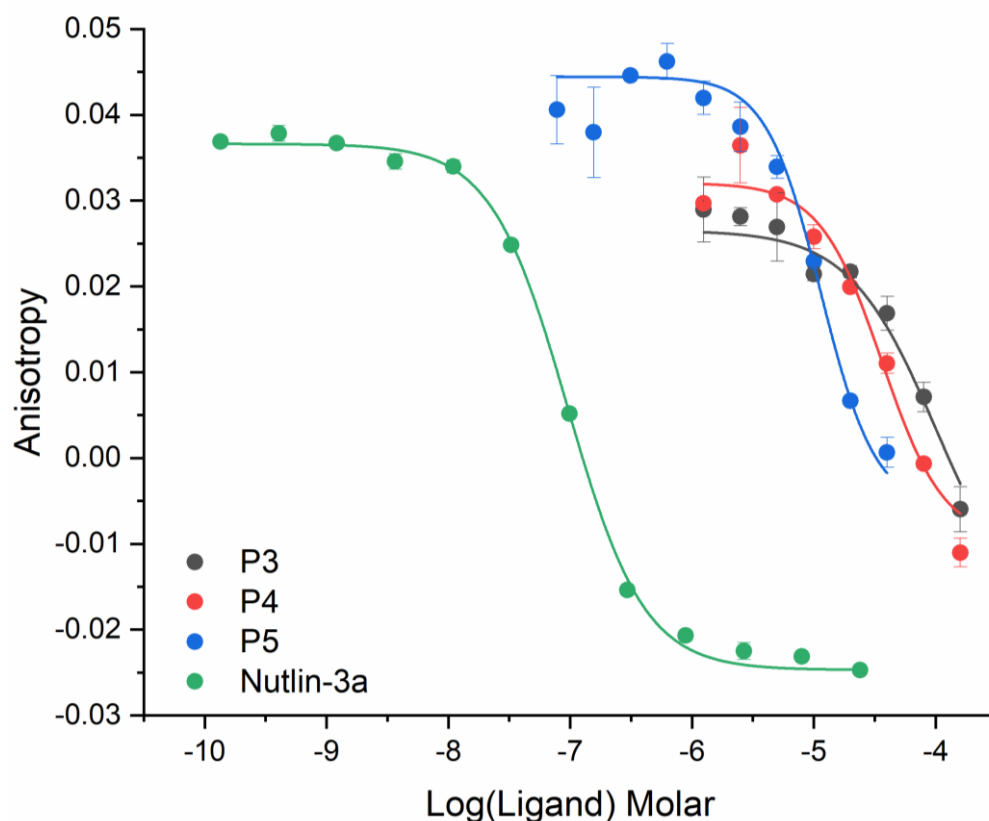


Figure 4.3. Evaluation of the dose-dependent activity of products **P3** (black, IC₅₀: 102 ± 20 µM), **P4** (red, IC₅₀: 36 ± 11 µM) and **P5** (blue, IC₅₀: 11.4 ± 1.2 µM) in the p53₁₅₋₃₁ Flu/*hDM2*₁₇₋₁₂₅ fluorescence polarisation assay, using aqueous phosphate buffer pH 7.5 containing 40 mM phosphate, 200 mM NaCl and 0.02 mg/mL Bovine Serum Albumin. Nutlin-3a (green, IC₅₀: 95 ± 1.6 nM) is shown for comparison.

Despite **P1** being apparently soluble throughout the dilution series, the anisotropy of the free tracer was not observed at high concentrations of **P1** (IC₅₀: 3.6 ± 0.4 µM) (Figure 4.4). This was indicative of more complex behaviour between *hDM2*₁₇₋₁₂₅, p53₁₅₋₃₁ Flu and **P1** and further experiments

using orthogonal assays may be able to validate **P1** as an MDM2 ligand (such experiments are described in section 4.2).

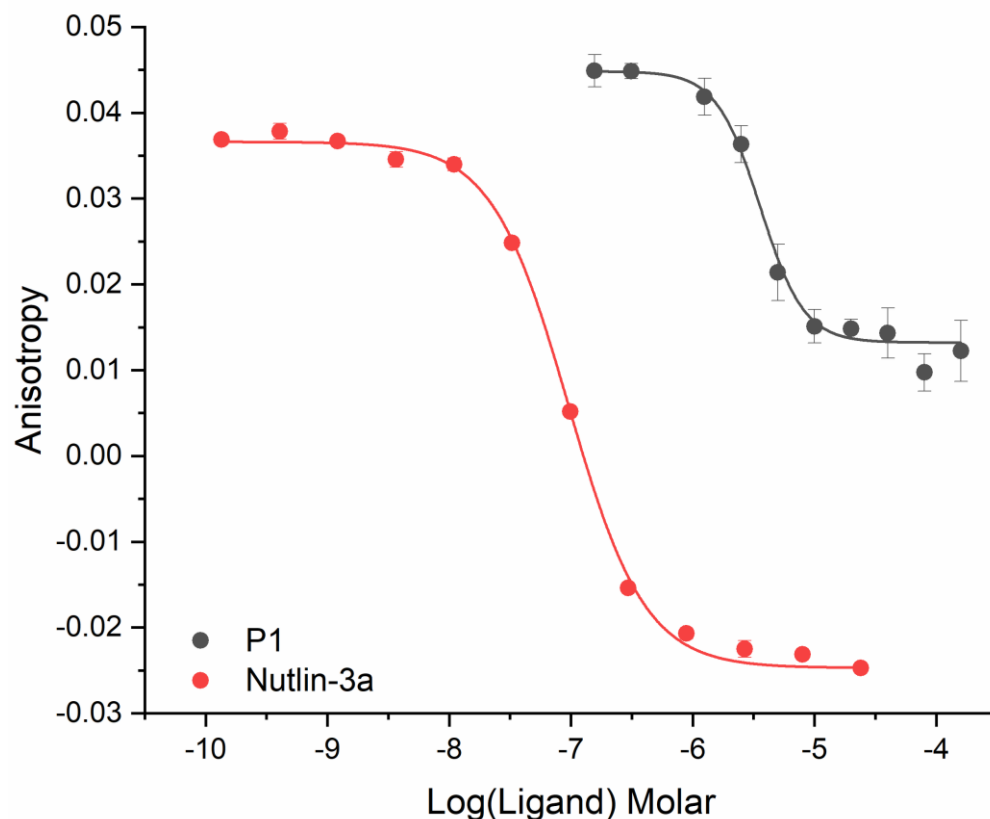


Figure 4.4. Evaluation of the dose-dependent activity of product **P1** (black, IC₅₀: 3.6 ± 0.4 μM) in the p53¹⁵⁻³¹ Flu/hDM2¹⁷⁻¹²⁵ fluorescence polarisation assay, using aqueous phosphate buffer pH 7.5 containing 40 mM phosphate, 200 mM NaCl and 0.02 mg/mL Bovine Serum Albumin. Nutlin-3a (red, IC₅₀: 95 ± 1.6 nM) is shown for comparison.

In summary, the seven pure products from the scale-up of crude reaction mixtures were tested in the fluorescence polarisation assay and were carried forward for testing in orthogonal biophysical assays. Five products (**P2** – **P6**) gave low micromolar IC₅₀ values, and the other two products (**P1** and **P7**) did not result in the expected dose-dependent decrease in anisotropy.

4.2 Characterisation of the Binding of Putative Inhibitors to Human-MDM2 by NMR Spectroscopy

Biophysical assays are widely used to authenticate the observed activity of small-molecule ligands in discovery workflows.^{7,30,243} All seven of the pure products (**P1** – **P7**) were tested in a protein-observed ¹H/¹⁵N-HSQC NMR experiment using ¹⁵N-labelled *hDM2*₁₇₋₁₂₅. The products that demonstrated binding to ¹⁵N-*hDM2*₁₇₋₁₂₅ were considered validated ligands.

Protein-observed ¹H/¹⁵N-HSQC NMR is a well-established technique for verifying interactions between a small-molecule ligand and a target protein.^{55,243} ¹H/¹⁵N-HSQC NMR is very sensitive and may be used to estimate the binding affinity of weak ligands (1 μ M to 1000 μ M typically).²⁴³ The technique can also give some low-level structural information on the binding site of a ligand and allows for detection of non-specific protein-ligand interactions such as protein unfolding or aggregation.^{55,243} As such, ¹H/¹⁵N-HSQC NMR experiments may be used to demonstrate dose-dependent binding of a ligand to a protein, and biophysically validate ligands.

4.2.1 Titration Experiments for the Characterisation of Putative Inhibitors

Increasing amounts of products **P1** – **P7** (25 – 300 μ M) were sequentially titrated into a fixed concentration of ¹⁵N-labelled *hDM2*₁₇₋₁₂₅ (50 μ M) and the ¹H/¹⁵N-HSQC spectrum recorded for each sample. Reporter peaks in the peptide binding cleft of ¹⁵N-labelled *hDM2*₁₇₋₁₂₅, identified by their N-H HSQC cross-peak resonances,^{239,244} were selected for the observation of chemical shift perturbations of the HSQC spectrum. The chemical shift perturbation for each reporter peak was then recorded and used to estimate K_d for the protein-ligand interaction (for full details, see Chapter 5.3.9).

Two products, **P3** and **P7**, did not give dose-dependent shifts of reporter cross-peaks in the ¹H/¹⁵N-HSQC spectrum. Both the spectral quality and intensity were significantly reduced as the molar ratio of **P3** increased through the titration series indicating that the product was causing protein unfolding or aggregation.²⁴³ The spectral quality remained consistent as the concentration of **P7** increased, however no chemical shift perturbation occurred, indicating that the product simply does not bind *hDM2*₁₇₋₁₂₅ under the conditions of the assay.

Products **P2** and **P4** – **P6** all gave dose-dependent chemical shift perturbation of the reporter cross-peaks, in fast-exchange kinetic regimes, and K_d could be estimated (Figure 4.5).²⁴³ The measured K_d values for **P2** and **P4** were broadly similar to the IC_{50} values measured in the fluorescence anisotropy assay ($35 \pm 13 \mu\text{M}$ and $35 \pm 16 \mu\text{M}$ respectively). The measured K_d values for **P5** and **P6** were found to be significantly lower than the concentration of ^{15}N -labelled *hDM2*₁₇₋₁₂₅ ($50 \mu\text{M}$ ^{15}N -*hDM2*₁₇₋₁₂₅ and K_d : $<10 \mu\text{M}$ each) in each titration experiment, so should not be considered accurate estimates of binding affinity, but still confirm a protein-ligand interaction in each case.

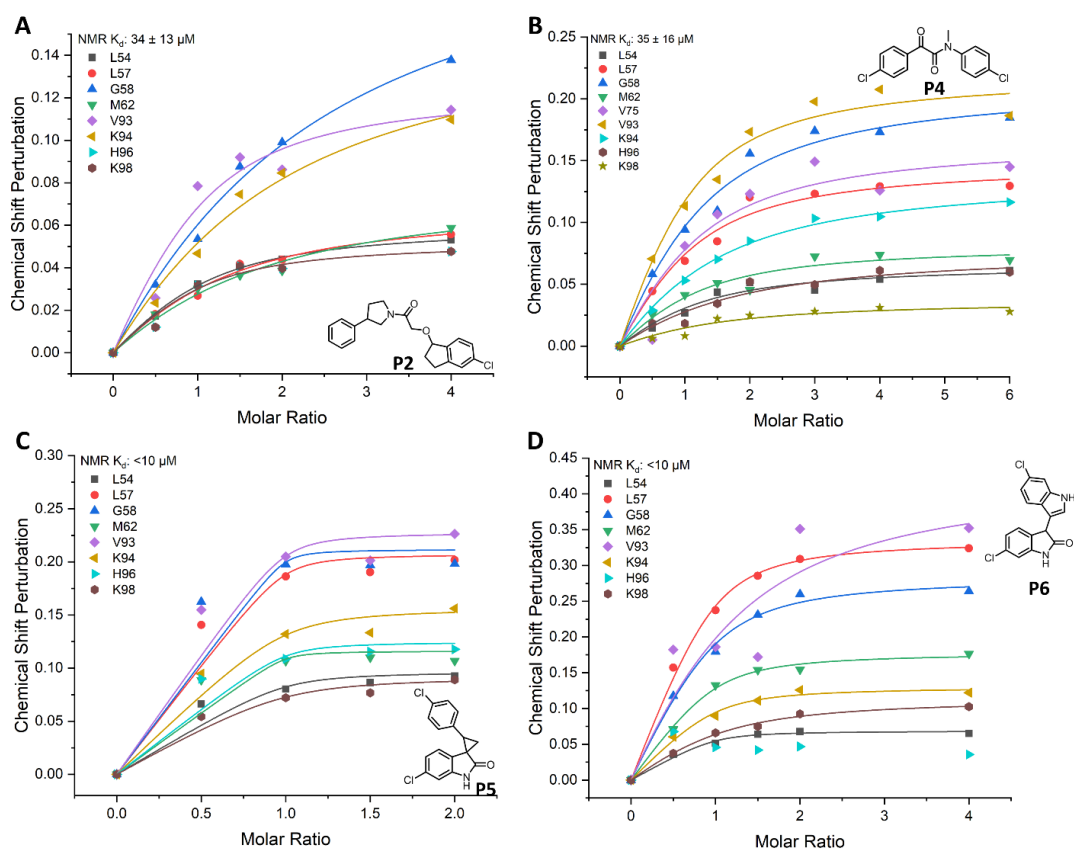


Figure 4.5. $^1\text{H}/^{15}\text{N}$ HSQC titration curves for products **P2** and **P4** – **P6** (25 – $300 \mu\text{M}$) with ^{15}N -labelled *hDM2*₁₇₋₁₂₅ ($50 \mu\text{M}$) in aqueous phosphate buffer pH 7.5 containing 100 mM phosphate, 1 mM DTT, 2.5% glycerol, 10% D_2O and 1% DMSO. K_d was estimated using established fitting procedures.²⁴³

The recorded $^1\text{H}/^{15}\text{N}$ -HSQC spectrum for the titration of **P1** gave chemical shift perturbations in the intermediate exchange kinetic regime, which means K_d cannot be straightforwardly estimated, and the overall spectrum quality was slightly reduced. To investigate if **P1** was competitive with p53₁₅₋₃₁ for *hDM2*₁₇₋₁₂₅ binding, a separate $^1\text{H}/^{15}\text{N}$ -HSQC experiment was devised

(Figure 4.6). Firstly, the ^{15}N -labelled $h\text{DM2}_{17-125}/\text{p53}_{15-31}$ complex was formed in a one-to-one molar ratio (50 μM , Figure 4.6 panel A) and four molar equivalents of **P1** were added (200 μM , Figure 4.6 panel B). No subsequent chemical shift perturbation of the HSQC spectrum was observed which suggests that **P1** does not form ternary or multimeric complexes with the $h\text{DM2}_{17-125}/\text{p53}_{15-31}$ complex or displace the p53_{15-31} peptide. This is an expected result as the p53_{15-31} transactivation domain has a tight binding interaction and should not be displaced by a weak small-molecule ligand at one-to-one $h\text{DM2}_{17-125}/\text{p53}_{15-31}$ molar ratios.²³⁸

Secondly, the ^{15}N -labelled $h\text{DM2}_{17-125}/\text{P1}$ complex was formed in a one-to-one molar ratio (50 μM , Figure 4.6 panel C) and p53_{15-31} was titrated into the sample in 0.25 molar equivalents (12.5 – 50 μM , Figure 4.6 panel D). A dose-dependent perturbation of HSQC cross-peaks from the **P1** bound ^{15}N -labelled $h\text{DM2}_{17-125}$ complex to the p53_{15-31} bound complex was observed, demonstrating that **P1** binds to the $h\text{DM2}_{17-125}$ peptide binding cleft reversibly and competitively with p53_{15-31} .

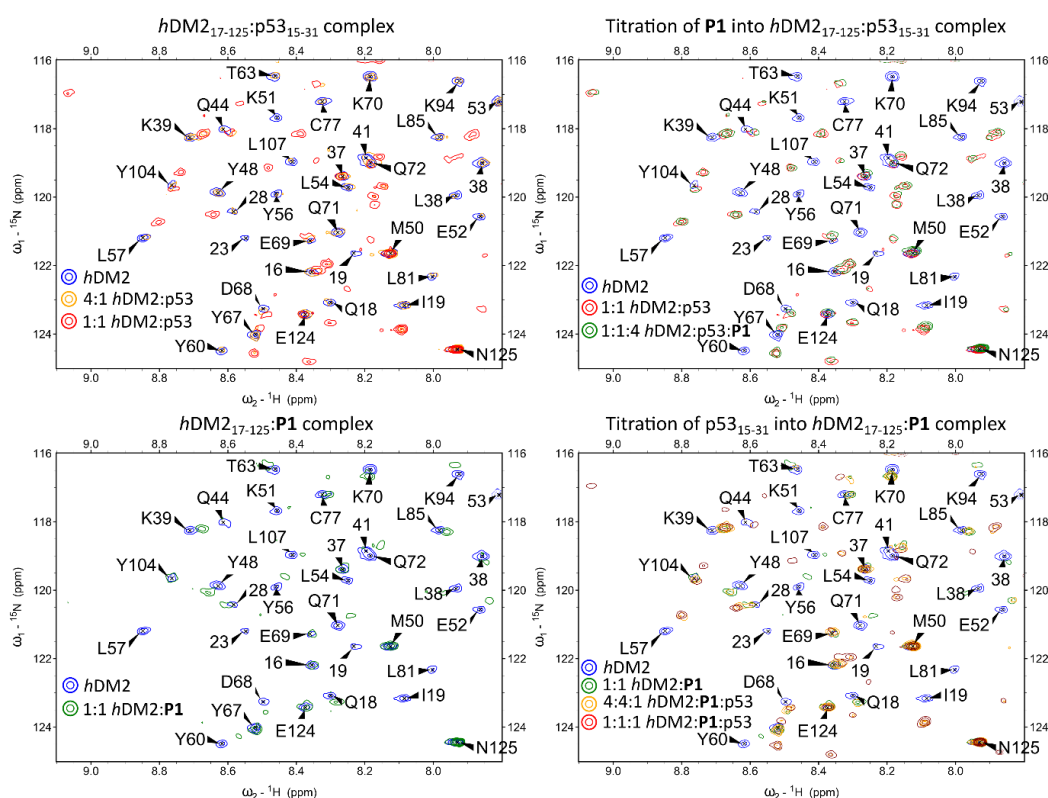


Figure 4.6. $^1\text{H}/^{15}\text{N}$ -HSQC NMR competition assay to displace $h\text{DM2}_{17-125}/\text{ligand}$ complexes with competitor. Experiments were conducted with ^{15}N -labelled $h\text{DM2}_{17-125}$ (50 μM), p53_{15-31} (12.5 or 50 μM) and **P1** (50 or 200 μM), using aqueous phosphate buffer pH 7.5 containing 100 mM phosphate, 1 mM DTT, 2.5% glycerol, 10% D_2O and 1% DMSO. Spectra were recorded

for the displacement of the ^{15}N -labelled *hDM2/P1* complex with p53 after 24 hours and no additional chemical shift was observed.

Due to the unexpectedly high anisotropy observed at high concentrations of **P1** in the fluorescence anisotropy assay (Section 4.1, Figure 4.4) a third, ligand-observed NMR experiment was conducted to determine the binding affinity of **P1** to *hDM2*₁₇₋₁₂₅. Product **P1** contains a trifluoromethyl functional group meaning a ^{19}F -ligand observed NMR experiment was possible and could be used to determine the K_d of the **P1**/*hDM2*₁₇₋₁₂₅ interaction. Carr-Purcell-Meiboom-Gill (CPMG) pulse sequences are used in fragment-based screening to detect binding of a fragment to a protein by measuring changes in chemical shift anisotropy.²⁴⁵ CPMG experiments are reliant on measuring the transverse or spin-spin T_2 relaxation times for an observed nucleus and record the resonance for molecules with longer relaxation times.^{245,246} The relationship between the T_2 relaxation of a molecule and its molecular weight is inverse meaning that large molecules have short relaxation times.^{245,247} Therefore, a small molecule ligand will give a clear resonance as an unbound monomer in solution and its signal will decrease in intensity when a target protein is added as a function of the increasing fraction of bound ligand.^{245,246}

Three spectra for the free **P1** ligand (100 μM) in aqueous phosphate buffer pH 7.5 were initially recorded, firstly a standard ^{19}F NMR spectrum, and then two ^{19}F -CPMG spectra at 50 ms and 300 ms delay times. Varying the delay times in the CPMG pulse sequence allows for filtering of complexes of varying sizes, with shorter delay times removing larger complexes from the spectrum.^{245,248} Free ligands that are sensitive to CPMG delay times, primarily through line-broadening or decreasing intensity, are usually exhibiting behaviour typical of aggregation mechanisms which are the primary cause of false positive hits in high-throughput screening.^{41,248-252} All three spectra showed a clear ^{19}F -signal corresponding to the trifluoromethyl group of **P1** at -119 ppm demonstrating that the free ligand is a monomer in aqueous solution.²⁴⁸

Aliquots of *hDM2*₁₇₋₁₂₅ (0.25 – 10 μM) were then titrated into the **P1** sample (100 μM) and a dose-dependent decrease in ^{19}F -signal was observed (Figure 4.7) indicating a protein-ligand interaction. The spectrum was not sensitive to a decrease in delay time (50 and 300 ms) indicating that the protein is not aggregating upon ligand binding. The addition of 2 μM *hDM2*₁₇₋₁₂₅ completely suppressed the 100 μM **P1** signal indicating that **P1** binds

tightly to *hDM2*₁₇₋₁₂₅. Compounds with low micro-molar binding affinity (**P1** FA IC₅₀: 3.6 ± 0.4 μM and ¹H/¹⁵N-HSQC NMR K_d: <10 μM) are at the edge of what can be reliably determined using NMR-based assays, and tight-binding compounds often quickly reach the limit for determining K_d.^{243,245} The sharp suppression of the ¹⁹F-CPMG signal could also indicate that the interaction between **P1**/*hDM2*₁₇₋₁₂₅ is super-stoichiometric.²⁴⁵ CPMG experiments are not sensitive enough to determine the exact multiplicity of a protein-ligand complex, but it is likely that **P1** is initiating the formation of a complex that involves multiple molecules of *hDM2*₁₇₋₁₂₅ and, or, multiple molecules of **P1**.²⁵³ The nature of protein multimerization may be complex, particularly of multimers mediated by small-molecule ligands, and given that this specific multimeric complex does not form in the presence of p53₁₅₋₃₁, product **P1** was not investigated further.

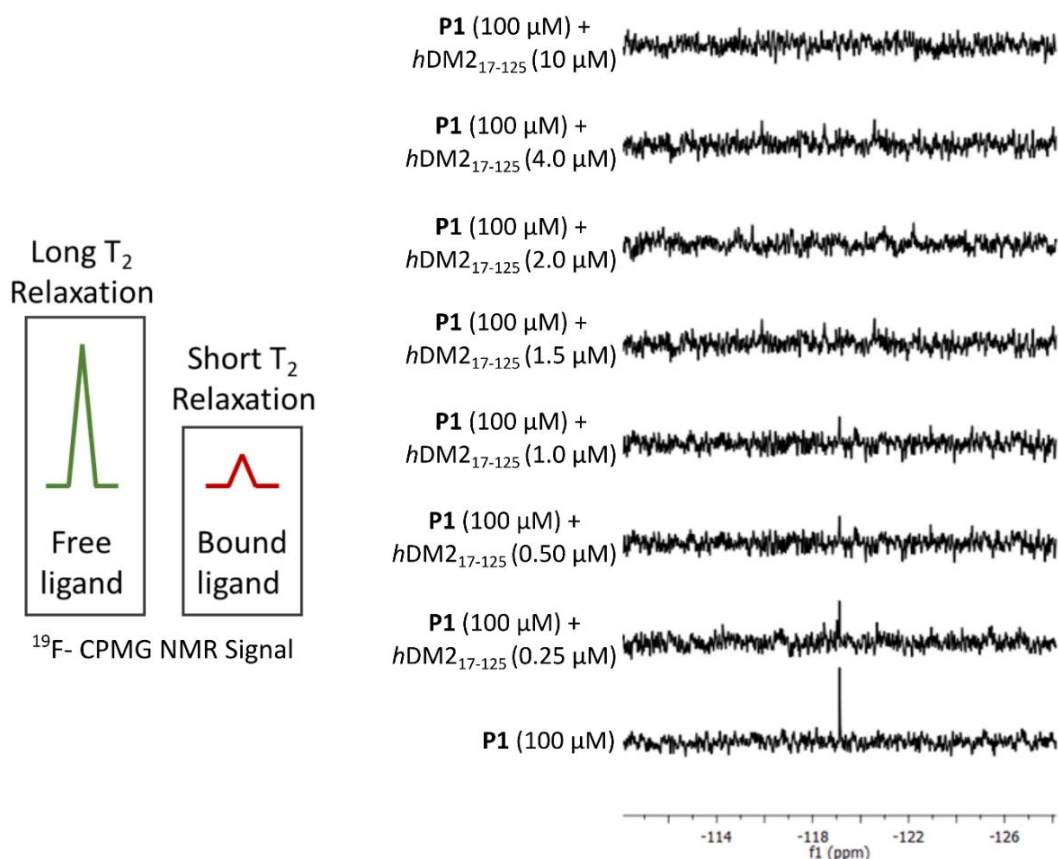


Figure 4.7. Ligand-observed ¹⁹F-CPMG NMR titration experiment with **P1** (100 μM) and *hDM2*₁₇₋₁₂₅ using aqueous phosphate buffer pH 7.5 containing 100 mM phosphate, 1 mM DTT, 2.5% glycerol, 10% D₂O and 1% DMSO.

Of the five products taken forward for testing in orthogonal biological assays four products (**P2** and **P4 – P6**) demonstrated dose-dependent binding to the

peptide binding cleft of *hDM2*₁₇₋₁₂₅. **P2** and **P4 – P6** were also shown to be competitive with the p53₁₅₋₃₁ peptide in the fluorescence anisotropy assay. Positive results that corroborate the findings of two independent experiments provides ample evidence to conclude that **P2** and **P4 – P6** are validated ligands of *hDM2*₁₇₋₁₂₅ and can be analysed in greater detail to gain structural insights.

4.2.2 Structural Insights into the Binding of Putative Inhibitors to Human-MDM2

Protein-observed ¹H/¹⁵N-HSQC NMR experiments can give some low-level structural information for a protein-ligand interaction through the chemical-shift perturbation of specific residues and can aid in identifying the binding site of a ligand.^{243,254} The ¹H/¹⁵N-HSQC spectrum for *hDM2*₁₇₋₁₂₅ has been partially assigned^{239,244} and several of the key residues in the *hDM2*₁₇₋₁₂₅ peptide binding cleft (L57, G58, I74, L81 and V93) can be used to infer structural information for products **P2** and **P4 – P6** (Figure 4.8).

Residues L57 and L81 are important surface residues found in the tryptophan sub-pocket on *hDM2*₁₇₋₁₂₅. Addition of ligands **P2** and **P4 – P6** give clear chemical shift perturbations of the corresponding L57 and L81 cross-peaks, indicating that all four products make a key interaction with the Tryptophan-hotspot residues which have been shown to be essential for binding (highlighted in Figure 4.8).¹⁵⁸ Residue G58 is found on the edge of the tryptophan and phenylalanine sub-pockets and the corresponding cross-peak is also perturbed by addition of the products. Residue V93 is on the upper edge of the leucine and tryptophan hotspots and the corresponding cross-peak is robustly perturbed by **P2** and **P4 – P6** demonstrating that these products are primarily binding to the tryptophan and leucine sub-pockets (Figure 4.9). Interestingly, **P6** appears to demonstrate a secondary, flipped, binding mode where it can bind either a combination of leucine and tryptophan sub-pockets or tryptophan and phenylalanine sub-pockets. Perturbation of the cross-peak for residue I74, which is part of a structural β -sheet near the phenylalanine sub-pocket, shows significant chemical shift perturbation in the presence of **P6** indicating binding to the phenylalanine sub-pocket (highlighted in Figure 4.8). However, the other key surface residues in the phenylalanine sub-pocket have not been assigned and further structural detail cannot be extracted.

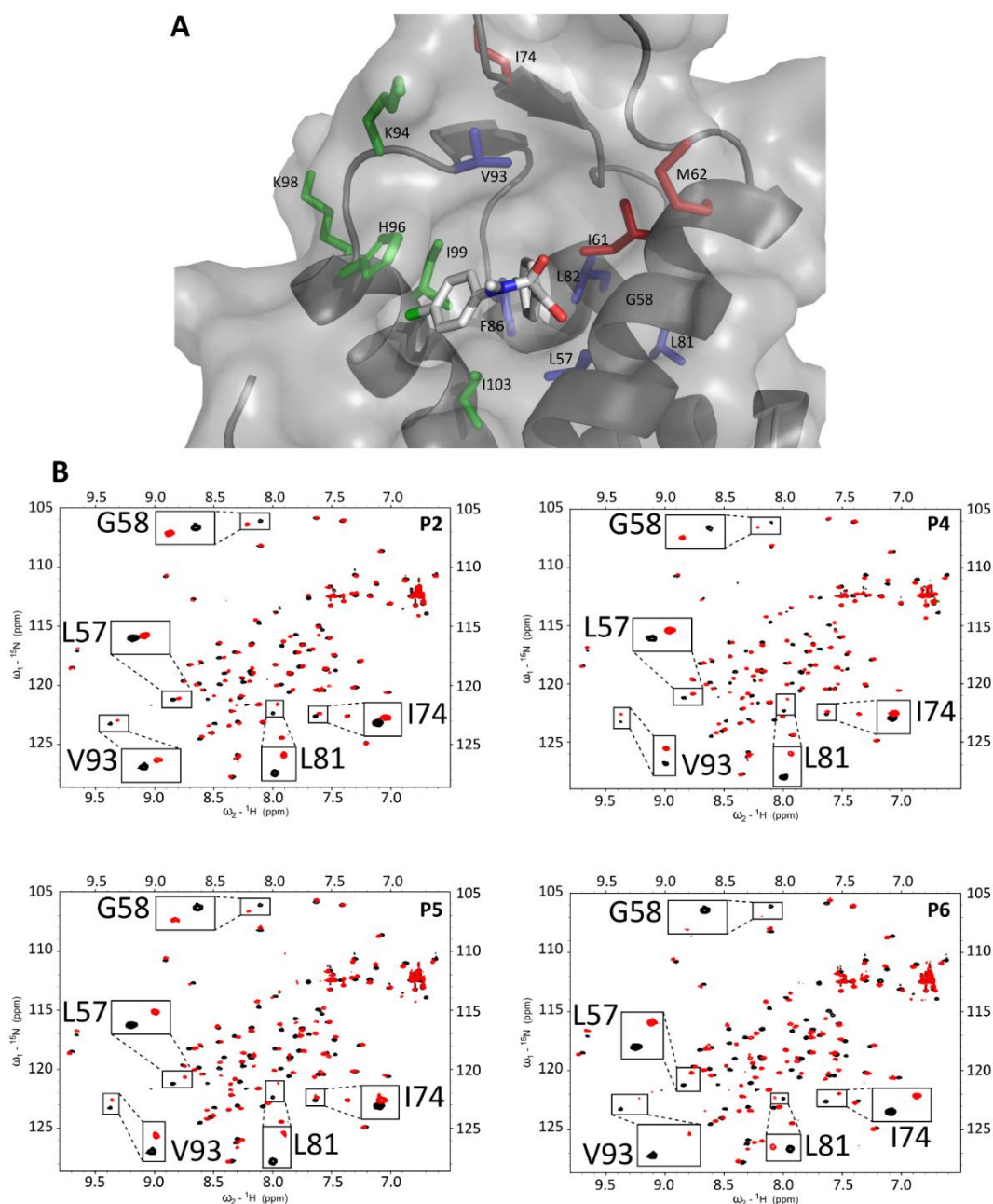


Figure 4.8. **A**: Highlighted residues. **B**: $^1\text{H}/^{15}\text{N}$ -HSQC NMR spectra for products **P2** and **P4 – P6**, where the black cross-peaks represent free *hDM2*₁₇₋₁₂₅ (50 μM) and the red cross-peaks *hDM2*₁₇₋₁₂₅ fully bound to a product ligand (100, 200 or 300 μM). Important residues involved in small-molecule binding to the *hDM2*₁₇₋₁₂₅ peptide binding cleft are highlighted and the respective perturbation of each $^1\text{H}/^{15}\text{N}$ cross-peak is expanded.

The products **P2** and **P4 – P6** were docked into an X-ray crystal structure of Human-MDM2 (*hDM2*) bound to a small molecule inhibitor (PDB: 6Q9H) using Autodock Vina (for experimental details, see: Chapter 5).²⁵⁵ The products were docked as single stereoisomers and the top ten docked

poses, scored by Vina, were then compared to the $^{15}\text{N}/^1\text{H}$ HSQC chemical shift perturbation data, and representative structures were chosen (Figure 4.10). Docking studies were also performed by Dr Chris Tinworth using MOE²⁵⁶ that identified similar predicted binding poses. Products **P2**, **P4** and **P5** showed good correlation with the experimental NMR data, forming a key interaction with the tryptophan and leucine sub-pockets in the top ranked structures. Products **P2** and **P4** demonstrate similar binding poses, rotating the bonds between the two phenyl rings to engage the peptide binding cleft and pointing the amide-carbonyl towards the solvent. Each enantiomer of **P2** gave similar predicted binding poses and rankings suggesting that the absolute configuration of the ligand is not important for *hDM2* binding.

Interestingly the (*R*)- and (*S*)-enantiomers of **P6** have different predicted binding poses, with (*R*)-**P6** predicting only interactions with the tryptophan and leucine sub-pockets, and (*S*)-**P6** predicting interactions with either the tryptophan and leucine, or tryptophan and phenylalanine sub-pockets. This result mirrors observations in the $^{15}\text{N}/^1\text{H}$ HSQC experimental data, in that a flipped binding mode was possible for **P6**, however the individual enantiomers of **P6** have not been isolated so this conclusion cannot be attributed to stereochemistry alone.

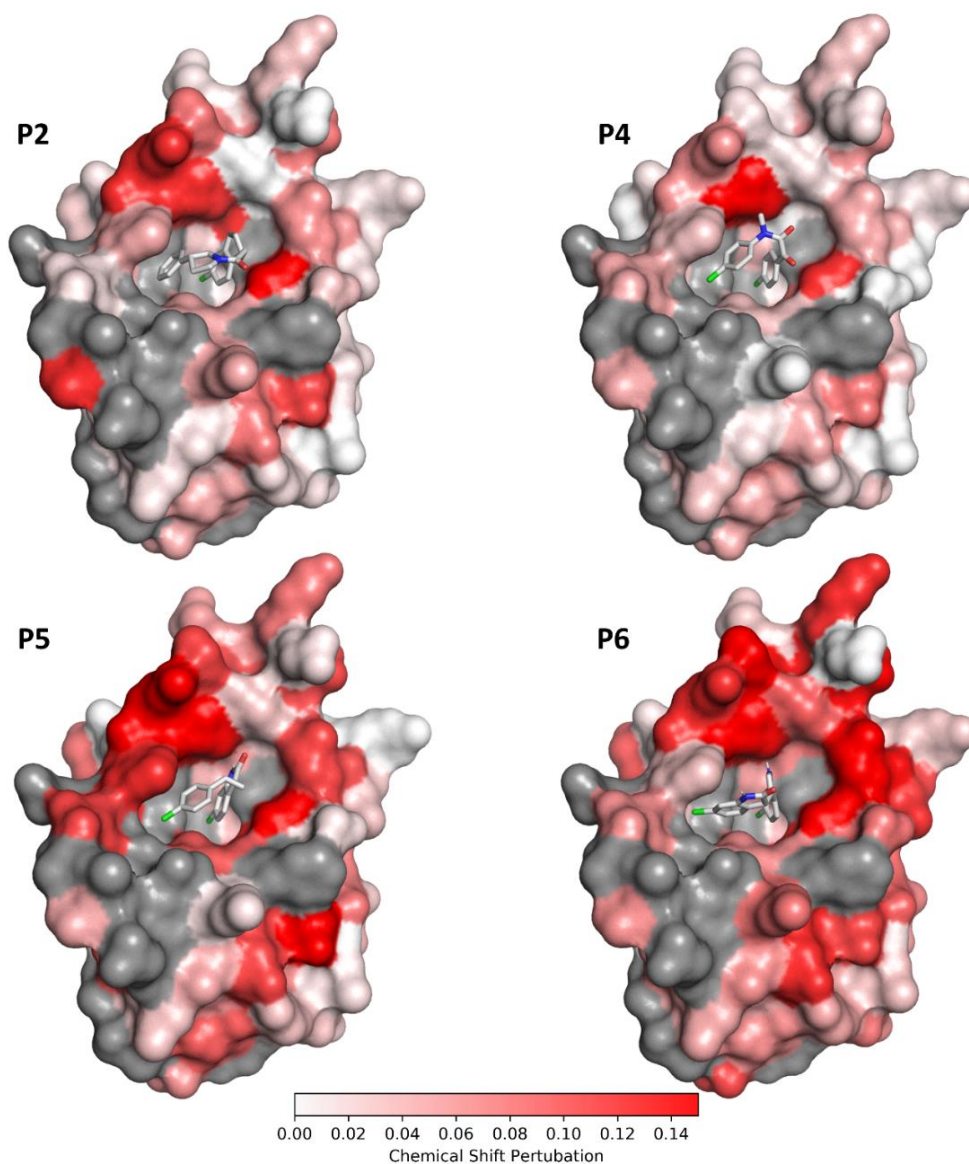


Figure 4.9. $^1\text{H}/^{15}\text{N}$ HSQC chemical shift perturbation of assigned peaks for $50\ \mu\text{M}$ ^{15}N -labelled *hDM2*₁₇₋₁₂₅ on addition of ligand (A: $200\ \mu\text{M}$ **P2**; B: $300\ \mu\text{M}$ **P4**; C: $100\ \mu\text{M}$ **P5**; D: $200\ \mu\text{M}$ **P6**) (unassigned residues highlighted in grey). The ligands (**P2** and **P4** – **P6**) are docked into *hDM2* (PDB: 6Q9H) using Autodock Vina.

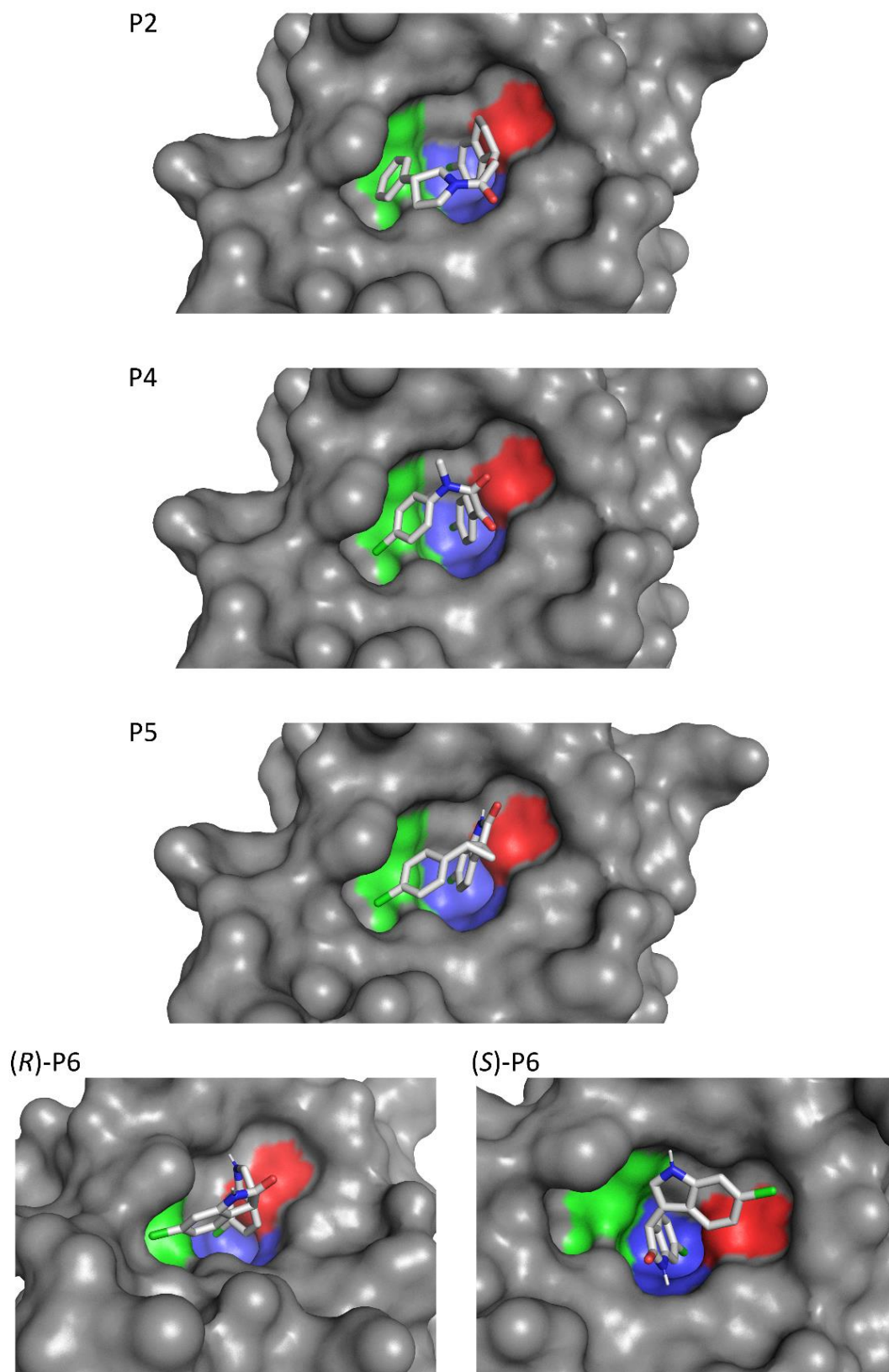


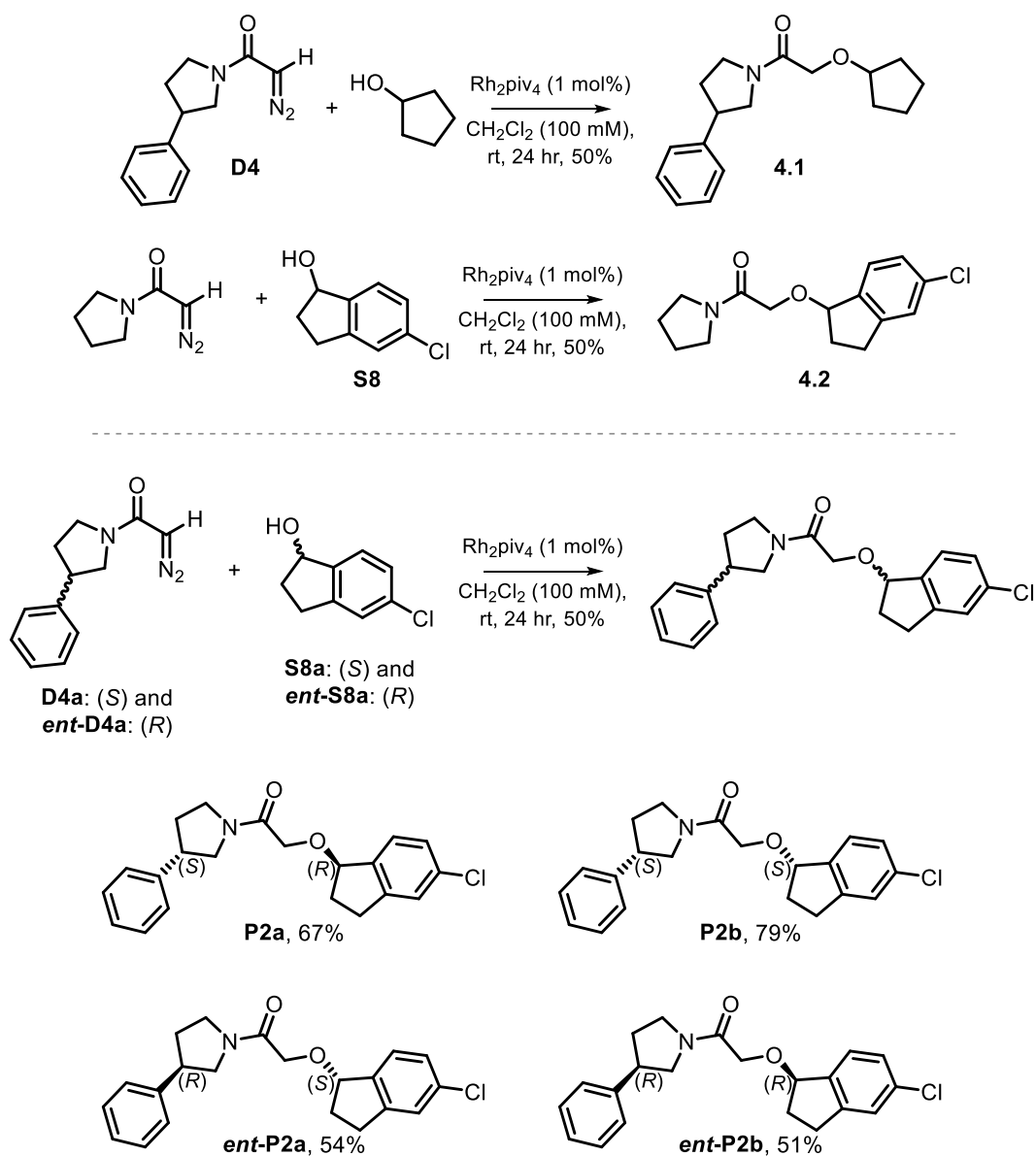
Figure 4.10. Products **P2**, **P4**, **P5**, (*R*)-**P6** and (*S*)-**P6** docked into the MDM2 peptide binding cleft (PDB: 6Q9H); the sub-pockets targeted by p53 hotspot residues F19 (red), W23 (blue) and L26 (green) are shown.

In summary, analysis of the $^1\text{H}/^{15}\text{N}$ -HSQC titration experiments elucidated some low-level structural information that confirmed the product ligands were interacting with the peptide binding cleft of *hDM2*₁₇₋₁₂₅ and were primarily occupying the tryptophan and leucine sub-pockets. A docking study predicted that **P2** and **P4 – P6** were preferentially interacting with the tryptophan and leucine sub-pockets yielding further structural information.

4.3 Development of Structure-Activity Relationships for Selected Human-MDM2 Inhibitor Classes

Development of structure-activity relationships are key in medicinal chemistry for the identification of functional groups that are important for the observed binding between a ligand and a protein target.^{7,55,257} A range of analogues for **P2**, **P5** and **P6**, that maintained the core scaffold of each ligand, was subsequently synthesised to elucidate structure-activity relationships between the chemotypes found in the validated product ligands. The analogues were synthesised using similar conditions to the scale-up of hit reaction mixtures described in chapter 3, using either Rh_2piv_4 or Rh_2pfb_4 as a catalyst (1 mol%), a diazo substrate (100 mM), and a co-substrate (Scheme 4.1).

Initially two analogues of **P2**, in which one of the aromatic rings had been removed, **4.1** and **4.2**, were synthesised (Scheme 4.1) and tested in the fluorescence anisotropy assay for inhibition of the p53₁₅₋₃₁ Flu/*hDM2*₁₇₋₁₂₅ protein-protein interaction (Figure 4.11) Neither analogue displayed a significant decrease in anisotropy up to 320 μM of **4.1** or **4.2** indicating that both aromatic rings are essential for the observed activity of **P2**. All four stereoisomers of **P2** were also individually prepared from enantiomerically pure starting materials (Scheme 4.1), tested in the fluorescence anisotropy assay and found to have similar activity: **P2a** (S,R) IC_{50} $28.2 \pm 3.0 \mu\text{M}$; **ent-P2a** (R,S) IC_{50} $36.5 \pm 3.5 \mu\text{M}$; **P2b** (S,S) IC_{50} $26.0 \pm 3.8 \mu\text{M}$; and **ent-P2b** (S,R) IC_{50} $28.6 \pm 2.9 \mu\text{M}$ (Figure 4.11). This result is consistent with the docking pose predictions (Chapter 4, section 4.2.2) where each stereoisomer of **P2** was able to adopt a similar binding pose that places the chlorophenyl aromatic ring in the tryptophan sub-pocket and the phenyl ring in the leucine sub-pocket.



Scheme 4.1. Synthesis of **P2** analogues.

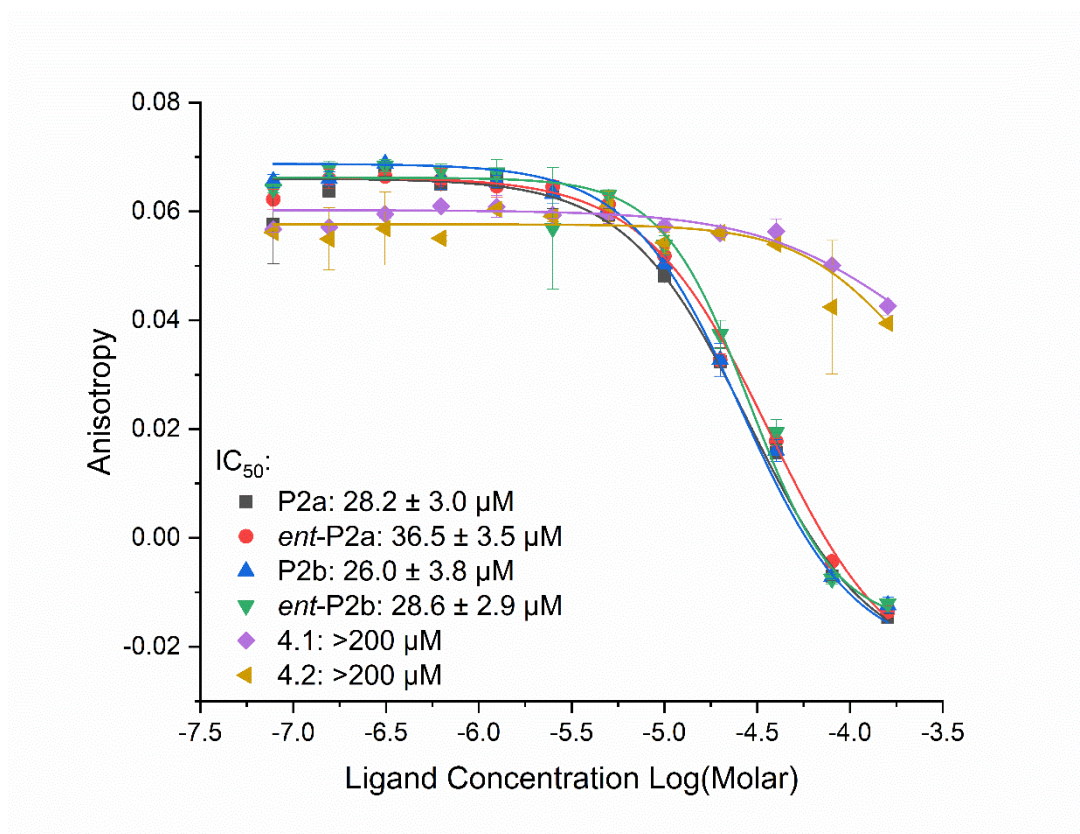
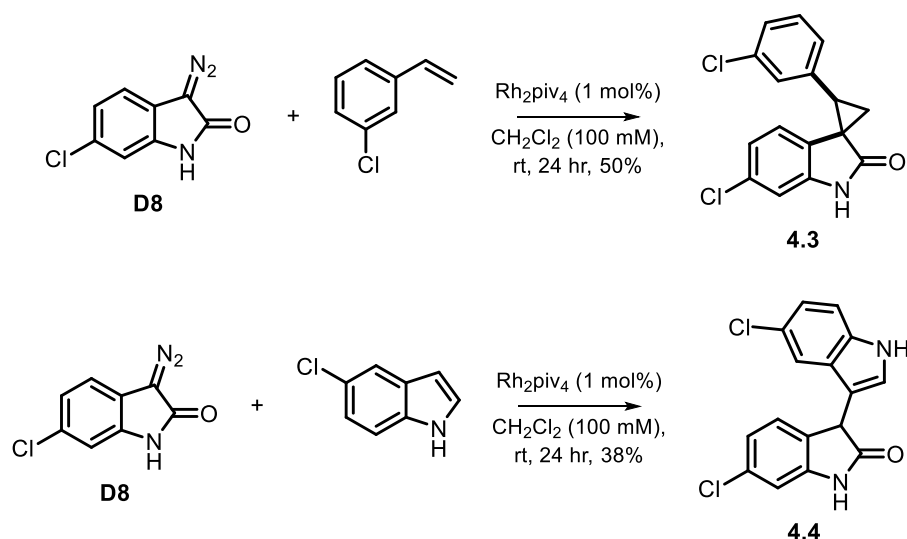


Figure 4.11. Dose-response experiments for analogues **4.1** (purple, IC_{50} : >200 μ M) and **4.2** (gold, IC_{50} : >200 μ M), and **P2** single enantiomers **P2a** (black, IC_{50} : 28.2 ± 3.0 μ M), **ent-P2a** (red, IC_{50} : 36.5 ± 3.5 μ M), **P2b** (blue, IC_{50} : 26.0 ± 3.8 μ M) and **ent-P2b** (green, IC_{50} : 28.6 ± 2.9 μ M) in the p53₁₅₋₃₁ FlU/hDM2₁₇₋₁₂₅ fluorescence polarisation assay, using aqueous phosphate buffer pH 7.5 containing 40 mM phosphate, 200 mM NaCl, 0.02 mg/mL Bovine Serum Albumin and 1% DMSO.

Analogues of **P5** and **P6** with chlorine atoms in different positions were also prepared and tested in the fluorescence anisotropy assay (Scheme 4.2). Analogue **4.3** was found to have reduced activity compared with **P5** and a full dose-response curve could not be obtained due to poor solubility in aqueous phosphate buffer pH 7.5 (**4.3**: >50 μ M and **P5**: 11.4 ± 1.2 μ M). Analogue **4.4** was also found to have reduced activity compared to **P6** (**4.4**: 3.5 ± 1.1 μ M and **P6**: 0.94 ± 0.03 μ M) confirming that the position of the chlorine atom was important for binding in both series of ligands.



Scheme 4.2. Synthesis of **P5** and **P6** analogues.

In summary, several analogues of products **P2**, **P5** and **P6** were synthesised (**4.1 – 4.4**, **P2a**, *ent*-**P2a**, **P2b** and *ent*-**P2b**) and tested in the fluorescence anisotropy assay. The observed activity of the analogues highlighted some simple structure-activity relationships demonstrating that the chlorophenyl and phenyl rings are key for inhibition of the p53¹⁵⁻³¹ Flu/*hDM2*¹⁷⁻¹²⁵ interaction.

4.4 Molecular Similarity and Comparisons with Known MDM2 Ligands

One of the central aims of an activity-directed synthesis experiment is to discover new, active molecules that are highly dissimilar to molecules discovered using more traditional discovery workflows. To assess the distinctiveness of the products, **P2** and **P4 – P6** were compared to 1314 ligands with annotated activity against MDM2 in the ChEMBL database.²⁵⁸ The compounds included in the ChEMBL database cover at least eight discovery campaigns for clinical candidates, including RG7112,¹⁴⁹ RG7388,¹⁵³ MI-77301,¹⁵⁰ AMG232,¹⁵⁴ DS-3032,²⁵⁹ HDM201,¹⁵⁵ CGM097¹⁵⁵ and MK-8242,²⁶⁰ and compounds from over 300 scientific papers. Comparison of the products with compounds deposited into the ChEMBL database, therefore, represents a robust assessment of the diversity of products that can emerge from activity-directed synthesis experiments.

The Morgan fingerprint (generated in RDKit) of each product was compared in a pair-wise manner to the Morgan fingerprint of each of the ligands deposited into ChEMBL using the Tanimoto similarity coefficient. The coefficient scores the similarity between a pair of molecular fingerprints where the maximum score indicating an identical molecule is one, and the minimum score indicating no similarity is zero.²⁶¹ For the analysis of organic small molecules, scores >0.7 indicate a pair of molecules with considerable similarity and scores <0.5 indicate molecules with little similarity.^{261,262} When compared in a pair-wise manner, each of the products gave very low similarity scores with known MDM2 ligands, with an average score of 0.37 across all four products, demonstrating that **P2** and **P4 – P6** are novel MDM2 ligands (Figure 4.13).

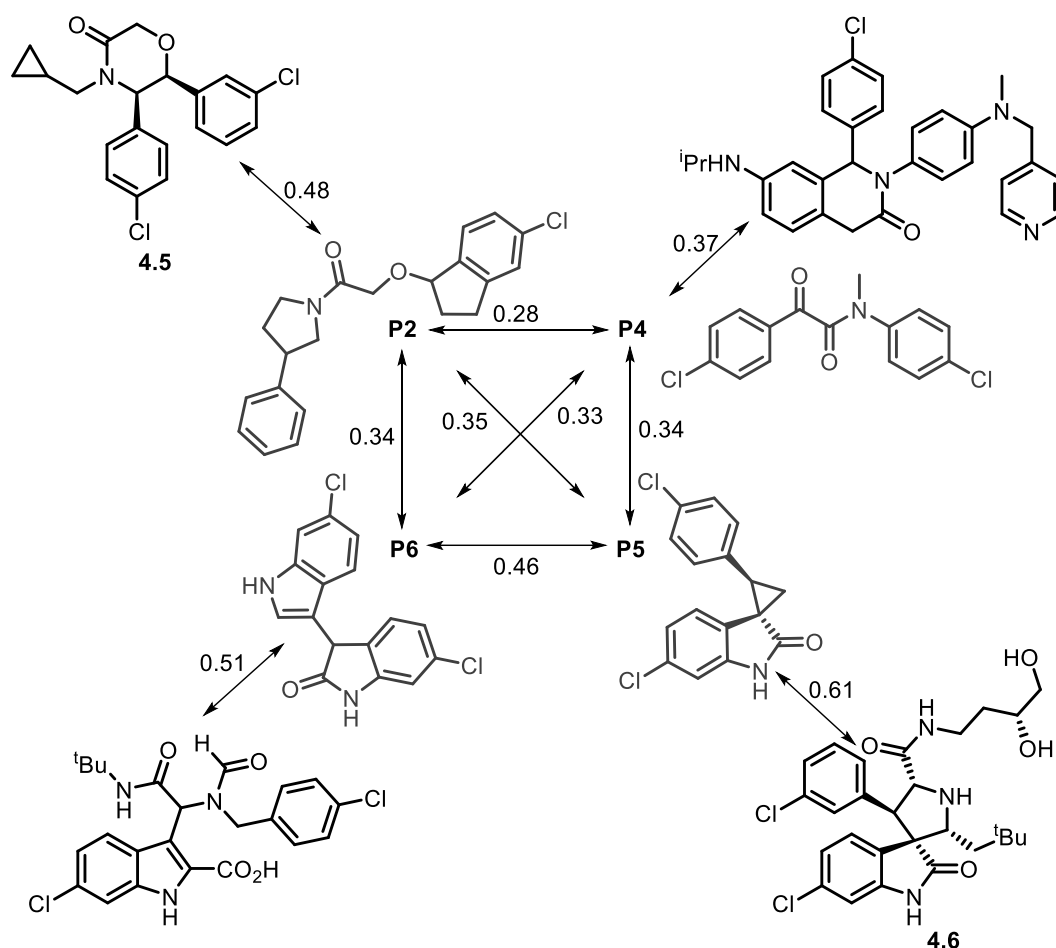


Figure 4.13. Molecular similarities (Tanimoto coefficient) of the products **P2** and **P4 – P6** and their nearest neighbour MDM2 ligands in ChEMBL.

The products also have low similarity scores with each other, a feature driven by the diversity in core scaffolding functionality between **P2** and **P4 – P6** (Figure 4.13). This is a result of the different fragment optimisation

strategies that have emerged from each of the product forming reactions. The O-H insertion reaction forming **P2** is comparable to fragment linking where two fragments are tethered by a flexible linker, while the cyclopropanation and C-H insertion reactions forming **P5** and **P6** are comparable to fragment merging where two fragments are joined through a rigid core.^{57,263,264} Finally, the O-H insertion and oxidation reactions that form **P4** act as a method of fragment growth where an existing ligand is altered to improve its activity.^{57,263} Discovering multiple series of new ligands from a single set of experiments with such diversity in scaffold functionality is uncommon and demonstrates the advantages of conducting activity-directed experiments for initial molecular discovery.

To further compare the similarity of the products with known MDM2 ligands, the top predicted docking poses for **P2** and **P4 – P6** (Section 4.2.2) were compared with X-ray crystal structures from the most similar ligands bound to MDM2 (Figure 4.14). All four of the products show significant complementarity between their docked binding poses and the X-ray structures of their nearest neighbours. Each product places the key aromatic moieties in the leucine and tryptophan sub-pockets in similar orientations to the known ligands, indicating that activity-directed experiments can be used as a method of experimental scaffold hopping.^{265,266}

Each overlay also indicates potential vectors for elaboration of the products. For example, substitution of the **P2** pyrrolidine-C2 position with an ethyl, (iso)propyl or sec-butyl group could engage the phenylalanine hotspot and create a ligand that interacts with all three MDM2 sub-pockets. Substitution of the **P5** cyclopropyl ring with a polar functional group at the C1' position, could engage the solvent facing edge of the peptide binding cleft and addition of an alkyl group at the C2' position could interact with the phenylalanine hotspot. This could create an analogue of **P5** that interacts with all three hot-spots and the solvent facing part of the peptide binding cleft similarly to the most potent inhibitors of the p53/MDM2 interaction. Clear growth vectors for **P4** and **P6** are also suggested by the overlays, specifically: substitution of the **P4** amide methyl group with a larger alkyl group; and addition of a polar functional group at the **P6** indole-C2 position could also create new interactions with the MDM2 peptide binding cleft.

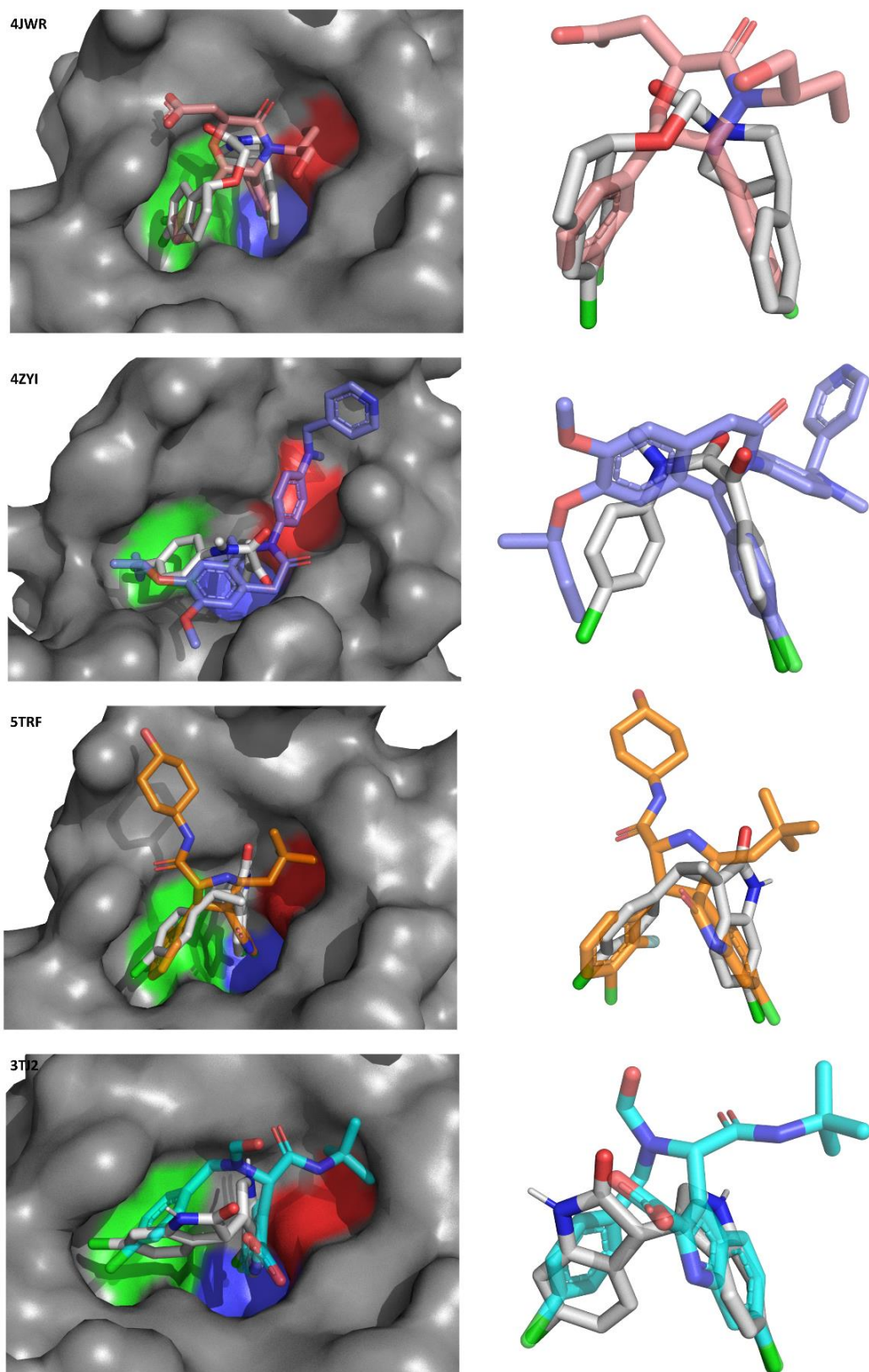


Figure 4.14. Overlays of the predicted binding poses for products **P2** and **P4** – **P6**, and crystal structures of their nearest neighbour MDM2 ligand in the ChEMBL database.^{151,155,267}

The ligands discovered by activity-directed synthesis (**P2** and **P4 – P6**) have favourable molecular properties with ligand efficiencies ranging from 0.28 to 0.4, molecular weights below 400, and fewer than three rotatable bonds. A common method of assessing the quality of a small molecule ligand is to compare the ligand efficiency (LE) and lipophilic ligand efficiency (LLE).^{35,58} LLE is a useful measure of a ligand's affinity towards its protein target compared to its affinity towards octanol, meaning that scores >0 indicate favourable binding to a protein. Ligands identified during early-stage drug discovery for protein-protein interactions typically have LE >0.2 due to the challenging nature of identifying ligands for proteins that do not have an evolved small molecule binding site.³⁵ Ligands with LE >0.3 and LLE >0 are considered desirable starting points for a hit-to-lead medicinal chemistry effort.^{35,57,58} Often when hit molecules are optimised into lead compounds LE is maintained and LLE is improved, and clinical candidates typically have high LLE values (>5).³⁵

Products **P4** and **P6** both have LE >0.3 and LLE >0 (LE: 0.31 and LLE: 0.32, and LE: 0.39 and LLE: 1.15 respectively) making them highly tractable starting points for further development as inhibitors regardless of target or discovery method (Figure 4.15). Products **P2** and **P5** also have favourable LE and borderline LLE (LE: 0.26 and LE: -0.14, and LE: 0.34 and LLE: -0.78 respectively) and may also be starting points for the discovery of inhibitors of protein-protein interactions. Interestingly, only 8 ligands in the analysis, including **P6**, have LE \geq 0.4 and \leq 21 heavy atoms which reflects on the main discovery methods employed against MDM2 (HTS and virtual screening) and the fact that no fragment screening campaigns have been reported. Comparing the LE and LLE of the products with the hit molecules that were identified at the start of the discovery campaigns yielding the clinical candidates AMG-232,¹⁵⁴ CGM097,^{155,268} MI-77301¹⁵⁰ and RG7388¹⁵³ shows that the products occupy the same ideal molecular property space as the original hit molecules (Figure 4.15).

The hit molecule that initiated the discovery of AMG-232 (**4.5**) is also the most similar ligand to **P2**, and surprisingly both molecules have similar potency in their respective fluorescence anisotropy assays (**4.5**: 11.7 ± 1.8 μ M and **P2**: 15.0 ± 0.1 μ M). The most similar ligand to **P5** is the hit molecule that was synthesised early in the structure-based design approach that was used for the discovery of MI-77301, and has very similar LE/LLE values (**4.6**, LE: 0.31 and LLE: -1.3, and **P5**, LE: 0.34 and LLE: -0.78). The most similar ligand to **P4** is a molecule that was synthesised during the hit-to-lead

campaign for the discovery of CGM097 and **P4** has superior ligand efficiency compared to the initial hit (hit, LE: 0.22 and LLE: 1.62, and **P4**: LE: 0.31 and LLE: 0.32). The ligands discovered by activity-directed synthesis, therefore, occupy a section of chemical space that is highly relevant to the p53/MDM2 protein-protein interaction and to medicinal chemists undertaking drug discovery projects. Ultimately, each of the ADS products could have been used to initiate a drug discovery campaign against this PPI.

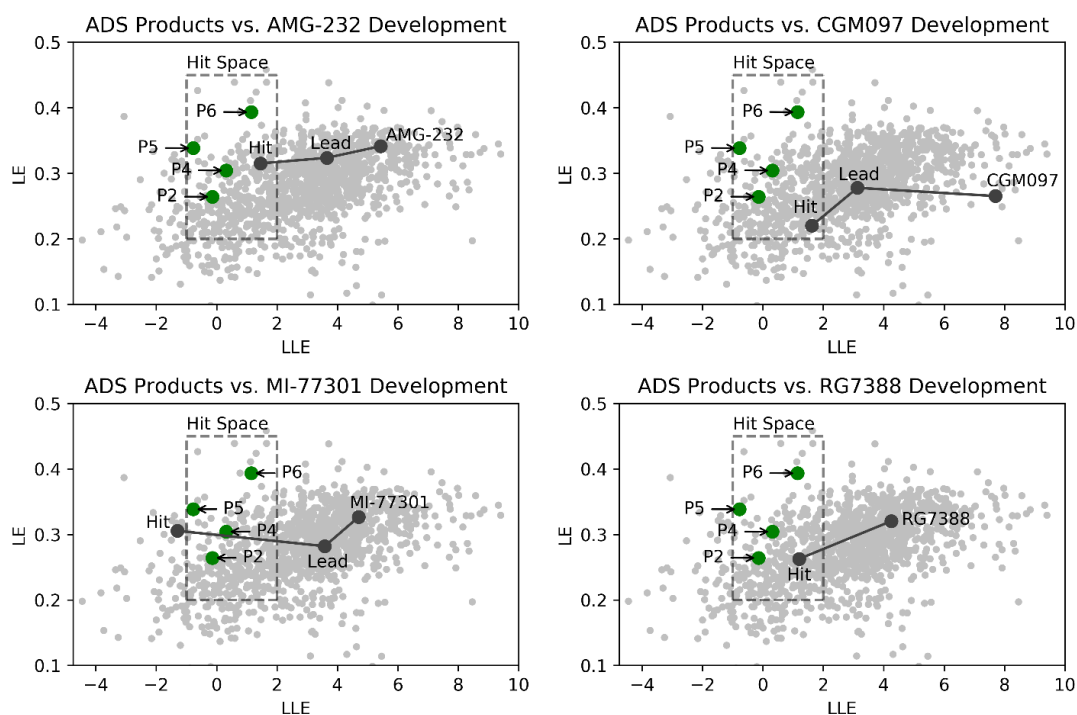


Figure 4.15. Comparison of lipophilic ligand efficiency (LLE) and ligand efficiency (LE), and the change in both values as drug development progresses from hit to lead to clinical candidate. Product ligands **P2** and **P4** – **P6** – **P6** are highlighted in green, clinical candidates AMG-232, CGM097, MI-77301 and RG7388 are in black, and ligands deposited to ChEMBL are in grey.

In summary, products **P2** and **P4** – **P6** were found to be unique starting points for molecular discovery campaigns due to their low similarity to known MDM2 ligands and favourable LE/LLE metrics. The activity-directed workflow has also enabled an experimental scaffold hopping approach where the product ligands can make similar interactions with the MDM2 peptide binding cleft but with very different core scaffolds. The four-product series each represent significant chemical matter that could be used to form the basis of a drug discovery effort to make novel inhibitors of the p53/MDM2 protein-protein interaction.

4.5 Conclusions

In summary, activity-directed synthesis has been demonstrated as an effective tool for the discovery of inhibitors against a challenging biological target that does not have an evolved small molecule binding site. The efficiency of activity-directed synthesis allows for large-scale exploration of chemical space with a remarkably low requirement for multi-step chemical synthesis. This is highlighted by the low number of reactions conducted in conventional labware. For the two high-throughput reaction arrays, totalling 346 microscale reactions, ten diazo substrates and one co-substrate were synthesised, and nine hit reaction mixtures were scaled up. The total number of reactions completed in conventional labware to discover the four product ligands was 20, which represents a time and resource efficient method for the identification of bioactive small molecules.

In traditional early-stage drug discovery, fragment or HTS compound libraries are meticulously curated and optimised for diversity in parameters such as shape, molecular properties and novelty to give hits favoured by medicinal chemists. Through the utilisation of highly reactive and promiscuous rhodium(II)-carbene chemistry each of the ligand series discovered by activity-directed synthesis have distinct scaffolds and high novelty, without the need for predetermined engineering of the reaction products. This could enable the discovery of new ligands for targets that do not have elucidated structural information for the biological interaction of interest, or ligands for targets where many molecules have been reported and a new starting point is desired.

The observed diversity of reaction products has been driven by two factors. Firstly, by using diazo substrates and catalysts with different reactivity multiple types of scaffold forming reactions can occur, and additionally these reaction types are underrepresented in the synthesis of screening collections. Secondly, by utilising α -diazo amide substrates the product molecules have low similarity with known ligands, as carbon-carbon bond forming reactions α -to amide carbonyls are not very common. Also, the fragment-like properties of the diazo substrates and co-substrates gives the products favourable molecular properties. As a result, each of the product ligand series fall within limits imposed during the evaluation of hits commonly identified by high-throughput screening.

The active products also have comparable potency and efficiency to the hit molecules that started drug discovery campaigns for several clinical

candidate molecules targeting the p53/MDM2 protein-protein interaction. The products from activity-directed synthesis experiments could have been selected for further investigation in a pharmaceutical setting and used as starting points for drug discovery. Retrospective analysis of the discovery of the nutlin ligands showed that fragment-based drug discovery could have been used to identify the critical chemical matter for drug discovery, but that suitable fragment libraries would not have been available at the time. Activity-directed synthesis could alternatively form the basis of an approach for the discovery of these ligand classes without the need for extensive investment in chemical synthesis. This could allow for a reduction in the size of the screening collections maintained by pharmaceutical companies as early screening can be conducted with low requirements for synthetic chemistry. Activity-directed synthesis, therefore, enables a more dynamic approach to small molecule discovery as the chemical matter screened is not fixed at the time of library design and could allow for target specific information to be better incorporated into early work.

Future applications of activity-directed synthesis could focus on the application of late-stage functionalisation chemistry for the elaboration of hit molecules for the p53/MDM2 protein-protein interaction (Figure 4.16). Employing activity-directed synthesis in a hit-to-lead paradigm would demonstrate that the approach could also be used in the optimisation of a known ligand, as well as for hit discovery. Late-stage functionalisation such as hydroxylation, fluorination, C-H activation and C-H amination, and decarboxylative cross-couplings are well established techniques for the site-selective modification of complex molecules and could be used in an activity-directed synthesis workflow for the optimisation of hit molecules.²⁶⁹

Photoredox-based Minisci chemistry could be used for the late-stage modification of hit molecules and the chemistry has been established in activity-directed synthesis format by Dr Andrew Gomm. The photoredox-mediated process is cross-dehydrogenative between a saturated N-heterocycle and a heteroarene and is ideal for the site-specific elaboration of a ligand, for example 4.7 (Figure 4.17). The primary p53¹⁵⁻³¹ Flu/*hDM2*¹⁷⁻¹²⁵ fluorescence anisotropy assay has also been shown to tolerate a number of different reaction conditions for these transformations and, as such, the approach could be implemented within the established high-throughput screening protocol outlined in Chapter 3 (Appendix A).

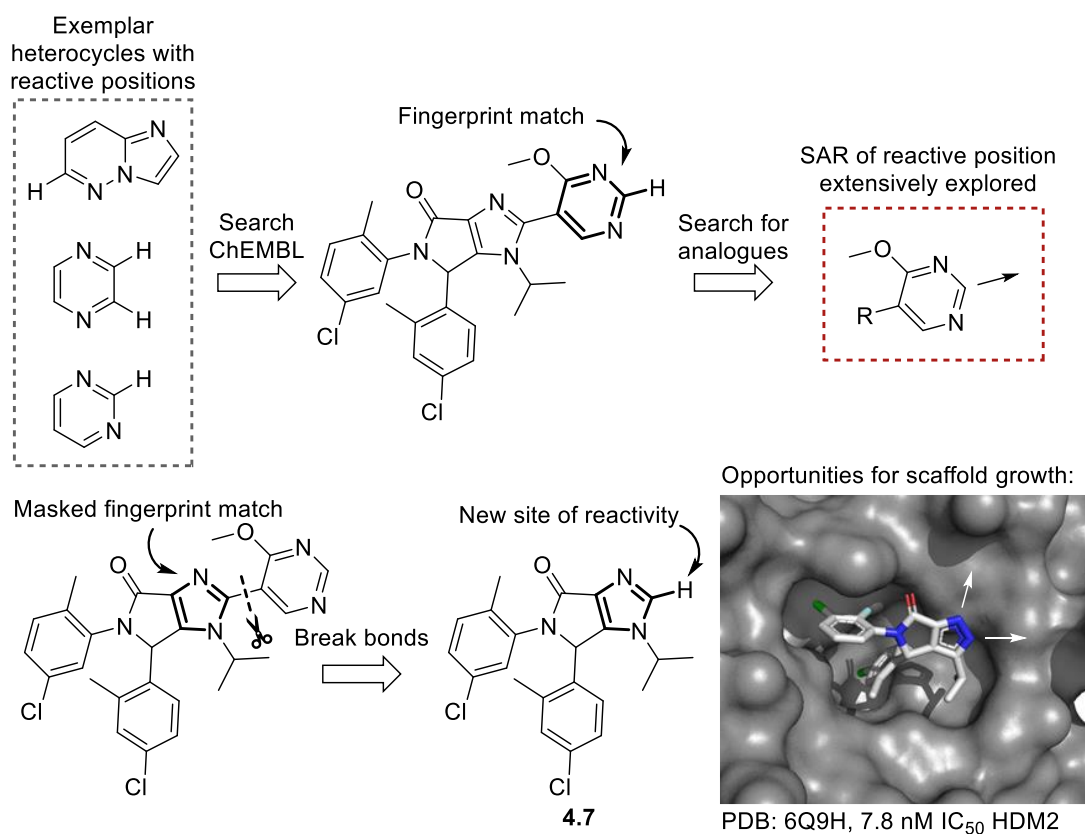


Figure 4.16. Workflow for identifying a potential *hDM2* ligand, **4.7**, as a photoredox substrate. A closely related analogue was also reported to have an IC₅₀ of 7.8 nM in a TR-FRET assay with 0.1 nM MDM2 and 10 nM labelled-p53 peptide.²⁷⁰

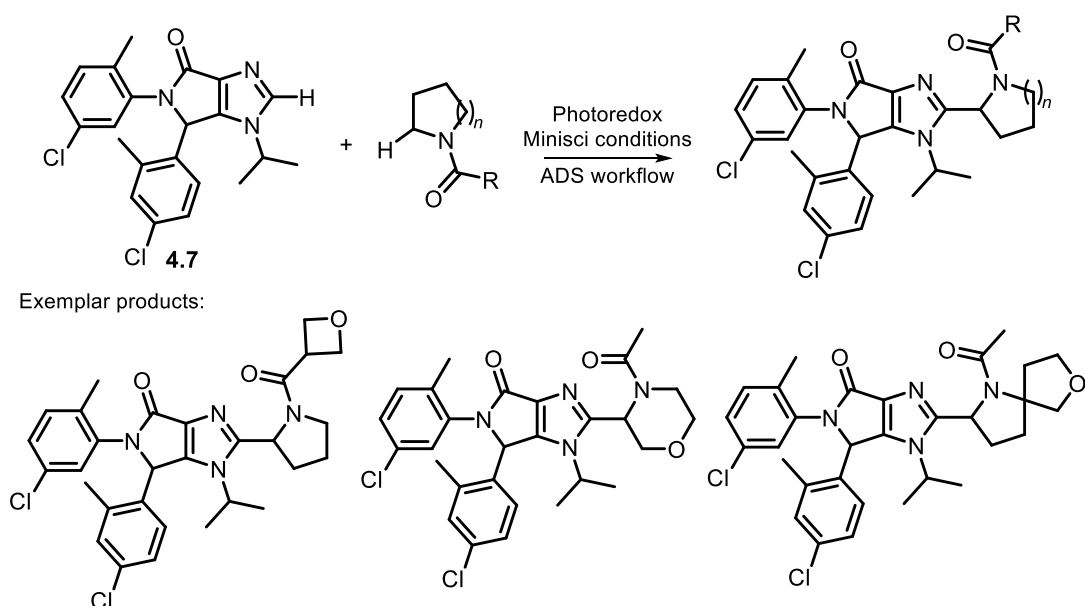


Figure 4.17. Hypothetical ADS reaction products from a reaction array using the photoredox Minisci reaction.

Activity-directed synthesis experiments work robustly for the identification of new hit molecules, and future work focussing on developing a workflow for the hit-to-lead optimisation of a ligand would demonstrate a well-rounded platform for molecular discovery. Ultimately, the work described in this thesis has shown that activity-directed synthesis can drive molecular discovery, while also augmenting other well-established methods, and could be a highly productive addition to current pharmaceutical screening technologies.

Chapter 5 Experimental

5.1 General Information and Instrumentation

Reagents were purchased from commercial suppliers and used without further purification (Sigma-Aldrich, Fluorochem, Strem Chemicals, Enamine BB (EU), and Scientific Laboratory Supplies). Rhodium catalysts were purchased from Sigma-Aldrich or Strem Chemicals and used as supplied. $\text{Rh}_2(5R\text{-MEPY})_4$ was synthesised using a reported literature procedure.²⁷¹ Water-sensitive reactions were performed in anhydrous solvents obtained from a PureSolv MD5 Purification System. Solvents were removed under reduced pressure using a Büchi rotary evaporator with a Vacuubrand PC2001 Vario diaphragm pump, or under N_2 blowdown at 40 °C. Flash column chromatography was carried out using silica gel 60 (35-70 μm particles) supplied by Merck or using RediSep *Rf* silica cartridges and a Teledyne ISCO CombiFlash automatic chromatography companion. Thin-layer chromatography was conducted with Macherey-Nagel Polygram SIL G/UV254 0.2mm silica gel 60 with fluorescent indicator plates.

Ultra-high performance liquid chromatography (UPLC) was performed using an Agilent 1260 Infinity System (5-minute method, 1 mL/min, 0.5 μL injection) equipped with an X-select Waters Charged Surface Hybrid (CSH) C18 2.5 μm (30 x 2.1 mm) C18 column (XP) or an Agilent 1290 Infinity II system (1-minute method, 2.2 mL/min, 0.5 μL injection) equipped with an X-select Waters Charged Surface Hybrid (CSH) C18 2.5 μm (30 x 2.1 mm) C18 column (XP) at 40 °C and 220 nm wavelength. Analytical LC-MS was performed using several systems. Firstly, a system comprising an Ultimate3000 HPLC instrument with a Bruker Amazon Speed MS detector with electrospray ionisation. The system ran with a positive and negative switching mode and UV diode array detector using a Phenomenex Kinetex C18 (50 mm x 2.1 mm x 2.6 μm) column and gradient elution with two binary solvent systems: MeCN/ H_2O or MeCN/ H_2O plus 0.1% formic acid. Secondly, analytical LC-MS was conducted on a Waters Acquity UPLC instrument equipped with an Acquity UPLC CSH C18 column and Waters micromass ZQ MS using alternate-scan positive and negative electrospray. Analytes were detected as the summed UV wavelength of 210–350 nm. Accurate mass spectrometry (HRMS) was performed using electrospray ionisation on

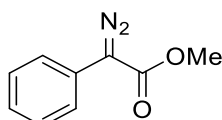
a Bruker MaXis Impact spectrometer equipped with a Phenomenex Luna C18 50 x 2 mm 5 micron column with two solvent systems: MeCN/H₂O + 0.1% formic acid or MeCN/H₂O. Mass-directed automatic purification (MDAP) was conducted on either an Xbridge C18 column or Sunfire C18 column at ambient temperature using a modified acetonitrile-water gradient (as specified in the experimental section). MS analysis was carried out using a Waters ZQ MS using alternate-scan positive and negative electrospray and a summed UV wavelength of 210–350 nm.

NMR analysis was conducted using a Bruker AV-400 spectrometer (¹H = 400 MHz, ¹³C = 100 MHz and ¹⁹F = 376 MHz C-F decoupled), Bruker AV-500(Cyroprobe) spectrometer (¹H = 500 MHz and ¹³C = 125 MHz), JEOL ECA600ii 14.1 T spectrometer (¹H = 600 MHz and ¹³C = 150 MHz), 750 MHz Oxford Magnet spectrometer (TCI-Cyroprobe, ¹H optimized triple resonance NMR 'inverse' probe) (¹H = 750 MHz and ¹⁵N = 76 MHz), or a 600 MHz Oxford Magnet spectrometer (QCI-P-Cryoprobe, ¹H optimised quadruple resonance NMR 'inverse' probe) (¹H = 600 MHz and ¹⁵N = 61 MHz) using an internal deuterium lock. Chemical shifts are quoted in parts per million (ppm) and coupling constants are given in Hz. Splitting patterns have been abbreviated as follows: s (singlet), d (doublet), dd (doublet of doublets), t (triplet), q (quartet) and m (multiplet). NMR data is reported in the format: ppm (number of protons, splitting pattern, coupling constant). Infrared spectra were recorded on a Bruker Alpha ATR FR-IR spectrometer; absorptions are reported in wavenumber (cm⁻¹).

5.2 Materials and Methods for the Catalyst Knowledge Base

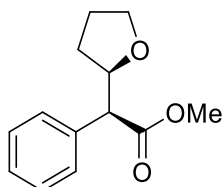
5.2.1 Synthesis of Compounds for the High-Throughput Screening of Model Reactions

Methyl 2-diazo-2-phenylacetate, **5**²³¹



DBU (7.0 mL, 46.6 mmol) was added slowly to a stirred solution of methyl 2-phenylacetate (4.7 mL, 33.3 mmol) and 4-acetamidobenzenesulfonyl azide (9.6 g, 40.0 mmol) in acetonitrile (60 mL) at 0 °C. The reaction was allowed to warm to room temperature overnight, and after 20 hours of stirring gave a bright orange solution. The solvent was then removed under reduced pressure, and the residue dissolved in CH₂Cl₂ (25 mL) and partitioned with distilled water (75 mL). The product was subsequently extracted using Et₂O (3 x 75 mL), and the organics collected and sequentially washed with 10% NH₄Cl (3 x 30 mL) and brine (3 x 30 mL). The organics were collected and dried (Na₂SO₄), and the solvent was removed under reduced pressure to give a bright orange oil. The product was purified by silica gel chromatography eluting 5:95 tert-butylmethylether/cyclohexane to afford the *diazo* **5**²³¹ as a bright orange oil (4.3 g, 73%). *R*_f (90:10 Cyclohexane/Et₂O) 0.49. δ_H (500 MHz, Acetone-d₆) 7.54 (2H, app. dd, *J* 8.5, 1.1 Hz, phenyl-2H and -6H), 7.41 (2H, app. dd, *J* 8.5, 7.6 Hz, phenyl-3H and -5H), 7.22 – 7.19 (1H, m, phenyl-4H) and 3.85 (3H, s, methyl-H). LCMS (ESI): C₉H₈N₂O₂ requires [M+H]⁺, calculated 177.1, found 177.1.

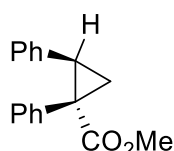
Methyl 2-(oxolan-2-yl)-2-phenylacetate, **7**²⁷²



A three-neck 50 mL round-bottom flask, equipped with a dropping funnel, was charged with Rh₂oct₄ (2.5 mg, 3.2 μmol) and degassed under a stream of N₂. THF (0.5 mL) was added and the dropping funnel charged with methyl 2-diazo-2-phenylacetate (112.8 mg, 0.6 mmol) in heptane (4.5 mL) and degassed by bubbling a stream of N₂ through the solution. After 5 minutes of

degassing, the stream of N₂ was slowed and removed from the bulk solvent, and the diazo solution added dropwise over 105 minutes. Heptane (1 mL) was then added to the dropping funnel to wash through any remaining diazo. After 16 hours the solvent was removed under reduced pressure to give a blue residue. The residue was purified by MDAP (10 mM ammonium bicarbonate modifier gradient), and the appropriate fractions were combined and dried under a stream of nitrogen to afford the *acetate* **7**²⁷² as a colourless oil (74 mg, 52%, dr 2.7:1). δ_{H} major (500 MHz, CDCl₃) 7.31 (2H, d, *J* 8.2 Hz, phenyl-2H and -6H), 7.27 – 7.25 (2H, m, phenyl-3H and -5H), 7.22 – 7.19 (1H, m, phenyl-4H), 4.38 (1H, dt, *J* 8.3, 7.0 Hz, oxolanyl-2H), 3.73 (1H, dt, *J* 8.2, 7.0 Hz, oxolanyl-5H_a), 3.66 – 3.62 (1H, m, oxolanyl-5H_b), 3.60 (3H, s, methyl-H), 3.56 (1H, m, acetate-2H), 2.08 – 2.02 (1H, m, oxolanyl-3H_a), 1.82 – 1.78 (2H, m, oxolanyl-4H₂) and 1.60 (1H, app. ddd, *J* 15.8, 12.3, 8.0 Hz, oxolanyl-3H_b). δ_{H} minor (500 MHz, CDCl₃) 7.29 – 7.19 (5H, m, phenyl), 4.45 (1H, dt, *J* 9.9, 7.0 Hz, oxolanyl-2H), 3.85 (1H, dt, *J* 8.4, 7.0 Hz, oxolanyl-5H_a), 3.78 – 3.72 (1H, m, oxolanyl-5H_b), 3.63 (3H, s, methyl-H), 3.46 (1H, m, acetate-2H), 1.82 – 1.73 (2H, m, oxolanyl-3H_a and -4H_a), 1.67 – 1.56 (1H, m, oxolanyl-4H_b) and 1.36 (ddd, 1H, *J* 16, 12.6, 7.6 Hz, oxolanyl-3H_b). HRMS (ESI): C₁₃H₁₆O₃ requires [M+Na]⁺, calculated 243.0997, found 243.0993.

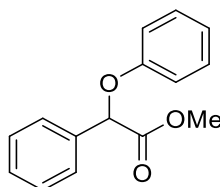
Methyl-1,2-diphenylcyclopropane-1-carboxylate, **8**^{273,274}



A three-neck round bottom flask, equipped with a dropping funnel, was charged with Rh₂tfa₄ (2.1 mg, 3.2 μ mol) and degassed with a stream of N₂. Styrene (0.2 mL, 2.8 mmol) was added and the dropping funnel charged with methyl 2-phenyldiazoacetate (100 mg, 0.6 mmol) in CH₂Cl₂ (4.1 mL). The diazo solution was then degassed by bubbling a stream of N₂ through the solution and after 5 minutes the stream of N₂ was slowed and removed from the bulk solvent. The diazo was added dropwise to the solution over 2.5 hours, and a further aliquot of CH₂Cl₂ (1.0 mL) was added to the dropping funnel to wash through any remaining diazo. After 16 hours of stirring the solvent was removed under reduced pressure to give a blue residue and was purified by MDAP (+0.1% TFA modifier gradient). The appropriate fractions were combined and dried under a stream of nitrogen to afford the

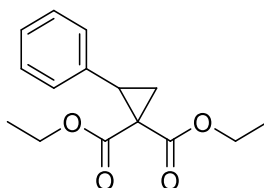
cyclopropane **8**^{273,274} as a thick colourless oil (113 mg, 70%, dr 95:5). δ_{H} (500 MHz, CDCl_3) 7.15 – 7.13 (3H, m, 1-phenyl-2H, -4H and -6H), 7.08 – 7.03 (5H, m, 1-phenyl-3H and -5H, and 2-phenyl-2H, -4H and -6H), 6.79 – 6.77 (2H, m, 2-phenyl-3H and 5H), 3.67 (3H, s, methyl), 3.13 (1H, dd, J 9.3, 7.3 Hz, cyclopropane-2H), 2.15 (1H, dd, J 9.3, 4.9 Hz, cyclopropane-3H_a) and 1.90 (1H, dd, J 7.3, 4.9 Hz, cyclopropane-3H_b). HRMS (ESI): $\text{C}_{17}\text{H}_{16}\text{O}_2$ requires $[\text{M}+\text{Na}]^+$, calculated 275.1048, found 275.1042.

Methyl 2-phenoxy-2-phenylacetate, **9**²⁷⁵



A three-neck round bottom flask, equipped with a dropping funnel, was charged with Rh_2pfb_4 (3.4 mg, 3.2 μmol) and phenol (602.3 mg, 6.4 mmol), and degassed with a stream of N_2 . CH_2Cl_2 (1.0 mL) was added and degassed by bubbling a stream of N_2 through the stirred solution. The dropping funnel was then charged with methyl 2-diazo-2-phenylacetate (112.8 mg, 0.7 mmol) in CH_2Cl_2 (4.5 mL) and degassed by bubbling a stream of N_2 through the solution. After 5 minutes of degassing, the stream of N_2 was slowed and removed from the bulk solvent, and the diazo solution added dropwise over 105 minutes. CH_2Cl_2 (1.0 mL) was then added to the dropping funnel to wash through any remaining diazo. After 16 hours the solvent was removed under reduced pressure to give a blue residue and was purified by MDAP (10 mM ammonium bicarbonate modifier gradient). The appropriate fractions were combined and dried under a stream of nitrogen to afford the acetate **8**²⁷⁵ as a colourless oil (89 mg, 57%). δ_{H} (500 MHz, CDCl_3) 7.50 (2H, m, phenoxy-2H and -6H), 7.34 – 7.28 (3H, m, phenoxy-3H, -4H and -5H), 7.21-7.17 (2H, m, phenyl-2H and -6H), 6.91 – 6.86 (3H, m, phenyl-3H, -4H and -5H), 5.57 (1H, s, phenylacetate-2H) and 3.66 (3H, s, methyl-H). HRMS (ESI): $\text{C}_{15}\text{H}_{14}\text{O}_3$ requires $[\text{M}+\text{Na}]^+$, calculated 265.0840, found 265.0834.

1,1-Diethyl 2-phenylcyclopropane-1,1-dicarboxylate, **10²⁷⁶**



A three-neck round bottom flask, equipped with a dropping funnel, was charged with Rh_2tfa_4 (1.8 mg, 2.7 μmol) and degassed with a stream of N_2 . Heptane (6.0 mL) was added and degassed by bubbling a stream of N_2 through the stirred solution, followed by the addition of styrene (90 μL , 1.3 mmol). The dropping funnel was charged with diethyl 2-diazomalonate (50 mg, 0.3 mmol) in heptane (2.1 mL) and degassed by bubbling a stream of N_2 through the solution. After 5 minutes of degassing, the stream of N_2 was slowed and removed from the bulk solvent, and the diazo solution added dropwise over 105 minutes. Heptane (1.0 mL) was then added to the dropping funnel to wash through any remaining diazo. After 16 hours of stirring the solvent was removed under reduced pressure to give a blue residue which was purified by MDAP (10 mM ammonium bicarbonate modifier gradient). The appropriate fractions were combined and dried under a stream of nitrogen to afford the *cyclopropane* **10**²⁷⁶ as a colourless oil (10 mg, 14%). δ_{H} (400 MHz, CDCl_3) 7.28 – 7.22 (5H, m, 2-phenyl-H), 4.32 – 4.21 (2H, m, ethyl-1H₂), 3.89 – 3.84 (2H, m, ethyl-1'H₂), 3.24 (1H, app. t, J 8.6, cyclopropane-2H), 2.19 (1H, dd, J 8.0 and 5.2, cyclopropane-3H_a), 1.72 (1H, dd, J 9.2 and 5.0, cyclopropane-3H_b), 1.32 (3H, t, J 7.1, ethyl-2H₃) and 0.89 (3H, t, J 7.1, ethyl-2'H₃).

5.2.2 Protocols for the Screening of Model Reactions

Reactions were carried out in 0.75 mL shell vials (Chemglass CV-2100-0830) equipped with a teflon-coated stir bar (Biotage 0.2-0.5 mL magnetic stir bar #355545) and sealed using either a Freeslate 96-well reaction block or a Sigma-Aldrich Kitalysis 24-well reaction block (Z742107 Aldrich). Prior to the assembly of each reaction array the following stock solutions were made: diazo (**5** or **6**) in reaction solvent (1.25 M); catalyst in THF (25 mM); and a reactant (phenol) in reaction solvent (12.5 M). THF and styrene were added neat to reaction wells. Quantitative UPLC methods were developed by generating response factors for standard samples of products **7**, **8**, **9**, **10** against the external standard 4,4'-di-¹Bubiphenyl at 220 nM UV wavelength. All subsequent UPLC analysis was conducted at 220 nM detection wavelength and 0.5 mg/mL sample concentration relative to the initial mass of **5** or **6**.

Reaction assembly workflow (see general procedures for further detail):

1. Dispense catalyst stock to reaction vial and dry under vacuum
2. Add reaction solvent to each well for 200 μ L total volume
3. Stir for 2 minutes
4. Add reactant to each well (THF, styrene or phenol)
5. Add methyl 2-diazo 2-phenylacetate stock to each vial (3.5 mg per well)
6. Seal reaction block and stir for 24 hours at room temperature
7. Remove solvent by N₂ blowdown at 40 °C (16-20 hours)
8. Add 0.5 mL 4,4'-di-¹Bubiphenyl solution in MeCN (1.66 mg/mL standard, 0.83 mg per well)
9. Dilute 100 μ L sample into 700 μ L 80:20 MeCN/H₂O (0.5 mg/mL reactants, 0.025 mg/mL standard) and analyse reaction sample by UPLC at 220 nM UV wavelength

The mass of product was calculated using the following formula:

Product Mass

$$= \text{Product Response Factor} \times \text{Mass of External Standard} \\ \times \left(\frac{\text{Product Peak Area}}{\text{External Standard Peak Area}} \right)$$

UPLC Gradients for Reaction Screening

UPLC Gradient 5 minute (1 mL/min):

Mobile Phase A (Aqueous) - 0.05% TFA in water

Mobile Phase B (Organic) – 0.05% TFA in acetonitrile

Time (mins)	A(%)	B(%)
0	97.00	3.00
3.7	5.00	95.00
4.0	5.00	95.00
4.1	97.00	3.00
5.5	97.00	3.00

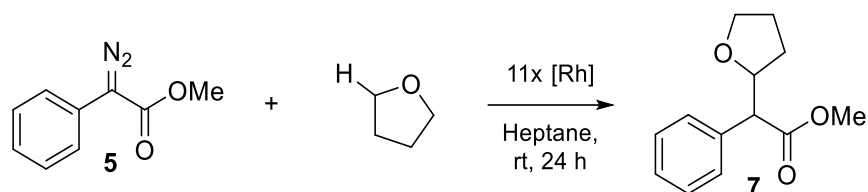
UPLC Gradient 1 minute (2.2 mL/min):

Mobile Phase A (Aqueous) - 0.05% TFA in water

Mobile Phase B (Organic) – 0.05% TFA in acetonitrile

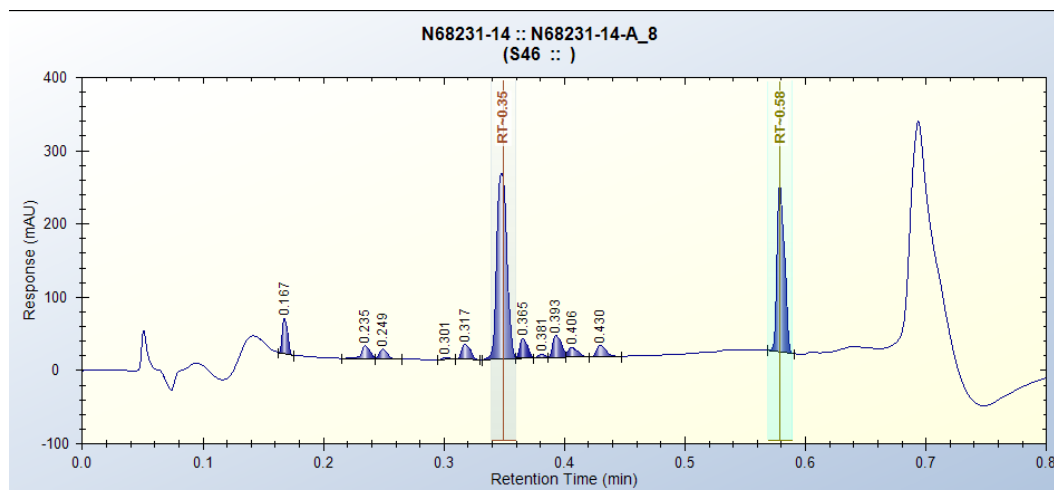
Time (mins)	A(%)	B(%)
0.00	97.00	3.00
0.50	2.00	98.00
0.60	2.00	98.00
0.61	97.00	3.00

Procedure for the UPLC Screening of a C–H Insertion Reaction



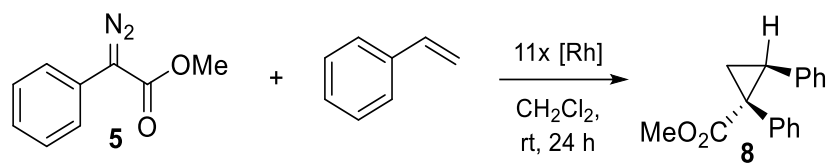
Catalyst stock (8 μ L, 25 mM) for each screened catalyst was added to an individual vial and the solvent removed under vacuum. Heptane (168 μ L) was then added to each vial and stirred for 2 minutes, followed by the sequential addition of THF (16 μ L) and methyl 2-diazo 2-phenylacetate (16 μ L, 1.25 M). The reaction block was then immediately sealed and stirred at room temperature for 24 hours. After removal of solvent under a stream of N_2 at 40 $^{\circ}C$ the reaction was analysed by UPLC.

1 min UPLC trace:



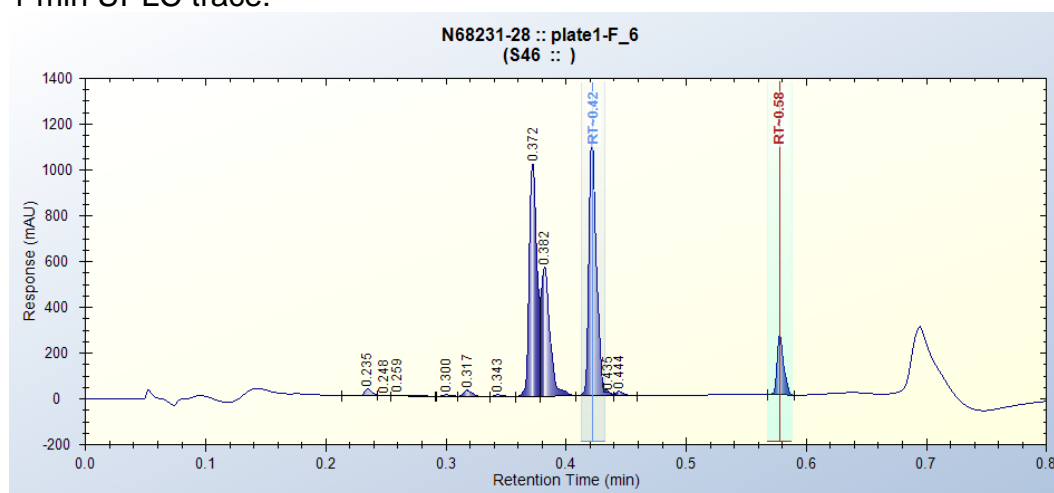
Compound	Retention time for 1 min UPLC gradient	Response Factor
	0.35	2.01
	0.58	-

Procedure for the UPLC Screening of a Cyclopropanation Reaction

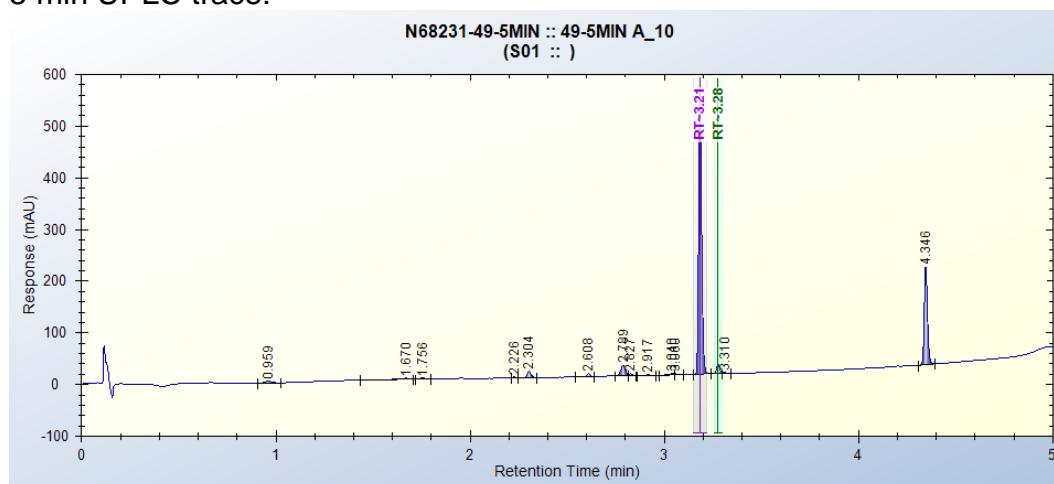


Catalyst stock (8.0 μ L, 25 mM) for each screened catalyst was added to an individual vial and the solvent removed under vacuum. CH_2Cl_2 (161.1 μ L) was then added to each vial and stirred for 2 minutes, followed by the sequential addition of styrene (22.9 μ L) and methyl 2-diazo 2-phenylacetate (16.0 μ L, 1.25 M). The reaction block was then immediately sealed and stirred at room temperature for 24 hours. After removal of solvent under a stream of N_2 at 40 $^{\circ}C$ the reaction was analysed by UPLC. Ratio of diastereomers was qualitatively measured using 5 min UPLC analysis to give a ratio of major/minor products.

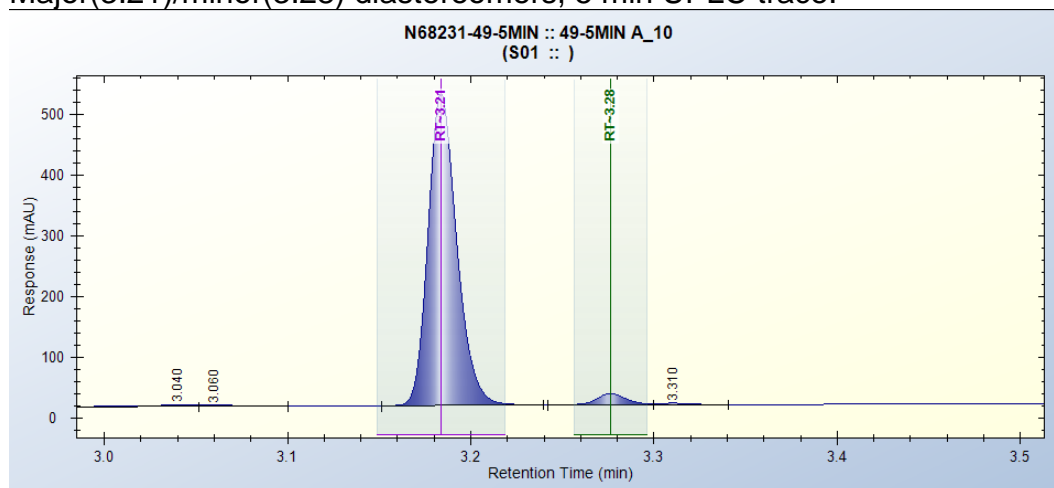
1 min UPLC trace:

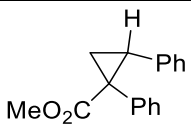
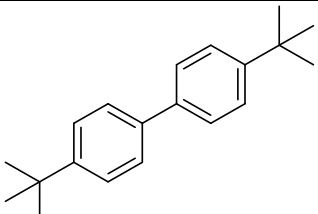


5 min UPLC trace:

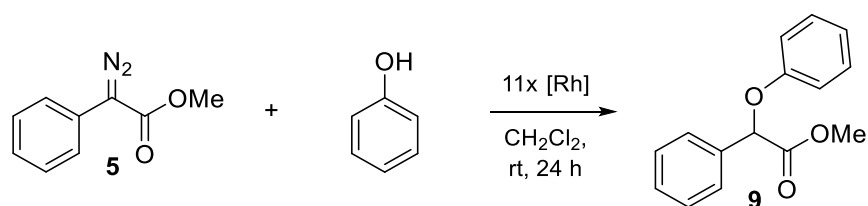


Major(3.21)/minor(3.28) diastereomers, 5 min UPLC trace:



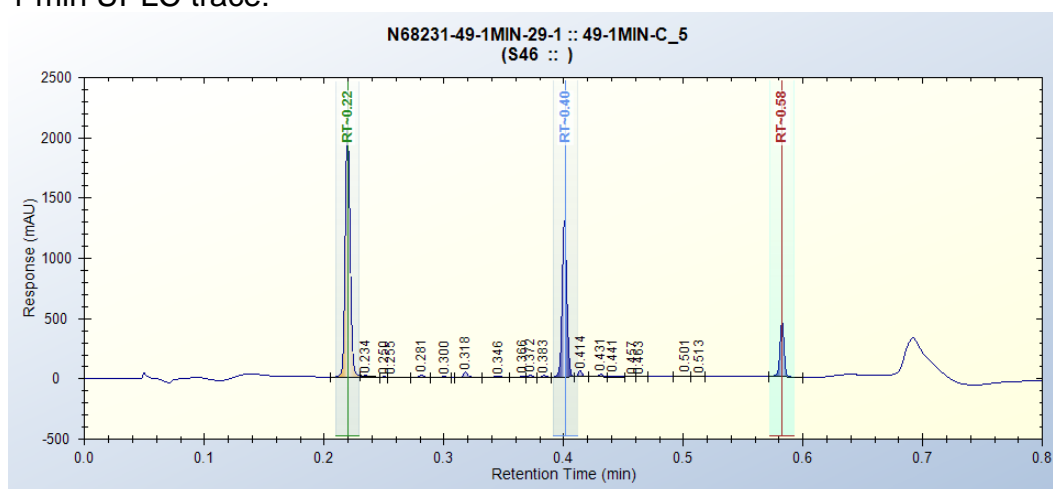
Compound	Retention time for 1 min UPLC gradient	Response Factor	Retention time for 5 min UPLC gradient
	0.42	0.89	3.21/3.28 Major/minor
	0.58	-	-

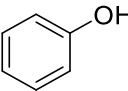
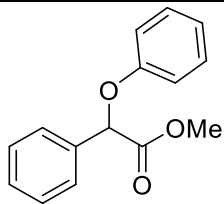
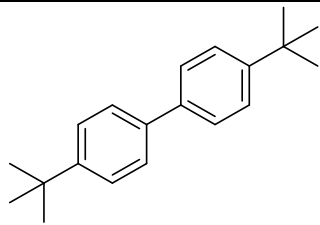
Procedure for the UPLC Screening of an O–H Insertion Reaction



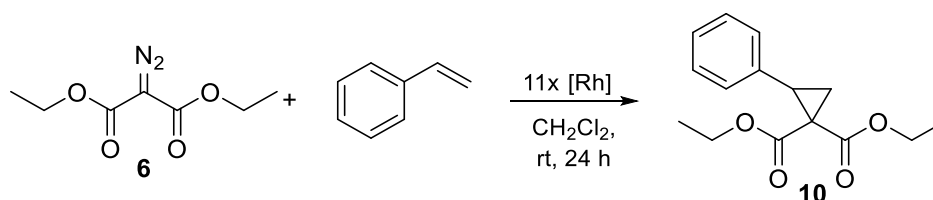
Catalyst stock (8 μ L, 25 mM) for each screened catalyst was added to an individual vial and the solvent removed under vacuum. CH₂Cl₂ (168 μ L) was then added to each vial and stirred for 2 minutes, followed by the sequential addition of phenol in CH₂Cl₂ (16 μ L, 12.5 M) and methyl 2-diazo 2-phenylacetate (16 μ L, 1.25 M). The reaction block was then immediately sealed and stirred at room temperature for 24 hours. After removal of solvent under a stream of N₂ at 40 °C the reaction was analysed by UPLC.

1 min UPLC trace:



Compound	Retention time for 1 min UPLC gradient	Response Factor
	0.22	-
	0.40	0.79
	0.58	-

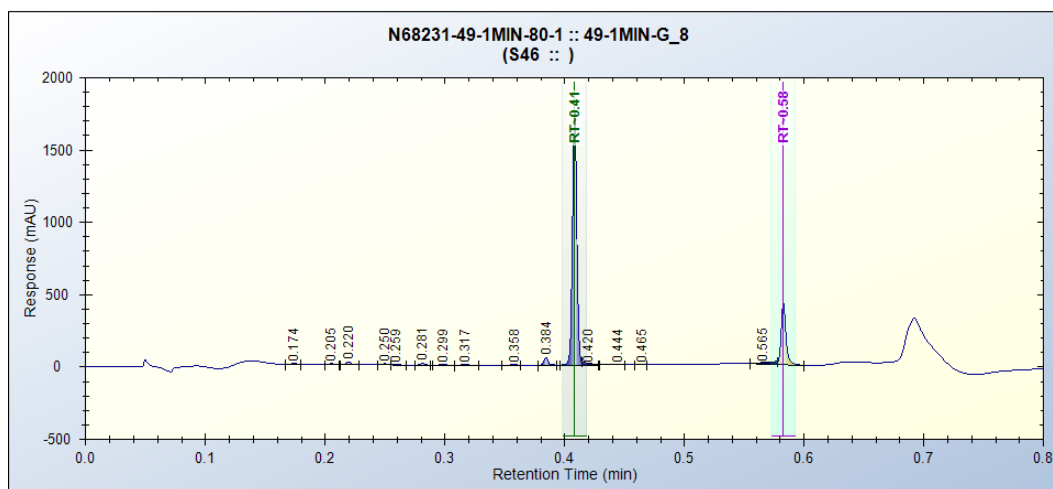
Procedure for the UPLC Screening of an Alternative Cyclopropanation Reaction

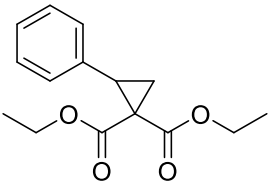
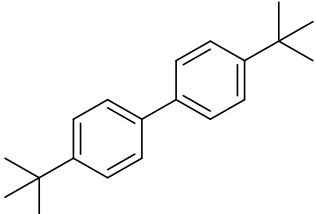


Catalyst stock (8.0 μ L, 25 mM) for each screened catalyst was added to an individual vial and the solvent removed under vacuum. CH₂Cl₂ (161.1 μ L) was then added to each vial and stirred for 2 minutes, followed by the sequential addition of styrene (22.9 μ L) and diethyl 2-diazomalonate (16.0 μ L, 1.25 M). The reaction block was then immediately sealed and stirred at

room temperature for 24 hours. After removal of solvent under a stream of N₂ at 40 °C the reaction was analysed by UPLC.

1 min UPLC trace:



Compound	Retention time for 1 min UPLC gradient	Response Factor
	0.41	0.79
	0.58	-

Solvent Screen for a C-H insertion Reaction

Catalyst stock (8 μL , 25 mM) for each screened catalyst was added to an individual vial and the solvent removed under vacuum. A reaction solvent (168 μL) was then added to an individual vial and stirred for 2 minutes, followed by the sequential addition of THF (16 μL) and methyl 2-diazo 2-phenylacetate (16 μL , 1.25 M). The reaction block was then immediately sealed and stirred at room temperature for 24 hours. After removal of solvent under a stream of N_2 at 40 $^\circ\text{C}$ the reaction was analysed by UPLC.

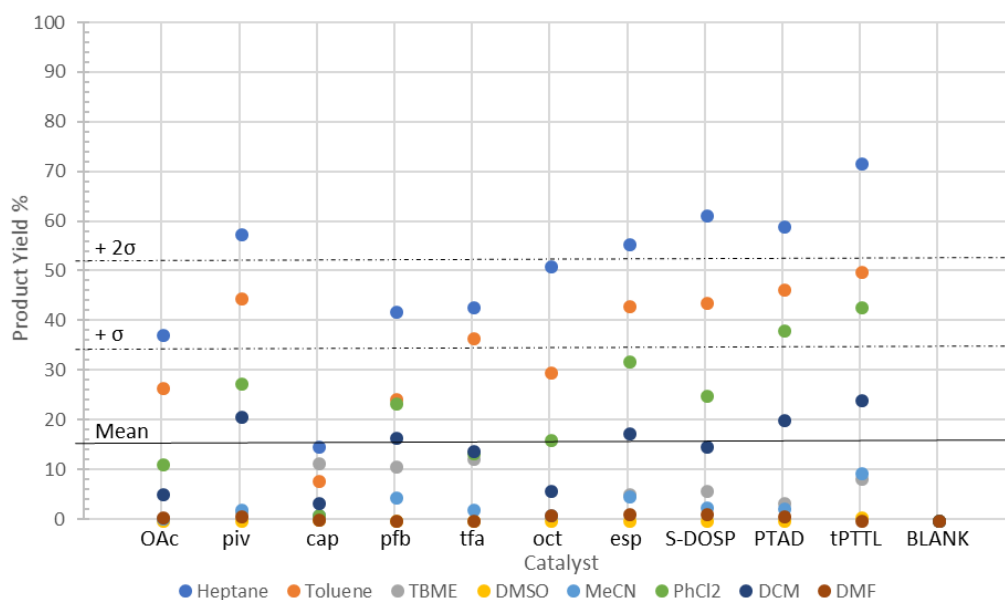
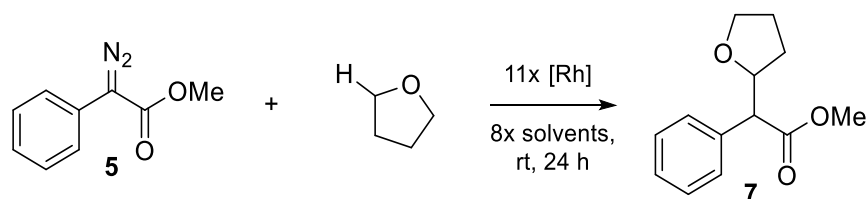


Figure 5.1. Solvent screen for C-H insertion of methyl 2-diazo 2-phenylacetate into THF ($n = 1$).

Table 5.1 Summary of solvent screen yields for the formation of **7**.

Catalyst	Yield of 7 % for each reaction solvent							
	Heptane	Toluene	TBME	DMSO	MeCN	PhCl ₂	DCM	DMF
Rh ₂ OAc ₄ , 3a	37.4	26.8	0.5	0.0	0.3	11.3	5.4	0.7
Rh ₂ piv ₄ , 3b	57.6	44.7	1.7	0.0	2.2	27.6	20.9	0.9
Rh ₂ oct ₄ , 3c	51.2	29.9	0.6	0.0	1.2	16.2	6.1	1.2
Rh ₂ tfa ₄ , 3e	43.0	36.8	12.4	0.0	2.3	13.5	14.1	0.0
Rh ₂ pf ₄ , 3h	42.2	24.4	11.0	0.0	4.6	23.7	16.7	0.0
Rh ₂ esp ₂ , 3q	55.6	43.3	5.4	0.0	4.9	32.1	17.7	1.3
Rh ₂ S-DOSP ₄ , 3r	61.4	44.0	5.9	0.0	2.7	25.1	14.9	1.3
Rh ₂ S-tertPTTL ₄ , 3u	72.0	50.1	8.4	0.7	9.5	42.9	24.2	0.0
Rh ₂ S-PTAD ₄ , 3s	59.3	46.5	3.4	0.0	2.4	38.3	20.4	0.8
Rh ₂ cap ₄ , 4a	15.0	7.9	11.5	0.3	1.2	1.1	3.6	0.3

Solvent screen for C-H insertion of methyl 2-diazo 2-phenylacetate into THF (n = 1).

5.2.3 Computational Details

Optimised geometries for all rhodium(II) complexes were calculated with the Gaussian09²⁷⁷ Software package in isolation using the standard BP86^{213,214} density functional as implemented in Gaussian with the DZP basis set 6-31G(d)^{208,278–280} on all atoms apart from rhodium where the Stuttgart/Dresden effective core potential MWB289 was used. Optimisations used ‘tight’ convergence criteria. Vibrational frequencies were not computed, and the energetic data does not include a correction for zero-point energy. In the absence of frequency calculations, stationary points have not been verified as minima. Geometry optimisations were started from crystal structure geometries of the complex of interest, where a structure was available (Table 5.4), or by careful structural modification of related complexes.

Conformational searches were performed by Dr Natalie Fey (University of Bristol) to enhance conformational sampling. The description of the calculations below, and the results outlined in Tables 5.2 and 5.3, are the work of Dr Natalie Fey and the candidate jointly. Conformational searches used the default MMX force field in PCModel.²⁸¹ GMMX was used for stochastic conformational searches, generally with default settings. 500 iteration conformational searches were performed on the dirhodium complex as well as its carbene complex (stop criteria defined as Emin found 10 times and duplicates found 50 times). These conformational searches were hampered by missing parameters (circumvented by replacing atoms with

elements where parameters existed and suitable restraints) and difficulties with convergence, most likely due to the large number of connected rings arising from the dirhodium core. A selection of conformers was then re-optimised fully at the DFT level, as described above. These DFT calculations were performed by the candidate. A full re-parameterisation of the force field lay outside the scope of this project, but for **3d** and **4g**, the impact of conformational change on the descriptors was explored. These are summarised in Tables **5.2** and **5.3** below. For complex **3d**, the crystal structure geometry led to the lowest energy conformer. This was not the case for **4g**, however, the range of energies found for **4g** was small and the descriptors for the XRD-derived conformer and the lowest energy species found are reasonably similar. Inspection of descriptors shows limited variation in structural and energetic parameters, but a larger range for the steric descriptors (He_8 and $|wV|$) as well as for the energy for Diazo precursor to form the carbene complex ($\Delta E(\text{coord})$), as might be expected.¹⁷⁷ While conformational change can have a large impact for the prediction of selectivity, reactivity and dynamic behaviour of some of the catalysts, in view of the computational problems with sampling conformer space reliably, it was decided that Boltzmann-averaged descriptors would not be included in the present version of this database.

Table 5.2. 3d conformers (Rh₂TBSP₄)

Parameter/Complex	TSBP (XRD)	TBSP_ m1_c1	TBSP_ m1b_c1	TBSP_ m1b_c5	TBSP_ m1b_c10	range	TSBP_carb (XRD)	TBSP_ carb_c1	TBSP_ carb_c3	range
Rel. E, kcal/mol	0.00	3.78	3.74	4.35	2.73	4.35	0.00	10.00	8.12	10.00
r(Rh-Rh), 1	2.392	2.400	2.400	2.394	2.392	0.008	-	-	-	-
r(Rh-L), 1	2.058	2.059	2.060	2.057	2.057	0.003	-	-	-	-
∠(Rh-Rh-L), 1	88.6	88.5	88.5	88.6	88.6	0.1	-	-	-	-
∠(O-C-X)	126.5	126.5	126.5	126.5	126.5	0.0	-	-	-	-
r(αC-R1)	1.537	1.538	1.538	1.537	1.537	0.001	-	-	-	-
r(Rh-Rh), 2	-	-	-	-	-	-	2.466	2.464	2.464	0.002
r(Rh-L), 2	-	-	-	-	-	-	2.069	2.069	2.069	0.000
∠(Rh-Rh-L), 2	-	-	-	-	-	-	87.7	87.7	87.7	0.0
r(Rh-C)	-	-	-	-	-	-	1.979	1.975	1.978	0.004
∠(C-C-C)	-	-	-	-	-	-	113.8	114.4	115.6	1.7
ΔE(coord)	-	-	-	-	-	-	22.67	12.66	14.55	10.00
He8	-	-	-	-	-	-	33.25	37.49	33.19	4.29
wV	-	-	-	-	-	-	26.70	26.70	19.40	7.30
HOMO, 1	-0.1665	-0.1699	-0.1696	-0.1701	-0.1695	0.0036	-	-	-	-
LUMO, 1	-0.1310	-0.1304	-0.1302	-0.1347	-0.1340	0.0045	-	-	-	-
ΔE(FMO)	-0.0355	-0.0395	-0.0394	-0.0354	-0.0355	0.0040	-	-	-	-
Q Rh, 1	0.8072	0.8164	1.0406	1.1353	1.1318	0.3281	-	-	-	-
Q(Donor Atoms, mean), 1	-4.2841	-4.2822	-4.3875	-4.4222	-4.4212	0.1400	-	-	-	-
Q(L, mean), 1	-1.0253	-1.1395	-1.0406	-1.1353	-1.1319	0.1142	-	-	-	-

Table 5.3. 4g conformers (Rh₂MEPY₄)

Parameter/Complex	MEPY	MEPY_ c1	MEPY_ c42	range	MEPY_ carb (XRD)	MEPY_ carb_c1	MEPY_ carb_c35	MEPY_ carb_c75	range
Rel. E, kcal/mol	1.50	0.00	5.40	5.40	1.68	0.23	0.50	0.00	1.68
r(Rh-Rh), 1	2.466	2.463	2.463	0.004	-	-	-	-	-
r(Rh-L), 1	2.064	2.068	2.061	0.006	-	-	-	-	-
∠(Rh-Rh-L), 1	87.8	88.0	88.1	0.2	-	-	-	-	-
∠(O-C-X)	125.6	125.6	125.6	0.0	-	-	-	-	-
r(αC-R1)	1.527	1.525	1.524	0.002	-	-	-	-	-
r(Rh-Rh), 2	-	-	-	-	2.540	2.522	2.536	2.533	0.017
r(Rh-L), 2	-	-	-	-	2.070	2.072	2.071	2.071	0.002
∠(Rh-Rh-L), 2	-	-	-	-	87.0	87.2	87.1	87.0	0.2
r(Rh-C)	-	-	-	-	1.992	1.974	1.996	1.994	0.022
∠(C-C-C)	-	-	-	-	110.4	110.7	111.6	110.3	1.3
ΔE(coord)	-	-	-	-	8.73	10.18	9.91	10.41	1.68
He8	-	-	-	-	58.71	76.19	54.45	60.84	21.74
wV	-	-	-	-	14.70	19.80	18.00	13.70	6.10
HOMO, 1	-0.1261	-0.1337	-0.1330	0.0076	-	-	-	-	-
LUMO, 1	-0.0866	-0.1175	-0.1088	0.0309	-	-	-	-	-
ΔE(FMO)	-0.0396	-0.0162	-0.0242	0.0233	-	-	-	-	-
Q Rh, 1	0.6234	0.6197	0.6106	0.0128	-	-	-	-	-
Q(Donor Atoms, mean), 1	-4.1089	-4.1169	-4.1078	0.0091	-	-	-	-	-
Q(L, mean), 1	-0.8028	-0.7340	-0.7828	0.0687	-	-	-	-	-

5.2.4 Design of a Descriptor Database

A range of steric and electronic quantum chemical descriptors were then captured from optimised geometries of each rhodium(II) complex. The coordination energy of the carbene generated from a symmetrical α -diazo malonamide precursor was calculated from converged energies (a.u.) using the equation below.

$$\Delta E(\text{coord}) \text{ kcal mol}^{-1} = 627.5095 \times ((E1 + E_{\text{Diazo}}) - (E2 + E_{N_2}))$$

Natural population analysis (NBO) was calculated using the converged complex and the energy difference between the HOMO and LUMO molecular orbitals was calculated in atomic units using the equation below.

$$\Delta E(\text{FMO}) \text{ a.u.} = \text{LUMO} - \text{HOMO}$$

Other parameters captured from the NBO analysis include the charge on the rhodium atoms, charge on the ligand donor atoms and the average charge on the ligands.

For the PCA model, 48 ligands (Table 5.4) were optimised as complexes **1** and **2** ($n = 96$) and 19 chemical descriptors were extracted from these optimised complex geometries. Each descriptor (x_i) was scaled to dimensionless values (x) using the equation below.

$$x = \frac{(x_i - u)}{s}$$

Where $u = \frac{1}{N} \sum_{i=1}^N x_i$ (sample mean of x), N = number of data points

$$s = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - u)^2}$$
 (sample standard deviation of x).

Chemical descriptors considered:

1. All bond lengths and angles (including dihedral angles) for both complexes **1** and **2** calculated at BP86/6-31G(d)/MWB28 level of theory in Gaussian09. Lengths and angles were assessed for their response to carbene formation and the following descriptors were selected: $r(\text{Rh-Rh})$, $r(\text{Rh-L})$, $\angle(\text{Rh-Rh-L})$, $r(\alpha\text{C-R1})$, $\angle(\text{O-C-X})$, $r(\text{Rh-C})$ and $\angle(\text{C-C-C})$ as these showed a clear response to changes in ligand properties.
2. Binding energies of ancillary ligands, including diazo malonamide (BP86/6-31G(d)/MWB28), methyl 2-diazo-2-phenylacetate (BP86/6-31G(d)/MWB28), and acetonitrile (B3LYP/6-31G(d)/MWB28). Complexes bearing acetonitrile ligands were found to possess extreme bond angles for complexes with

carboxamidate ligands and were thus removed from subsequent analyses. Acetonitrile complexes were initially calculated using the B3LYP functional, however, due to an inability to obtain convergence on several carboxamidate complexes, the functional was changed to BP86 for all other calculations.

3. Several steric descriptors were considered, including an adapted version of the He₈ ring used in Bristol's LKBs,¹⁷⁸⁻¹⁸² Distance-Weighted Volume²¹⁵ and first-generation Sterimol parameters.¹⁶⁵ He₈ ring interaction energies were calculated as single-point energies at BP86/6-31G(d)/MWB28 level of theory, where the He₈ ring was aligned 1.9 Å from the rhodium core (average r(Rh-C) bond length). Distance-Weighted Volume was derived from the MolQuO web app (<http://rodi.urv.es/~carbo/quadrants/index.html>), aligning the quadrants with the Rh-C bond and then removing the carbene ligand from the optimised geometry. First-generation Sterimol parameters were calculated by aligning the L vector with the Rh-Rh bond and using a python script available here: <https://github.com/bobbypaton/Sterimol>. Sterimol parameters were found to be prone to outliers due to the extreme size of many of the ligands, and less capable of describing the steric environment around the Rh atoms, so were not used in subsequent analysis. A recently updated version of this descriptor, which explores conformational variation (wSterimol),²⁸² may be able to address this limitation.
4. Descriptors derived from the Natural Population Analysis of the converged complexes including: the HOMO and LUMO energies for complexes 1 and 2; charges on atoms including the core Rhodium atoms, ligand donor atoms and the average charge on the ligand for complexes 1 and 2; and the difference in energy between the frontier molecular orbitals.

Distance weighted volume, or quadrant occupation (see discussion above for details), gives a measure of the steric bulkiness of the ligand and its influence over the metal centre, and was calculated using the formula below.

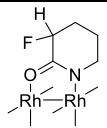
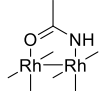
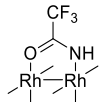
$$V_{W,k,l} = \sum_{i=1}^n \frac{r_i^k}{d_i^l}$$

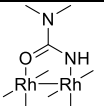
where $k = 3$ and $l = 1$, d^l = distance of atom to metal centre and r^k = van der Waals radius of atom

Table 5.4. Modelled complexes. Cambridge Crystallographic Database accession codes are shown below catalyst names.

No.	Catalyst	Structure
3a	Rh ₂ OAc ₄ BARPIT	
3b	Rh ₂ piv ₄ HUWKOB	
3c	Rh ₂ oct ₄ XUZNOW	
3d	Rh ₂ TBSP ₄ LOPKOQ	
3e	Rh ₂ tfa ₄ SAXYAS	
3f	Rh ₂ dfa ₄	
3g	Rh ₂ mfa ₄	
3h	Rh ₂ pfb ₄ IGAFUR	
3i	Rh ₂ FPiV ₄	
3j	Rh ₂ bnz ₄	
3k	Rh ₂ DABN ₄	
3l	Rh ₂ POMB ₄	
3m	Rh ₂ FBNZ ₄	
3n	Rh ₂ TFBN ₄	
3o	Rh ₂ PFBN ₄	
3p	Rh ₂ TFMB ₄	
3q	Rh ₂ esp ₂ KUTXAA	
3r	Rh ₂ DOSP ₄	
3s	Rh ₂ PTAD ₄ WAJBAN	

3t	Rh ₂ PTTL ₄	
3u	Rh ₂ tPTTL ₄	
4a	Rh ₂ cap ₄ QEZMOY	
4b	Rh ₂ pip ₄	
4c	Rh ₂ MPDO ₄	
4d	Rh ₂ MEOX ₄ PIZBAC	
4e	Rh ₂ pyr ₄	
4f	Rh ₂ PYOH ₄	
4g	Rh ₂ MEPY ₄ HAWVAC	
4h	Rh ₂ OXAZ ₄	
4i	Rh ₂ dhpO ₄	
4j	-	
4k	Rh ₂ OXAL ₄	
4l	Rh ₂ pyO ₄	
4m	Rh ₂ MPPIM ₄	
4n	Rh ₂ OXPN ₄	
4o	Rh ₂ dhpD ₄	
4p	-	
4q	Rh ₂ MIDO ₄	
4r	Rh ₂ dhpY ₄	
4s	Rh ₂ MACIM ₄	
4t	Rh ₂ dfpyr ₄	
4u	Rh ₂ dfpip ₄	
4v	Rh ₂ DIZO ₄	
4w	Rh ₂ mfpyr ₄	

4x	Rh ₂ mfpip ₄	
4y	Rh ₂ acam ₄	
4z	Rh ₂ tfam ₄	

4aa	Rh ₂ DMU ₄	
-----	----------------------------------	---

5.2.5 Exemplar Submission Script for Gaussian09

The ARC3 high-performance computing cluster, at the University of Leeds, was used to calculate optimised geometries for all the rhodium(II) complexes shown in Table 5.4. An example submission script applicable to ARC3 is shown below.

```
#$ -V -cwd
#$ -l h_rt=24:00:00
#$ -l h_vmem=4G
#$ -l disk=4G
#$ -pe smp 8

module load gaussian
export OMP_NUM_THREADS=1
export GAUSS_SCRDIR=$TMPDIR
g09 rhodium_complex.com
```

The Gaussian09 file header for the corresponding submission script is also shown below.

```
%nprocshared=8
%mem=2000MB
%NoSave
%chk=rhodium_complex.chk
# opt BP86/gen 5d 7f pseudo=cards
```

5.2.6 Exemplar Script for the Automated Extraction of Descriptors

A python script for the automated extraction of parameters from the Guassian09 .log file of converged complexes was used to build the final version of the knowledge base, and the corresponding code is shown below.

```
from __future__ import print_function
import argparse
parser = argparse.ArgumentParser()
parser.add_argument("filenames", nargs="+")
parser.add_argument("-p", "--parameters", nargs="*", default=[])
args = parser.parse_args()

def get_gaussian_results(filename, parameter_list):
    file = open(filename, "r")
    energy = None
    optg_complete = False # default that the optimisation is NOT complete
    parameters = {key: None for key in parameter_list}
    for line in file:
        if "SCF Done" in line:
            energy = float(line.strip().split()[4])
        if "Optimized Parameters" in line:
            optg_complete = True
        if optg_complete:
            for parameter in parameters.keys():
                if parameter in line:
                    value = float(line.strip().split()[3])
                    parameters[parameter] = value
    file.close()
    return energy, parameters, optg_complete
print("filename energy "+" ".join(args.parameters))
for filename in args.filenames:
    #print(filename)
    energy, parameters, optg_complete = get_gaussian_results(filename,
args.parameters)
    assert(optg_complete) # Check that optimisation IS complete then display result
    #print(parameters)
    #print(energy)
    print(filename + " " + str(energy) + " " + " ".join([str(parameters[key]) for key in
args.parameters]))

get_gaussian_results(rhodium_complex.log, R(1,2)) # Pseudo-code for extracting the Rh-
Rh bond length assuming the rhodium atoms are atoms 1 and 2 in the gaussian file.
```

5.2.7 Data Extracted from DFT-Converged Rhodium(II) Complexes

Table 5.5 Descriptor data

Complex	No.	r(Rh-Rh), 1 (Å)	r(Rh-L), 1 (Å)	\angle (Rh-Rh-L), 1	\angle (O-C-X)	r(α C1-R1) (Å)	r(Rh-Rh), 2 (Å)	r(Rh-L), 2 (Å)	\angle (Rh-Rh-L), 2	r(Rh-C) (Å)	\angle (C-C-C)
Rh2OAc4	3a	2.3904	2.0556	88.69225	126.1627	1.515	2.4648	2.065525	87.69724	1.9729	115.5716
Rh2tfa4	3e	2.4037	2.053688	88.60913	128.6942	1.5522	2.4921	2.064513	87.44825	1.9832	117.0577
Rh2pfb4	3h	2.4084	2.0536	88.52184	128.7221	1.5533	2.4915	2.064563	87.45353	1.9851	117.54
Rh2cap4	4a	2.4091	2.061563	88.60183	123.1859	1.52625	2.4624	2.077113	87.4265	1.9972	110.3102
Rh2esp2	3q	2.3862	2.058725	88.64534	125.3802	1.54355	2.4551	2.0691	87.73745	1.9763	114.8787
Rh2TBSP4	3d	2.3922	2.058038	88.64276	126.4889	1.536775	2.4661	2.069413	87.66359	1.9786	113.8305
Rh2piv4	3b	2.3873	2.056675	88.64755	125.4675	1.5422	2.4593	2.06735	87.67254	1.9724	114.6697
Rh2oct4	3c	2.3902	2.05645	88.6592	125.8777	1.5227	2.4608	2.067025	87.73165	1.9723	114.7489
Rh2MEOX4	4d	2.4711	2.063213	87.92615	127.7151	1.369975	2.5479	2.072625	86.95375	2.0006	111.3165
Rh2MEPY4	4g	2.4662	2.0637	87.84624	125.5897	1.526675	2.5396	2.070075	87.02478	1.9924	110.4379
Rh2PTAD4	3s	2.3876	2.053638	88.63128	126.1072	1.547875	2.4578	2.0672	87.70965	1.9728	114.2438
Rh2PTTL4	3t	2.3874	2.053725	88.63924	126.1469	1.54705	2.4581	2.066075	87.71286	1.9735	114.2253
Rh2DOSP4	3r	2.3898	2.057938	88.68329	126.5004	1.536475	2.4645	2.069513	87.6764	1.9854	112.5911
Rh2tertPTTL4	3u	2.387	2.05465	88.64523	126.1291	1.54715	2.4577	2.066013	87.71779	1.9731	114.2204
Rh2pip4	4b	2.4082	2.060575	88.56604	123.2836	1.5276	2.4782	2.0751	87.63741	1.9917	107.7132
Rh2pyr4	4e	2.4522	2.058213	88.31283	126.0343	1.5253	2.5288	2.071138	87.15415	1.9797	109.2122
Rh2OXAZ4	4h	2.4143	2.061425	88.43944	125.0821	1.37215	2.4883	2.074538	87.5471	1.9989	109.4973
Rh2OXAL4	4k	2.465	2.060988	88.32644	128.3155	1.3724	2.5429	2.0725	87.03189	1.9809	110.5352
Rh2OXPN4	4n	2.428	2.063	88.54245	125.6473	1.3812	2.4929	2.075813	87.19973	1.9978	111.6562
Rh2MIDO4	4q	2.4627	2.064025	88.40193	126.9106	1.40215	2.5415	2.075088	87.39979	1.9742	108.7325
Rh2MPDO4	4c	2.4003	2.06295	88.5657	122.9188	1.401225	2.466	2.077913	87.5959	1.9932	106.1646
Rh2PYOH4	4f	2.4048	2.0511	89.13498	121.3814	1.4303	2.4549	2.066513	87.29741	1.9959	111.4973
Rh2dhp4	4i	2.4603	2.055	88.3918	126.0434	1.4825	2.5414	2.064963	87.32451	1.981	112.0302
Rh2pyo4	4l	2.4578	2.05685	88.38496	125.9489	1.5262	2.5383	2.068588	87.13825	1.9844	110.4622
Rh2dhp4	4o	2.4171	2.059275	88.61054	123.4697	1.5234	2.4739	2.073661	87.2797	1.9933	111.821
Rh2dhpy4	4r	2.4178	2.05625	88.55289	123.7707	1.48325	2.4924	2.069713	87.61476	1.9901	108.5639
Rh2dfpyr4	4t	2.4615	2.0532	88.28734	127.6318	1.536	2.5469	2.065088	87.00008	1.9875	111.4793
Rh2mfpyr4	4w	2.4552	2.054375	88.34173	126.7989	1.5302	2.5368	2.066725	87.05054	1.9822	110.2788
Rh2dfpip4	4u	2.4146	2.05555	88.56323	124.8482	1.54375	2.4957	2.069513	87.5174	1.9994	108.7959
Rh2mfpip4	4x	2.4094	2.05675	88.49751	124.0973	1.53125	2.4829	2.07255	87.38653	1.9961	108.7212
Rh2MACIM4	4p	2.4523	2.061175	88.42836	126.8245	1.41275	2.5252	2.076313	87.15031	1.98	107.6853
Rh2MACIM4_A	4s	2.4711	2.0657	87.94701	126.4995	1.414175	2.5345	2.073588	86.92713	2.0015	110.3956
Rh2MPPIM4	4j	2.4515	2.060813	88.43143	126.8015	1.4134	2.5293	2.077325	87.3054	1.9848	107.5012
Rh2MPPIM4_A	4m	2.4707	2.064838	87.93806	126.5221	1.41505	2.5315	2.075563	86.84775	2.0041	109.4536
Rh2DIZO4	4v	2.4155	2.057663	88.63804	124.5496	1.430175	2.4705	2.076325	87.44414	2.0118	107.645
Rh2acam4	4y	2.4314	2.05445	88.23983	123.319	1.5193	2.4943	2.063113	87.37613	1.9551	117.3778
Rh2tfam4	4z	2.4442	2.0509	88.26348	126.1411	1.541	2.5168	2.058513	87.30193	1.9628	118.1972
Rh2DMU4	4aa	2.4206	2.058425	88.47038	123.0672	1.399125	2.4734	2.066438	87.4447	1.9553	115.8328
Rh2bnz4	3j	2.3871	2.0506	88.7349	125.4768	1.4929	2.4615	2.060625	87.74085	1.9708	114.4433
Rh2DABN4	3k	2.3855	2.050938	88.80255	125.2522	1.482025	2.4533	2.059425	87.89366	1.97	113.9495
Rh2POMB4	3l	2.3883	2.051313	88.7262	125.3599	1.48625	2.4537	2.059513	87.86411	1.9704	114.6449
Rh2TFMB4	3p	2.3886	2.0503	88.71806	125.6467	1.494	2.4595	2.059925	87.64995	1.9739	114.5271
Rh2FBNZ4	3m	2.3874	2.050525	88.74549	125.4981	1.4898	2.4554	2.065975	87.83151	1.9718	114.8362
Rh2TFBN4	3n	2.3942	2.054125	88.64741	126.5726	1.4924	2.4729	2.061125	87.65856	1.979	115.6387

Rh2PFBN4	3o	2.3852	2.0489	88.723	125.808	1.4957	2.4737	2.065	87.64573	1.9817	116.05
Rh2dfa4	3f	2.4036	2.053625	88.58965	128.1126	1.5416	2.4836	2.065388	87.55831	1.9794	116.869
Rh2mfa4	3g	2.3994	2.05775	88.62931	127.5868	1.529225	2.477	2.068275	87.58025	1.9793	117.6103
Rh2FPIV4	3i	2.3944	2.051663	88.53656	127.1141	1.55695	2.4779	2.067038	87.4557	1.9948	115.8411

Table 5.5 continued.

No	$\Delta E(\text{coord})$	Sterimol L	SterimolB1	SterimolB5	He_8	$ \text{wV} $	HOMO, 1	LUMO, 1	$\Delta E(\text{FMO})$
3a	22.26686	4.79	4.85	5.85	20.63117	0	-0.16939	-0.13257	0.03682
3e	29.40413	4.8	5.42	6.2	19.99772	0	-0.22456	-0.18604	0.03852
3h	29.794	4.89	6.83	8.57	20.30875	2.4	-0.22206	-0.1841	0.03796
4a	12.23489	6.55	4.93	7.13	55.12895	15.2	-0.12038	-0.10225	0.01813
3q	21.16797	4.79	6.14	8.35	21.26676	5.4	-0.16755	-0.13127	0.03628
3d	22.66718	11.93	7.62	13.65	33.25419	26.7	-0.16652	-0.13099	0.03553
3b	21.40432	4.79	6.05	6.99	21.47971	4.8	-0.16605	-0.12978	0.03627
3c	21.55542	7.77	6.66	12.68	20.96546	7.8	-0.1665	-0.13009	0.03641
4d	6.939207	8.96	5.16	6.62	60.30271	15.8	-0.14875	-0.12189	0.02686
4g	8.726304	9.18	5.14	6.71	58.70671	14.7	-0.12613	-0.08658	0.03955
3s	21.23446	9.09	7.82	9.64	22.61748	27.8	-0.1659	-0.13075	0.03515
3t	21.47939	8.93	7.41	8.32	22.55454	28	-0.16636	-0.13104	0.03532
3r	17.74799	23.89	8.21	25.03	31.69143	36.9	-0.16615	-0.13041	0.03574
3u	21.45875	11.23	7.88	10.41	22.50282	40.2	-0.16223	-0.12691	0.03532
4b	12.2192	5.99	5.32	7.08	53.03076	11.4	-0.12178	-0.10327	0.01851
4e	16.30173	5.74	5.1	6.29	43.15772	8.5	-0.12889	-0.11513	0.01376
4h	13.20737	5.77	5.26	6.91	51.69668	11.1	-0.12924	-0.11119	0.01805
4k	19.22107	5.73	5.04	6.28	37.39957	8.2	-0.1429	-0.12104	0.02186
4n	14.94868	6.51	4.83	7.02	44.66466	12.4	-0.13166	-0.11333	0.01833
4q	17.58666	5.78	5.56	6.98	43.08334	10.2	-0.12058	-0.10892	0.01166
4c	11.68626	5.89	5.56	6.98	54.14027	14.9	-0.10863	-0.09285	0.01578
4f	15.79384	6.01	5.27	7.03	47.8954	10.8	-0.14863	-0.12872	0.01991
4i	19.43113	5.55	4.66	6.41	38.71222	10.1	-0.13087	-0.12025	0.01062
4l	18.81432	5.87	4.98	6.3	51.43823	7.8	-0.13861	-0.13147	0.00714
4o	15.52796	6.04	5.08	7.02	46.78839	10.7	-0.13731	-0.1228	0.01451
4r	13.79536	5.88	5.15	7.08	50.99556	10.7	-0.12572	-0.11039	0.01533
4t	20.06076	5.73	5.44	6.37	39.91734	8.6	-0.16117	-0.14613	0.01504
4w	17.60651	5.76	5.21	6.29	42.0534	8.3	-0.14679	-0.1319	0.01489
4u	16.33769	5.99	5.43	7.07	51.85375	12.5	-0.14951	-0.13022	0.01929
4x	11.40141	5.98	5.42	7.06	52.16438	12.5	-0.13864	-0.119	0.01964
4p	16.1105	5.77	5.9	7.77	47.19848	11.1	-0.15853	-0.14624	0.01229
4s	7.096756	9.1	5.84	7.8	64.71005	15.7	-0.15433	-0.1214	0.03293
4j	15.19324	11.23	6.57	7.8	47.80398	21.5	-0.15762	-0.14472	0.0129
4m	6.64862	11.25	6.56	8.26	53.56466	27	-0.15334	-0.11916	0.03418
4v	13.2335	5.95	5.78	8.07	52.42294	12.9	-0.13221	-0.11508	0.01713
4y	26.84968	4.83	4.91	5.81	22.07075	1	-0.13031	-0.11369	0.01662
4z	30.73594	4.84	5.49	6.25	21.11823	1.7	-0.18067	-0.16188	0.01879
4aa	26.33938	4.92	5.52	7	22.9179	5.5	-0.1129	-0.09669	0.01621
3j	21.85532	4.85	6.78	9.13	21.57902	2	-0.16764	-0.13199	0.03565
3k	20.6465	4.86	8.28	11.27	21.66365	3.2	-0.13792	-0.10356	0.03436

3l	20.5143	4.86	8.26	11.23	21.84006	2.3	-0.15525	-0.1204	0.03485
3p	22.51241	4.84	8.56	10.47	21.33658	2	-0.18829	-0.15202	0.03627
3m	21.47649	4.85	7.32	9.74	21.3375	2	-0.17349	-0.13785	0.03564
3n	24.13533	4.95	7.28	9.76	21.18246	1.6	-0.17735	-0.14063	0.03672
3o	28.75752	5.47	7.36	9.79	16.95986	4.6	-0.1891	-0.15127	0.03783
3f	27.42214	4.8	5.35	6.2	20.06181	0	-0.21009	-0.17266	0.03743
3g	26.32712	4.8	4.89	6.17	19.8209	1.4	-0.19209	-0.15556	0.03653
3i	29.10249	5.49	6.59	7.66	22.4572	6.4	-0.22239	-0.18322	0.03917

Table 5.5 continued.

No.	Q Rh	Q(Donor Atoms, mean)	Q(L, mean)
3a	0.80629	-4.31789	-1.00425
3e	0.79276	-4.10036	-1.13426
3h	0.78939	-4.08076	-1.12332
4a	0.6522	-4.14731	-0.7166
3q	0.83497	-4.33155	-1.0183
3d	0.80715	-4.28413	-1.02527
3b	0.82605	-4.33186	-1.00907
3c	0.80827	-4.32914	-1.24281
4d	0.6407	-4.33444	-0.83666
4g	0.62339	-4.10886	-0.80275
3s	0.89223	-4.41617	-1.11186
3t	0.8317	-4.31028	-1.0441
3r	0.82854	-4.28021	-1.01252
3u	0.83201	-4.31165	-1.03985
4b	0.62913	-4.1665	-0.73117
4e	0.62541	-4.13542	-0.77352
4h	0.65093	-4.403	-0.79159
4k	0.63686	-4.37843	-0.83492
4n	0.64834	-4.28228	-0.74896
4q	0.64281	-4.38313	-0.76787
4c	0.67141	-4.44688	-0.72964
4f	0.63973	-3.90413	-0.79312
4i	0.65438	-4.08522	-0.78647
4l	0.64311	-4.06333	-0.81341
4o	0.64013	-4.07627	-0.78234
4r	0.63195	-4.08795	-0.76295
4t	0.60243	-3.98149	-0.8322
4w	0.60802	-4.03045	-0.80031
4u	0.60649	-4.01033	-0.79638
4x	0.61204	-4.0752	-0.76478
4p	0.65817	-4.27235	-0.83569
4s	0.65699	-4.25213	-0.86174
4j	0.65467	-4.26562	-0.85483
4m	0.65094	-4.23983	-0.86727
4v	0.66967	-4.16472	-0.78785
4y	0.60918	-4.90265	-0.70204

4z	0.59348	-4.71809	-0.82173
4aa	0.65056	-5.20242	-0.69174
3j	0.83369	-4.3023	-1.03784
3k	0.83651	-4.36041	-0.99226
3l	0.83654	-4.33368	-1.01773
3p	0.8366	-4.28181	-1.0663
3m	0.83729	-4.31196	-1.04071
3n	0.8021	-4.14494	-1.04506
3o	0.8102	-4.14632	-1.08255
3f	0.78576	-4.16766	-1.09024
3g	0.79037	-4.26683	-1.06525
3i	0.81255	-4.11378	-1.1405

Table 5.6. PCA Eigenvalues for PC1/PC2/PC3

Complex	PC1	PC2	PC3
3a	2.765124	0.394087	0.484166
3b	2.540759	0.826934	0.045106
3c	3.063959	0.547643	-0.55779
3d	1.964327	0.611458	-1.67993
3e	4.135164	-3.07732	0.48039
3f	3.709259	-2.22617	0.528501
3g	3.415276	-1.35453	0.537182
3h	3.952224	-3.17644	0.358793
3i	3.981831	-2.19594	-0.03057
3j	2.816006	0.904917	0.259393
3k	2.262869	2.254427	0.206891
3l	2.618915	1.529706	0.209988
3m	3.126386	0.931589	0.256251
3n	2.910799	-0.25628	0.367238
3o	3.692856	-0.27468	0.545152
3p	3.271487	0.143183	0.224148
3q	2.732144	0.955249	-0.02999
3r	1.690642	1.041756	-2.77287
3s	2.757774	0.918259	-2.03934
3t	2.429197	0.918512	-1.74944
3u	2.235113	1.202976	-2.59396
4a	-2.31806	2.673761	0.080818
4aa	-0.90338	1.722572	2.42053
4b	-2.42068	2.666118	0.217413
4c	-2.70736	3.442609	-0.17431
4d	-3.75862	-2.63471	-1.64474
4e	-3.08175	-0.70591	0.577585
4f	-0.73693	2.934292	1.212598
4g	-4.12887	-1.14019	-1.6823
4h	-2.05698	1.294935	-0.03218
4i	-2.40733	-0.89312	0.97583
4j	-2.34399	-1.14344	-0.74184
4k	-2.29413	-2.33117	0.200509
4l	-2.87215	-1.21016	0.982162
4m	-3.61938	-2.05115	-2.67747
4n	-2.06685	0.469549	0.283543
4o	-1.69553	1.427121	0.866271
4p	-2.32495	-1.46413	0.096215
4q	-3.05464	-0.58895	0.491496
4r	-2.31969	1.94737	0.45642
4s	-3.58574	-2.15338	-1.88819
4t	-2.20451	-2.84644	0.730663
4u	-1.69042	0.81985	0.236748
4v	-1.97567	1.848575	-0.10632
4w	-2.63137	-1.73973	0.658611
4x	-2.29055	1.450535	-0.01318
4y	-0.97945	0.235402	2.846985
4z	0.396875	-2.64954	2.576822

5.2.8 Exemplar Python Code for Principal Component Analysis

PCA (LAPACK implementation, single-value decomposition) was then performed with the sklearn (0.20.3) python (3.7.3) package for all the captured and scaled descriptors using the python code outlined below.

```
# Required packages for PCA decomposition
from sklearn.decomposition import PCA
# PCA decomposition
pca = PCA(n_components = 3)
principal_components = pca.fit_transform(data)
```

Outlined below is the python script used to find the top ranked solutions shown in Figures 2.10, 2.11 and 2.12, and Table 2.2. The script below selects the PCA model with the smallest mean squared error for the loss of reprojection (note: more than one solution will have a low value).

```
# data is an array for the number of descriptors and x a DataFrame containing the
knowledge base data
def PCA_solution(x, data):
    solution = {}
    for L in range(min(x), max(x)+1):
        res = []
        # The result of each combination of descriptors is appended here
        for i in itertools.combinations(features, L):
            # loop that uses itertools to compute all combinations of descriptors
            descriptors = data.loc[:, i].values # Extract descriptors from DataFrame
            X_train = StandardScaler().fit_transform(descriptors)
            # Scaling data for PCA decomposition
            pca = PCA(n_components=3)
            # Calling PCA and setting number of principal components to be computed
            principal_components = pca.fit_transform(X_train)
            X_projected = pca.inverse_transform(principal_components)
            # Reprojection of PCA decomposition
            loss = ((X_train - X_projected) ** 2).mean()
            # Calculating mean squared error for loss of reprojection
            var = sum(pca.explained_variance_ratio_)
            # Variation captured for each solution
            res.append([loss, var, i]) # Appending results for each combination
        std_loss = np.std([res[i][0] for i in range(len(res))])
        # Calculating standard deviation for the loss of all solutions
        std_var = np.std([res[i][1] for i in range(len(res))])
```

```

# Calculating standard deviation for variation
    solution["solution{0}".format(L)] = min(res), std_loss, std_var
# Adding result to dictionary
    return solution

PCA_solution(x = [10, 11, 12, 13, 14], data = knowledgebase)
# Pseudo-code for computing solutions using 10, 11, 12, 13, and 14 descriptors

```

5.3.9 Overlay of Alternative Model Reactions onto the PCA Map

The yields for the C–H and O–H insertion reactions between **5** and THF (giving **7** as a product) or phenol (giving **9** as a product), and the cyclopropanation of styrene with an acceptor/acceptor diazo, **6**, (giving **10** as a product) are projected onto the PCA plots below (Figures 5.2, 5.3, and 5.4). For the overlay of the donor/acceptor diazo cyclopropanation of styrene, yielding **8**, see Figure 2.17. All four model reactions gave a range of yields, with O,O catalysts generally giving higher yields than N,O catalysts. Product yields clustered based on PC1, PC2 and PC3 values demonstrating that both steric and electronic factors contribute to outcomes.

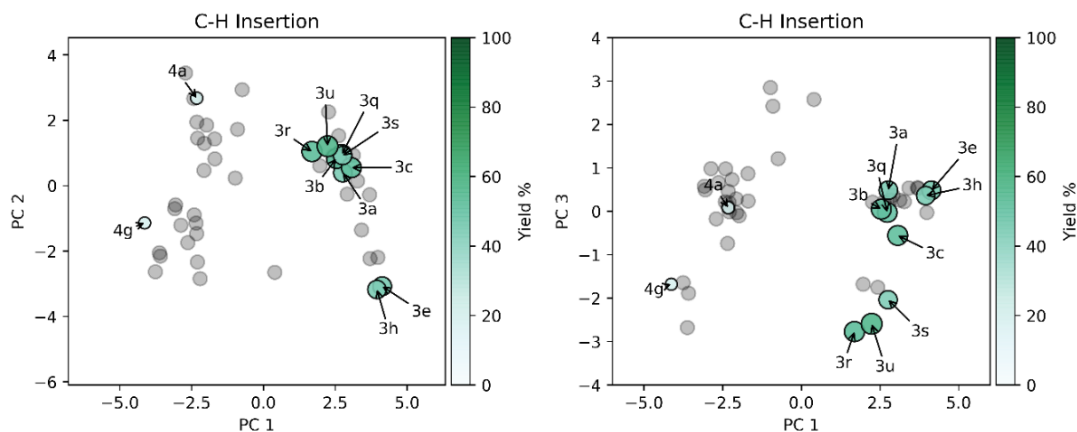


Figure 5.2. PC1/PC2 and PC1/PC3 plots with yield overlay for C-H insertion.

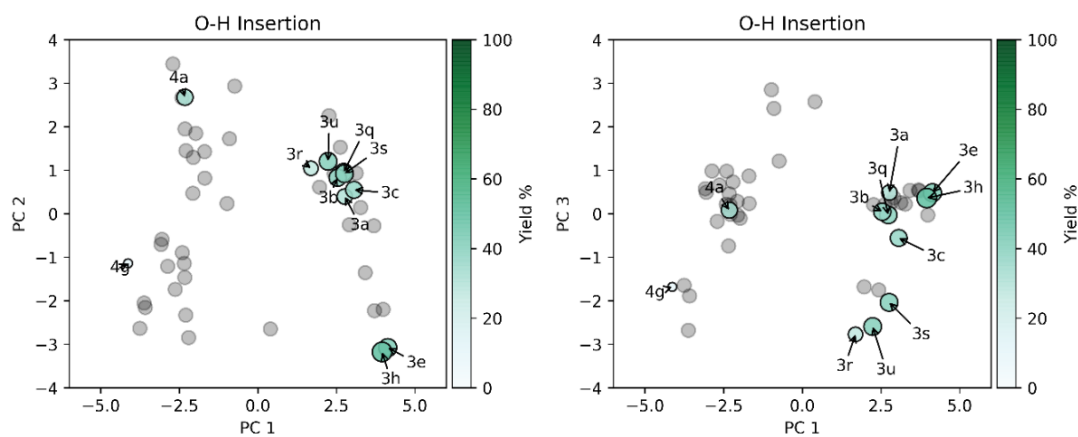


Figure 5.3. PC1/PC2 and PC1/PC3 plots with yield overlay for O-H insertion.

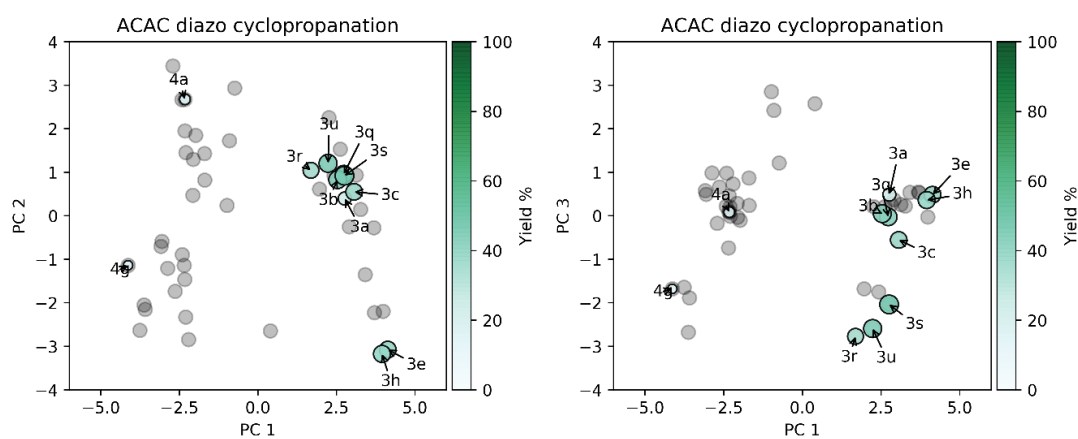


Figure 5.4. PC1/PC2 and PC1/PC3 plots with yield overlay for the ACAC diazo cyclopropanation.

5.3 Materials and Methods for Activity-Directed Synthesis

5.3.1 Synthesis of Diazo Compounds

CAUTION: All diazo compounds (excluding those isolated as solid material) described below appear to be volatile at room temperature under reduced pressure. Gradual loss of mass was observed when left under high vacuum. Diazo compounds are potentially explosive on contact and should be treated with caution, although no adverse events occurred during the experiments described in this thesis.

General procedure for diazo transfer, A

4-Acetamidobenzenesulfonyl azide (1.2 eq) was added portion-wise to a solution of benzylic ester or tertiary amide (1 eq) and 1,8-Diazabicyclo[5.4.0]undec-7-ene (1.4 eq) in acetonitrile (substrate concentration: 0.01 M) at 0 °C using an ice-bath. After 16 hours, the solvent was removed under reduced pressure to give a crude product that was dissolved in diethyl ether (40 mL) and washed sequentially with 10% w/w citric acid (2 x 10 mL), brine (2 x 10 mL), 10% w/v ammonium chloride (2 x 10 mL) and brine (2 x 10 mL). The organics were passed through a phase separation frit and concentrated under reduced pressure to give a crude product.

General procedure for one-pot acylation and diazo transfer, B

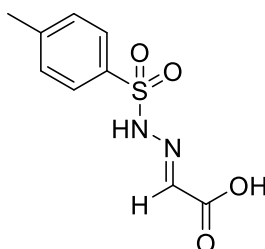
A secondary amine (1 eq) and 2,2,2,6-trimethyl-1,3-dioxin-4-one (1.5 eq) were added sequentially to a 20 mL microwave vial, dissolved in toluene (substrate concentration: 1.4 M) and heated under microwave irradiation at 110 °C for 30 minutes. The solvent was then removed under reduced pressure and the crude material dissolved in acetonitrile (substrate concentration: 0.01 M). The solution was then cooled to 0 °C and 1,8-Diazabicyclo[5.4.0]undec-7-ene (1.4 eq) and 4-Acetamidobenzenesulfonyl azide (1.2 eq) were added sequentially. After 16 hours, the solvent was removed under reduced pressure to give a crude product that was dissolved in diethyl ether (40 mL) and washed sequentially with 10% w/w citric acid (2 x 10 mL), brine (2 x 10 mL), 10% w/v ammonium chloride (2 x 10 mL) and brine (2 x 10 mL). The organics were passed through a phase separation frit and concentrated under reduced pressure to give a crude product.

General procedure for hydrazone decomposition, C

Compound **3.17** (1 eq) was suspended in Toluene (substrate concentration: 0.2 M) and stirred. Thionyl chloride (2 eq) was added and the reaction

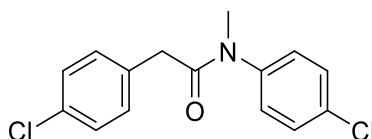
heated to 90 °C for 3 hours, then the solvent removed under vacuum to give an orange solid. The solid was then dissolved in DCM (substrate concentration: 0.1 M) and cooled to 0 °C using an ice-bath. A solution of primary or secondary amine (3.4 mmol) and *N,N*-dimethylaniline (1 eq) in DCM (0.3 M) was subsequently added slowly to the stirred solution over 5 minutes. After 1-hour triethylamine (5 eq) was added and the reaction allowed to warm to room temperature overnight. The organics were then washed sequentially with 10% w/w citric acid (2 x 20 mL), brine (2 x 20 mL), 10% w/v ammonium chloride (2x 20 mL) and brine (2 x 20 mL), passed through a phase separation frit and concentrated under reduced pressure to give a crude product.

2-(4-Methylbenzenesulfonamido)imino Acetic Acid, **3.17**



Glyoxylic acid (7.4g, 80 mmol) and *p*-toluenesulfonylhydrazide (10 g, 50 mmol) were dissolved in THF and the mixture stirred vigorously. After 24 hours, the solvent was removed under reduced pressure to give a viscous oil which was triturated with distilled water until a fine white powder was obtained. The solid was then air-dried overnight and then recrystallised in EtOAc to afford the *hydrazone* **3.17** as a white amorphous solid (10.9 g, 90%). δ_{H} (400 MHz, Acetone- d_6) 11.01 (1H, s, acetic acid-H), 7.82 (2H, d, *J* 8.0, phenyl-2H and -6H), 7.43 (2H, d, *J* 8.0, phenyl-3H and -5H), 7.32 (1H, s, imino-2H) and 2.42 (3H, s, methyl). δ_{C} (100 MHz) 206.3, 163.8, 145.3, 137.3, 130.6, 128.5 and 21.4.

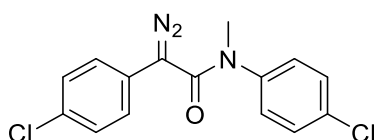
***N*,2-bis(4-chlorophenyl)-*N*-methylacetamide**



4-Chlorophenyl acetic acid (1.00 g, 5.9 mmol) was dissolved in DCM (8 mL), then 4-chloro-*N*-methyl aniline (707 μL , 5.9 mmol) and DCC (1.33 g, 6.5 mmol) were added sequentially. After 16 hours of stirring the reaction was

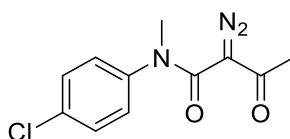
cooled to 0 °C using an ice-bath and a white precipitate formed. The by-product precipitate was then removed by filtration and the solvent removed under reduced pressure to give a light-yellow oil, which passed through a silica plug eluting 9:1 Pentane/Et₂O then 100% Et₂O to give the *amide* as a crude colourless oil (885 mg, 30%). δ_{H} (400 MHz, Acetone-d₆) 7.48 (2H, d, *J* 8.3, *N*-chlorophenyl-2H and -6H), 7.35 – 7.33 (2H, m, chlorophenyl-2H and 6H), 7.26 (2H, d, *J* 8.3, *N*-chlorophenyl-3H and -5H), 7.15 – 7.07 (2H, m, chlorophenyl-3H and -5H), 3.46 (2H, s, 2H₂) and 3.22 (3H, s, methyl).

***N*,2-Bis(4-chlorophenyl)-2-diazo-*N*-methylacetamide, D3**



According to general procedure **A**, *N*,2-bis(4-chlorophenyl)-*N*-methylacetamide (300 mg, 1.0 mmol) gave a crude material. The viscous orange oil was purified by flash column chromatography eluting with 8:2 Pentane/Et₂O to yield the *diazo* **D3** as an orange oil (114 mg, 35%), *R_f* 0.14 (1:1 Pentane/Et₂O); δ_{H} (400 MHz, Acetone-d₆) 7.42 – 7.30 (8H, m, phenyl-H) and 3.36 (3H, s, methyl). δ_{C} (100 MHz) 164.9, 144.2, 132.7, 131.7, 130.8, 130.0, 129.7, 128.3, 127.7, 127.2 and 38.7. IR ν_{max} (CH₂Cl₂ film/cm⁻¹) 2066 (diazo), 1635, 1490 and 1313. HRMS (ESI): C₁₅H₁₁³⁵Cl₂N₃O requires [M+Na]⁺, calculated 342.0177, found 342.0165.

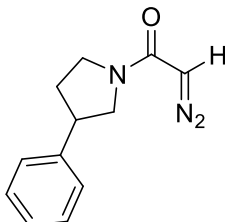
***N*-(4-Chlorophenyl)-2-diazo-*N*-methyl-3-oxobutanamide, D1²⁸³**



According to general procedure **B**, 4-Chloro-*N*-methyl aniline (1.00 mL, 8.3 mmol) gave a crude material. During the reaction a white precipitate formed, which was removed by filtration, and the solvent then removed under reduced pressure to give a dark red oil which was purified by flash column chromatography eluting with 100% DCM, then 9:1 DCM/Et₂O, to yield the *diazo* **D1** as a yellow oil that solidified on standing (1.34 g, 64%), *R_f* 0.10 (DCM); δ_{H} (400 MHz, CDCl₃) 7.39 (2H, d, *J* 8.7, chlorophenyl-2H and -6H), 7.14 (2H, d, *J* 8.7, chlorophenyl-3H and -5H), 3.34 (3H, s, *N*-methyl) and 2.46 (3H, s, oxobutanamide-H₃). δ_{C} (100 MHz) 191.4, 161.1, 157.1, 141.7,

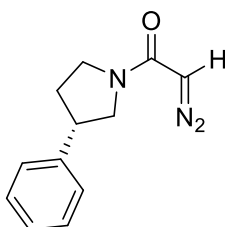
133.9, 130.6, 127.5, 38.6 and 28.5. IR ν_{\max} (CH₂Cl₂ film)/cm⁻¹ 2184 (diazo), 1636, 1590, 1487, 1356. HRMS (ESI): C₁₁H₁₀³⁵ClN₃O₂ requires [M+Na]⁺, calculated 274.0359, found 274.0350.

2-Diazo-1-(3-phenylpyrrolidin-1-yl)ethan-1-one, D4



According to general procedure **C**, 3-phenyl pyrrolidine (500 mg, 3.4 mmol) gave a crude material. The dark, viscous orange oil which was purified by flash column chromatography eluting with 100% DCM for 3 column volumes then 90:10 DCM/Et₂O to afford the *diazo* **D4** as a bright orange oil (574 mg, 78%), R_F 0.32 (Et₂O); δ_H (500 MHz, CD₂Cl₂) 7.33 (2H, dd, J 8.0 and 7.0, phenyl-3H and -5H), 7.26 (3H, m, phenyl-2H, -4H and -6H), 4.89 (1H, s, ethenone-2H), 4.02 – 3.81 (1H, m, pyrrolidinyl-3H), 3.61 – 3.23 (4H, m, pyrrolidinyl-4H₂ and -5H₂), 2.35 – 2.29 (1H, m, pyrrolidinyl-2H_a) and 2.07 – 2.00 (1H, m, pyrrolidinyl-2H_b). δ_C (125 MHz) 163.8 (rot-A and rot-B), 141.4 (rot-B), 141.1 (rot-A), 129.9 (rot-B), 129.8 (rot-A), 128.7 (rot-A and rot-B), 127.14 (rot-A and rot-B), 52.5 (rot-A and rot-B), 52.3 (rot-A and rot-B), 46.2 (rot-A), 45.8 (rot-B), 44.6 (rot-A), 42.9 (rot-B), 33.4 (rot-A) and 33.1 (rot-B). IR ν_{\max} (CH₂Cl₂ film)/cm⁻¹ 2095 (diazo), 1600, 1417 and 1166. HRMS (ESI): C₁₂H₁₃N₃O requires [M+H]⁺, calculated 216.1137, found 216.1127.

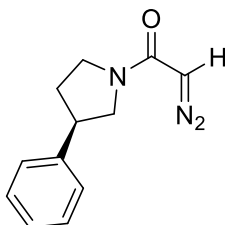
2-Diazo-1-((S)-3-phenylpyrrolidin-1-yl)ethan-1-one, D4a



According to general procedure **C**, (S)-3-phenyl pyrrolidine (147 mg, 1.0 mmol) gave a crude material. The dark, viscous orange oil which was purified by flash column chromatography eluting with 100% DCM for 3 column volumes then 90:10 DCM/Et₂O to afford the *diazo* **D4a** as a bright orange oil (58.1 mg, 27%), R_F 0.46 (Et₂O); $[\alpha]_D^{20} = -35$ degrees cm² g⁻¹ (c

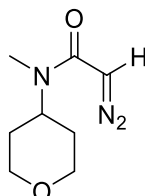
1.00, CH₂Cl₂); spectroscopically identical to 2-diazo-1-(3-phenylpyrrolidin-1-yl)ethan-1-one.

2-Diazo-1-((*R*)-3-phenylpyrrolidin-1-yl)ethan-1-one, *ent*-D4a



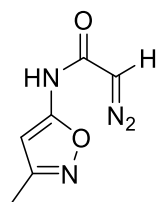
According to general procedure **C**, (*R*)-3-phenyl pyrrolidine (147 mg, 1.0 mmol) gave a crude material. The dark, viscous orange oil which was purified by flash column chromatography eluting with 100% DCM for 3 column volumes then 90:10 DCM/Et₂O to afford the *diazo ent*-D4a as a bright orange oil (143 mg, 67%), *R*_F 0.54 (Et₂O); [α]_D²⁰ = +28 degrees cm² g⁻¹ (c 1.00, CH₂Cl₂); spectroscopically identical to 2-diazo-1-(3-phenylpyrrolidin-1-yl)ethan-1-one.

2-Diazo-*N*-methyl-*N*-(oxan-4-yl)acetamide, D5



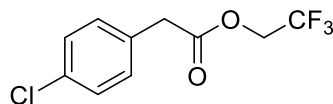
According to general procedure **C**, *N*-methyl-*N*-tetrahydro-2H-pyran-4-ylamine (500 mg, 4.3 mmol) gave a crude material. The dark orange oil was purified by flash column chromatography eluting with 90:10 DCM/Et₂O for 3 column volumes then 100% Et₂O to afford *diazo* D5 as a bright orange oil (708 mg, 90%), *R*_F 0.10 (Et₂O); δ _H (500 MHz, CD₂Cl₂) 5.01 (1H, s, acetamide-2H), 3.97 (2H, dd, *J* 11.2, 4.6, oxanyl-3H_a and -5H_a), 3.44 (2H, app. t, *J* 11.4, oxanyl-3H_b and -5H_b), 2.7 (3H, s, *N*-methyl), 1.77 – 1.73 (3H, m, oxanyl-1H and 2H₂) and 1.55 – 1.52 (2H, m, oxanyl-6H₂). δ _C (125 MHz) 165.9, 104.6, 67.8, 64.2, 46.8 and 30.7. IR ν _{max} (CH₂Cl₂ film)/cm⁻¹ 2098 (diazo), 1600, 1410, 1256, 1144 and 1086. HRMS (ESI): C₈H₁₃N₃O₂ requires [M+Na]⁺, calculated 206.0905, found 206.0892.

2-Diazo-*N*-(3-methyl-1,2-oxazol-5-yl)acetamide, D7



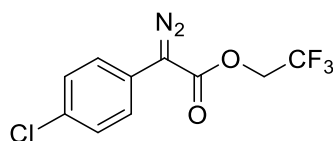
According to general procedure **C**, 5-amino-3-methylisoxazole (422 mg, 4.3 mmol) gave a crude material. The dark orange oil was purified by flash column chromatography eluting with 90:10 DCM/Et₂O to afford *diazo* **D7** as a bright orange oil (170 mg, 24%), *R*_F 0.49 (Et₂O); δ_H (500 MHz, Acetone-d₆) 10.17 (1H, broad s, *N*-acetamide-H), 6.13 (1H, s, oxazolyl-4H), 5.99 (1H, s, acetamide-2H) and 2.19 (3H, s, methyl). δ_C (125 MHz) 162.7 (rot-A), 162.6 (rot-B), 162.2 (rot-A), 162.1 (rot-B), 161.6 (rot-A and rot-B), 88.7 (rot-A), 88.6 (rot-B), 49.3 (rot-A and rot-B) and 11.6 (rot-A and rot-B). IR ν_{max} (CH₂Cl₂ film)/cm⁻¹ 3205, 2109 (diazo), 1648, 1551, 1381, 1364, 1197 and 1155. HRMS (ESI): C₆H₆N₄O₂ requires [M+H]⁺, calculated 189.0388, found 189.0379.

2,2,2-Trifluoroethyl 2-(4-chlorophenyl)acetate



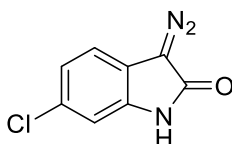
4-Chlorophenyl acetic acid (1.0 g, 5.9 mmol) was suspended in 2,2,2-trifluoroethanol (60 mL) and concentrated H₂SO₄ (0.4 mL) was added, then the suspension heated to reflux. After 16 hours the reaction was cooled to room temperature and neutralised to pH 7 with 1M NaOH, and the solvent removed under vacuum. The oil was then dissolved in DCM (75 mL) and washed sequentially with brine (3 x 30 mL). The organics were then collected, passed through a phase separation frit, and dried under vacuum to afford the *ester* as a crude colourless oil (1.4 g, 98%). δ_H (500 MHz, Chloroform-d) 7.32 (2H, d, *J* 8.5, chlorophenyl-2H and -6H), 7.22 (2H, d, *J* 8.5, chlorophenyl-2H and -6H), 4.48 (2H, q, *J* 8.4, ethyl-1H₂), 3.70 (2H, s, acetate-2H). δ_C (100 MHz) 169.7, 133.7, 131.3, 130.8, 129.0, 123.0 (q, *J*_{C-F} 277 Hz), 60.8 (q, *J*_{C-F} 37 Hz), 40.0.

2,2,2-Trifluoroethyl 2-(4-chlorophenyl)-2-diazoacetate, **D2**²²⁰



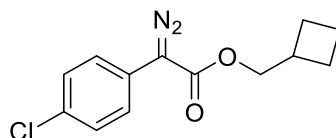
According to general procedure **A**, 2,2,2-Trifluoroethyl 2-(4-chlorophenyl)acetate (500 mg, 2.0 mmol) gave a crude material which was purified by flash column chromatography eluting with 20:1 Pentane/Et₂O to give the *diazo* **D2** as an orange oil (235 mg, 43%), *R*_F 0.25 (20:1 Pentane/Et₂O); (235 mg, 43%). δ_H (500 MHz, Chloroform-d) 7.39 (4H, m, chlorophenyl) and 4.65 (2H, q, *J* 8.3, ethyl-1H₂). δ_C (125 MHz) 163.1, 132.2, 129.4, 124.3, 123.0 (q, *J*_{C-F} 277), 121.6 and 60.5 (q, *J*_{C-F} 37). IR ν_{max} (CH₂Cl₂ film)/cm⁻¹ 2093 (diazo), 1715, 1494, 1283, 1168 and 1141.

6-Chloro-3-diazo-1H-indol-2-one, **D8**²⁸⁴



According to general procedure **A**, 6-chloro-2-oxindol (336 mg, 2.0 mmol) gave a crude material that was purified by flash column chromatography eluting with 100% DCM then 95:5 DCM/Et₂O to yield the *diazo* **D8** as an orange oil (256 mg, 72%), *R*_F 0.19 (95:5 DCM/Et₂O); δ_H (500 MHz, Acetone-d₆) 9.72 (1H, broad s, indolone-NH), 7.39 (1H, d, *J* 8.2, indolone-4H), 7.06 (1H, dd, *J* 8.2 and 1.9, indolone-5H) and 7.02 (1H, d, *J* 1.9, indolone-7H). δ_C (125 MHz) 168.2, 134.8, 131.0, 122.2, 120.9, 117.2, 111.0, 111.0 and 61.4. HRMS (ESI): C₈H₄³⁵ClN₃O requires [M+Na]⁺, calculated 215.9940, found 215.9929.

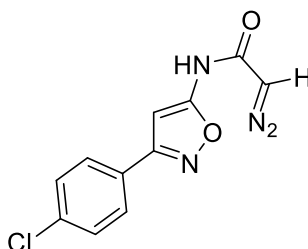
Cyclobutylmethyl 2-(4-chlorophenyl)-2-diazoacetate, **D9**



N,N-diisopropylethylamine (2.9 mL, 17 mmol) was added to a stirred suspension of EDC.HCl (882 mg, 4.6 mmol) and 4-dimethylaminopyridine (51 mg, 0.40 mmol) in DCM (50 mL), followed by the sequential addition of cyclobutane methanol (0.4 mL, 4.6 mmol) and 4-chlorophenyl acetic acid (717 mg, 4.2 mmol). After 20 hours half the solvent was removed under

vacuum and the reaction mixture washed with 10% w/v aqueous citric acid (2 x 50 mL), brine (1 x 50 mL), 10% v/v aqueous NaHCO₃ (2 x 50 mL), brine (1 x 50 mL) and distilled water (3 x 100 mL). The combined organic layer was then passed through a phase separation frit and concentrated under reduced pressure to give a crude oil. The oil was then dissolved in acetonitrile (42 mL, 0.1 M) and cooled to 0 °C using an ice-bath, 1,8-Diazabicyclo[5.4.0]undec-7-ene (0.9 mL, 1.4 eq) was then added followed by the portion-wise addition of 4-Acetamidobenzenesulfonyl azide (1.1 g, 1.2 eq). After 16 hours, the solvent was removed under reduced pressure to give a crude product that was dissolved in diethyl ether (40 mL) and washed sequentially with 10% w/w citric acid (2 x 10 mL), brine (2 x 10 mL), 10% w/v ammonium chloride (2 x 10 mL) and brine (2 x 10 mL). The organics were passed through a phase separation frit and concentrated under reduced pressure to give a crude material that was purified by flash column chromatography eluting with 20:1 Pentane/Et₂O to give the *diazo* **D9** as an orange oil (1.1 g, 99%), *R_F* 0.46 (20:1 Pentane/Et₂O); δ_H (500 MHz, Acetone-d₆) 7.57 (2H, d, *J* 8.7, chlorophenyl-2H and -6H), 7.43 (2H, d, *J* 8.7, chlorophenyl-3H and -5H), 4.24 (2H, d, *J* 6.6, cyclobutylmethyl-4H₂), 2.71 (1H, dt, *J* 14.9 and 7.3, cyclobutylmethyl-3H), 2.12 – 2.06 (2H, m, cyclobutylmethyl-1H₂) and 1.97 – 1.81 (4H, m, cyclobutylmethyl-2H₂ and -2'H₂). δ_C (125 MHz) 165.3, 131.6, 129.7, 126.2, 126.0, 69.1, 63.7, 35.1, 25.1 and 18.9. IR ν_{max} (CH₂Cl₂ film)/cm⁻¹ 2084 (diazo), 1682 and 1490. HRMS (ESI): C₁₃H₁₃³⁵ClN₂O₂ requires [2M+H -N₂]⁺, calculated 501.1348, found 501.1343.

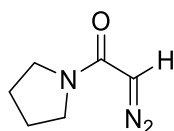
***N*-[3-(4-Chlorophenyl)-1,2-oxazol-5-yl]-2-diazoacetamide, D10**



According to general procedure **C**, 3-(4-chlorophenyl)isoxazole-5-amine (668 mg, 3.4 mmol) gave a crude material that was purified by flash column chromatography eluting with 9:1 DCM/Et₂O to afford *diazo* **D10** as a bright orange oil (165 mg, 19%), *R_F* 0.09 (9:1 DCM/Et₂O); δ_H (500 MHz, Acetonitrile-d₃) 9.14 (1H, s, diazoacetamide-NH), 7.82 (2H, d, *J* 8.7, phenyl-2H and -6H), 7.49 (2H, d, *J* 8.7, phenyl-3H and -5H), 6.65 (1H, s, oxazolyl-

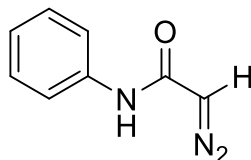
4H) and 5.31 (1H, s, diazoacetamide-2H). δ_{C} (125 MHz) 163.3, 163.2, 163.1, 136.5, 130.0, 129.2, 129.1, 86.3, and 49.9. IR ν_{max} (CH_2Cl_2 film)/ cm^{-1} 2994, 2113 (diazo), 1678 and 1364. HRMS (ESI): $\text{C}_{11}\text{H}_7^{35}\text{ClN}_4\text{O}_2$ requires $[\text{M}+\text{H}]^+$, calculated 263.0336, found 263.0323.

2-Diazo-1-(pyrrolidin-1-yl)ethenone, 3.14



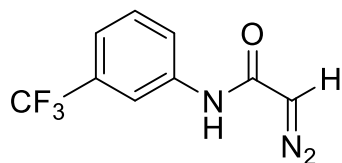
According to general procedure **C**, pyrrolidine (2 mL, 24 mmol) gave a crude material that was purified by flash column chromatography eluting with 100% Et_2O to afford the *diazo* **3.14** as a bright orange oil (828 mg, 50%), R_{F} 0.2 (Et_2O); δ_{H} (500 MHz, Acetone- d_6) 5.34 (1H, s, ethenone-2H), 3.41 – 3.27 (4H, m, pyrrolidiny-2H₂ and -2'H₂) and 1.95 – 1.80 (4H, m, pyrrolidiny-3H₂ and -3'H₂). δ_{C} (125 MHz) 164.0, 57.7, 46.5 and 18.9. HRMS (ESI): $\text{C}_6\text{H}_9\text{N}_3\text{O}$ requires $[\text{M}+\text{Na}]^+$, calculated 162.0643, found 162.0638.

2-Diazo-N-phenylacetamide, 3.16



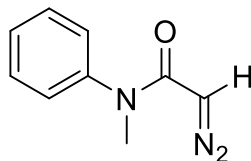
According to general procedure **C**, aniline (600 μL , 6.6 mmol) gave a crude material that was purified by flash column chromatography eluting with 9:1 DCM/ Et_2O to afford the *diazo* **3.14** as an orange solid (682 mg, 72%), R_{F} 0.52 (9:1 DCM/ Et_2O); δ_{H} (400 MHz, Acetone- d_6) 8.88 (1H, s, acetamide-NH), 7.62 – 7.58 (2H, m, phenyl-2H and -6H), 7.30 – 7.24 (2H, m, phenyl-3H and -5H), 7.04 – 6.99 (1H, m, phenyl-4H) and 5.40 (1H, s, acetamide-2H). δ_{C} (100 MHz) 164.4, 140.7, 129.6, 123.7, 119.7, 119.6 and 48.6. IR ν_{max} (CH_2Cl_2 film)/ cm^{-1} 3085, 2093 (diazo), 1631, 1548, 1442 and 1369. HRMS (ESI): $\text{C}_8\text{H}_8\text{N}_3\text{O}$ requires $[\text{M}+\text{H}]^+$, calculated 162.0667, found 162.0655.

2-diazo-*N*-[3-(trifluoromethyl)phenyl]acetamide, 3.1



According to general procedure **C**, 3-trifluoromethyl aniline (378 μ L, 3 mmol) gave a crude material that was purified by flash column chromatography eluting with 9:1 DCM/Et₂O to afford the *diazo* **3.14** as a pale yellow solid (145 mg, 21%), R_F 0.30 (9:1 DCM/Et₂O); δ_H (400 MHz, Acetone-d₆) 9.21 (1H, s, acetamide-MH), 8.13 (1H, s, phenyl-2H), 7.75 (1H, broad app. d, J 8.0, phenyl-3H), 7.50 (1H, t, J 8.0, phenyl-5H), 7.35 (1H, broad app. d, J 8.0, phenyl-6H), 5.46 (1H, s, acetamide-2H). δ_C (100 MHz) 165.0 (rot A), 164.9 (rot B), 141.5 (rot A), 141.4 (rot B), 131.4 (q, J_{C-F} 32), 130.6, 125.2 (q, J_{C-F} 272), 122.9 (rot A), 122.8 (rot B), 120.0 (rot A), 120.0 (rot B), 116.0 (q, J_{C-F} 4), 49.0.

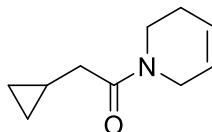
2-Diazo-*N*-methyl-*N*-phenylacetamide, 3.12



According to general procedure **C**, *N*-methylaniline (650 μ L, 6 mmol) gave a crude material that was purified by flash column chromatography eluting with 9:1 DCM/Et₂O to afford the *diazo* **3.12** as an orange oil (419 mg, 80%), R_F 0.45 (1:1 pentane/Et₂O). δ_H (400 MHz, Acetone-d₆) 7.48 – 7.28 (2H, m, phenyl-2H and -6H), 7.28 – 7.18 (1H, m, phenyl-4H), 7.20 – 7.16 (2H, m, phenyl-3H and -5H), 4.81 (1H, s, acetamide-2H) and 3.12 (3H, s, methyl). δ_C (100 MHz) 165.7, 144.4, 130.5, 128.4, 128.2, 47.4 and 37.1. IR ν_{max} (CH₂Cl₂ film)/cm⁻¹ 3118, 3061, 2934, 2101 (diazo) and 1621.

5.3.2 Synthesis of Co-Substrates

2-Cyclopropyl-1-(1,2,3,6-tetrahydropyridin-1-yl)ethan-1-one, **S5**



Cyclopropylacetic acid (0.5 g, 5.0 mmol) and carbonylimidazole (0.8 g, 5.0 mmol) were dissolved in THF (20 mL) and stirred for 30 minutes, followed by dropwise addition of 1,2,3,6-tetrahydropyridine (0.4 g, 5.0 mmol) in THF (4 mL). After 16 hours 1M HCl (10 mL) was added and the mixture stirred vigorously for 10 minutes. The solvent was then reduced to a minimum under vacuum and partitioned with DCM (30 mL). The organics were washed sequentially with 20% v/v NaHCO₃ (1 x 20 mL) and brine (1 x 20 mL), passed through a phase separation filter, and dried under vacuum to afford *amide* **S5** as a colourless oil (0.7 g, 86%). δ_{H} (500 MHz, Chloroform-d) 5.89 – 5.77 (1H, m, *THP-5H), 5.70 – 5.61 (1H, m, THP-4H), 4.05 – 3.90 (2H, m, THP-6H₂), 3.67 (1H, t, *J* 5.8, THP-2H_a), 3.49 (1H, t, *J* 5.8, THP-2H_b), 2.26 (2H, dd, *J* 6.8 and 12.0, ethanone-2H₂), 2.18 – 2.11 (2H, m, THP-3H₂), 1.07 – 1.01 (1H, m, cyclopropyl-1H), 0.56 – 0.51 (2H, m, cyclopropyl-2H_a and 2'H_a) and 0.18 – 0.13 (2H, m, cyclopropyl-2H_b and 2'H_b). δ_{C} (125 MHz) 171.5 (rot-A), 171.5 (rot-B), 126.8 (rot-A), 125.1 (rot-B), 124.6 (rot-A), 123.4 (rot-B), 45.1 (rot-A), 42.7 (rot-A), 42.0 (rot-B), 39.1 (rot-B), 38.8 (rot-A), 38.3 (rot-B), 26.0 (rot-A), 25.0 (rot-B), 7.4 (rot-A), 7.2 (rot-B), 4.6 (rot-A), 4.5 (rot-B). IR ν_{max} (CH₂Cl₂ film)/cm⁻¹ 3079, 2918, 1622 and 1431. HRMS (ESI): C₁₀H₁₅NO requires [M+H]⁺, calculated 166.1231, found 166.1226. *THP = tetrahydropyridin-1-yl.

5.3.3 Synthesis of MDM2 Ligands

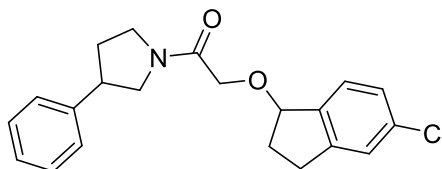
General procedure for the scale-up of ADS hits, D

A crimp vial (10 or 20 mL) was sequentially charged with solutions of rhodium(II) catalyst in CH₂Cl₂ (240 μ L, 12.5 mM) and co-substrate in CH₂Cl₂ (240 μ L, 6.25 M) and stirred. A solution of diazo in CH₂Cl₂ (240 μ L, 1.25 M) was added and the vial capped. After 24 hours 900 mg of Quadrapure TUTM resin was added, followed by a further 720 μ L CH₂Cl₂. After a further 24 hours the resin was removed by filtration and the solvent evaporated under reduced pressure to yield the crude reaction product.

General procedure for the scale-up of ADS hits/analogues using a syringe pump, E

A 20 mL vial was charged with Rhodium(II) catalyst (1 mol%) and degassed under N₂ atmosphere, followed by the addition of co-substrate (6.25 M) in CH₂Cl₂. Diazo (1.25 M) in CH₂Cl₂ was then added dropwise to the stirred solution over 6 hours using a syringe pump. After 24 hours Quadrapure TU™ resin was then added and the reaction left for a further 24 hours. The resin was then removed by filtration and the solvent removed under reduced pressure to give a crude product.

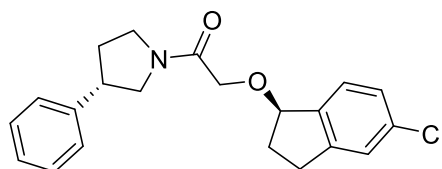
2-[(5-Chloro-2,3-dihydro-1H-inden-1-yl)oxy]-1-(3-phenylpyrrolidin-1-yl)ethanone, P2



According to general procedure **D**, Rh₂piv₄ (2.8 mg, 4.6 μmol), 2-diazo-1-(3-phenylpyrrolidin-1-yl)ethan-1-one (100 mg, 0.46 mmol) and 5-chloro-2,3-dihydro-1H-inden-1-ol (391 mg, 2.32 mmol) gave an orange oil. The crude oil was then purified by flash column chromatography eluting with 100% Et₂O to afford *ether* **P2** as a colourless oil (127 mg, 78%), *R*_F 0.18 (100% Et₂O); δ_H (600 MHz, d⁶-DMSO, 1:1 mixture of diastereomers; 1:1 mixture of rotamers): 7.48 – 7.45 (1H, m, indenyl-6H), 7.35 – 7.23 (7H, m, phenyl-H, and indenyl-4H and -7H), 4.97 – 4.93 (1H, m, indenyloxy-1H), 4.21 – 4.14 (2H, m, ethanone-2H), 3.89 – 3.83 (1H, m, pyrrolidiny-3H), 3.66 – 3.59 (1H, m, pyrrolidiny-2H_a), 3.50 – 3.39 (1H, m, pyrrolidiny-2H_b), 3.37 – 3.21 (2H, m, pyrrolidiny-5H), 3.00 – 2.74 (2H, m, pyrrolidiny-4H), 2.33 – 2.219 (2H, m, 2,3-dihydroindenyl-3H) and 2.03 – 1.87 (2H, m, 2,3-dihydroindenyl-2H). δ_C (150 MHz, d⁶-DMSO): 167.3 (broad s, major), 167.3 (rot-A, minor), 167.2 (rot-B, minor), 146.3 (broad s, major), 146.3 (broad s, minor), 141.6 (broad s, major), 141.6 (broad s, minor), 141.2 (broad s, major), 141.1 (broad s, minor), 132.9 (broad s, major+minor), 128.5 (broad s, major+minor), 127.1 (major), 127.0 (minor), 126.8 (major), 126.8 (minor), 126.7 (broad s, major), 126.6 (broad s, minor), 126.2 (broad s, major), 126.2 (broad s, minor), 124.7 (broad s, major+minor), 82.0 (broad s, major), 81.9 (broad s, minor), 67.7 (rot-A, major), 67.7 (rot-B, major), 67.6 (broad s, minor), 51.6 (broad s, major), 51.2 (broad s, minor), 45.4 (broad s, major), 45.0 (broad s, minor), 43.7 (broad s, major), 41.6 (broad s, minor), 33.0 (broad s, major), 33.0

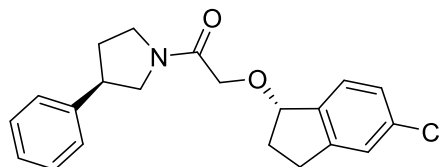
(broad s, minor), 32.0 (broad s, major), 32.0 (broad s, minor), 30.8 (broad s, major) and 30.0 (broad s, minor). HRMS (ESI): $C_{21}H_{22}^{35}ClNO_2$ requires $[M+H]^+$, calculated 356.1417, found 356.1425.

2-(((1*R*)-5-Chloro-2,3-dihydro-1H-inden-1-yl]oxy)-1-(((3*S*)-3-phenylpyrrolidin-1-yl]ethenone, **P2a**



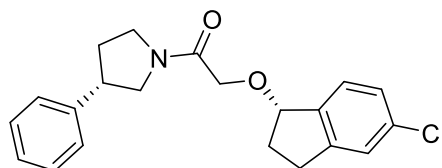
According to general procedure **D**, Rh_2piv_4 (0.6 mg, 1.0 μ mol), 2-diazo-1-((*S*)-3-phenylpyrrolidin-1-yl)ethan-1-one (21.5 mg, 0.1 mmol) and (*1R*)-5-chloro-2,3-dihydro-1H-inden-1-ol (84.0 mg, 0.5 mmol) gave an orange oil. The crude oil was then purified by flash column chromatography eluting with 100% Et_2O to afford *ether* **P2a** as a colourless oil (24 mg, 67%), R_f 0.48 (100% Et_2O); δ_H (500 MHz, Chloroform-*d*): 7.32 – 7.09 (16H, m, Ar, rot-A and rot-B), 4.93 (1H, ddd, J 17.6, 6.4 and 3.7, inden-1-yloxy-1H, rot-A and rot-B), 4.11 – 4.09 (2H, m, ethenone-2H₂, rot-A and rot-B), 3.98 – 3.93 (1H, m, pyrrolidiny-2H_a, rot-A), 3.80 – 3.72 (2H, m, pyrrolidiny-3H, rot-A and rot-B), 3.65 – 3.60 (1H, m, pyrrolidiny-2H_a, rot-B), 3.50 – 3.25 (6H, m, pyrrolidiny-2H_a and -5H₂, rot-A and rot-B), 3.01 – 2.96 (2H, m, 2,3-dihydroindenyl-3H_a, rot-A and rot-B), 2.74 – 2.71 (2H, m, 2,3-dihydroindenyl-3H_b, rot-A and rot-B), 2.33 – 2.20 (4H, m, pyrrolidiny-4H₂, rot-A and rot-B), 2.13 – 2.05 (2H, m, 2,3-dihydroindenyl-2H_a, rot-A and rot-B) and 2.00 – 1.88 (2H, m, 2,3-dihydroindenyl-2H_b, rot-A and rot-B). δ_C (125 MHz, Chloroform-*d*): 168.3 (rot A), 168.2 (rot B), 146.4 (rot A), 146.4 (rot B), 141.0 (rot-A and rot-B), 140.8 (rot A), 140.6 (rot B), 134.6 (rot-A and rot-B), 128.9 (rot A), 128.8 (rot B), 127.2 (rot A), 127.1 (rot B), 126.8 (rot A), 126.7 (rot B), 126.6 (rot A), 126.6 (rot B), 125.3 (rot-A and rot-B), 83.1 (rot-A and rot-B), 68.6 (rot A), 68.5 (rot B), 52.6 (rot A), 52.1 (rot B), 46.1 (rot A), 46.0 (rot B), 44.7 (rot-A and rot-B), 42.3 (rot-A and rot-B), 33.8 (rot-A and rot-B), 32.6 (rot-A and rot-B), 31.3 (rot-A and rot-B) and 30.3 (rot-A and rot-B). HRMS (ESI): $C_{21}H_{22}^{35}ClNO_2$ requires $[M+Na]^+$, calculated 378.1237, found 378.1237.

2-(((1S)-5-Chloro-2,3-dihydro-1H-inden-1-yl]oxy)-1-((3R)-3-phenylpyrrolidin-1-yl]ethenone, *ent*-P2a



According to general procedure **D**, Rh₂piv₄ (0.6 mg, 1.0 μmol), 2-diazo-1-((*R*)-3-phenylpyrrolidin-1-yl)ethan-1-one (21.5 mg, 0.1 mmol) and (1*S*)-5-chloro-2,3-dihydro-1H-inden-1-ol (84.0 mg, 0.5 mmol) gave an orange oil. The crude oil was then purified by flash column chromatography eluting with 100% Et₂O to afford *ether ent*-**P2a** as a colourless oil (19 mg, 54%), *R*_F 0.32 (100% Et₂O); spectroscopically identical to compound **P2a**.

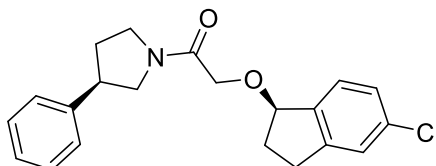
2-(((1S)-5-Chloro-2,3-dihydro-1H-inden-1-yl]oxy)-1-((3S)-3-phenylpyrrolidin-1-yl]ethenone, **P2b**



According to general procedure **D**, Rh₂piv₄ (0.6 mg, 1.0 μmol), 2-diazo-1-((*S*)-3-phenylpyrrolidin-1-yl)ethan-1-one (21.5 mg, 0.1 mmol) and (1*S*)-5-chloro-2,3-dihydro-1H-inden-1-ol (84.0 mg, 0.5 mmol) gave an orange oil. The crude oil was then purified by flash column chromatography eluting with 100% Et₂O to afford *ether* **P2b** as a colourless oil (28 mg, 79%), *R*_F 0.47 (100% Et₂O); δ_H (500 MHz, Chloroform-*d*): 7.33 – 7.08 (16H, m, rot-A and rot-B), 4.94 (2H, ddd, *J* 12.4, 6.5 and 3.7, inden-1-yloxy-1H, rot-A and rot-B), 4.14 – 4.09 (4H, m, ethenone-2H₂, rot-A and rot-B), 3.97 – 3.93 (1H, m, pyrrolidiny-2H_a, rot-A), 3.83 – 3.72 (2H, m, pyrrolidiny-3H, rot-A and rot-B), 3.65 – 3.61 (1H, m, pyrrolidiny-2H_b, rot-B), 3.48 – 3.24 (6H, m, pyrrolidiny-2H_a and -5H₂, rot-A and rot-B), 3.02 – 2.95 (2H, m, 2,3-dihydroindenyl-3H_a, rot-A and rot-B), 2.76 – 2.68 (2H, m, 2,3-dihydroindenyl-3H_b, rot-A and rot-B), 2.32 – 2.18 (4H, m, pyrrolidiny-4H₂, rot-A and rot-B), 2.11 – 2.06 (2H, m, 2,3-dihydroindenyl-2H_a, rot-A and rot-B) and 2.01 – 1.84 (2H, m, 2,3-dihydroindenyl-2H_b, rot-A and rot B). δ_C (125 MHz, Chloroform-*d*): 168.3 (rot A), 168.3 (rot B), 146.4 (rot-A and rot-B), 141.0 (rot A), 140.7 (rot-A and rot-B), 140.6 (rot B), 134.6 (rot-A and rot-B), 128.9 (rot A), 128.8 (rot B), 127.2 (rot-A and rot-B), 127.1 (rot A), 127.1 (rot B), 126.8 (rot A), 126.7 (rot B), 126.6 (rot A), 126.6 (rot B), 125.3 (rot A), 125.3 (rot B), 83.2 (rot A), 83.1 (rot

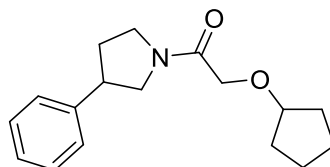
B), 68.7 (rot A), 68.5 (rot B), 52.6 (rot A), 52.1 (rot B), 46.1 (rot A), 46.0 (rot B), 44.7 (rot-A and rot-B), 42.3 (rot-A and rot-B), 33.7 (rot A), 32.6 (rot A and rot-B), 31.3 (rot B), 30.3 (rot-A) and 30.3 (rot B). HRMS (ESI): $C_{21}H_{22}^{35}ClNO_2$ requires $[M+Na]^+$, calculated 378.1237, found 378.1230.

2-(((1*R*)-5-Chloro-2,3-dihydro-1*H*-inden-1-yl)oxy)-1-((3*R*)-3-phenylpyrrolidin-1-yl)ethanone, *ent*-**P2b**



According to general procedure **D**, Rh_2piv_4 (0.6 mg, 1.0 μ mol), 2-diazo-1-((*R*)-3-phenylpyrrolidin-1-yl)ethan-1-one (21.5 mg, 0.1 mmol) and (1*R*)-5-chloro-2,3-dihydro-1*H*-inden-1-ol (84.0 mg, 0.5 mmol) gave an orange oil. The crude oil was then purified by flash column chromatography eluting with 100% Et_2O to afford *ether ent*-**P2b** as a colourless oil (18 mg, 51%), R_F 0.59 (100% Et_2O); spectroscopically identical to compound **P2b**.

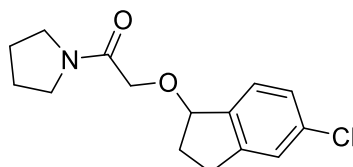
2-(Cyclopentyloxy)-1-(3-phenylpyrrolidin-1-yl)ethanone, **4.1**



According to general procedure **E**, Rh_2piv_4 (1.4 mg, 2.5 μ mol), 2-diazo-1-(3-phenylpyrrolidin-1-yl)ethan-1-one (50 mg, 0.25 mmol) and cyclopentanol (105 μ L, 1.2 mmol) gave an orange oil. The crude oil was then purified by flash column chromatography eluting with 100% Et_2O to afford *ether* **4.1** as a colourless oil (34 mg, 50%), R_F 0.29 (100% Et_2O); δ_H (500 MHz, $CDCl_3$): 7.29 – 7.16 (10H, m, phenyl-H, rot-A and rot-B), 4.00 (4H, d, J 11.4, ethenone-2H₂, rot-A and rot-B), 3.98 – 3.86 (4H, m, cyclopentyloxy-1H and pyrrolidinyl -2H_a, rot-A and rot-B), 3.74 (2H, ddd, J 10.9, 8.2 and 2.9, pyrrolidinyl-2H_b, rot-A), 3.68 (1H, ddd, J 10.9, 8.2 and 2.9, pyrrolidinyl-2H_b, rot-B), 3.54 – 3.25 (6H, m, pyrrolidinyl-3H and -5H₂, rot-A and rot-B), 2.33 – 2.19 (2H, m, pyrrolidinyl-4H_a, rot-A and rot-B), 2.04 – 1.87 (2H, m, pyrrolidinyl-4H_b, rot-A and rot-B) and 1.67 – 1.62 (16H, m, cyclopentyloxy-2H and -3H, rot-A and rot-B). δ_C (125 MHz, $CDCl_3$): 168.6 (rot A), 168.5 (rot B), 141.1 (rot A), 140.9 (rot B), 128.9 (rot A), 128.8 (rot B), 127.2 (rot A), 127.1

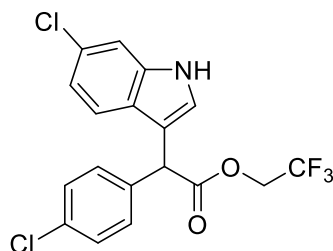
(rot B), 127.1 (rot A), 127.0 (rot B), 82.4 (rot A), 82.3 (rot B), 69.4 (rot A), 69.2 (rot B), 52.6 (rot A), 52.0 (rot B), 46.1 (rot A), 46.0 (rot B), 44.8 (rot A), 42.3 (rot B), 33.8 (rot A), 32.2 (rot A), 32.2 (rot B), 31.4 (rot B), 23.6 (rot A) and 23.6 (rot B). HRMS (ESI): $C_{17}H_{23}NO_2$ requires $[M+H]^+$, calculated 274.1807, found 274.1803.

2-[(5-Chloro-2,3-dihydro-1H-inden-1-yl)oxy]-1-(pyrrolidin-1-yl)ethenone, 4.2



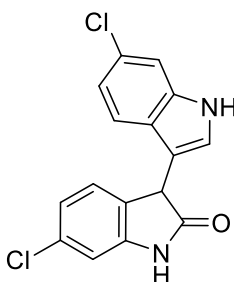
According to general procedure **E**, Rh_2piv_4 (2.2 mg, 3.6 μ mol), 2-diazo-1-(pyrrolidin-1-yl)ethenone (50 mg, 0.36 mmol) and 5-chloro-2,3-dihydro-1H-inden-1-ol (120 mg, 0.72 mmol) gave an orange oil. The crude oil was then purified by flash column chromatography eluting with 100% Et_2O to afford *ether* **4.2** as a colourless oil (51 mg, 50%), R_F 0.09 (100% Et_2O); δ_H (500 MHz, $CDCl_3$): 7.37 (1H, d, J 8.0, indenyl-oxy-7H), 7.22 (1H, broad s, indenyl-oxy-4H), 7.18 – 7.16 (1H, m, indenyl-oxy-6H), 4.99 (1H, dd, J 6.5 and 3.7, indenyl-oxy-1H), 4.14 (2H, m, ethenone-1H), 3.49 (2H, app. t, J 6.9, pyrrolidinyl-2H_a), 3.45 – 3.38 (2H, m, pyrrolidinyl-2H_b), 3.09 – 3.03 (1H, m, indenyl-oxy-3H_a), 2.81 – 2.76 (1H, m, indenyl-oxy-3H_b), 2.35 (1H, ddt, J 13.0, 8.5 and 6.4, indenyl-oxy-2H_a), 2.15 (1H, dddd, J 13.3, 8.4, 4.8 and 3.8, indenyl-oxy-2H_b), 1.95 – 1.90 (2H, m, pyrrolidinyl-3H_a) and 1.86 – 1.81 (2H, m, pyrrolidinyl-3H_b). δ_C (125 MHz, $CDCl_3$) 168.3, 146.4, 140.7, 134.5, 126.7, 126.6, 125.2, 83.0, 68.5, 46.2 (rot-A), 46.1 (rot-B), 32.5, 30.3, 26.3 (rot-A) and 24.0 (rot-B). HRMS (ESI): $C_{15}H_{18}^{35}ClNO_2$ requires $[M+Na]^+$, calculated 302.0924, found 302.0926.

2,2,2-Trifluoroethyl 2-(6-chloro-1H-indol-3-yl)-2-(4-chlorophenyl)acetate, P1



According to general procedure **E**, Rh₂tpfb₄ (3.8 mg, 3.6 μmol), 2,2,2-trifluoroethyl 2-(4-chlorophenyl)-2-diazoacetate (100 mg, 0.36 mmol) and 6-chloroindole (272 mg, 1.8 mmol) gave an orange oil. The crude oil was then purified by flash column chromatography eluting with 9:1 pentane/Et₂O to afford *indole* **P1** as a colourless oil (20 mg, 14%), *R_F* 0.19 (3:1 Pentane/Et₂O); δ_H (500 MHz, CDCl₃): 8.10 (1H, s, 6-chloroindolyl-NH), 7.30 (1H, d, *J* 1.8, 6-chloroindolyl-4H), 7.28 – 7.24 (4H, m, 4-chlorophenyl-2H₂ and -3H₂), 7.21 (1H, d, *J* 8.5, 6-chloroindolyl-2H), 7.14 (1H, dd, *J* 2.5 and 0.8, 6-chloroindolyl-7H), 6.99 (1H, dd, *J* 8.5 and 1.8, 6-chloroindolyl-5H), 5.24 (1H, s, acetate-2H) and 4.49 (2H, qq, *J* 12.7 and 8.4, trifluoroethyl-1H). δ_C (125 MHz, CDCl₃): 170.9, 136.8, 135.8, 133.9, 129.8, 129.1, 128.8, 124.9, 124.0, 121.8, 121.0, 119.9, 112.5, 111.5, 61.0 (q, *J_{C-F}* 36.7), 47.9. HRMS (ESI): C₁₈H₁₂³⁵Cl₂F₃NO₂ requires [M+H]⁺, calculated 402.0275, found 402.0270.

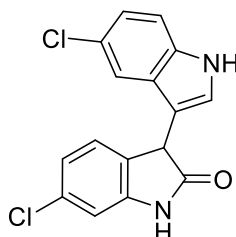
6,6'-Dichloro-1H,1'H,3H-[3,3'-biindol]-2-one, P6



According to general procedure **D**, Rh₂tpfb₄ (2.8 mg, 2.6 μmol), 6-chloro-3-diazo-1H-indol-2-one (50 mg, 0.26 mmol) and 6-chloroindole (254 mg, 1.3 mmol) gave an orange oil. The crude oil was then purified by flash column chromatography eluting with 9:1 DCM/Et₂O to afford *oxindole* **P6** as a colourless oil (44 mg, 53%), *R_F* 0.17 (9:1 DCM/Et₂O); δ_H (500 MHz, CDCl₃): 8.54 (1H, broad s, biindol-2-one-NH), 8.29 (1H, broad s, biindol-NH), 7.31 (1H, d, *J* 1.8 Hz, biindol-4H), 7.14 (1H, d, *J* 8.4, biindol-2H), 7.06 – 7.04 (2H, m, biindol-2-one-4H and -7H), 6.99 (2H, td, *J* 8.4 and 1.8, biindol-2-one-5H

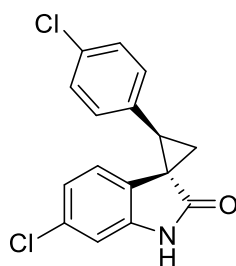
and biindol-5H), 6.94 (1H, d, J 1.8, biindol-7H) and 4.83 (1H, s, biindol-2-one-3H). δ_c (125 MHz, $CDCl_3$): 178.5, 142.4, 137.1, 134.2, 128.7, 127.8, 126.1, 124.7, 124.3, 122.9, 121.0, 120.1, 111.5, 110.7, 110.4 and 44.5. HRMS (ESI): $C_{16}H_{10}^{35}Cl_2N_2O$ requires $[M+H]^+$, calculated 317.0248, found 317.0226.

5',6-Dichloro-1H,1'H,3H-[3,3'-biindol]-2-one, 4.4



According to general procedure **E**, Rh_2pfb_4 (2.8 mg, 2.6 μ mol), 6-chloro-3-diazo-1H-indol-2-one (50 mg, 0.26 mmol) and 5-chloroindole (254 mg, 1.3 mmol) gave an orange oil. The crude oil was then purified by flash column chromatography eluting with 9:1 DCM/ Et_2O to afford *oxindole* **4.4** as a colourless oil (31.2 mg, 38%), R_F 0.17 (9:1 DCM/ Et_2O); δ_H (500 MHz, $DMSO-d_6$): 11.27 (1H, s, biindol-2-one-NH), 10.71 (1H, s, biindol-NH), 7.39 (1H, dd, J 8.6 and 0.5, biindol-2-one-4H), 7.31 (1H, d, J 2.5, biindol-4H), 7.11 – 7.04 (3H, m, biindol-2-one-5H, biindol-2H and -5H), 6.98 – 6.95 (2H, m, biindol-2-one-7H and biindol-7H) and 4.98 (1H, s, biindol-2-one-3H). δ_c (125 MHz, $DMSO-d_6$): 177.4, 144.0, 134.9, 132.1, 128.9, 127.1, 126.2, 125.9, 123.3, 121.3, 121.2, 117.6, 113.3, 109.4, 109.4 and 43.6. HRMS (ESI): $C_{16}H_{10}^{35}Cl_2N_2O$ requires $[M+H]^+$, calculated 317.0248, found 317.0237.

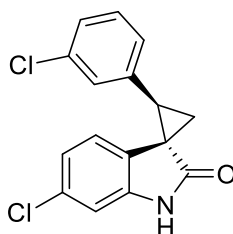
6'-Chloro-3-(4-chlorophenyl)-1'H-spiro[cyclopropane-1,3'-indol]-2'-one, P5



According to general procedure **D**, Rh_2piv_4 (1.6 mg, 2.6 μ mol), 6-chloro-3-diazo-1H-indol-2-one (50 mg, 0.26 mmol) and 4-chlorostyrene (156 μ L, 1.3 mmol) gave an orange oil. The crude oil was then purified by flash column

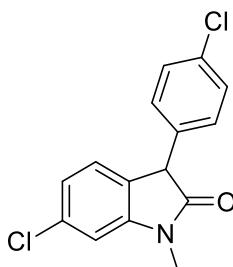
chromatography eluting with 9:1 DCM/Et₂O to afford *oxindole P5* as a colourless oil (46 mg, 58%), *R_F* 0.17 (9:1 DCM/Et₂O); δ_H (500 MHz, CDCl₃): 8.57 (1H, broad s, indolone-NH), 7.28 (2H, d, *J* 8.2, 4-chlorophenyl-3H₂), 7.11 (2H, d, *J* 8.2, 4-chlorophenyl-2H₂), 6.96 (1H, d, *J* 1.8, indol-2-one-7H), 6.69 (1H, dd, *J* 8.1 and 1.9, indol-2-one-5H) 5.84 (1H, d, *J* 8.1, indol-2-one-4H), 3.28 (1H, app. t, *J* 8.6, cyclopropane-1H), 2.23 (1H, dd, *J* 9.2 and 4.8, cyclopropane-2H_a), 1.97 (1H, dd, *J* 8.0 and 4.8, cyclopropane-2H_b). δ_C (125 MHz, CDCl₃): 178.3, 142.0, 133.7, 133.3, 132.8, 131.4, 128.9, 126.0, 121.9, 121.8, 110.5, 35.7, 33.5 and 22.9. HRMS (ESI): C₁₆H₁₁³⁵Cl₂NO requires [M+H]⁺, calculated 304.0296, found 304.0286.

6'-Chloro-3-(3-chlorophenyl)-1'H-spiro[cyclopropane-1,3'-indol]-2'-one, 4.3



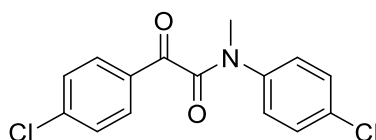
According to general procedure **E**, Rh₂piv₄ (1.6 mg, 2.6 μmol), 6-chloro-3-diazo-1H-indol-2-one (50 mg, 0.26 mmol) and 3-chlorostyrene (165 μL, 1.3 mmol) gave an orange oil. The crude oil was then purified by flash column chromatography eluting with 9:1 DCM/Et₂O to afford *oxindole 4.3* as a colourless oil (39 mg, 49%), *R_F* 0.38 (9:1 DCM/Et₂O); δ_H (500 MHz, CDCl₃): 8.42 (1H, br s, indol-2-one-NH), 7.27 – 7.22 (3H, m, 3-chlorophenyl-4H, -5H and -6H), 7.04 (1H, broad dd, *J* 7.3 and 0.6, 3-chlorophenyl-2H), 6.96 (1H, d, *J* 1.8, indol-2-one-7H), 6.69 (1H, dd, *J* 8.1 and 1.9, indol-2-one-5H), 5.87 (1H, d, *J* 8.1, indol-2-one-4H), 3.29 (1H, app. t, *J* 8.6, cyclopropane-1H), 2.23 (1H, dd, *J* 9.2 and 4.8, cyclopropane-2H_a) and 1.99 (1H, dd, *J* 8.0 and 4.8, cyclopropane-2H_b). δ_C (125 MHz, CDCl₃): 178.0, 142.0, 136.9, 134.6, 132.9, 130.0, 129.9, 129.9, 128.3, 128.1, 125.9, 121.9, 110.5, 35.8, 33.5 and 22.7. HRMS (ESI): C₁₆H₁₁³⁵Cl₂NO requires [M+H]⁺, calculated 304.0296, found 304.0285.

6-Chloro-3-(4-chlorophenyl)-1-methyl-3H-indol-2-one, P3



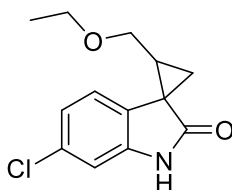
According to general procedure **D**, Rh₂piv₄ (1.9 mg, 3.0 μmol) and N,2-bis(4-chlorophenyl)-2-diazo-N-methylacetamide (100 mg, 0.3 mmol) gave an orange oil. The crude oil was then purified by flash column chromatography initially eluting with 100% DCM, which gave a mixture of products, and the oil was re-purified eluting 8:2 pentane/EtOAc to afford *oxindole* **P3** as a colourless oil (3.5 mg, 4%), *R*_F 0.15 (8:2 pentane/EtOAc); δ_H (500 MHz, CDCl₃): 7.33 – 7.31 (3H, m, 4-chlorophenyl-3H₂ and indol-2-one-4H), 7.14 – 7.11 (3H, m, 4-chlorophenyl-2H₂ and indol-2-one-5H), 6.82 (1H, d, *J* 8.3, indol-2-one-4H), 4.57 (1H, s, indol-2-one-3H) and 3.24 (3H, s, indol-2-one-NCH₃). δ_C (125 MHz, CDCl₃): 175.2, 143.2, 134.4, 134.0, 130.0, 129.9, 129.3, 128.8, 128.4, 125.6, 109.4, 51.5 and 26.8. HRMS (ESI): C₁₅H₁₁³⁵Cl₂NO requires [M+H]⁺, calculated 292.0295, found 292.0279.

N,2-Bis(4-chlorophenyl)-N-methyl-2-oxoacetamide, P4



According to general procedure **D**, Rh₂piv₄ (1.9 mg, 3.0 μmol) and N,2-bis(4-chlorophenyl)-2-diazo-N-methylacetamide (100 mg, 0.3 mmol) gave an orange oil. The crude oil was then purified by flash column chromatography initially eluting with 100% DCM, which gave a mixture of products, and the oil was re-purified eluting 8:2 pentane/EtOAc to afford *oxoacetamide* **P4** as a colourless oil (3.6 mg, 4%), *R*_F 0.38 (8:2 pentane/EtOAc); δ_H (500 MHz, CDCl₃): 7.81 (2H, d, *J* 8.4, acetamide Ar-3H), 7.44 (2H, d, *J* 8.4, acetamide Ar-2H), 7.24 (2H, d, *J* 8.5, oxo Ar-3H), 7.06 (2H, d, *J* 8.5, oxo Ar-2H) and 3.45 (3H, s, oxoacetamide-NCH₃). δ_C (125 MHz, CDCl₃): 189.3, 166.6, 141.3, 139.8, 134.3, 131.9, 130.9, 130.0, 129.5, 128.2 and 36.5. HRMS (ESI): C₁₅H₁₁³⁵Cl₂NO₂ requires [M+Na]⁺, calculated 330.0064, found 330.0059.

3-(Ethoxymethyl)-1'H-spiro[cyclopropane-1,3'-indol]-2'-one, P7



According to general procedure **D**, Rh₂piv₄ (1.3 mg, 2.1 μmol), 6-chloro-3-diazo-1H-indol-2-one (40 mg, 0.21 mmol) and 3-methoxy-1-propene (113 μL, 1.0 mmol) gave an orange oil. The crude oil was then purified by flash column chromatography eluting with 9:1 DCM/Et₂O to afford *oxindole* **P7** as a colourless oil (7.3 mg, 14%), *R_F* 0.08 (9:1 DCM/Et₂O); δ_H (500 MHz, CDCl₃): 8.31 (1H, s, indol-2-one-NH), 6.98 (1H, dd, *J* 8.0 and 1.9, indol-2-one-5H), 6.94 (1H, d, *J* 1.5, indol-2-one-7H), 6.90 (1H, d, *J* 8.0, indol-2-one-4H), 3.77 (1H, dd, *J* 11.2 and 5.5, methyl-H_a), 3.64 (1H, dd, *J* 11.2 and 7.6, methyl-H_b), 3.45 (2H, q, *J* 7.0, ethoxy-1H₂), 2.24 (1H, dtd, *J* 13.1, 7.6 and 5.5, cyclopropane-3H), 1.95 (1H, dd, *J* 9.4 and 4.5, cyclopropane-2H_a), 1.57 (1H, dd, *J* 7.8 and 4.5, cyclopropane-2H_b) and 1.16 (3H, t, *J* 7.0, ethoxy-2H₃). δ_C (125 MHz, CDCl₃): 178.3, 142.3, 132.7, 126.9, 122.2, 121.8, 110.6, 67.4, 66.3, 32.0, 31.6, 22.0 and 15.2. HRMS (ESI): C₁₃H₁₄³⁵CINO₂ requires [M+H]⁺, calculated 252.0791, found 252.0777.

5.3.4 Implementation of High-Throughput Chemistry for Activity-Directed Synthesis Reaction Arrays

Activity-Directed Synthesis reactions were carried out in 0.75 mL shell vials (Chemglass CV-2100-0830) equipped with a teflon-coated stir bar (Biotage 0.2-0.5 mL magnetic stir bar #355545) and sealed using either a Freeslate 96-well reaction block or a Sigma-Aldrich Kitalysis 24-well reaction block (Z742107 Aldrich). Prior to the assembly of each reaction array the following stock solutions were made: diazo reaction solvent (1.25 M); catalyst in THF (25 mM); and co-substrate in CH₂Cl₂ (6.25 M). Each reaction vial was charged with catalyst stock (8 μL) and the solvent allowed to evaporate to dryness, then CH₂Cl₂ (84 μL) was added and the reaction block placed on a magnetic stirring plate. Each reaction vial was then sequentially charged with co-substrate stock (8 μL) and diazo stock (8 μL), then the plate sealed using a Teflon film and stirred. After 24 hours Quadrapure TU resin (30 mg) was added to each vial and left overnight to scavenge the catalyst. The solvent was then evaporated under a stream of nitrogen gas and the crude material dissolved in molecular biology grade DMSO (200 μL) to create a 50 mM total product concentration biological screening master stock that was passed through a 96-well filter plate (Agilent Technologies: #200933-100) and stored at -20 °C.

Example 96-well reaction plate layout:

		1	2	3	4	5	6	7	8	9	10	11	12	Catalyst
D1	A	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	D1 Control	Rh ₂ piv ₄ Control	PIV
D2	B	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	D2 Control	BLANK	PIV
D3	C	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	D3 Control	BLANK	PIV
D4	D	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	D4 Control	BLANK	PIV
D1	E	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	D1 Control	Rh ₂ pf ₆ Control	PFB
D2	F	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	D2 Control	BLANK	PFB
D3	G	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	D3 Control	BLANK	PFB
D4	H	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	D4 Control	BLANK	PFB

5.3.4.1 LC/MS Analysis of Round One Reaction Mixtures

All 154 reactions from the round 1 reaction array were analysed by LC-MS to investigate how many combinations had produced a desired product (Table 5.7). All samples were diluted to 1 mg/mL concentrations from the original 50 mM DMSO master stock, with respect to the initial diazo starting concentration. Reaction wells containing diazo and substrate were analysed for intermolecular product(s) and blank control wells were analysed for intramolecular product(s). Dark green squares indicate clear m/z for the desired product(s) and a clear corresponding UV peak(s). Light green squares indicate m/z for the desired product(s) and either weak or no corresponding UV peak(s). Blank squares indicate that no m/z was observed for the desired product(s).

Table 5.7. LC-MS data from the round 1 reaction array.

Diazo/ Substrate	Catalyst	Formula	Adduct	Expected	Found	Peak Intensity	UV peak?
D1S1	Rh ₂ piv ₄	C19H16Cl ₂ N ₂ O ₂	-	374.06	-	-	N
D1S2	Rh ₂ piv ₄	C17H15Cl ₂ NO ₃	-	351.04	-	-	N
D1S3	Rh ₂ piv ₄	C25H22Cl ₃ NO ₂	-	473.07	-	-	N
D1S4	Rh ₂ piv ₄	C18H21ClN ₂ O ₄	-	364.12	-	-	N
D1S5	Rh ₂ piv ₄	C21H25ClN ₂ O ₃	H	389.16	389.16	5x10 ⁶	N
D1S6	Rh ₂ piv ₄	C20H18ClNO ₃	-	355.10	-	-	N
D1S7	Rh ₂ piv ₄	C22H23ClN ₂ O ₂	H	383.15	383.15	1x10 ⁷	N
D1S8	Rh ₂ piv ₄	C20H19Cl ₂ NO ₃	H	392.07	392.08	5x10 ⁶	N
D1S9	Rh ₂ piv ₄	C20H17Cl ₂ NO ₂	H	374.06	374.07	5x10 ⁶	N
D1S10	Rh ₂ piv ₄	C15H18ClNO ₄ S	H	344.07	344.07	3x10 ⁶	N
D1blank	Rh ₂ piv ₄	C11H10ClNO ₂	2M + 2H	448.10	448.64	7.5x10 ⁷	Y
D1S1	Rh ₂ pf ₄	C19H16Cl ₂ N ₂ O ₂	H	375.06	375.07	1x10 ⁷	N
D1S2	Rh ₂ pf ₄	C17H15Cl ₂ NO ₃	H	352.04	352.24	4x10 ⁷	Y
D1S3	Rh ₂ pf ₄	C25H22Cl ₃ NO ₂	Na	496.06	498.1	4x10 ⁷	Y
D1S4	Rh ₂ pf ₄	C18H21ClN ₂ O ₄	H	365.12	365.03	1x10 ⁷	Y
D1S5	Rh ₂ pf ₄	C21H25ClN ₂ O ₃	-	388.16	-	-	N
D1S6	Rh ₂ pf ₄	C20H18ClNO ₃	-	355.10	-	-	N
D1S7	Rh ₂ pf ₄	C22H23ClN ₂ O ₂	-	382.15	-	-	N
D1S8	Rh ₂ pf ₄	C20H19Cl ₂ NO ₃	Na	414.06	413.97	3x10 ⁶	Y
D1S9	Rh ₂ pf ₄	C20H17Cl ₂ NO ₂	H	374.06	374.07	2x10 ⁷	N
D1S10	Rh ₂ pf ₄	C15H18ClNO ₄ S	-	343.07	-	-	N
D1blank	Rh ₂ pf ₄	C11H10ClNO ₂	-	223.04	-	-	N
D2S1	Rh ₂ piv ₄	C18H12Cl ₂ F ₃ NO ₂	H	402.02	401.98	3x10 ⁶	N
D2S2	Rh ₂ piv ₄	C16H11Cl ₂ F ₃ O ₃	H	379.00	379.01	7.5x10 ⁶	N
D2S3	Rh ₂ piv ₄	C24H18Cl ₃ F ₃ O ₂	-H	499.02	498.86	7.5x10 ⁷	Y
D2S4	Rh ₂ piv ₄	C17H17ClF ₃ NO ₄	H	392.08	392.02	3x10 ⁷	Y
D2S5	Rh ₂ piv ₄	C20H21ClF ₃ NO ₃	-H	414.11	414.11	6x10 ⁷	Y

D2S6	Rh ₂ piv ₄	C19H14ClF3O3	Na	405.04	404.80	1x10 ⁷	Y
D2S7	Rh ₂ piv ₄	C21H19ClF3NO2	H	409.11	410.02	7.5x10 ⁸	Y
D2S8	Rh ₂ piv ₄	C19H15Cl2F3O3	Na	441.02	440.97	6x10 ⁶	Y
D2S9	Rh ₂ piv ₄	C19H13Cl2F3O2	H	401.02	400.93	2x10 ⁶	N
D2S10	Rh ₂ piv ₄	C14H14ClF3O4S	-H	369.01	368.85	2x10 ⁷	Y
D2blank	Rh ₂ piv ₄	C20H12Cl2F6O4	-H	498.99	498.99	1x10 ⁶	N
D2S1	Rh ₂ pf ₄	C18H12Cl2F3NO2	-H	400.01	399.92	1x10 ⁸	Y
D2S2	Rh ₂ pf ₄	C16H11Cl2F3O3	Na	400.99	400.89	1.5x10 ⁶	N
D2S3	Rh ₂ pf ₄	C24H18Cl3F3O2	-	500.03	-	-	N
D2S4	Rh ₂ pf ₄	C17H17ClF3NO4	H	392.08	392.09	5x10 ⁷	Y
D2S5	Rh ₂ pf ₄	C20H21ClF3NO3	-	415.12	-	-	N
D2S6	Rh ₂ pf ₄	C19H14ClF3O3	-	382.06	-	-	N
D2S7	Rh ₂ pf ₄	C21H19ClF3NO2	H	410.11	410.02	7.5x10 ⁸	Y
D2S8	Rh ₂ pf ₄	C19H15Cl2F3O3	Na	441.02	440.96	2x10 ⁶	Y
D2S9	Rh ₂ pf ₄	C19H13Cl2F3O2	H	401.02	400.92	1x10 ⁷	Y
D2S10	Rh ₂ pf ₄	C14H14ClF3O4S	-	370.03	-	-	N
D2blank	Rh ₂ pf ₄	C20H12Cl2F6O4	-	500.00	-	-	N
D3S1	Rh ₂ piv ₄	C23H17Cl3N2O	H	445.04	444.93	2.5x10 ⁷	Y
D3S2	Rh ₂ piv ₄	C21H16Cl3NO2	H	422.02	421.90	2x10 ⁸	Y
D3S3	Rh ₂ piv ₄	C29H23Cl4NO	-	541.05	-	-	N
D3S4	Rh ₂ piv ₄	C22H22Cl2N2O3	H	433.10	432.98	2x10 ⁸	Y
D3S5	Rh ₂ piv ₄	C25H26Cl2N2O2	K	495.10	495.07	1.25x10 ⁷	N
D3S6	Rh ₂ piv ₄	C24H19Cl2NO2	-	423.08	-	-	N
D3S7	Rh ₂ piv ₄	C26H24Cl2N2O	H	451.13	451.13	2x10 ⁷	N
D3S8	Rh ₂ piv ₄	C24H20Cl3NO2	-	459.06	-	-	N
D3S9	Rh ₂ piv ₄	C24H18Cl3NO	-	441.05	-	-	N
D3S10	Rh ₂ piv ₄	C19H19Cl2NO3S	-H	410.05	410.04	2x10 ⁵	N
D3blank	Rh ₂ piv ₄	C15H11Cl2NO	-H	290.01	289.71	5x10 ⁶	Y
D3S1	Rh ₂ pf ₄	C23H17Cl3N2O	H	445.04	444.95	7.5x10 ⁷	Y
D3S2	Rh ₂ pf ₄	C21H16Cl3NO2	-	419.02	-	-	N
D3S3	Rh ₂ pf ₄	C29H23Cl4NO	-	541.05	-	-	N
D3S4	Rh ₂ pf ₄	C22H22Cl2N2O3	H	433.10	432.97	4x10 ⁸	Y
D3S5	Rh ₂ pf ₄	C25H26Cl2N2O2	-	456.14	-	-	N
D3S6	Rh ₂ pf ₄	C24H19Cl2NO2	-	423.08	-	-	N
D3S7	Rh ₂ pf ₄	C26H24Cl2N2O	-	450.13	-	-	N
D3S8	Rh ₂ pf ₄	C24H20Cl3NO2	-	459.06	-	-	N
D3S9	Rh ₂ pf ₄	C24H18Cl3NO	-	441.05	-	-	N
D3S10	Rh ₂ pf ₄	C19H19Cl2NO3S	-	411.05	-	-	N
D3blank	Rh ₂ pf ₄	C15H11Cl2NO	H	292.02	292.02	4x10 ⁷	Y
D3S1	Rh ₂ piv ₄	C20H19ClN2O	H	339.12	339.02	1x10 ⁸	Y
D3S2	Rh ₂ piv ₄	C18H18ClNO2	H	316.10	315.96	1.5x10 ⁸	Y
D3S3	Rh ₂ piv ₄	C26H25Cl2NO	-	437.13	-	-	N
D4S4	Rh ₂ piv ₄	C19H24N2O3	NH ₄	347.18	347.08	1x10 ⁷	Y
D4S5	Rh ₂ piv ₄	C22H28N2O2	-	352.2151	-	-	N
D4S6	Rh ₂ piv ₄	C21H21NO2	H	320.16	320.01	7x10 ⁷	Y

D4S7	Rh ₂ piv ₄	C23H26N2O	-	346.21	-	-	N
D4S8	Rh ₂ piv ₄	C21H22ClNO2	2M + Na	733.26	733.19	2x10 ⁸	Y
D4S9	Rh ₂ piv ₄	C21H20ClNO	H	338.12	338.03	1.7x10 ⁸	Y
D4S10	Rh ₂ piv ₄	C16H21NO3S	H	308.12	307.97	2x10 ⁷	Y
D4blank	Rh ₂ piv ₄	C24H26N2O2	H	375.20	375.1	3x10 ⁸	Y
D4S1	Rh ₂ pf ₄	C20H19ClNO2	H	339.12	339.01	1x10 ⁸	Y
D4S2	Rh ₂ pf ₄	C18H18ClNO2	-	315.10	-	-	N
D4S3	Rh ₂ pf ₄	C26H25Cl2NO	Na	460.12	460.23	1x10 ⁷	Y
D4S4	Rh ₂ pf ₄	C19H24N2O3	H	329.18	329.19	3x10 ⁷	N
D4S5	Rh ₂ pf ₄	C22H28N2O2	-H	351.22	351.21	5x10 ⁵	N
D4S6	Rh ₂ pf ₄	C21H21NO2	H	320.16	320.05	8x10 ⁶	Y
D4S7	Rh ₂ pf ₄	C23H26N2O	-	346.20	-	-	N
D4S8	Rh ₂ pf ₄	C21H22ClNO2	2M + Na	733.26	733.21	2x10 ⁷	N
D4S9	Rh ₂ pf ₄	C21H20ClNO	H	338.12	338.04	4x10 ⁷	Y
D4S10	Rh ₂ pf ₄	C16H21NO3S	H	308.12	308.13	1x10 ⁷	N
D4blank	Rh ₂ pf ₄	C24H26N2O2	H	375.20	375.08	3x10 ⁸	Y
D5S1	Rh ₂ piv ₄	C16H19ClNO2	H	307.11	306.94	1x10 ⁸	Y
D5S2	Rh ₂ piv ₄	C14H18ClNO3	H	284.10	283.88	3x10 ⁷	Y
D5S3	Rh ₂ piv ₄	C22H25Cl2NO2	H	406.13	406.13	5x10 ⁶	N
D5S4	Rh ₂ piv ₄	C15H24N2O4	H	297.17	297.18	4x10 ⁶	N
D5S5	Rh ₂ piv ₄	C18H28N2O3	NH ₄	338.24	338.23	1x10 ⁷	N
D5S6	Rh ₂ piv ₄	C17H21NO3	H	288.15	287.94	1x10 ⁷	Y
D5S7	Rh ₂ piv ₄	C19H26N2O2	-	314.30	-	-	N
D5S8	Rh ₂ piv ₄	C17H22ClNO3	2M + Na	669.25	669.24	1x10 ⁸	Y
D5S9	Rh ₂ piv ₄	C17H20ClNO2	H	306.12	305.94	3x10 ⁷	Y
D5S10	Rh ₂ piv ₄	C12H21NO4S	-	275.12	-	-	N
D5blank	Rh ₂ piv ₄	C8H13NO2	-	155.09	-	-	N
D5S1	Rh ₂ pf ₄	C16H19ClNO2	H	307.11	306.92	3x10 ⁷	Y
D5S2	Rh ₂ pf ₄	C14H18ClNO3	H	284.10	283.86	2x10 ⁷	Y
D5S3	Rh ₂ pf ₄	C22H25Cl2NO2	Na	428.12	428.22	2x10 ⁶	Y
D5S4	Rh ₂ pf ₄	C15H24N2O4	H	297.17	297.18	5x10 ⁶	N
D5S5	Rh ₂ pf ₄	C18H28N2O3	-	320.21	-	-	N
D5S6	Rh ₂ pf ₄	C17H21NO3	H	288.15	288.16	5x10 ⁶	N
D5S7	Rh ₂ pf ₄	C19H26N2O2	H	315.20	315.21	1x10 ⁷	N
D5S8	Rh ₂ pf ₄	C17H22ClNO3	2M + Na	669.25	669.24	5x10 ⁷	Y
D5S9	Rh ₂ pf ₄	C17H20ClNO2	H	306.12	305.93	2.5x10 ⁶	Y
D5S10	Rh ₂ pf ₄	C12H21NO4S	-	275.12	-	-	N
D5blank	Rh ₂ pf ₄	C8H13NO2	2M + H	311.20	311.04	1.5x10 ⁷	N
D6S1	Rh ₂ piv ₄	C17H14ClNO2	-H	298.07	298.06	1x10 ⁶	N
D6S2	Rh ₂ piv ₄	C15H13ClO3	H	277.06	277.06	3x10 ⁶	N
D6S3	Rh ₂ piv ₄	C23H20Cl2O2	-	398.08	-	-	N
D6S4	Rh ₂ piv ₄	C16H19NO4	H	290.13	290.14	5x10 ⁶	N
D6S5	Rh ₂ piv ₄	C19H24NO3	K	353.14	353.05	2x10 ⁷	Y
D6S6	Rh ₂ piv ₄	C18H16O3	H	281.11	281.12	7.5x10 ⁶	N
D6S7	Rh ₂ piv ₄	C20H22NO2	H	309.17	309.17	1.5x10 ⁸	Y

D6S8	Rh ₂ piv ₄	C18H17ClO ₃	H	317.09	317.06	3x10 ⁷	Y
D6S9	Rh ₂ piv ₄	C18H15ClO ₂	H	299.08	299.08	7.5x10 ⁶	N
D6S10	Rh ₂ piv ₄	C13H16O ₄ S	H	269.08	269.08	3x10 ⁷	N
D6blank	Rh ₂ piv ₄	C18H16O ₄	H	297.11	297.11	7.5x10 ⁷	Y
D6S1	Rh ₂ pf ₄	C17H14ClNO ₂	H	300.07	299.94	1x10 ⁷	Y
D6S2	Rh ₂ pf ₄	C15H13ClO ₃	-	276.06	-	-	N
D6S3	Rh ₂ pf ₄	C23H20Cl ₂ O ₂	-	398.08	-	-	N
D6S4	Rh ₂ pf ₄	C16H19NO ₄	H	290.13	290.14	5x10 ⁶	N
D6S5	Rh ₂ pf ₄	C19H24NO ₃	K	353.14	353.11	1x10 ⁷	N
D6S6	Rh ₂ pf ₄	C18H16O ₃	-	280.11	-	-	N
D6S7	Rh ₂ pf ₄	C20H22NO ₂	H	309.17	309.17	1.5x10 ⁸	Y
D6S8	Rh ₂ pf ₄	C18H17ClO ₃	2M + Na	655.16	655.05	6x10 ⁷	Y
D6S9	Rh ₂ pf ₄	C18H15ClO ₂	H	299.08	299.08	4x10 ⁷	N
D6S10	Rh ₂ pf ₄	C13H16O ₄ S	H	269.08	269.00	3x10 ⁷	Y
D6blank	Rh ₂ pf ₄	C18H16O ₄	-	296.10	-	-	N
D7S1	Rh ₂ piv ₄	C14H12ClN ₃ O ₂	H	290.06	289.93	2x10 ⁷	Y
D7S2	Rh ₂ piv ₄	C12H11ClN ₂ O ₃	H	267.05	266.89	6x10 ⁶	Y
D7S3	Rh ₂ piv ₄	C20H18Cl ₂ N ₂ O ₂	H	389.08	389.08	6x10 ⁶	Y
D7S4	Rh ₂ piv ₄	C13H17N ₃ O ₄	Na	302.11	302.01	2x10 ⁶	Y
D7S5	Rh ₂ piv ₄	C16H21N ₃ O ₃	-H	302.16	302.01	1x10 ⁶	Y
D7S6	Rh ₂ piv ₄	C15H14N ₂ O ₃	H	271.10	270.93	3x10 ⁷	Y
D7S7	Rh ₂ piv ₄	C17H20N ₃ O ₂	NH ₄	316.19	316.10	2x10 ⁶	N
D7S8	Rh ₂ piv ₄	C15H15ClN ₂ O ₃	H	307.08	306.96	5x10 ⁷	Y
D7S9	Rh ₂ piv ₄	C15H13ClN ₂ O ₂	H	289.07	288.93	4x10 ⁷	Y
D7S10	Rh ₂ piv ₄	C10H14N ₂ O ₄ S	H	259.07	258.87	6x10 ⁶	Y
D7blank	Rh ₂ piv ₄	C12H12N ₄ O ₄	H	276.09	276.95	8x10 ⁶	Y
D7S1	Rh ₂ pf ₄	C14H12ClN ₃ O ₂	H	290.06	289.92	1.5x10 ⁷	Y
D7S2	Rh ₂ pf ₄	C12H11ClN ₂ O ₃	H	267.05	266.89	1x10 ⁷	Y
D7S3	Rh ₂ pf ₄	C20H18Cl ₂ N ₂ O ₂	H	389.07	389.06	1.5x10 ⁷	Y
D7S4	Rh ₂ pf ₄	C13H17N ₃ O ₄	-H	278.12	278.11	4x10 ⁶	N
D7S5	Rh ₂ pf ₄	C16H21N ₃ O ₃	H	304.16	304.17	4x10 ⁶	N
D7S6	Rh ₂ pf ₄	C15H14N ₂ O ₃	H	271.10	270.91	1x10 ⁷	Y
D7S7	Rh ₂ pf ₄	C17H19N ₃ O ₂	H	298.15	298.06	1.5x10 ⁷	Y
D7S8	Rh ₂ pf ₄	C15H15ClN ₂ O ₃	2M + Na	635.14	635.04	1x10 ⁷	Y
D7S9	Rh ₂ pf ₄	C15H13ClN ₂ O ₂	H	289.07	288.92	1.5x10 ⁷	Y
D7S10	Rh ₂ pf ₄	C10H14N ₂ O ₄ S	H	259.07	258.91	3x10 ⁶	Y
D7blank	Rh ₂ pf ₄	C12H12N ₄ O ₄	H	277.09	276.95	7x10 ⁷	Y

5.3.5 Fluorescence Anisotropy Assay for the Inhibition of the p53/MDM2 Protein-Protein Interaction

The fluorescein-labelled p53¹⁵⁻³¹ Flu transactivation domain peptide (Ac-SQETFSDLWKLLPENNVCF(Flu)-NH₂) was purchased from Peptide Synthetics. The assay was carried out using Perkin-Elmer 384-well Optiplate assay plates (6007270). Fluorescence anisotropy assays were performed in a buffer containing 40 mM phosphate pH 7.5, 200 mM NaCl and 0.02 mg/mL bovine serum albumin (PBSA).

Results were collected using a Perkin-Elmer Envision 2103 Multilabel Reader using a 431 nm mirror, 480(104) nm excitation filter, and 535(208) and 535(209) nm emission filters after 2.5 or 24 hours of incubation at room temperature. Test well anisotropy values were then calculated using the blank corrected S and P channel values using the following formula:

$$\text{Eq. 5.1:} \quad \text{Intensity} = (2 \times P_{\text{corrected}} \times G \text{ factor}) + S_{\text{corrected}}$$

$$\text{Eq. 5.2:} \quad \text{Anisotropy} = \frac{S_{\text{corrected}} - G \times P_{\text{corrected}}}{\text{Intensity}}$$

The fraction of bound tracer was calculated using the following formula:

$$\text{Eq. 5.3:} \quad \text{Fraction Bound} = \frac{(r - r_{\text{min}})}{(\lambda(r - r_{\text{max}}) - (r - r_{\text{min}}))}$$

Where λ is the intensity of bound/unbound tracer ($\lambda = I_{\text{bound}}/I_{\text{unbound}}$) and r is anisotropy.

5.3.5.1 Binding of p53¹⁵⁻³¹ Flu to hDM2¹⁷⁻¹²⁵

A serial dilution of hDM2¹⁷⁻¹²⁵ (0.0006 μ M to 20.75 μ M, final concentration) was added to a fixed concentration of p53¹⁵⁻³¹ Flu (54.5 nM) and PBSA buffer (20 μ L), to give a 60 μ L total volume per assay well. Each dilution was performed in triplicate and the measured intensity of each well calculated using equation 5.1, then anisotropy was calculated using equation 5.2. Fraction bound of the p53¹⁵⁻³¹ Flu tracer could also be determined using equation 5.3 (Figure 5.5).

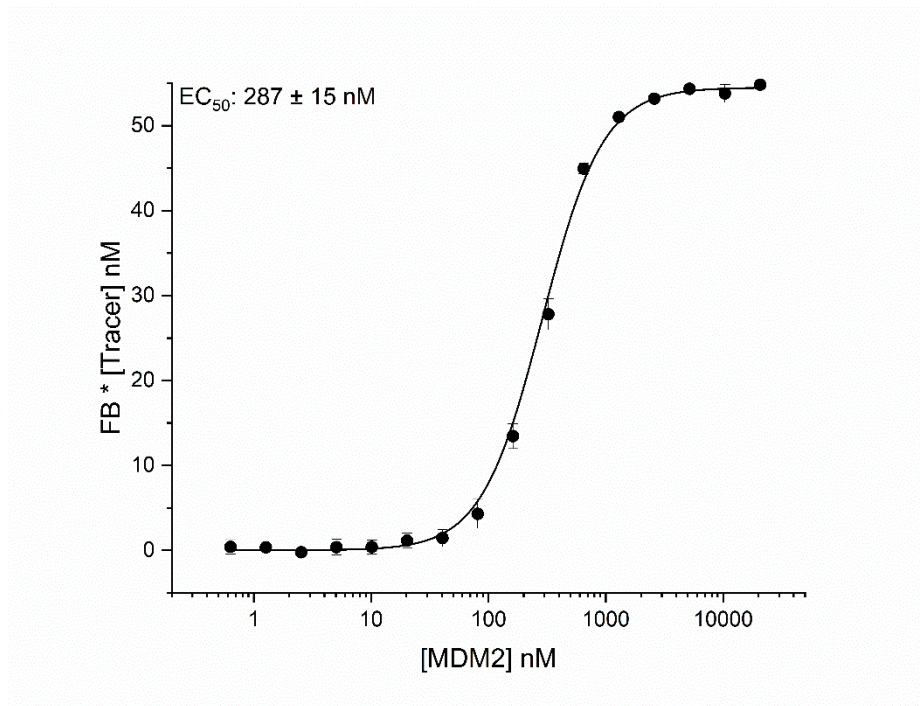
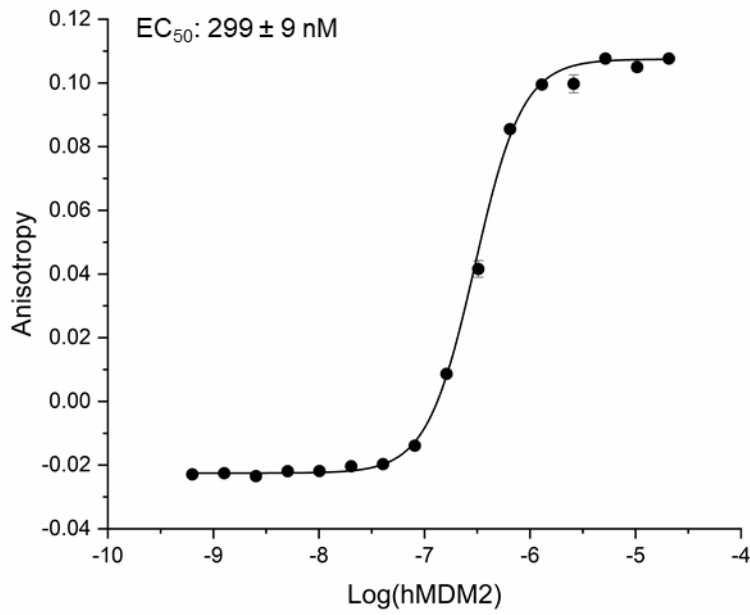


Figure 5.5. Fluorescence anisotropy titration of human-MDM2 (0.0006 μ M to 20.75 μ M) into fixed concentration p53-IAF tracer (54.5 nM) in aqueous phosphate buffer (pH 7.5, 40 mM phosphate, 200 mM NaCl and 0.02 mg/mL Bovine Serum Albumin).

5.3.5.2 Inhibition of the p53¹⁵⁻³¹ Flu/*hDM2*¹⁷⁻¹²⁵ protein-protein interaction with Nutlin-3a

Nutlin-3a was serially diluted in DMSO and then diluted 33-fold in PBSA to give effective concentrations between 73 μ M and 0.4 nM in 3% DMSO/PBSA. Each serial dilution was repeated in triplicate. An aliquot of each point (20 μ L) was then added to a 384-well assay plate, followed by *hDM2* (150 nM) and p53¹⁵⁻³¹ Flu (25 nM), to give final concentrations of Nutlin-3a between 24 μ M and 0.08 nM (Figure S2).

EC₅₀ values were determined and curves were fit in Origin Pro 2019b using a non-linear curve fitting with the dose response fitting procedure and Levenberg Marquardt iteration algorithm.

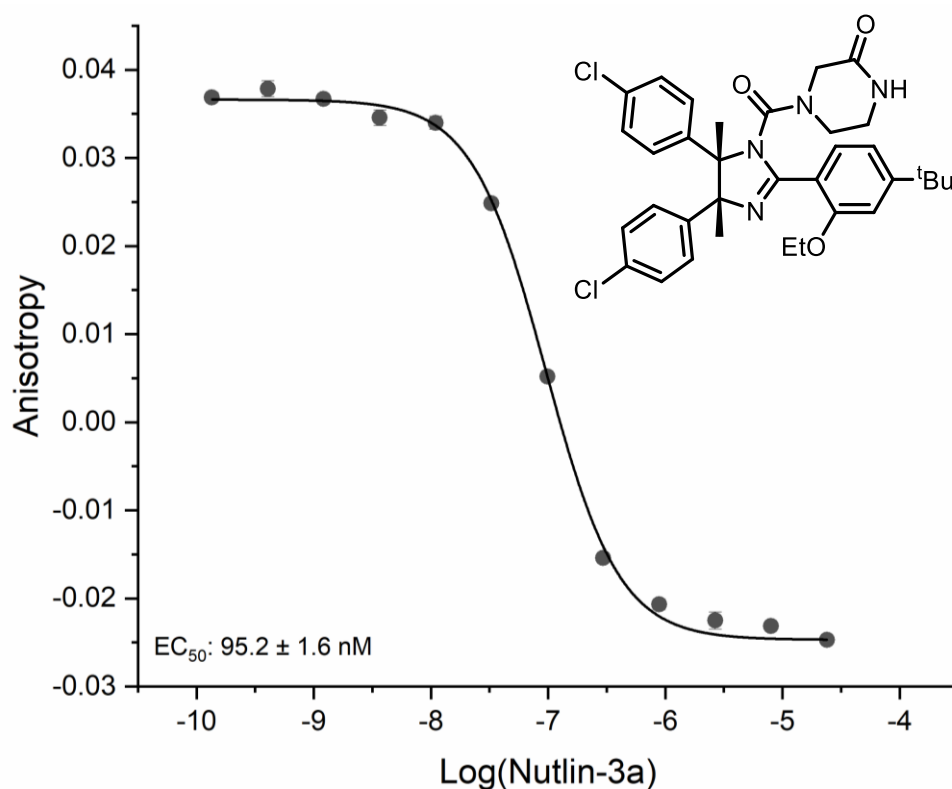


Figure 5.6. Dose response of Nutlin-3a (positive control) in our p53/MDM2 fluorescence anisotropy assay in pH 7.5 aqueous phosphate buffer (40 mM phosphate, 200 mM NaCl and 0.02 mg/mL Bovine Serum Albumin). Observed EC₅₀: 95.2 ± 1.6 nM, reported EC₅₀: 90 nM.¹⁵⁷

5.3.6 Procedure for the Screening of Reaction Mixtures

An aliquot from the master stock of each reaction mixture (2 μL , 50 mM) was diluted into 48 μL DMSO to create a 2 mM total product concentration intermediate screening stock that was used for all subsequent reaction mixture screening. An aliquot of each 2 mM reaction mixture stock (4.9 μL) was then diluted into 155.1 μL PBSA buffer (pH 7.5, 40 mM phosphate, 200 mM NaCl and 0.02 mg/mL Bovine Serum Albumin, PBSA) to create a 60 μM 3% DMSO screening stock. 20 μL of each screening stock was then added to its corresponding well in a 384-well PerkinElmer Opti-plate (see example plate layouts above). Each test well was then charged sequentially with 20 μL 450 nM *hDM2* in pH 7.5 PBSA buffer and 20 μL 75 nM p53-tracer in pH 7.5 PBSA buffer. Each blank well was then charged with 20 μL 450 nM *hDM2* in pH 7.5 PBSA buffer and 20 μL pH 7.5 PBSA buffer. The total volume of each well was 60 μL and the final concentrations of each reagent were:

- Reaction mixture: 20 μM (Total Product Concentration)
- *hDM2*: 150 nM
- p53-tracer: 25 nM

Results were collected using a Perkin-Elmer Envision 2103 Multilabel Reader using a 431 nm mirror, 480 nm excitation filter and 535 nm emission filter after 2.5 or 24 hours of incubation at room temperature. Test well anisotropy values were then calculated using the blank corrected S and P channel values using the following equations 5.1 and 5.2.

Percentage inhibition values were then calculated using Nutlin-3a (10 μM) as the positive control reference and a 1% DMSO blank well containing 150 nM *hDM2* and 25 nM p53-tracer as the negative control reference.

$$\begin{aligned} & \% \text{ Inhibition relative to } 10 \mu\text{M Nutlin} - 3a \\ & = \frac{\text{DMSO Control Anisotropy} - \text{Sample Anisotropy}}{\text{Negative Control Anisotropy} - \text{Positive Control Anisotropy}} \times 100 \end{aligned}$$

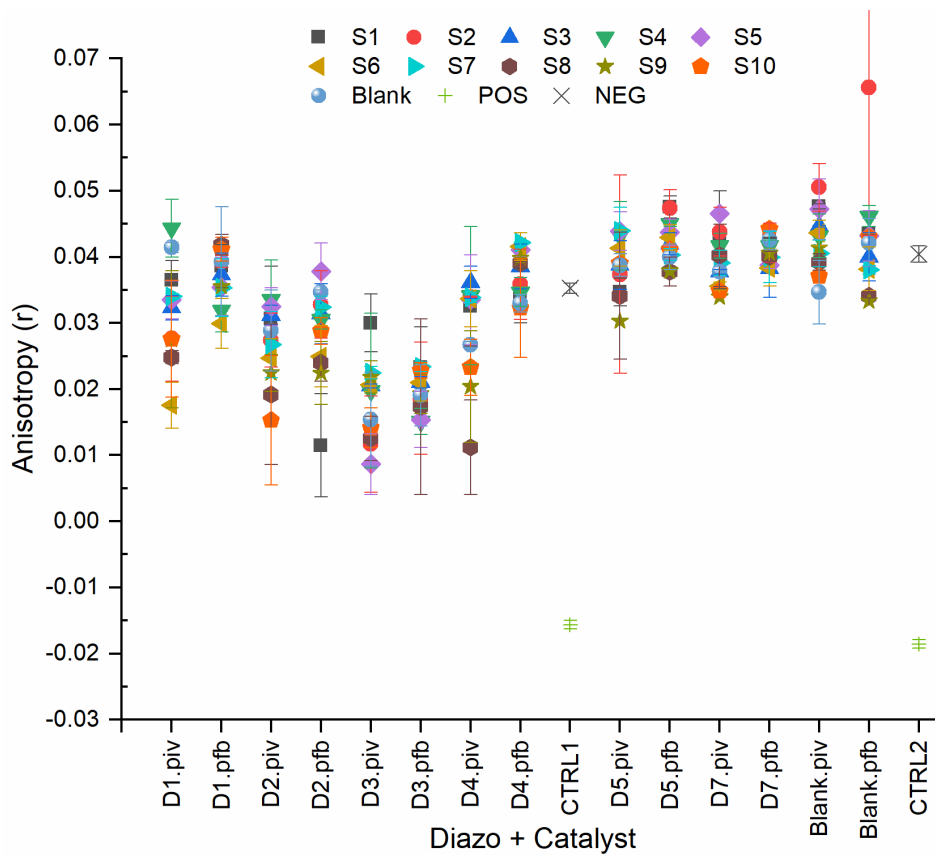
Total product concentration was used to standardise the effective screening concentrations for the high-throughput screening of reaction mixtures and is defined as the concentration of the limiting reagent in each well before the reaction took place (in all cases this is the diazo reagent).

Example 384-well biological assay plate layout:

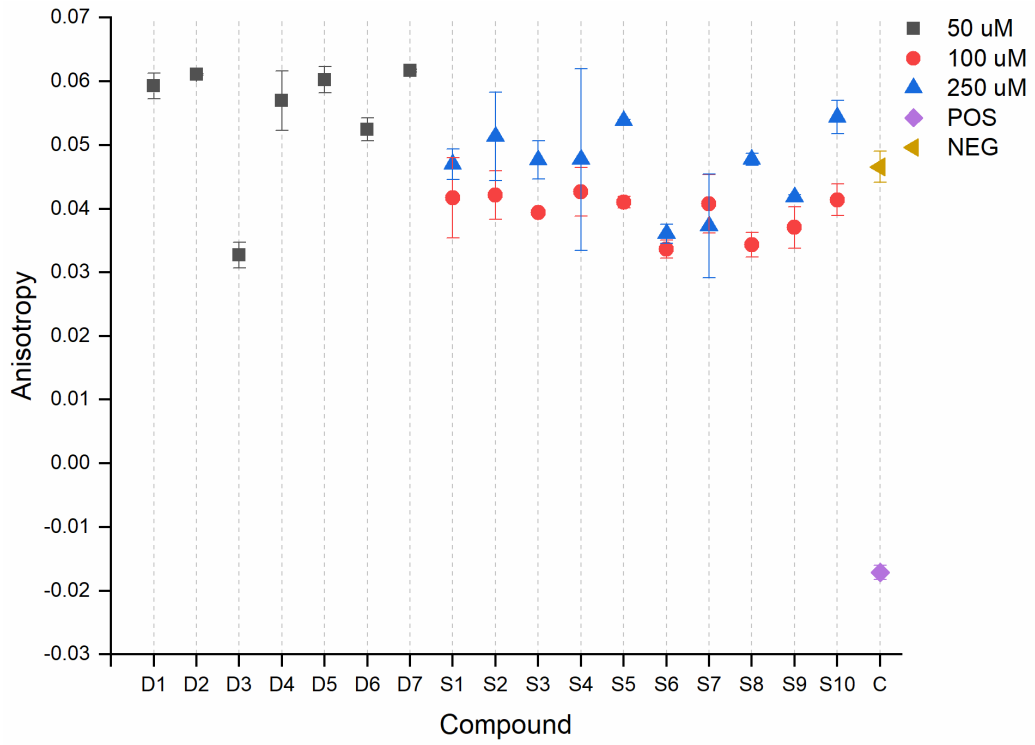
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	A1	A1	A2	A2	A3	A3	A4	A4	A5	A5	A6	A6	A7	A7	A8	A8	A9	A9	A10	A10	A11	A11	PIV	PIV
B	A1B	A1B	A2B	A2B	A3B	A3B	A4B	A4B	A5B	A5B	A6B	A6B	A7B	A7B	A8B	A8B	A9B	A9B	A10B	A10B	A11B	A11B	PIVB	PIVB
C	B1	B1	B2	B2	B3	B3	B4	B4	B5	B5	B6	B6	B7	B7	B8	B8	B9	B9	B10	B10	B11	B11	neg	neg
D	B1B	B1B	B2B	B2B	B3B	B3B	B4B	B4B	B5B	B5B	B6B	B6B	B7B	B7B	B8B	B8B	B9B	B9B	B10B	B10B	B11B	B11B	B	B
E	C1	C1	C2	C2	C3	C3	C4	C4	C5	C5	C6	C6	C7	C7	C8	C8	C9	C9	C10	C10	C11	C11	neg	neg
F	C1B	C1B	C2B	C2B	C3B	C3B	C4B	C4B	C5B	C5B	C6B	C6B	C7B	C7B	C8B	C8B	C9B	C9B	C10B	C10B	C11B	C11B	B	B
G	D1	D1	D2	D2	D3	D3	D4	D4	D5	D5	D6	D6	D7	D7	D8	D8	D9	D9	D10	D10	D11	D11	neg	neg
H	D1B	D1B	D2B	D2B	D3B	D3B	D4B	D4B	D5B	D5B	D6B	D6B	D7B	D7B	D8B	D8B	D9B	D9B	D10B	D10B	D11B	D11B	B	B
I	E1	E1	E2	E2	E3	E3	E4	E4	E5	E5	E6	E6	E7	E7	E8	E8	E9	E9	E10	E10	E11	E11	PFB	PFB
J	E1B	E1B	E2B	E2B	E3B	E3B	E4B	E4B	E5B	E5B	E6B	E6B	E7B	E7B	E8B	E8B	E9B	E9B	E10B	E10B	E11B	E11B	PFB B	PFB B
K	F1	F1	F2	F2	F3	F3	F4	F4	F5	F5	F6	F6	F7	F7	F8	F8	F9	F9	F10	F10	F11	F11	pos	pos
L	F1B	F1B	F2B	F2B	F3B	F3B	F4B	F4B	F5B	F5B	F6B	F6B	F7B	F7B	F8B	F8B	F9B	F9B	F10B	F10B	F11B	F11B	pos B	pos B
M	G1	G1	G2	G2	G3	G3	G4	G4	G5	G5	G6	G6	G7	G7	G8	G8	G9	G9	G10	G10	G11	G11	pos	pos
N	G1B	G1B	G2B	G2B	G3B	G3B	G4B	G4B	G5B	G5B	G6B	G6B	G7B	G7B	G8B	G8B	G9B	G9B	G10B	G10B	G11B	G11B	pos B	pos B
O	H1	H1	H2	H2	H3	H3	H4	H4	H5	H5	H6	H6	H7	H7	H8	H8	H9	H9	H10	H10	H11	H11	pos	pos
P	H1B	H1B	H2B	H2B	H3B	H3B	H4B	H4B	H5B	H5B	H6B	H6B	H7B	H7B	H8B	H8B	H9B	H9B	H10B	H10B	H11B	H11B	pos B	pos B

5.3.7 Fluorescence Anisotropy Data for Reaction Mixtures from Rounds One and Two

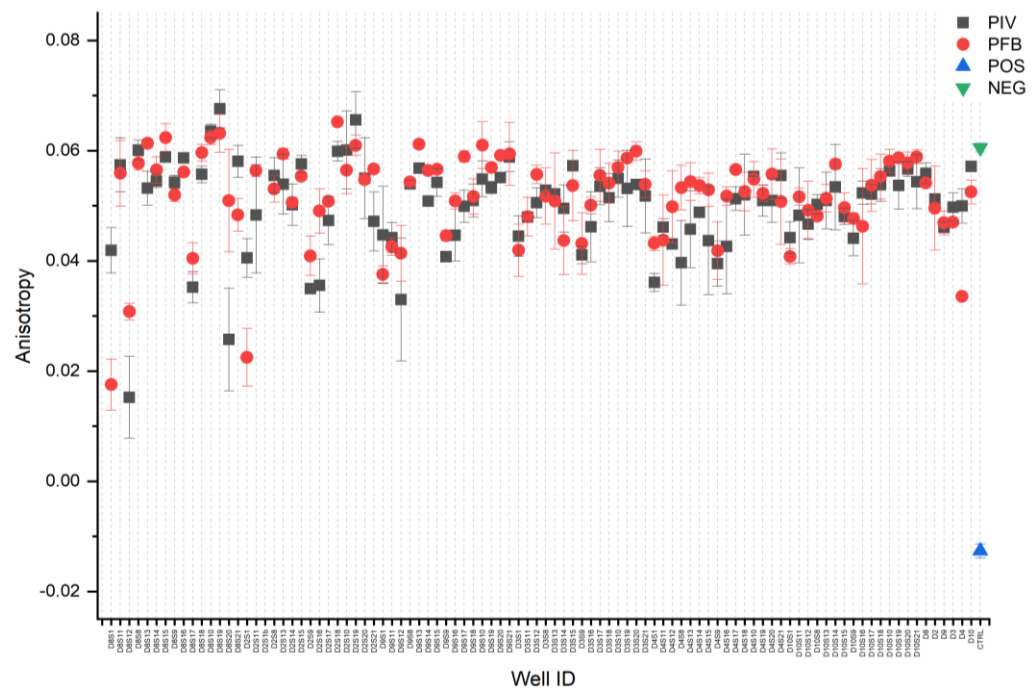
Round 1 HTS at 20 μ M total product concentration:



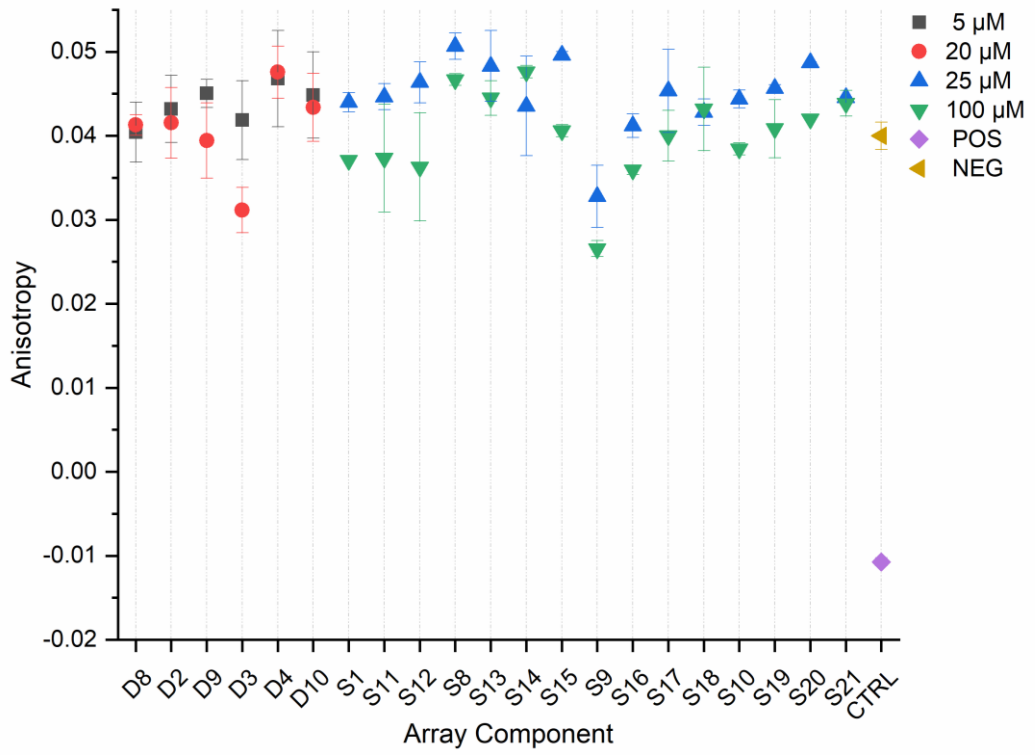
Reaction array 1 reagents:



Round 2 HTS at 20 uM:



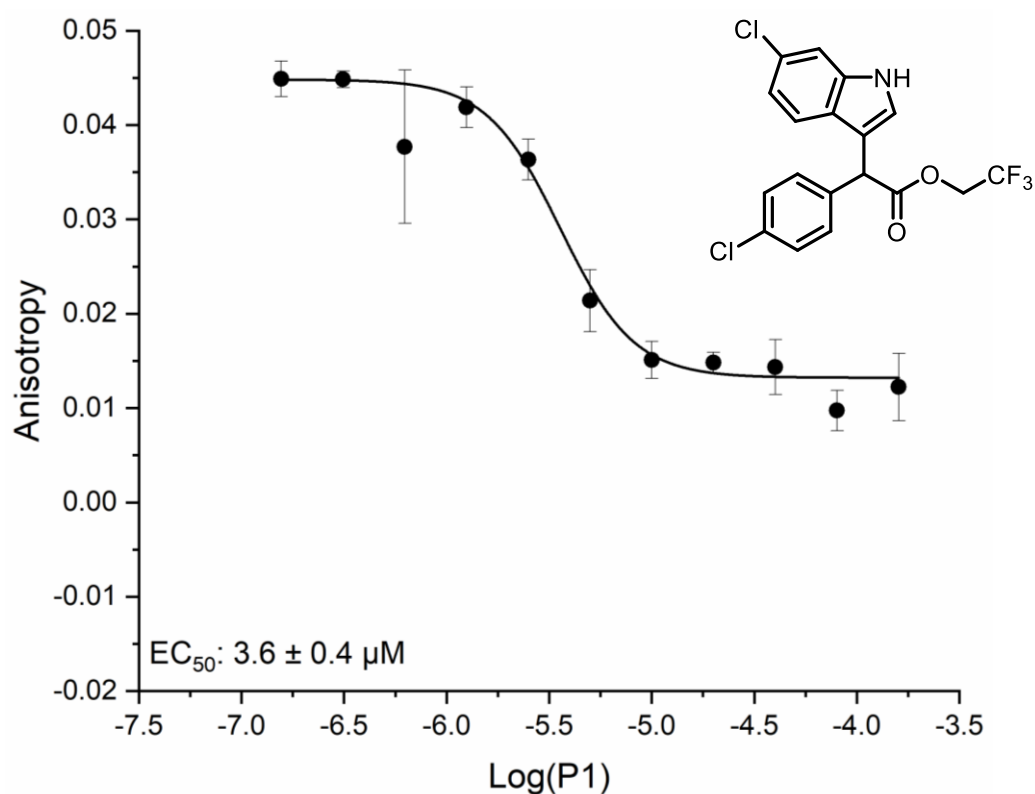
Reaction array 2 reagents:

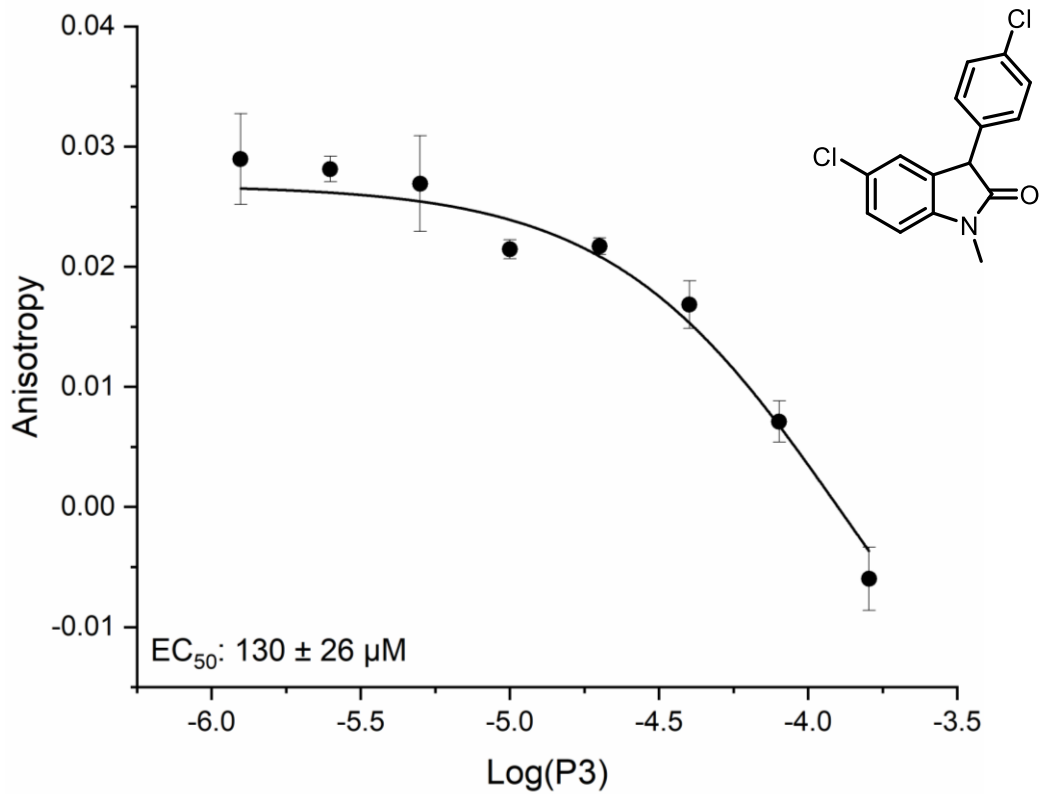
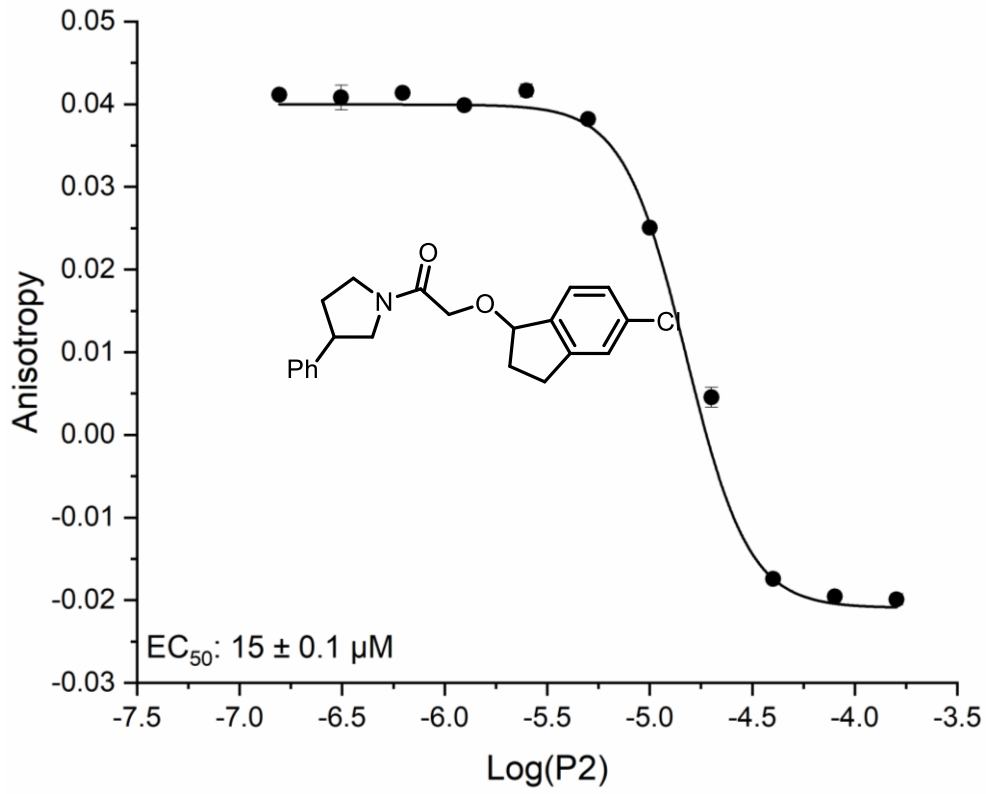


5.3.8 Determining EC₅₀ Values for Isolated Compounds

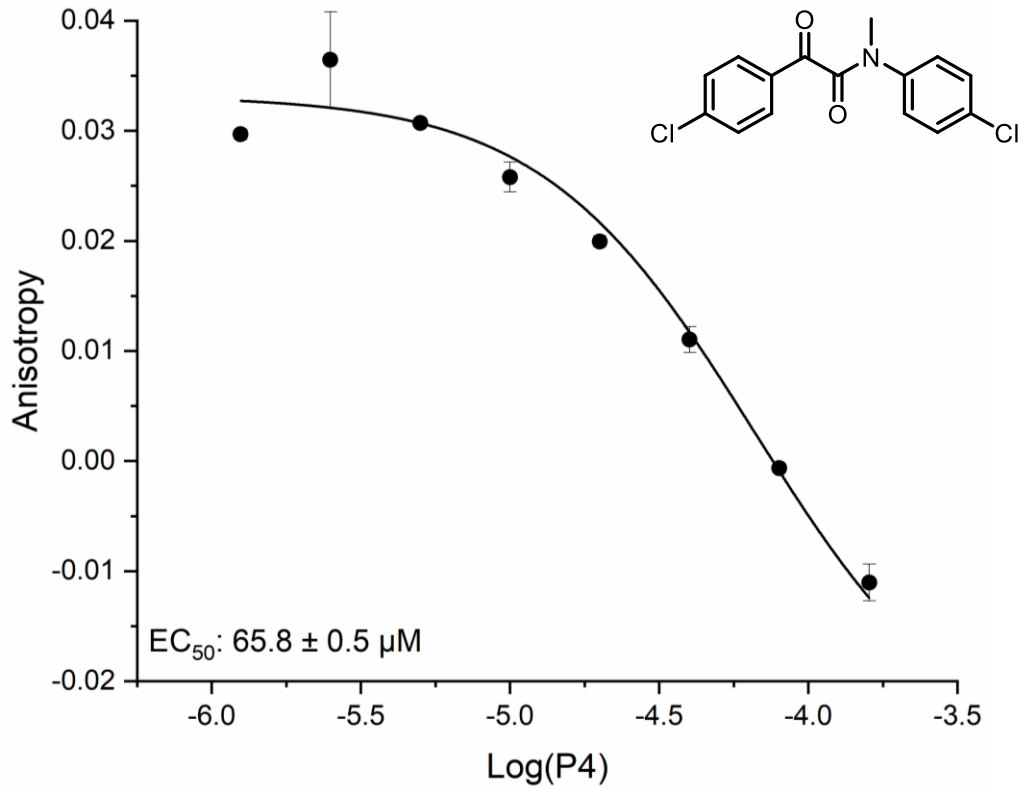
Pure compounds (**P1** – **P7** and **4.1** – **4.4**) were serially diluted in 100% DMSO (using 12 two-fold dilution steps) to achieve the correct effective concentrations (15.8 – 0.007 mM), then diluted 33-fold in pH 7.5 aqueous phosphate buffer (40 mM phosphate, 200 mM NaCl and 0.02 mg/mL Bovine Serum Albumin) to achieve a 3% DMSO intermediate stock solution (480 – 0.23 μ M). The assay was then implemented similarly to the examples above to give final compound concentrations between 0.08 and 160 μ M.

EC₅₀ values were determined and curves were fit in Origin Pro 2019b using a non-linear curve fitting with the dose response fitting procedure and Levenberg Marquardt iteration algorithm.

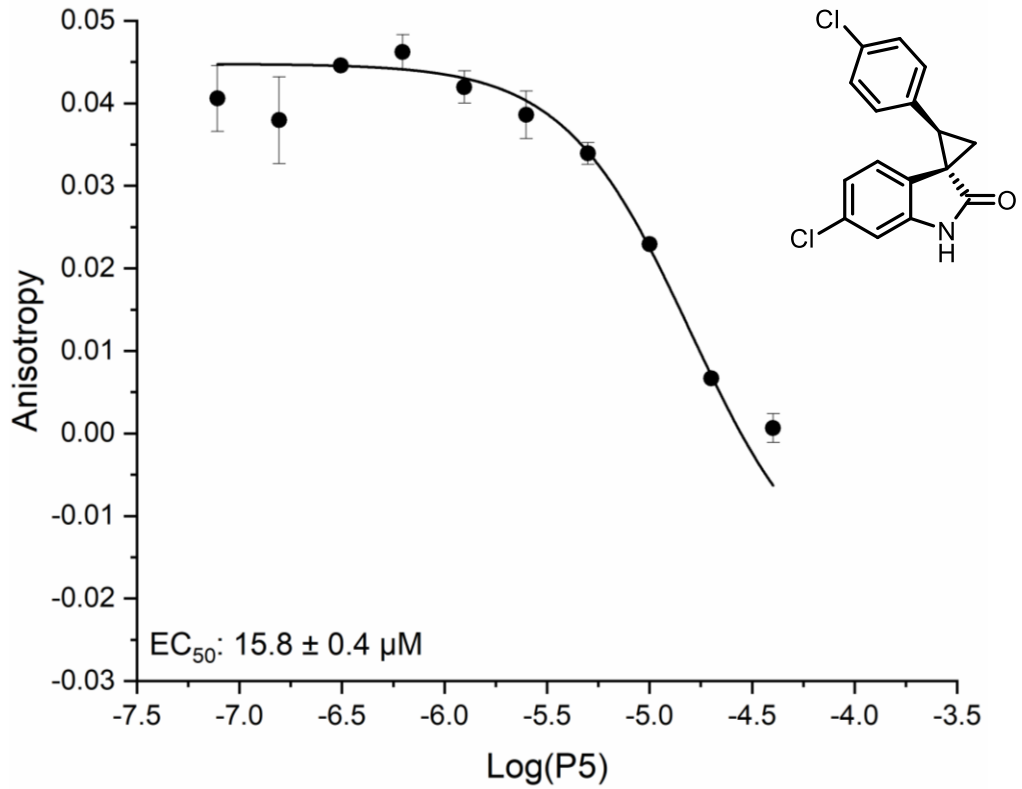




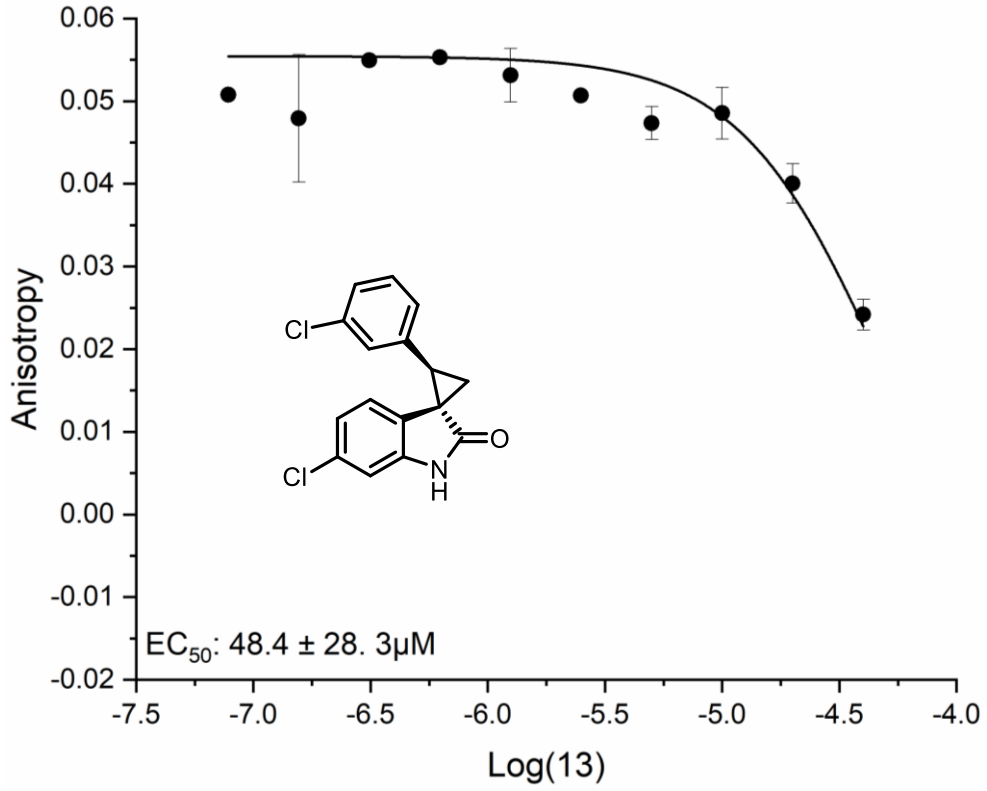
Due to poor compound solubility a full dose-response curve could not be obtained.



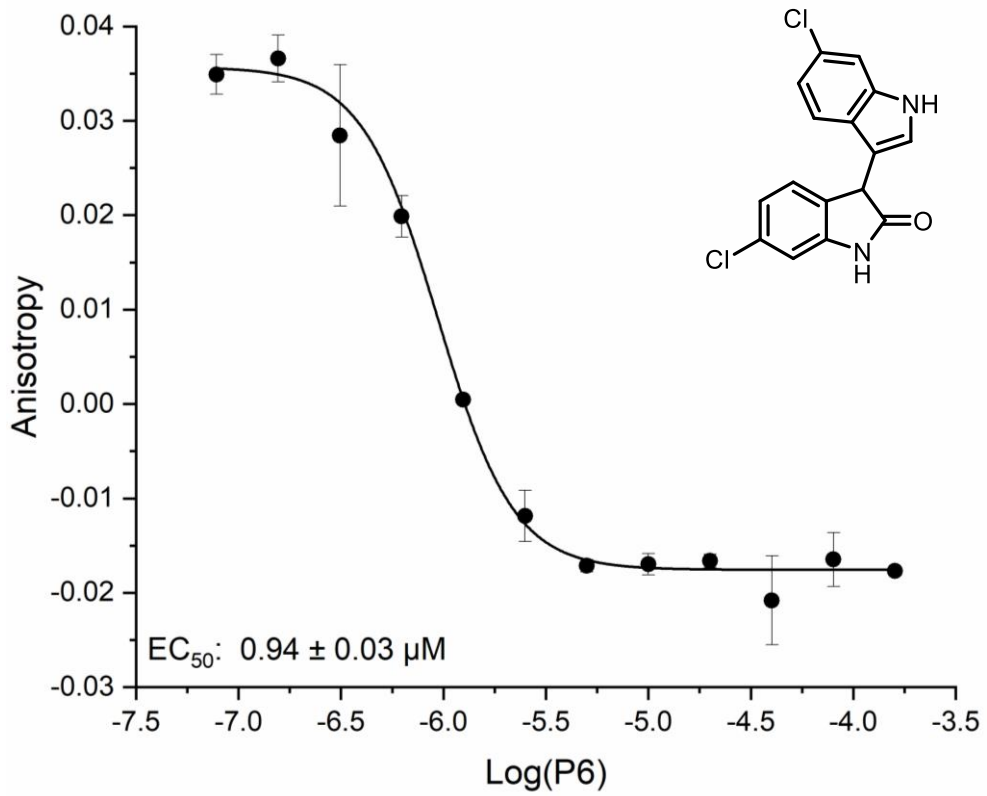
Due to poor compound solubility a full dose-response curve could not be obtained.

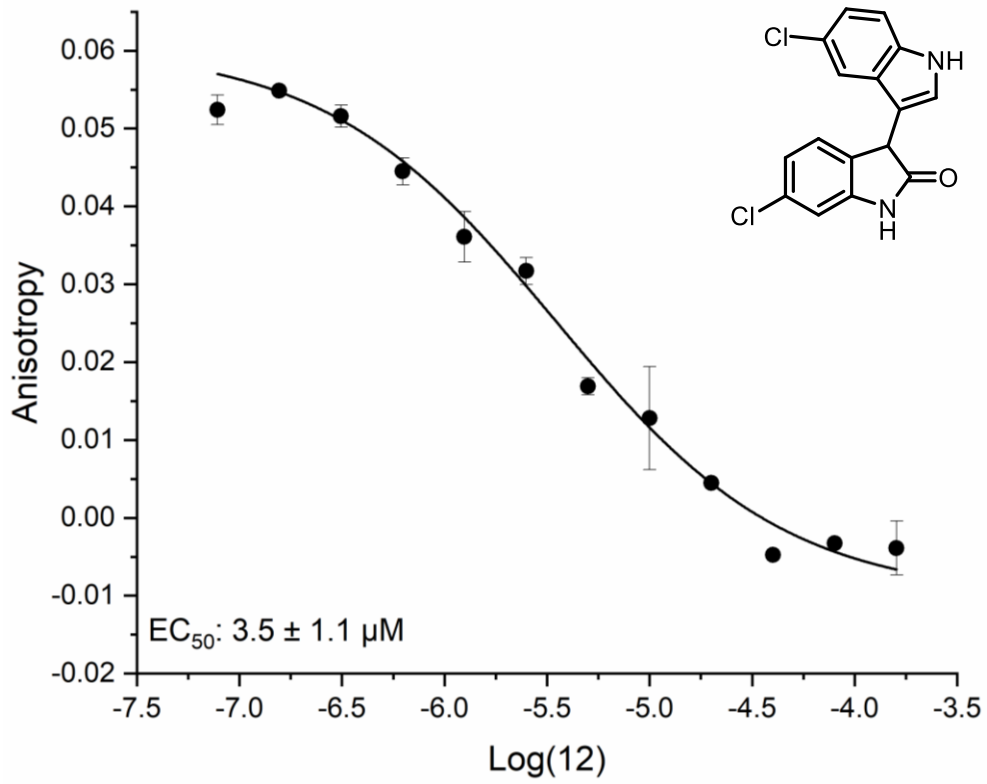


Due to poor compound solubility a full dose-response curve could not be obtained.

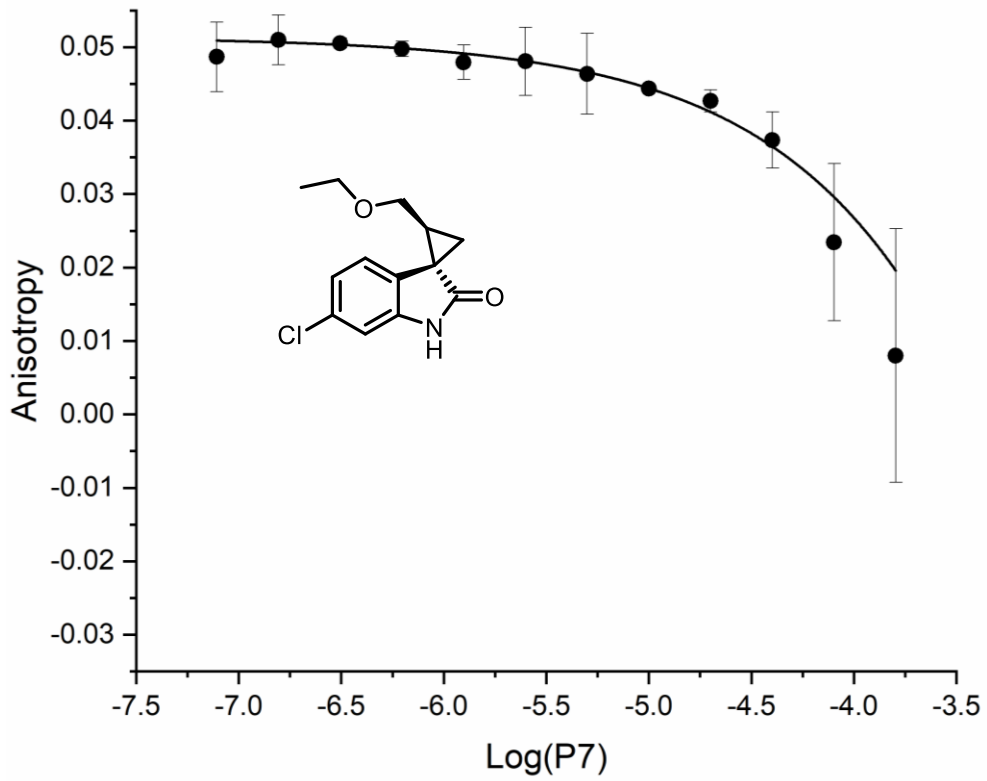


Due to poor compound solubility a full dose-response curve could not be obtained.





P7



Due to poor compound solubility a full dose-response curve could not be obtained. Curve fitting failed and EC₅₀ could not be determined.

5.3.9 NMR Measurements for K_d Estimation

5.3.9.1 K_d Estimation Using $^1\text{H}/^{15}\text{N}$ -HSQC NMR

NMR titrations were performed by recording a series of $^1\text{H}/^{15}\text{N}$ -HSQC experiments on a 750 MHz Oxford Magnet spectrometer (TCI-Cyproprobe, ^1H optimized triple resonance NMR 'inverse' probe) (^1H = 750 MHz and ^{15}N = 76 MHz) in pH 7.5 aqueous phosphate buffer containing 100 mM phosphate, 1 mM DTT and 2.5% glycerol with 50 μM ^{15}N -labelled hDM2₁₇₋₁₂₅, 10% D₂O and 1% DMSO. Temperature was maintained at 298 K throughout the experiments. Pure compounds were titrated into the ^{15}N -MDM2 sample in 0.5-, 1-, 1.5- and 2-molar equivalents relative to ^{15}N -MDM2 as standard and further molar equivalents of 4- and 6-times compound-to-MDM2 were added if the protein was not fully saturated. Data was processed using Topspin and analysed with Sparky.²⁸⁵

K_d values were obtained by plotting the observed chemical shift perturbation (csp) of the reporter peaks L54, L57, G58, M62, V75, V93, K94, H96 and K98 against the molar ratio of ligand. The csp of each reporter peak was calculated as the deviation from the free protein resonances using the equation:

$$csp = \sqrt{(\omega_{2\ free} - \omega_{2\ bound})^2 + \frac{(\omega_{1\ free} - \omega_{1\ bound})^2}{10}}$$

Where ω_1 is the ^{15}N chemical shift and ω_2 is the ^1H chemical shift corresponding to the observed HSQC cross-peak for a given reporter residue.

K_d values for each reporter peak were then obtained by solving the equation:

$$\Delta = \Delta_o \frac{(K_d + [L] + [P]) - \sqrt{((K_d + [L] + [P])^2 - 4[P][L])}}{2[P]}$$

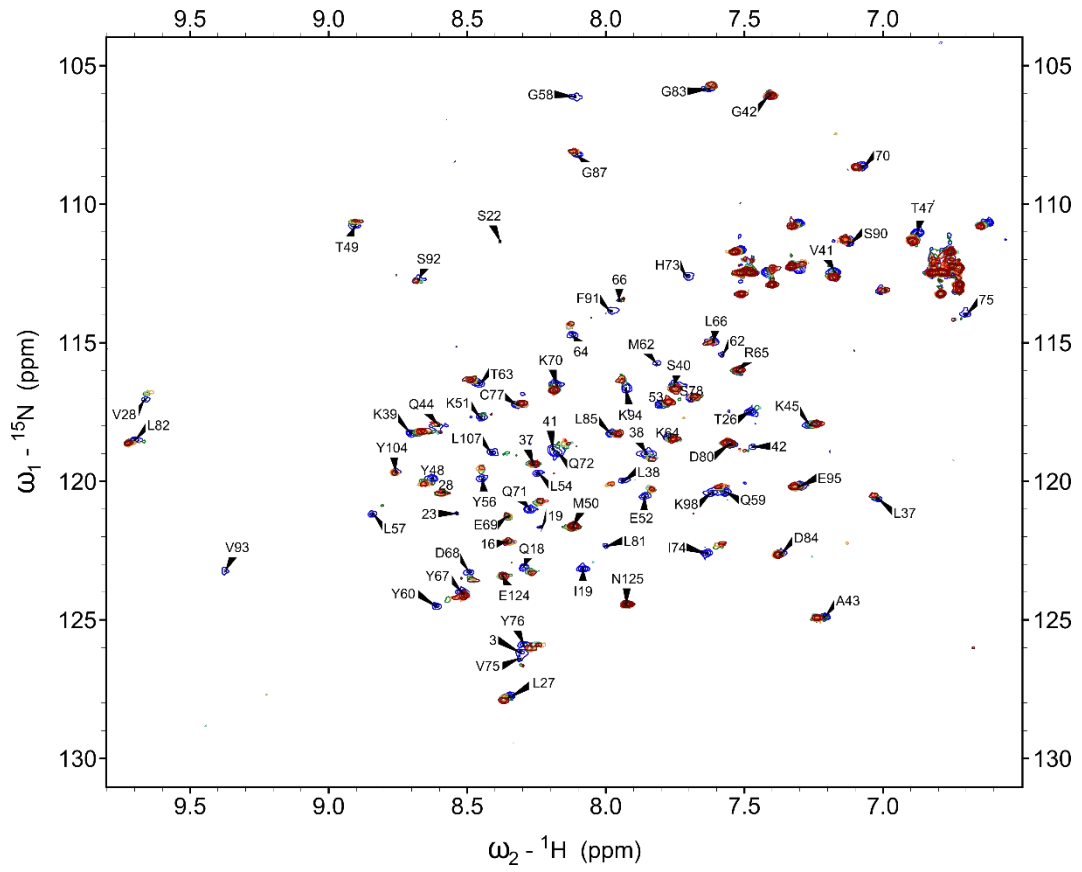
Where Δ is the observed csp, Δ_o is the maximum csp, and [P] and [L] are the protein and ligand concentrations respectively. The global K_d was then obtained from the average K_d for the combined reporter peaks:

$$Global\ K_d = \frac{\sum \text{Log}_{10}(iK_d)}{n}$$

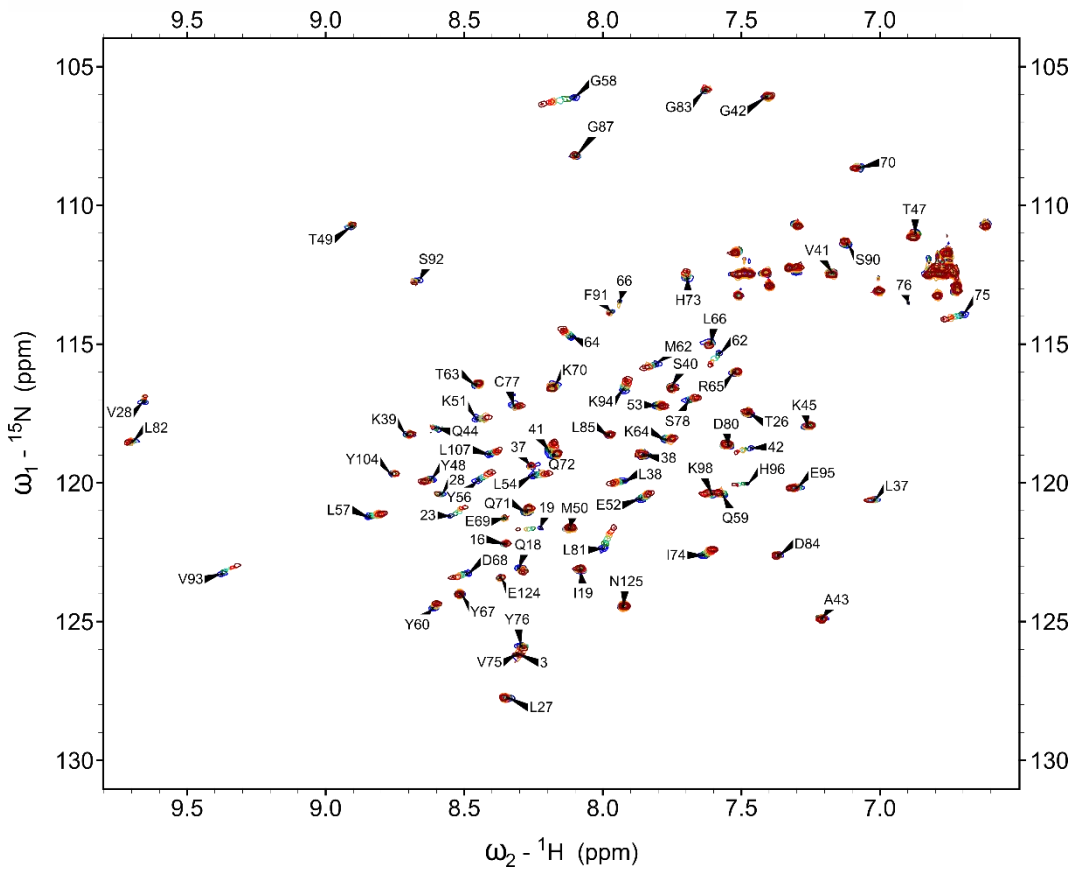
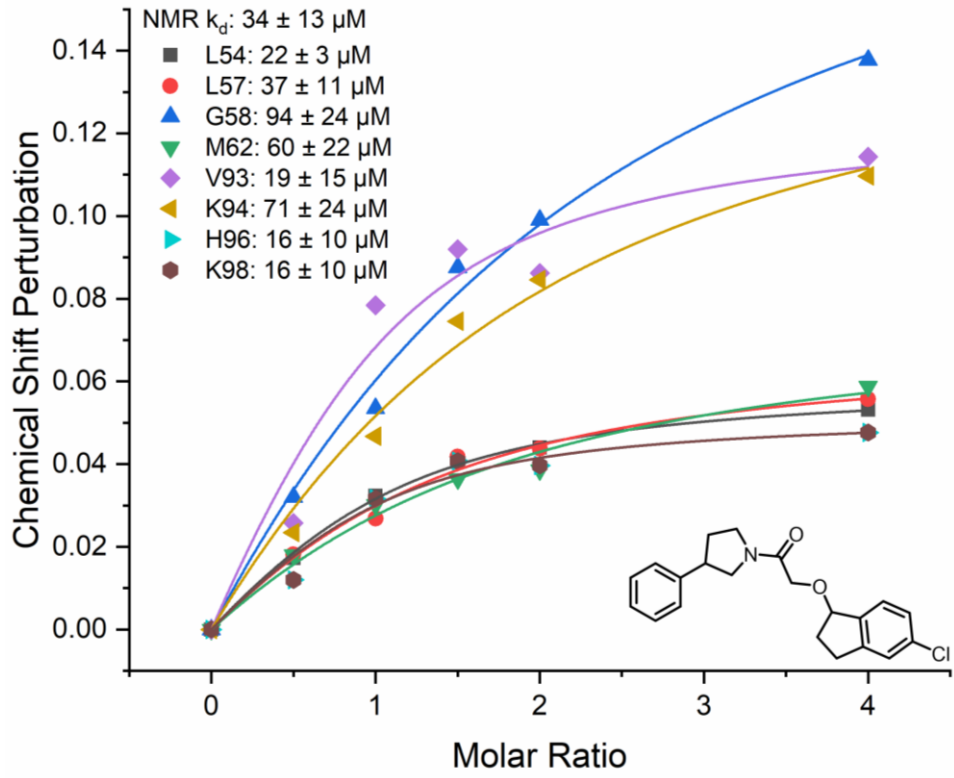
Where i is the reporter peak and n is the number of reporter peaks.

5.3.9.1 $^1\text{H}/^{15}\text{N}$ -HSQC Spectra and Fitting

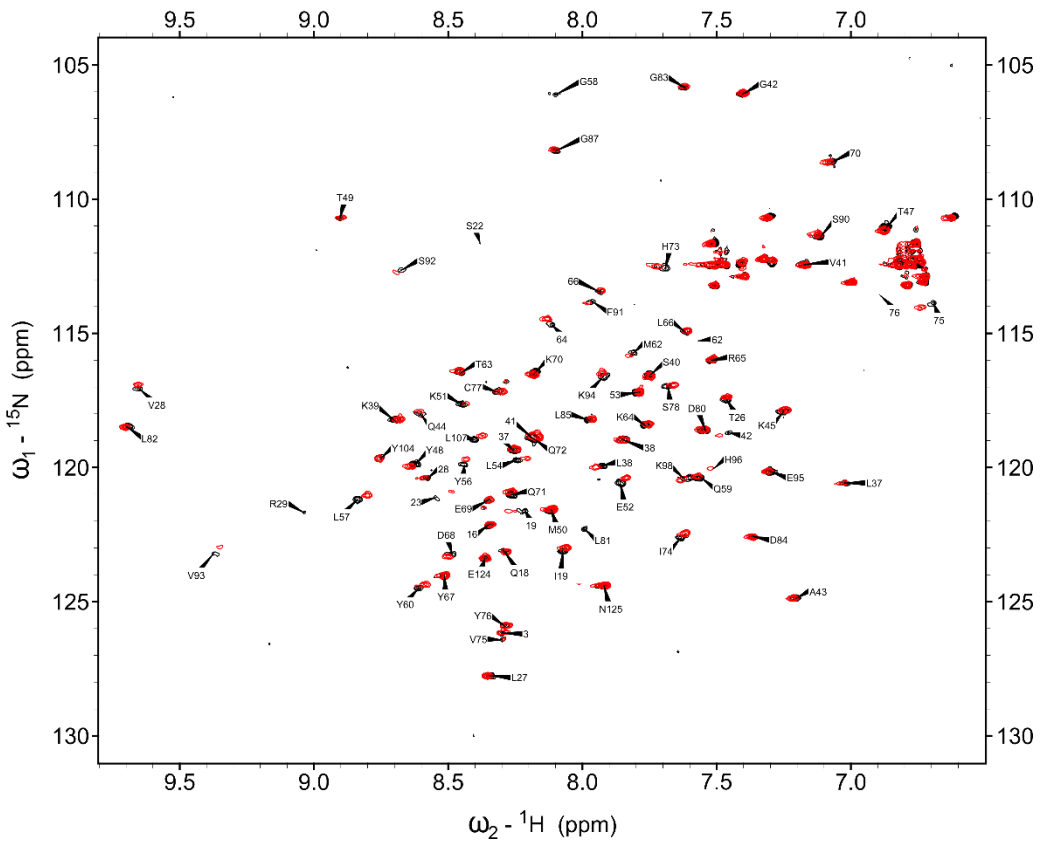
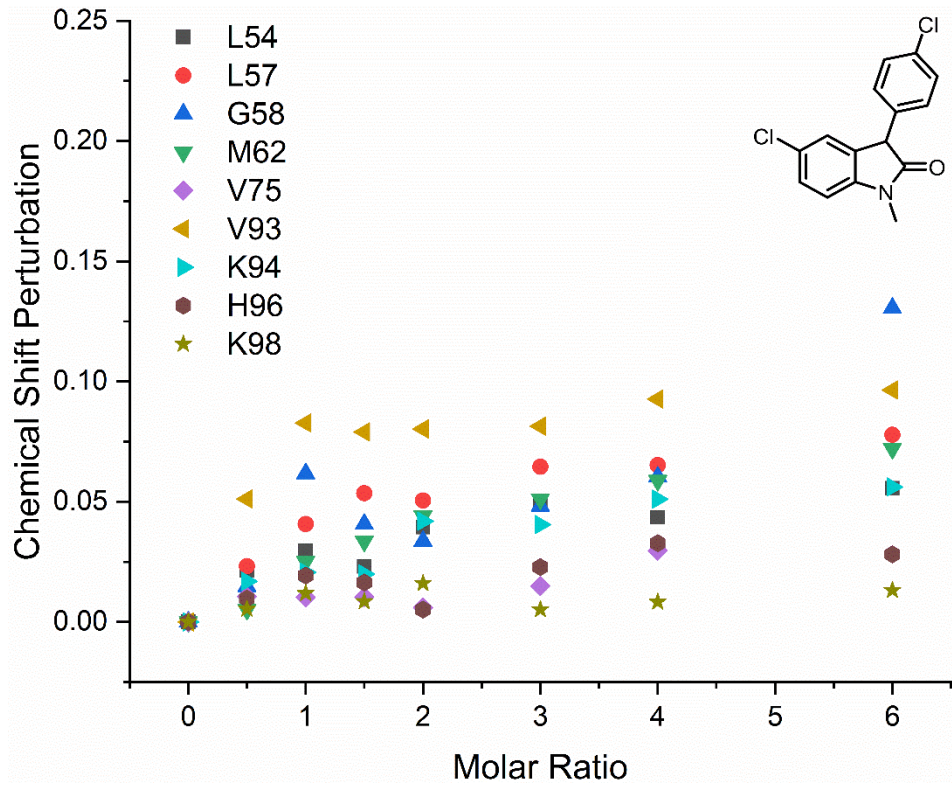
P1 – due to intermediate and slow exchange chemical shift perturbation K_d could not be estimated using the reporter peaks outlined above.



P2



P3

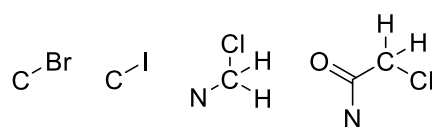


5.3.9.2 ¹⁹F-CPMG Ligand-Observed NMR

NMR titrations were performed by recording a series of ¹⁹F-CPMG experiments on a Bruker AV-400 NMR spectrometer (¹H = 400 MHz, ¹³C = 100 MHz and ¹⁹F = 376 MHz C-F decoupled) in pH 7.5 aqueous phosphate buffer containing 100 mM phosphate and 1 mM DTT with 100 μM **P1**, 10% D₂O and 5% d⁶-DMSO. Temperature was maintained at 298 K throughout the experiments. *hDM2*₁₇₋₁₂₅ was titrated into the **P1** sample in 0.0025, 0.005, 0.01, 0.015, 0.02, 0.04, and 0.1 molar equivalents and the CPMG spectra²⁸⁶ was recorded at 50 ms and 300 ms delay times. Data was processed using Topspin.

5.4 Computational Methods for the Design and Evaluation of Reaction Arrays

Diazo substrates were selected during the round 1 reaction array design by filtering a database of commercially available compounds for primary and secondary amines. The filtering workflow was primarily conducted using Pipeline Pilot and amines with undesirable functional groups, more than 20 heavy atoms, or AlogP values <-2 and >5 were removed from the library (Scheme 5.1). The amines were then enumerated in RDKit, using diazotization workflows described in Chapter 3, to form a library of virtual diazo substrates (Figure 5.9). This library was then used to select amines for diazo synthesis in round 1.



Scheme 5.1 Examples of undesirable functional groups removed from the amine library.

After compiling libraries of potential diazo substrates and co-substrates, a virtual reaction array was enumerated, and the theoretical coverage of chemical space was investigated. The goal of virtual reaction array enumeration was to discard array designs with particularly poor coverage of chemical space, rather than select the best array design. The final round 1 array design, described in Chapter 3, was chosen to balance the diversity in reactants with the ability to synthesise diazo substrates (Figure 3.3 and 3.4).

5.4.1 Exemplar Reaction SMARTS for the Enumeration of Virtual Products from Reaction Arrays

Described below are reaction SMARTS for common rhodium(II) catalysed reactions with diazo compounds (Table 5.8). The SMART strings were used to enumerate libraries of virtual products for hypothetical ADS reaction arrays. The libraries were used to assess potential product diversity within a reaction array and were used to aid in the selection of substrates and co-substrates.

Table 5.8 Reaction SMARTS for enumeration of ADS arrays using rhodium(II) chemistry

Reaction	SMARTS
Cyclopropanation	[#6:1]=[N+:2]=[#7-:3].[#6:4]=[#6:5]>>[#6:5]-1-[#6:4]-[#6:1]-1
O–H insertion	[#6:1]=[N+:2]=[#7-:3].[#6:5]-[#8:4]>>[#6:5]-[#8:4]-[#6:1]
Isomunchnone isoxazole formation	[#7-:3]=[N+:2]=[#6:1]-[#6:4]=[O:5].[C:7]#[N:6]>>[#8:5]-1-[#6:4]=[#6:1]-[#7:6]=[#6:7]-1')
C–H insertion α -to a nitrogen atom	[#6:1]=[N+:2]=[#7-:3].[#6:4]-[#7:5]>>[#6:1]-[#6:4]-[#7:5]
C–H insertion α -to an oxygen atom	[#6:1]=[N+:2]=[#7-:3].[#6:4]-[#8:5]>>[#6:1]-[#6:4]-[#8:5]
C–H insertion α -to a nitrile group	[#6:1]=[N+:2]=[#7-:3].[#6:4][C:5]#[N:6]>>[#6:1]-[#6:4][C:5]#[N:6]
Benzylic C–H insertion	[#6:1]=[N+:2]=[#7-:3].[#6:4]-[c:5]:[c:6]>>[#6:1]-[#6:4]-[c:5]:[c:6]
C–H insertion at a C3 indole position	[#6:1]=[N+:2]=[#7-:3].[c:4]1[c:8][c:7][n:6][c:5]1>>[#6:1]-[c:4]1[c:8][c:7][n:6][c:5]1
C–H insertion at a C2 indole position	[#6:1]=[N+:2]=[#7-:3].[c:4]1[c:8][c:7][n:6][c:5]1>>[#6:1]-[c:5]1[c:4][c:8][c:7][n:6]1')

5.4.2 A Python Script for the Enumeration of Virtual Reaction Arrays using RDKit

Described below is a python function for the enumeration of virtual libraries using RDKit. The function was used to compare the chemical space covered by a hypothetical ADS reaction array to the chemical space covered by MDM2 ligands deposited into the ChEMBL database.²⁵⁸

```
# RDKit Modules
from rdkit import rdBase
from rdkit import Chem
from rdkit.Chem import AllChem
from rdkit.Chem import rdChemReactions
from rdkit.Chem import PandasTools

# Nested Functions for Reaction Enumeration (Remove '#' From Writer to Save
Excel File of Results)
# rxn is nested into multi_rxn -> only need to run multi_rxn for concatenated results
# Single Reaction Enumerator -> DOES NOT GIVE MOLECULES OBJECTS,
# ONLY SMILES
# Give SUBSTRATES and CO_SUBSTRATES as SMILES
def rxn(i):      # i = reaction SMART string
    r = AllChem.EnumerateLibraryFromReaction(i,[SUBSTRATES,
                                                CO_SUBSTRATES])
    r = pd.DataFrame([Chem.MolToSmiles(x[0]) for x in list(r)], columns=['SMILES'])
    r = r.drop_duplicates()
    return r
```

```
# Function for Concatenating Results of Multiple reactions
# Takes RXN Output and Concatenates into Complete DataFrame
# Generates Molecules on Complete DataFrame and Removes Duplicates
def multi_rxn(*i):      # i = array of SMARTS strings
    r = pd.DataFrame()
    for x in i:
        r = r.append([r, rxn(x)])
        PandasTools.AddMoleculeColumnToFrame(r, 'SMILES', 'Molecule',
                                                includeFingerprints=True)
    r = r.replace(to_replace='None',
                  value=np.nan).dropna().drop_duplicates(subset='SMILES').
                                                reset_index(drop=True)

    #writer = pd.ExcelWriter('virtual_library.xlsx')
    #r.to_excel(writer, 'Sheet1')
    #writer.save()
    return r
```

multi_rxn(RXN_SMARTS) # Pseudo code to using the reaction smarts in section
5.3.2

5.4.3 An Exemplar Virtual Reaction Array

Once a virtual reaction array was enumerated it was compared to the chemical space of some known MDM2 ligands (Figure 5.10). PCA or *t*-Distributed Stochastic Neighbour Embedding (*t*-SNE) methods were used to evaluate the molecular fingerprints (Morgan fingerprints) of the virtual product library and the library of known MDM2 ligands. Below is an example of this approach using the substrates and co-substrates from the round 1 ADS reaction array for the discovery of MDM2 ligands (Figure 5.11).

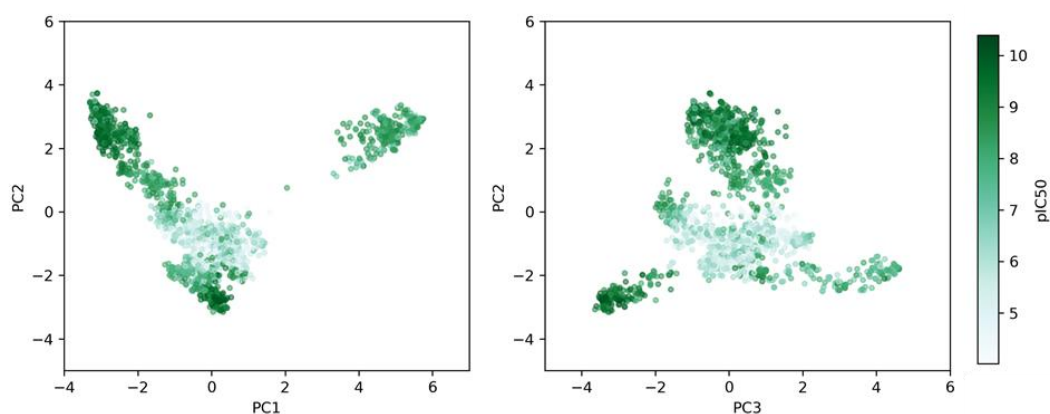


Figure 5.10. PCA decomposition of a library of known MDM2 ligands.

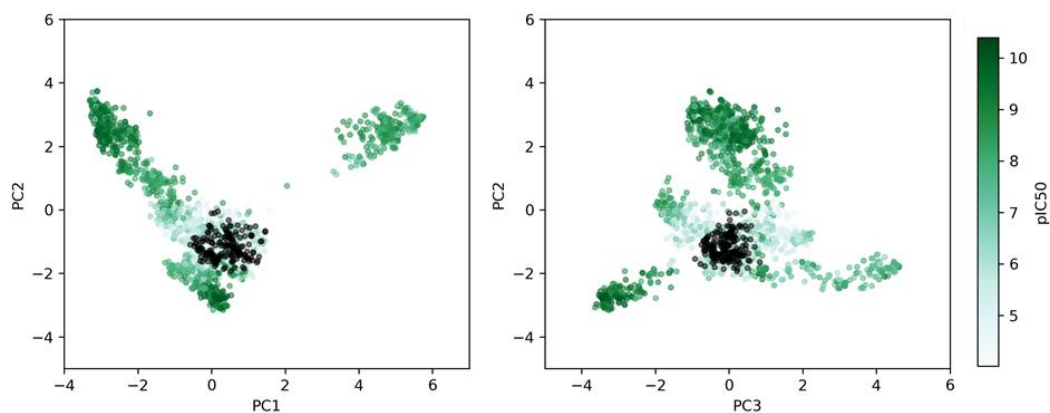


Figure 5.11. PCA decomposition of a library of known MDM2 ligands and virtual ADS products. ADS reaction products are represented in black.

5.5 Similarity Analysis of Hit Molecules

1769 compounds with annotated bioactivity towards MDM2 were obtained from the ChEMBL database (accessed: 16/01/2020). Subsequent processing gave 1314 compounds which were then used for further analysis. The Morgan molecular fingerprint was then computed for each molecule

using RDKit and the pairwise Tanimoto similarity scores were calculated (Table 5.9).

Table 5.9. Tanimoto similarity analysis comparing ADS products **P1** – **P6** to 1314 known hDM2 ligands from the ChEMBL database.

Metric	P1	P2	P3	P4	P5	P6
Mean	0.37	0.34	0.36	0.26	0.43	0.44
Median	0.38	0.34	0.36	0.25	0.44	0.45
Minimum similarity	0.12	0.12	0.13	0.13	0.13	0.12
Maximum similarity	0.50	0.48	0.46	0.37	0.61	0.51

Lipophilic ligand efficiency (LLE) was calculated using the equation below

$$LLE = pIC_{50} - cLogP$$

5.6 Molecular Docking

ADS products **P2**, **P4**, **P5** and **P6** were docked, using Autodock Vina,²⁵⁵ into an X-ray crystal structure of hDM2 (PDB: 6Q9H) with the small molecule ligand removed. Representative, DFT-optimised, 3-dimensional structures for **P2** and **P4** – **P6** were generated in Gaussian09²⁷⁷ at the B3LYP/6-31G(d) level of theory. Each ligand was docked as its individual stereoisomers. The cartesian coordinates for the DFT-optimised structures were then used as initial ligand geometries in Autodock Vina. The hDM2 crystal structure was then prepared by removing all hydrogen atoms and water molecules, and then the hydrogens for polar residues or residues that would be protonated under physiological conditions were re-added.

The following grid space (Å) was used for docking with PDB:6Q9H

center_x = 5.659

center_y = 15.381

center_z = -1.815

size_x = 22

size_y = 22

size_z = 22

Ligand files were then prepared by setting the rotatable bonds for each ligand, setting aromaticity, and constraining the rotation of amide bonds. The molecular docking of each ligand into *hDM2* was then computed with an exhaustiveness value of 64. Top ranked docking poses were then manually evaluated for the quality of the docked pose.

Example configuration file for Autodock Vina:

```
receptor = YOUR_STRUCTURE_HERE.pdbqt
ligand = YOUR_LIGAND_HERE.pdbqt
out = RESULT_HERE.pdbqt

center_x = 5.659
center_y = 15.381
center_z = -1.815

size_x = 22    # remember to convert from a.u. to Angstroms when setting this in
size_y = 22    # Autodock Tools otherwise it'll be very wrong (i.e. massive)
size_z = 22

exhaustiveness = 64    # 8 is the default and is pretty quick if docking many
                        # molecules. Use higher values if docking fewer molecules
                        # for greater accuracy/quality
```

List of References

- 1 S. M. Paul, D. S. Mytelka, C. T. Dunwiddie, C. C. Persinger, B. H. Munos, S. R. Lindborg and A. L. Schacht, *Nat Rev Drug Discov*, 2010, **9**, 203–214.
- 2 I. Khanna, *Drug Discovery Today*, 2012, **17**, 1088–1102.
- 3 F. Pammolli, L. Magazzini and M. Riccaboni, *Nat Rev Drug Discov*, 2011, **10**, 428–438.
- 4 A. Mullard, *Nat Rev Drug Discov*, 2020, **19**, 79–84.
- 5 I. Kola and J. Landis, *Nature Reviews Drug Discovery*, 2004, **3**, 711–716.
- 6 W. G. Kaelin, *Nat Rev Cancer*, 2017, **17**, 441–450.
- 7 J. P. Hughes, S. Rees, S. B. Kalindjian and K. L. Philpott, *British Journal of Pharmacology*, 2011, **162**, 1239–1249.
- 8 P. Morgan, D. G. Brown, S. Lennard, M. J. Anderton, J. C. Barrett, U. Eriksson, M. Fidock, B. Hamrén, A. Johnson, R. E. March, J. Matcham, J. Mettetal, D. J. Nicholls, S. Platz, S. Rees, M. A. Snowden and M. N. Pangalos, *Nat Rev Drug Discov*, 2018, **17**, 167–181.
- 9 M. E. Bunnage, *Nat Chem Biol*, 2011, **7**, 335–339.
- 10 J. Eder, R. Sedrani and C. Wiesmann, *Nat Rev Drug Discov*, 2014, **13**, 577–587.
- 11 I. Collins and P. Workman, *Nat Chem Biol*, 2006, **2**, 689–700.
- 12 G. M. Keserú and G. M. Makara, *Drug Discovery Today*, 2006, **11**, 741–748.
- 13 G. Siegal, E. AB and J. Schultz, *Drug Discovery Today*, 2007, **12**, 1032–1039.
- 14 D. A. Erlanson, B. J. Davis and W. Jahnke, *Cell Chemical Biology*, 2019, **26**, 9–15.
- 15 G. Wu and S. K. Doberstein, *Drug Discovery Today*, 2006, **11**, 718–724.
- 16 W. P. Walters and M. Namchuk, *Nat Rev Drug Discov*, 2003, **2**, 259–266.
- 17 D. Murray and M. Wigglesworth, 2016, 1–15.
- 18 K. H. Bleicher, H.-J. Böhm, K. Müller and A. I. Alanine, *Nat Rev Drug Discov*, 2003, **2**, 369–378.
- 19 B. J. Davis and D. A. Erlanson, *Bioorganic & Medicinal Chemistry Letters*, 2013, **23**, 2844–2852.
- 20 A. A. Shelat and R. K. Guy, *Nat Chem Biol*, 2007, **3**, 442–446.
- 21 H.-P. Shih, X. Zhang and A. M. Aronov, *Nat Rev Drug Discov*, 2018, **17**, 19–33.
- 22 S. L. Schreiber, J. D. Kotz, M. Li, J. Aubé, C. P. Austin, J. C. Reed, H. Rosen, E. L. White, L. A. Sklar, C. W. Lindsley, B. R. Alexander, J. A. Bittker, P. A. Clemons, A. de Souza, M. A. Foley, M. Palmer, A. F. Shamji, M. J. Wawer, O. McManus, M. Wu, B. Zou, H. Yu, J. E. Golden, F. J. Schoenen, A. Simeonov, A. Jadhav, M. R. Jackson, A. B. Pinkerton, T. D. Y. Chung, P. R. Griffin, B. F. Cravatt, P. S. Hodder, W. R. Roush, E. Roberts, D.-H. Chung, C. B. Jonsson, J. W. Noah, W. E. Severson, S. Ananthan, B. Edwards, T. I. Oprea, P. J. Conn, C. R. Hopkins, M. R. Wood, S. R. Stauffer, K. A. Emmitte, L. S. Brady, J. Driscoll, I. Y. Li, C. R.

- Loomis, R. N. Margolis, E. Michelotti, M. E. Perry, A. Pillai and Y. Yao, *Cell*, 2015, **161**, 1252–1265.
- 23 J. G. Moffat, J. Rudolph and D. Bailey, *Nat Rev Drug Discov*, 2014, **13**, 588–602.
- 24 J. Knowles and G. Gromo, *Nat Rev Drug Discov*, 2003, **2**, 63–69.
- 25 K. D. Barnash, L. I. James and S. V. Frye, *Nat Chem Biol*, 2017, **13**, 1053–1056.
- 26 I. Gashaw, P. Ellinghaus, A. Sommer and K. Asadullah, *Drug Discovery Today*, 2012, **17**, S24–S30.
- 27 A. M. Wassermann, E. Lounkine, D. Hoepfner, G. Le Goff, F. J. King, C. Studer, J. M. Peltier, M. L. Grippo, V. Prindle, J. Tao, A. Schuffenhauer, I. M. Wallace, S. Chen, P. Krastel, A. Cobos-Correa, C. N. Parker, J. W. Davies and M. Glick, *Nat Chem Biol*, 2015, **11**, 958–966.
- 28 A. Karawajczyk, F. Giordanetto, J. Benningshof, D. Hamza, T. Kalliokoski, K. Pouwer, R. Morgentin, A. Nelson, G. Müller, A. Piechot and D. Tzalis, *Drug Discovery Today*, 2015, **20**, 1310–1316.
- 29 R. Macarron, M. N. Banks, D. Bojanic, D. J. Burns, D. A. Cirovic, T. Garyantes, D. V. S. Green, R. P. Hertzberg, W. P. Janzen, J. W. Paslay, U. Schopfer and G. S. Sittampalam, *Nat Rev Drug Discov*, 2011, **10**, 188–195.
- 30 D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke and H. Jhoti, *Nat Rev Drug Discov*, 2016, **15**, 605–619.
- 31 G. Schneider, *Nat Rev Drug Discov*, 2018, **17**, 97–113.
- 32 D. A. Pereira and J. A. Williams, *British Journal of Pharmacology*, 2007, **152**, 53–61.
- 33 A. Nadin, C. Hattotuwigama and I. Churcher, *Angew. Chem. Int. Ed.*, 2012, **51**, 1114–1122.
- 34 C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Advanced Drug Delivery Reviews*, 1997, **23**, 3–25.
- 35 R. J. Young and P. D. Leeson, *J. Med. Chem.*, 2018, **61**, 6421–6467.
- 36 C. P. Tinworth and R. J. Young, *J. Med. Chem.*, 2020, [acs.jmedchem.9b01596](https://doi.org/10.1021/acs.jmedchem.9b01596).
- 37 M. D. Shultz, *J. Med. Chem.*, 2019, **62**, 1701–1714.
- 38 I. Rajman, *Drug Discovery Today*, 2008, **13**, 341–346.
- 39 D. C. Rees, M. Congreve, C. W. Murray and R. Carr, *Nature Reviews Drug Discovery*, 2004, **3**, 660–672.
- 40 S. L. McGovern, E. Caselli, N. Grigorieff and B. K. Shoichet, *J. Med. Chem.*, 2002, **45**, 1712–1722.
- 41 J. J. Irwin, D. Duan, H. Torosyan, A. K. Doak, K. T. Ziebart, T. Sterling, G. Tumanian and B. K. Shoichet, *J. Med. Chem.*, 2015, **58**, 7076–7087.
- 42 R. Carr and H. Jhoti, *Drug Discovery Today*, 2002, **7**, 522–527.
- 43 D. A. Erlanson, A. C. Braisted, D. R. Raphael, M. Randal, R. M. Stroud, E. M. Gordon and J. A. Wells, *Proceedings of the National Academy of Sciences*, 2000, **97**, 9367–9372.
- 44 D. Vetter, *Journal of Cellular Biochemistry*, 2002, **87**, 79–84.
- 45 M. Congreve, R. Carr, C. Murray and H. Jhoti, *Drug Discovery Today*, 2003, **8**, 876–877.
- 46 H. Jhoti, G. Williams, D. C. Rees and C. W. Murray, *Nat Rev Drug Discov*, 2013, **12**, 644–644.

- 47 N. Fuller, L. Spadola, S. Cowen, J. Patel, H. Schönherr, Q. Cao, A. McKenzie, F. Edfeldt, A. Rabow and R. Goodnow, *Drug Discovery Today*, 2016, **21**, 1272–1283.
- 48 P. C. Ray, M. Kiczun, M. Huggett, A. Lim, F. Prati, I. H. Gilbert and P. G. Wyatt, *Drug Discovery Today*, 2017, **22**, 43–56.
- 49 F. Giordanetto, C. Jin, L. Willmore, M. Feher and D. E. Shaw, *J. Med. Chem.*, 2019, **62**, 3381–3394.
- 50 A. D. Morley, A. Pugliese, K. Birchall, J. Bower, P. Brennan, N. Brown, T. Chapman, M. Drysdale, I. H. Gilbert, S. Hoelder, A. Jordan, S. V. Ley, A. Merritt, D. Miller, M. E. Swarbrick and P. G. Wyatt, *Drug Discovery Today*, 2013, **18**, 1221–1227.
- 51 J. A. Johnson, C. A. Nicolaou, S. E. Kirberger, A. K. Pandey, H. Hu and W. C. K. Pomerantz, *ACS Med. Chem. Lett.*, 2019, **10**, 1648–1654.
- 52 G. M. Keserü, D. A. Erlanson, G. G. Ferenczy, M. M. Hann, C. W. Murray and S. D. Pickett, *J. Med. Chem.*, 2016, **59**, 8189–8206.
- 53 E. H. Mashalidis, P. Śledź, S. Lang and C. Abell, *Nat Protoc*, 2013, **8**, 2309–2324.
- 54 C. W. Murray and D. C. Rees, *Nature Chem*, 2009, **1**, 187–192.
- 55 S. B. Shuker, P. J. Hajduk, R. P. Meadows and S. W. Fesik, *Science*, 1996, **274**, 1531–1534.
- 56 M. Saur, M. J. Hartshorn, J. Dong, J. Reeks, G. Bunkoczi, H. Jhoti and P. A. Williams, *Drug Discovery Today*, 2020, **25**, 485–490.
- 57 D. E. Scott, A. G. Coyne, S. A. Hudson and C. Abell, *Biochemistry*, 2012, **51**, 4990–5003.
- 58 A. L. Hopkins, G. M. Keserü, P. D. Leeson, D. C. Rees and C. H. Reynolds, *Nat Rev Drug Discov*, 2014, **13**, 105–121.
- 59 C. W. Murray and D. C. Rees, *Angew. Chem. Int. Ed.*, 2016, **55**, 488–492.
- 60 R. A. E. Carr, M. Congreve, C. W. Murray and D. C. Rees, *Drug Discovery Today*, 2005, **10**, 987–992.
- 61 J. Tsai, J. T. Lee, W. Wang, J. Zhang, H. Cho, S. Mamo, R. Bremer, S. Gillette, J. Kong, N. K. Haass, K. Sproesser, L. Li, K. S. M. Smalley, D. Fong, Y.-L. Zhu, A. Marimuthu, H. Nguyen, B. Lam, J. Liu, I. Cheung, J. Rice, Y. Suzuki, C. Luu, C. Settachatgul, R. Shellooe, J. Cantwell, S.-H. Kim, J. Schlessinger, K. Y. J. Zhang, B. L. West, B. Powell, G. Habets, C. Zhang, P. N. Ibrahim, P. Hirth, D. R. Artis, M. Herlyn and G. Bollag, *Proceedings of the National Academy of Sciences*, 2008, **105**, 3041–3046.
- 62 G. Bollag, P. Hirth, J. Tsai, J. Zhang, P. N. Ibrahim, H. Cho, W. Spevak, C. Zhang, Y. Zhang, G. Habets, E. A. Burton, B. Wong, G. Tsang, B. L. West, B. Powell, R. Shellooe, A. Marimuthu, H. Nguyen, K. Y. J. Zhang, D. R. Artis, J. Schlessinger, F. Su, B. Higgins, R. Iyer, K. D'Andrea, A. Koehler, M. Stumm, P. S. Lin, R. J. Lee, J. Grippo, I. Puzanov, K. B. Kim, A. Ribas, G. A. McArthur, J. A. Sosman, P. B. Chapman, K. T. Flaherty, X. Xu, K. L. Nathanson and K. Nolop, *Nature*, 2010, **467**, 596–599.
- 63 G. Bollag, J. Tsai, J. Zhang, C. Zhang, P. Ibrahim, K. Nolop and P. Hirth, *Nat Rev Drug Discov*, 2012, **11**, 873–886.
- 64 W. P. Walters, J. Green, J. R. Weiss and M. A. Murcko, *J. Med. Chem.*, 2011, **54**, 6405–6416.
- 65 D. G. Brown and J. Boström, *J. Med. Chem.*, 2016, **59**, 4443–4458.

- 66 T. W. J. Cooper, I. B. Campbell and S. J. F. Macdonald, *Angewandte Chemie International Edition*, 2010, **49**, 8082–8091.
- 67 S. D. Roughley and A. M. Jordan, *J. Med. Chem.*, 2011, **54**, 3451–3479.
- 68 N. Schneider, D. M. Lowe, R. A. Sayle, M. A. Tarselli and G. A. Landrum, *J. Med. Chem.*, 2016, **59**, 4385–4402.
- 69 J. Boström, D. G. Brown, R. J. Young and G. M. Keserü, *Nat Rev Drug Discov*, 2018, **17**, 709–727.
- 70 D. C. Blakemore, L. Castro, I. Churcher, D. C. Rees, A. W. Thomas, D. M. Wilson and A. Wood, *Nature Chem*, 2018, **10**, 383–394.
- 71 R. Zhang, G. Li, M. Wismer, P. Vachal, S. L. Colletti and Z.-C. Shi, *ACS Med. Chem. Lett.*, 2018, **9**, 773–777.
- 72 D. A. DiRocco, K. Dykstra, S. Krska, P. Vachal, D. V. Conway and M. Tudge, *Angew. Chem. Int. Ed.*, 2014, **53**, 4802–4806.
- 73 D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher and A. G. Doyle, *Science*, 2018, **360**, 186–190.
- 74 M. K. Nielsen, D. T. Ahneman, O. Riera and A. G. Doyle, *J. Am. Chem. Soc.*, 2018, **140**, 5004–5008.
- 75 S. Lin, S. Dikler, W. D. Blincoe, R. D. Ferguson, R. P. Sheridan, Z. Peng, D. V. Conway, K. Zawatzky, H. Wang, T. Cernak, I. W. Davies, D. A. DiRocco, H. Sheng, C. J. Welch and S. D. Dreher, *Science*, 2018, **361**, eaar6236.
- 76 R. A. Goodnow, C. E. Dumelin and A. D. Keefe, *Nat Rev Drug Discov*, 2017, **16**, 131–147.
- 77 N. Favalli, G. Bassi, J. Scheuermann and D. Neri, *FEBS Lett*, 2018, **592**, 2168–2180.
- 78 P. A. Harris, B. W. King, D. Bandyopadhyay, S. B. Berger, N. Campobasso, C. A. Capriotti, J. A. Cox, L. Dare, X. Dong, J. N. Finger, L. C. Grady, S. J. Hoffman, J. U. Jeong, J. Kang, V. Kasparcova, A. S. Lakdawala, R. Lehr, D. E. McNulty, R. Nagilla, M. T. Ouellette, C. S. Pao, A. R. Rendina, M. C. Schaeffer, J. D. Summerfield, B. A. Swift, R. D. Totoritis, P. Ward, A. Zhang, D. Zhang, R. W. Marquis, J. Bertin and P. J. Gough, *J. Med. Chem.*, 2016, **59**, 2163–2178.
- 79 S. Chow, S. Liver and A. Nelson, *Nat Rev Chem*, 2018, **2**, 174–183.
- 80 A. Buitrago Santanilla, E. L. Regalado, T. Pereira, M. Shevlin, K. Bateman, L.-C. Campeau, J. Schneeweis, S. Berritt, Z.-C. Shi, P. Nantermet, Y. Liu, R. Helmy, C. J. Welch, P. Vachal, I. W. Davies, T. Cernak and S. D. Dreher, *Science*, 2015, **347**, 49–53.
- 81 N. J. Gesmundo, B. Sauvagnat, P. J. Curran, M. P. Richards, C. L. Andrews, P. J. Dandliker and T. Cernak, *Nature*, 2018, **557**, 228–232.
- 82 G. Karageorgis, S. Warriner and A. Nelson, *Nature Chem*, 2014, **6**, 872–876.
- 83 G. Karageorgis, M. Dow, A. Aimon, S. Warriner and A. Nelson, *Angew. Chem. Int. Ed.*, 2015, **54**, 13538–13544.
- 84 D. T. Flood, C. Kingston, J. C. Vantourout, P. E. Dawson and P. S. Baran, *Isr. J. Chem.*, 2020, **60**, 268–280.
- 85 M. Song and G. T. Hwang, *J. Med. Chem.*, 2020, **63**, 6578–6599.
- 86 S. Brenner and R. A. Lerner, *Proceedings of the National Academy of Sciences*, 1992, **89**, 5381–5383.
- 87 J. Nielsen, S. Brenner and K. D. Janda, *J. Am. Chem. Soc.*, 1993, **115**, 9812–9813.

- 88 M. C. Needels, D. G. Jones, E. H. Tate, G. L. Heinkel, L. M. Kochersperger, W. J. Dower, R. W. Barrett and M. A. Gallop, *Proceedings of the National Academy of Sciences*, 1993, **90**, 10700–10704.
- 89 F. Buller, L. Mannocci, Y. Zhang, C. E. Dumelin, J. Scheuermann and D. Neri, *Bioorganic & Medicinal Chemistry Letters*, 2008, **18**, 5926–5931.
- 90 M. A. Clark, R. A. Acharya, C. C. Arico-Muendel, S. L. Belyanskaya, D. R. Benjamin, N. R. Carlson, P. A. Centrella, C. H. Chiu, S. P. Creaser, J. W. Cuzzo, C. P. Davie, Y. Ding, G. J. Franklin, K. D. Franzen, M. L. Geffer, S. P. Hale, N. J. V. Hansen, D. I. Israel, J. Jiang, M. J. Kavarana, M. S. Kelley, C. S. Kollmann, F. Li, K. Lind, S. Mataruse, P. F. Medeiros, J. A. Messer, P. Myers, H. O’Keefe, M. C. Oliff, C. E. Rise, A. L. Satz, S. R. Skinner, J. L. Svendsen, L. Tang, K. van Vloten, R. W. Wagner, G. Yao, B. Zhao and B. A. Morgan, *Nat Chem Biol*, 2009, **5**, 647–654.
- 91 L. Mannocci, Y. Zhang, J. Scheuermann, M. Leimbacher, G. De Bellis, E. Rizzi, C. Dumelin, S. Melkko and D. Neri, *Proceedings of the National Academy of Sciences*, 2008, **105**, 17670–17675.
- 92 Z. J. Gartner, B. N. Tse, R. Grubina, J. B. Doyon, T. M. Snyder and D. R. Liu, *Science*, 2004, **305**, 1601–1605.
- 93 M. H. Hansen, P. Blakskjær, L. K. Petersen, T. H. Hansen, J. W. Højfeldt, K. V. Gothelf and N. J. V. Hansen, *J. Am. Chem. Soc.*, 2009, **131**, 1322–1327.
- 94 L. K. Petersen, P. Blakskjær, A. Chaikuad, A. B. Christensen, J. Dietvorst, J. Holmkvist, S. Knapp, M. Kořínek, L. K. Larsen, A. E. Pedersen, S. Röhm, F. A. Sløk and N. J. V. Hansen, *Med. Chem. Commun.*, 2016, **7**, 1332–1339.
- 95 D. R. Halpin, J. A. Lee, S. J. Wrenn and P. B. Harbury, *PLoS Biol*, 2004, **2**, e175.
- 96 S. Melkko, J. Scheuermann, C. E. Dumelin and D. Neri, *Nat Biotechnol*, 2004, **22**, 568–574.
- 97 P. A. Harris, S. B. Berger, J. U. Jeong, R. Nagilla, D. Bandyopadhyay, N. Campobasso, C. A. Capriotti, J. A. Cox, L. Dare, X. Dong, P. M. Eidam, J. N. Finger, S. J. Hoffman, J. Kang, V. Kasparcova, B. W. King, R. Lehr, Y. Lan, L. K. Leister, J. D. Lich, T. T. MacDonald, N. A. Miller, M. T. Ouellette, C. S. Pao, A. Rahman, M. A. Reilly, A. R. Rendina, E. J. Rivera, M. C. Schaeffer, C. A. Sehon, R. R. Singhaus, H. H. Sun, B. A. Swift, R. D. Totoritis, A. Vossenkämper, P. Ward, D. D. Wisnoski, D. Zhang, R. W. Marquis, P. J. Gough and J. Bertin, *J. Med. Chem.*, 2017, **60**, 1247–1261.
- 98 P. A. Harris, D. Bandyopadhyay, S. B. Berger, N. Campobasso, C. A. Capriotti, J. A. Cox, L. Dare, J. N. Finger, S. J. Hoffman, K. M. Kahler, R. Lehr, J. D. Lich, R. Nagilla, R. T. Nolte, M. T. Ouellette, C. S. Pao, M. C. Schaeffer, A. Smallwood, H. H. Sun, B. A. Swift, R. D. Totoritis, P. Ward, R. W. Marquis, J. Bertin and P. J. Gough, *ACS Med. Chem. Lett.*, 2013, **4**, 1238–1243.
- 99 S. Berger, P. Harris, R. Nagilla, V. Kasparcova, S. Hoffman, B. Swift, L. Dare, M. Schaeffer, C. Capriotti, M. Ouellette, B. King, D. Wisnoski, J. Cox, M. Reilly, R. Marquis, J. Bertin and P. Gough, *Cell Death Discovery*, 2015, **7**.
- 100 K. Weisel, N. E. Scott, D. J. Tompson, B. J. Votta, S. Madhavan, K. Povey, A. Wolstenholme, M. Simeoni, T. Rudo, L. Richards-Peterson, T. Sahota, J. G. Wang, J. Lich, J. Finger, A. Verticelli, M. Reilly, P. J. Gough,

- P. A. Harris, J. Bertin and M.-L. Wang, *Pharmacol Res Perspect*, 2017, **5**, e00365.
- 101 H. Deng, J. Zhou, F. S. Sundersingh, J. Summerfield, D. Somers, J. A. Messer, A. L. Satz, N. Ancellin, C. C. Arico-Muendel, K. L. (Sargent) Bedard, A. Beljean, S. L. Belyanskaya, R. Bingham, S. E. Smith, E. Boursier, P. Carter, P. A. Centrella, M. A. Clark, C. Chung, C. P. Davie, J. L. Delorey, Y. Ding, G. J. Franklin, L. C. Grady, K. Herry, C. Hobbs, C. S. Kollmann, B. A. Morgan, L. J. (Pothier) Kaushansky and Q. Zhou, *ACS Med. Chem. Lett.*, 2015, **6**, 919–924.
- 102 P. Dickson and T. Kodadek, *Org. Biomol. Chem.*, 2019, **17**, 4676–4688.
- 103 R. M. Franzini and C. Randolph, *J. Med. Chem.*, 2016, **59**, 6629–6644.
- 104 D. Neri and R. A. Lerner, *Annu. Rev. Biochem.*, 2018, **87**, 479–502.
- 105 M. L. Malone and B. M. Paegel, *ACS Comb. Sci.*, 2016, **18**, 182–187.
- 106 E. M. Gordon, M. A. Gallop and D. V. Patel, *Acc. Chem. Res.*, 1996, **29**, 144–154.
- 107 D. T. Flood, S. Asai, X. Zhang, J. Wang, L. Yoon, Z. C. Adams, B. C. Dillingham, B. B. Sanchez, J. C. Vantourout, M. E. Flanagan, D. W. Piotrowski, P. Richardson, S. A. Green, R. A. Shenvi, J. S. Chen, P. S. Baran and P. E. Dawson, *J. Am. Chem. Soc.*, 2019, **141**, 9998–10006.
- 108 D. T. Flood, X. Zhang, X. Fu, Z. Zhao, S. Asai, B. B. Sanchez, E. J. Sturgell, J. C. Vantourout, P. Richardson, M. E. Flanagan, D. W. Piotrowski, D. K. Kölmel, J. Wan, M. Tsai, J. S. Chen, P. S. Baran and P. E. Dawson, *Angew. Chem. Int. Ed.*, 2020, **59**, 7377–7383.
- 109 R. M. Franzini, D. Neri and J. Scheuermann, *Acc. Chem. Res.*, 2014, **47**, 1247–1255.
- 110 C. J. Gerry, M. J. Wawer, P. A. Clemons and S. L. Schreiber, *J. Am. Chem. Soc.*, 2019, **141**, 10225–10235.
- 111 A. B. MacConnell, A. K. Price and B. M. Paegel, *ACS Comb. Sci.*, 2017, **19**, 181–192.
- 112 W. G. Cochrane, M. L. Malone, V. Q. Dang, V. Cavett, A. L. Satz and B. M. Paegel, *ACS Comb. Sci.*, 2019, **21**, 425–435.
- 113 C. S. Kollmann, X. Bai, C.-H. Tsai, H. Yang, K. E. Lind, S. R. Skinner, Z. Zhu, D. I. Israel, J. W. Cuzzo, B. A. Morgan, K. Yuki, C. Xie, T. A. Springer, M. Shimaoka and G. Eviendar, *Bioorganic & Medicinal Chemistry*, 2014, **22**, 2353–2365.
- 114 M. Shevlin, *ACS Med. Chem. Lett.*, 2017, **8**, 601–607.
- 115 S. W. Krska, D. A. DiRocco, S. D. Dreher and M. Shevlin, *Acc. Chem. Res.*, 2017, **50**, 2976–2985.
- 116 S. M. Pant, A. Mukonoweshuro, B. Desai, M. K. Ramjee, C. N. Selway, G. J. Tarver, A. G. Wright, K. Birchall, T. M. Chapman, T. A. Tervonen and J. Klefström, *J. Med. Chem.*, 2018, **61**, 4335–4347.
- 117 S. Y. F. Wong Hawkes, M. J. V. Chapela and M. Montembault, *QSAR Comb. Sci.*, 2005, **24**, 712–721.
- 118 J. Wang, G. Sui, V. P. Mocharla, R. J. Lin, M. E. Phelps, H. C. Kolb and H.-R. Tseng, *Angew. Chem. Int. Ed.*, 2006, **45**, 5276–5281.
- 119 A. Baranczak, N. P. Tu, J. Marjanovic, P. A. Searle, A. Vasudevan and S. W. Djuric, *ACS Med. Chem. Lett.*, 2017, **8**, 461–465.

- 120 W. Czechtizky, J. Dedio, B. Desai, K. Dixon, E. Farrant, Q. Feng, T. Morgan, D. M. Parry, M. K. Ramjee, C. N. Selway, T. Schmidt, G. J. Tarver and A. G. Wright, *ACS Med. Chem. Lett.*, 2013, **4**, 768–772.
- 121 M. Werner, C. Kuratli, R. E. Martin, R. Hochstrasser, D. Wechsler, T. Enderle, A. I. Alanine and H. Vogel, *Angew. Chem. Int. Ed.*, 2014, **53**, 1704–1708.
- 122 B. Desai, K. Dixon, E. Farrant, Q. Feng, K. R. Gibson, W. P. van Hoorn, J. Mills, T. Morgan, D. M. Parry, M. K. Ramjee, C. N. Selway, G. J. Tarver, G. Whitlock and A. G. Wright, *J. Med. Chem.*, 2013, **56**, 3033–3047.
- 123 A. Leggott, J. E. Clarke, S. Chow, S. L. Warriner, A. J. O'Neill and A. Nelson, *Chem. Commun.*, 2020, 10.1039.D0CC02361B.
- 124 H.-J. Böhm, A. Flohr and M. Stahl, *Drug Discovery Today: Technologies*, 2004, **1**, 217–224.
- 125 G. Lahav, N. Rosenfeld, A. Sigal, N. Geva-Zatorsky, A. J. Levine, M. B. Elowitz and U. Alon, *Nat Genet*, 2004, **36**, 147–150.
- 126 D. Michael and M. Oren, *Seminars in Cancer Biology*, 2003, **13**, 49–58.
- 127 P. Chène, *Nat Rev Cancer*, 2003, **3**, 102–109.
- 128 M. B. Kastan, Q. Zhan, W. El-Deiry, F. Carrier, T. Jacks, W. V. Walsh, B. S. Plunkett, B. Vogelstein and A. J. Fornace Jr., 1992, **71**, 587–589.
- 129 S. M. Cross, C. A. Sanchez, C. A. Morgan, M. K. Schimke, S. Ramel, R. L. Idzerda, W. H. Raskind and B. J. Reid, *Science*, 1995, **267**, 1353–1356.
- 130 H. Symonds, L. Krall, L. Remington, M. Saenz-Robles, S. Lowe, T. Jacks and T. Van Dyke, *Cell*, 1994, **78**, 703–711.
- 131 L. Cahilly-Snyder, T. Yang-Feng, U. Francke and D. L. George, *Somat Cell Mol Genet*, 1987, **13**, 235–244.
- 132 J. Momand, P. Zambetti, D. C. Olson, D. George and J. Levine, *Cell*, 1992, **69**, 1237–1245.
- 133 D. Oliner, K. W. Kinzler, D. L. George and B. Vogelstein, *Nature*, 1992, **358**, 80–83.
- 134 Y. Haupt, R. Maya, A. Kazaz and M. Oren, *Nature*, 1997, **387**, 296–299.
- 135 C. L. Brooks and W. Gu, *Molecular Cell*, 2006, **21**, 307–315.
- 136 P. H. Kussie, S. Gorina, V. Marechal, B. Elenbaas, J. Moreau, A. J. Levine and N. P. Pavletich, *Science*, 1996, **274**, 948–953.
- 137 C. Li, M. Pazgier, C. Li, W. Yuan, M. Liu, G. Wei, W.-Y. Lu and W. Lu, *Journal of Molecular Biology*, 2010, **398**, 200–213.
- 138 I. Massova and P. A. Kollman, *J. Am. Chem. Soc.*, 1999, **121**, 8133–8143.
- 139 T. Clackson and J. Wells, *Science*, 1995, **267**, 383–386.
- 140 N. London, B. Raveh and O. Schueler-Furman, *Current Opinion in Chemical Biology*, 2013, **17**, 952–959.
- 141 B. S. Zerbe, D. R. Hall, S. Vajda, A. Whitty and D. Kozakov, *J. Chem. Inf. Model.*, 2012, **52**, 2236–2244.
- 142 M. R. Arkin and J. A. Wells, *Nat Rev Drug Discov*, 2004, **3**, 301–317.
- 143 M. R. Arkin, Y. Tang and J. A. Wells, *Chemistry & Biology*, 2014, **21**, 1102–1114.
- 144 M. Pelay-Gimeno, A. Glas, O. Koch and T. N. Grossmann, *Angew. Chem. Int. Ed.*, 2015, **54**, 8896–8927.

- 145 D. E. Scott, A. R. Bayly, C. Abell and J. Skidmore, *Nat Rev Drug Discov*, 2016, **15**, 533–550.
- 146 L.-G. Milroy, T. N. Grossmann, S. Hennig, L. Brunsveld and C. Ottmann, *Chem. Rev.*, 2014, **114**, 4695–4748.
- 147 A. Czarna, B. Beck, S. Srivastava, G. M. Popowicz, S. Wolf, Y. Huang, M. Bista, T. A. Holak and A. Dömling, *Angewandte Chemie International Edition*, 2010, **49**, 5352–5356.
- 148 Y. Zhao, A. Aguilar, D. Bernard and S. Wang, *J. Med. Chem.*, 2015, **58**, 1038–1052.
- 149 B. Vu, P. Wovkulich, G. Pizzolato, A. Lovey, Q. Ding, N. Jiang, J.-J. Liu, C. Zhao, K. Glenn, Y. Wen, C. Tovar, K. Packman, L. Vassilev and B. Graves, *ACS Med. Chem. Lett.*, 2013, **4**, 466–469.
- 150 A. Aguilar, J. Lu, L. Liu, D. Du, D. Bernard, D. McEachern, S. Przybranowski, X. Li, R. Luo, B. Wen, D. Sun, H. Wang, J. Wen, G. Wang, Y. Zhai, M. Guo, D. Yang and S. Wang, *J. Med. Chem.*, 2017, **60**, 2819–2839.
- 151 S. Wang, W. Sun, Y. Zhao, D. McEachern, I. Meaux, C. Barriere, J. A. Stuckey, J. L. Meagher, L. Bai, L. Liu, C. G. Hoffman-Luca, J. Lu, S. Shangary, S. Yu, D. Bernard, A. Aguilar, O. Dos-Santos, L. Besret, S. Guerif, P. Pannier, D. Gorge-Bernat and L. Debussche, *Cancer Research*, 2014, **74**, 5855–5865.
- 152 A. Z. Gonzalez, J. Eksterowicz, M. D. Bartberger, H. P. Beck, J. Canon, A. Chen, D. Chow, J. Duquette, B. M. Fox, J. Fu, X. Huang, J. B. Houze, L. Jin, Y. Li, Z. Li, Y. Ling, M.-C. Lo, A. M. Long, L. R. McGee, J. McIntosh, D. L. McMinn, J. D. Oliner, T. Osgood, Y. Rew, A. Y. Saiki, P. Shaffer, S. Wortman, P. Yakowec, X. Yan, Q. Ye, D. Yu, X. Zhao, J. Zhou, S. H. Olson, J. C. Medina and D. Sun, *J. Med. Chem.*, 2014, **57**, 2472–2488.
- 153 Q. Ding, Z. Zhang, J.-J. Liu, N. Jiang, J. Zhang, T. M. Ross, X.-J. Chu, D. Bartkovitz, F. Podlaski, C. Janson, C. Tovar, Z. M. Filipovic, B. Higgins, K. Glenn, K. Packman, L. T. Vassilev and B. Graves, *J. Med. Chem.*, 2013, **56**, 5979–5983.
- 154 Y. Rew and D. Sun, *J. Med. Chem.*, 2014, **57**, 6332–6341.
- 155 P. Holzer, K. Masuya, P. Furet, J. Kallen, T. Valat-Stachyra, S. Ferretti, J. Berghausen, M. Bouisset-Leonard, N. Buschmann, C. Pissot-Soldermann, C. Rynn, S. Ruetz, S. Stutz, P. Chène, S. Jeay and F. Gessier, *J. Med. Chem.*, 2015, **58**, 6348–6358.
- 156 L. Fazal, M. Ahn, R. Bawn, T. Blackburn, L. Bevan, I. Buck, C. Cano, J. Castro, B. Cons, S. Cully, B. Golding, R. Griffin, I. Hardcastle, K. Hearn, C. Johnson, H. Newell, M. Noble, J. Reeks, E. Tamanini, H. Thomas, H. Walton, E. Willmore, P. Williams, Y. Zhao, S. Wedge and G. Chessari, *AACR Annual Meeting, Chicago*, 2018, poster presentation.
- 157 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.
- 158 D. C. Fry, C. Wartchow, B. Graves, C. Janson, C. Lukacs, U. Kammlott, C. Belunis, S. Palme, C. Klein and B. Vu, *ACS Med. Chem. Lett.*, 2013, **4**, 660–665.
- 159 D. C. Blakemore, L. Castro, I. Churcher, D. C. Rees, A. W. Thomas, D. M. Wilson and A. Wood, *Nature Chem*, 2018, **10**, 383–394.

- 160 J. S. Carey, D. Laffan, C. Thomson and M. T. Williams, *Org. Biomol. Chem.*, 2006, **4**, 2337.
- 161 S. M. Mennen, C. Alhambra, C. L. Allen, M. Barberis, S. Berritt, T. A. Brandt, A. D. Campbell, J. Castañón, A. H. Cherney, M. Christensen, D. B. Damon, J. Eugenio de Diego, S. García-Cerrada, P. García-Losada, R. Haro, J. Janey, D. C. Leitch, L. Li, F. Liu, P. C. Lobben, D. W. C. MacMillan, J. Magano, E. McInturff, S. Monfette, R. J. Post, D. Schultz, B. J. Sitter, J. M. Stevens, I. I. Strambeanu, J. Twilton, K. Wang and M. A. Zajac, *Org. Process Res. Dev.*, 2019, **23**, 1213–1242.
- 162 D. Perera, J. W. Tucker, S. Brahmabhatt, C. J. Helal, A. Chong, W. Farrell, P. Richardson and N. W. Sach, *Science*, 2018, **359**, 429–434.
- 163 C. B. Santiago, J.-Y. Guo and M. S. Sigman, *Chem. Sci.*, 2018, **9**, 2398–2412.
- 164 S. Zhao, T. Gensch, B. Murray, Z. L. Niemeyer, M. S. Sigman and M. R. Biscoe, *Science*, 2018, **362**, 670–674.
- 165 T. Piou, F. Romanov-Michailidis, M. A. Ashley, M. Romanova-Michaelides and T. Rovis, *J. Am. Chem. Soc.*, 2018, **140**, 9587–9593.
- 166 A. F. Zahrt, J. J. Henle, B. T. Rose, Y. Wang, W. T. Darrow and S. E. Denmark, *Science*, 2019, **363**, eaau5631.
- 167 D. J. Durand and N. Fey, *Chem. Rev.*, 2019, **119**, 6561–6594.
- 168 X. Yang, Y. Wang, R. Byrne, G. Schneider and S. Yang, *Chem. Rev.*, 2019, **119**, 10520–10594.
- 169 K. Wu and A. G. Doyle, *Nature Chem*, 2017, **9**, 779–784.
- 170 K. Liao, W. Liu, Z. L. Niemeyer, Z. Ren, J. Bacsá, D. G. Musaev, M. S. Sigman and H. M. L. Davies, *ACS Catal.*, 2018, **8**, 678–682.
- 171 Z. L. Niemeyer, A. Milo, D. P. Hickey and M. S. Sigman, *Nature Chem*, 2016, **8**, 610–617.
- 172 J.-Y. Guo, Y. Minko, C. B. Santiago and M. S. Sigman, *ACS Catal.*, 2017, **7**, 4144–4151.
- 173 C. W. Coley, N. S. Eyke and K. F. Jensen, *Angew. Chem. Int. Ed.*, 2020, anie.201909987.
- 174 T. J. Struble, J. C. Alvarez, S. P. Brown, M. Chytil, J. Cisar, R. L. DesJarlais, O. Engkvist, S. A. Frank, D. R. Greve, D. J. Griffin, X. Hou, J. W. Johannes, C. Kreatsoulas, B. Lahue, M. Mathea, G. Mogk, C. A. Nicolaou, A. D. Palmer, D. J. Price, R. I. Robinson, S. Salentin, L. Xing, T. Jaakkola, William. H. Green, R. Barzilay, C. W. Coley and K. F. Jensen, *J. Med. Chem.*, 2020, acs.jmedchem.9b02120.
- 175 D. V. S. Green, S. Pickett, C. Luscombe, S. Senger, D. Marcus, J. Meslamani, D. Brett, A. Powell and J. Masson, *J Comput Aided Mol Des*, 2020, **34**, 747–765.
- 176 Q. Peng, F. Duarte and R. S. Paton, *Chem. Soc. Rev.*, 2016, **45**, 6093–6107.
- 177 B. D. McLarney, S. Hanna, D. G. Musaev and S. France, *ACS Catal.*, 2019, **9**, 4526–4538.
- 178 N. Fey, A. C. Tshipis, S. E. Harris, J. N. Harvey, A. G. Orpen and R. A. Mansson, *Chem. Eur. J.*, 2006, **12**, 291–302.
- 179 N. Fey, J. N. Harvey, G. C. Lloyd-Jones, P. Murray, A. G. Orpen, R. Osborne and M. Purdie, *Organometallics*, 2008, **27**, 1372–1383.
- 180 J. Jover, N. Fey, J. N. Harvey, G. C. Lloyd-Jones, A. G. Orpen, G. J. J. Owen-Smith, P. Murray, D. R. J. Hose, R. Osborne and M. Purdie, *Organometallics*, 2010, **29**, 6245–6258.

- 181 J. Jover, N. Fey, J. N. Harvey, G. C. Lloyd-Jones, A. G. Orpen, G. J. J. Owen-Smith, P. Murray, D. R. J. Hose, R. Osborne and M. Purdie, *Organometallics*, 2012, **31**, 5302–5306.
- 182 N. Fey, A. Koumi, A. V. Malkov, J. D. Moseley, B. N. Nguyen, S. N. G. Tyler and C. E. Willans, *Dalton Trans.*, 2020, **49**, 8169–8178.
- 183 T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning*, Springer, Second Edition., 2009.
- 184 N. Fey, *Chemistry Central Journal*, 2015, **9**, 38.
- 185 H. M. L. Davies and D. Morton, *Chem. Soc. Rev.*, 2011, **40**, 1857.
- 186 H. M. L. Davies and J. R. Denton, *Chem. Soc. Rev.*, 2009, **38**, 3061.
- 187 A. Ford, H. Miel, A. Ring, C. N. Slattery, A. R. Maguire and M. A. McKerverey, *Chem. Rev.*, 2015, **115**, 9981–10080.
- 188 D. Gillingham and N. Fei, *Chem. Soc. Rev.*, 2013, **42**, 4918.
- 189 M. P. Doyle and D. C. Forbes, *Chem. Rev.*, 1998, **98**, 911–936.
- 190 A. Padwa and M. D. Weingarten, *Chem. Rev.*, 1996, **96**, 223–270.
- 191 A. Padwa, D. J. Austin, A. T. Price, M. A. Semones, M. P. Doyle, M. N. Protopopova, W. R. Winchester and A. Tran, *J. Am. Chem. Soc.*, 1993, **115**, 8669–8680.
- 192 A. Padwa and D. J. Austin, *Angewandte Chemie International Edition*, 1994, **33**, 1797–1815.
- 193 S. Miah, A. M. Z. Slawin, C. J. Moody, S. M. Sheehan, J. P. Marino, M. A. Semones, A. Padwa and I. C. Richards, *Tetrahedron*, 1996, **52**, 2489–2514.
- 194 K. Liao, S. Negretti, D. G. Musaev, J. Bacsá and H. M. L. Davies, *Nature*, 2016, **533**, 230–234.
- 195 K. Liao, T. C. Pickel, V. Boyarskikh, J. Bacsá, D. G. Musaev and H. M. L. Davies, *Nature*, 2017, **551**, 609–613.
- 196 C. Werlé, R. Goddard, P. Philipps, C. Farès and A. Fürstner, *J. Am. Chem. Soc.*, 2016, **138**, 3797–3805.
- 197 H. M. L. Davies, T. Hansen and M. R. Churchill, *J. Am. Chem. Soc.*, 2000, **122**, 3063–3070.
- 198 K. Liao, Y.-F. Yang, Y. Li, J. N. Sanders, K. N. Houk, D. G. Musaev and H. M. L. Davies, *Nature Chem*, 2018, **10**, 1048–1055.
- 199 T. Gensch, M. N. Hopkinson, F. Glorius and J. Wencel-Delord, *Chem. Soc. Rev.*, 2016, **45**, 2900–2936.
- 200 H. M. L. Davies, *J. Org. Chem.*, 2019, **84**, 12722–12745.
- 201 E. N. Bess, D. M. Guptill, H. M. L. Davies and M. S. Sigman, *Chem. Sci.*, 2015, **6**, 3057–3062.
- 202 S. R. Hare and D. J. Tantillo, *Chem. Sci.*, 2017, **8**, 1442–1449.
- 203 A. D. Becke, *The Journal of Chemical Physics*, 1993, **98**, 1372–1377.
- 204 C. Lee, W. Yang and R. G. Parr, *Physical Review B*, 1988, **37**, 785–789.
- 205 S. H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.*, 1980, **58**, 1200–1211.
- 206 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.
- 207 M. J. Frisch, J. A. Pople and J. S. Binkley, *The Journal of Chemical Physics*, 1984, **80**, 3265–3269.
- 208 P. C. Hariharan and J. A. Pople, *Theoret. Chim. Acta*, 1973, **28**, 213–222.

- 209 D. Andrae, U. Haussermann, M. Dolg, H. Stoll and H. Preuss, *Theoret. Chim. Acta*, 1990, **77**, 123–141.
- 210 M. C. Pirrung, H. Liu and A. T. Morehead, *J. Am. Chem. Soc.*, 2002, **124**, 1014–1023.
- 211 M. P. Doyle, *J. Org. Chem.*, 2006, **71**, 9253–9260.
- 212 M. P. Doyle, W. R. Winchester, J. A. A. Hoorn, V. Lynch, S. H. Simonsen and R. Ghosh, *J. Am. Chem. Soc.*, 1993, **115**, 9968–9978.
- 213 A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.
- 214 J. P. Perdew, *Phys. Rev. B*, 1986, **33**, 8822–8824.
- 215 S. Aguado-Ullate, S. Saureu, L. Guasch and J. J. Carbó, *Chem. Eur. J.*, 2012, **18**, 995–1005.
- 216 S. Aguado-Ullate, M. Urbano-Cuadrado, I. Villalba, E. Pires, J. I. García, C. Bo and J. J. Carbó, *Chem. Eur. J.*, 2012, **18**, 14026–14036.
- 217 H. M. L. Davies and L. Rusiniak, *Tetrahedron Letters*, 1998, **39**, 8811–8812.
- 218 D. T. Nowlan, T. M. Gregg, H. M. L. Davies and D. A. Singleton, *J. Am. Chem. Soc.*, 2003, **125**, 15902–15911.
- 219 D. J. Miller and C. J. Moody, *Tetrahedron*, 1995, **51**, 10811–10843.
- 220 Z. Yu, B. Ma, M. Chen, H.-H. Wu, L. Liu and J. Zhang, *J. Am. Chem. Soc.*, 2014, **136**, 6904–6907.
- 221 S. D. Dreher, *React. Chem. Eng.*, 2019, **4**, 1530–1535.
- 222 J. J. Henle, A. F. Zahrt, B. T. Rose, W. T. Darrow, Y. Wang and S. E. Denmark, *J. Am. Chem. Soc.*, 2020, **142**, 11578–11592.
- 223 G. A. Landrum, RDKit: Open-source cheminformatics, <http://www.rdkit.org>, (accessed 27 July 2020).
- 224 J. B. Baell and G. A. Holloway, *J. Med. Chem.*, 2010, **53**, 2719–2740.
- 225 J. B. Baell and J. W. M. Nissink, *ACS Chem. Biol.*, 2018, **13**, 36–44.
- 226 H. M. L. Davies and S. J. Hedley, *Chem. Soc. Rev.*, 2007, **36**, 1109.
- 227 H. M. L. Davies, T. Hansen and M. R. Churchill, *J. Am. Chem. Soc.*, 2000, **122**, 3063–3070.
- 228 T. Ye and M. A. McKerverey, *Chem. Rev.*, 1994, **94**, 1091–1160.
- 229 S. Chow, A. I. Green, C. Arter, S. Liver, A. Leggott, L. Trask, G. Karageorgis, S. Warriner and A. Nelson, *Synthesis*, 2020, 12.
- 230 M. Regitz, *Angew. Chem. Int. Ed. Engl.*, 1967, **6**, 733–749.
- 231 J. S. Baum, D. A. Shook, H. M. L. Davies and H. D. Smith, *Synthetic Communications*, 1987, **17**, 1709–1716.
- 232 J. B. Hendrickson and W. A. Wolf, *J. Org. Chem.*, 1968, **33**, 3610–3618.
- 233 M. H. A. Roehrl, J. Y. Wang and G. Wagner, *Biochemistry*, 2004, **43**, 16056–16066.
- 234 M. D. Hall, A. Yasgar, T. Peryea, J. C. Braisted, A. Jadhav, A. Simeonov and N. P. Coussens, *Methods Appl. Fluoresc.*, 2016, **4**, 022001.
- 235 G. Weber, *Biochemical Journal*, 1952, **51**, 145–155.
- 236 A. Pope, *Drug Discovery Today*, 1999, **4**, 350–362.
- 237 G. Weber, *The Journal of Chemical Physics*, 1971, **55**, 2399–2407.
- 238 J. P. Plante, T. Burnley, B. Malkova, M. E. Webb, S. L. Warriner, T. A. Edwards and A. J. Wilson, *Chem. Commun.*, 2009, 5091.
- 239 R. Fasan, R. L. A. Dias, K. Moehle, O. Zerbe, J. W. Vrijbloed, D. Obrecht and J. A. Robinson, *Angew. Chem. Int. Ed.*, 2004, **43**, 2109–2112.

- 240 B. Neises and W. Steglich, *Angew. Chem. Int. Ed. Engl.*, 1978, **17**, 522–524.
- 241 A. E. Tron, M. A. Belmonte, A. Adam, B. M. Aquila, L. H. Boise, E. Chiarparin, J. Cidado, K. J. Embrey, E. Gangl, F. D. Gibbons, G. P. Gregory, D. Hargreaves, J. A. Hendricks, J. W. Johannes, R. W. Johnstone, S. L. Kazmirski, J. G. Kettle, M. L. Lamb, S. M. Matulis, A. K. Nooka, M. J. Packer, B. Peng, P. B. Rawlins, D. W. Robbins, A. G. Schuller, N. Su, W. Yang, Q. Ye, X. Zheng, J. P. Secrist, E. A. Clark, D. M. Wilson, S. E. Fawell and A. W. Hird, *Nat Commun*, 2018, **9**, 5341.
- 242 M. D. Hall, A. Yasgar, T. Peryea, J. C. Braisted, A. Jadhav, A. Simeonov and N. P. Coussens, *Methods Appl. Fluoresc.*, 2016, **4**, 022001.
- 243 M. P. Williamson, *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2013, **73**, 1–16.
- 244 S. Uhrinova, D. Uhrin, H. Powers, K. Watt, D. Zheleva, P. Fischer, C. McInnes and P. N. Barlow, *Journal of Molecular Biology*, 2005, **350**, 587–598.
- 245 I. R. Kleckner and M. P. Foster, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 2011, **1814**, 942–968.
- 246 G. Vieira de Castro and A. Ciulli, *Chem. Commun.*, 2019, **55**, 1482–1485.
- 247 K. L. Richards, M. L. Rowe, P. B. Hudson, R. A. Williamson and M. J. Howard, *Sci Rep*, 2016, **6**, 19518.
- 248 Y. Ayotte, V. M. Marando, L. Vaillancourt, P. Bouchard, G. Heffron, P. W. Coote, S. T. Larda and S. R. LaPlante, *J. Med. Chem.*, 2019, **62**, 7885–7896.
- 249 A. N. Ganesh, E. N. Donders, B. K. Shoichet and M. S. Shoichet, *Nano Today*, 2018, **19**, 188–200.
- 250 O. Roche, P. Schneider, J. Zuegge, W. Guba, M. Kansy, A. Alanine, K. Bleicher, F. Danel, E.-M. Gutknecht, M. Rogers-Evans, W. Neidhart, H. Stalder, M. Dillon, E. Sjögren, N. Fotouhi, P. Gillespie, R. Goodnow, W. Harris, P. Jones, M. Taniguchi, S. Tsujii, W. von der Saal, G. Zimmermann and G. Schneider, *J. Med. Chem.*, 2002, **45**, 137–142.
- 251 B. Y. Feng, A. Simeonov, A. Jadhav, K. Babaoglu, J. Inglese, B. K. Shoichet and C. P. Austin, *J. Med. Chem.*, 2007, **50**, 2385–2390.
- 252 S. L. McGovern, B. T. Helfand, B. Feng and B. K. Shoichet, *J. Med. Chem.*, 2003, **46**, 4265–4272.
- 253 B. Graves, T. Thompson, M. Xia, C. Janson, C. Lukacs, D. Deo, P. Di Lello, D. Fry, C. Garvie, K.-S. Huang, L. Gao, C. Tovar, A. Lovey, J. Wanner and L. T. Vassilev, *Proceedings of the National Academy of Sciences*, 2012, **109**, 11788–11793.
- 254 B. Meyer and T. Peters, *Angew. Chem. Int. Ed.*, 2003, **42**, 864–890.
- 255 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2009, NA-NA.
- 256 *Molecular Operating Environment (MOE)*, University of Sherbrooke, Chemical Computing Group, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2019.
- 257 A. M. Wassermann, M. Wawer and J. Bajorath, *J. Med. Chem.*, 2010, **53**, 8209–8223.
- 258 ChEMBL, <https://www.ebi.ac.uk/chembl/>, (accessed 16 January 2020).
- 259 WO 2017069289 A1, 2015.

- 260 US 8,962,611 B2, 2013.
- 261 D. R. Flower, *J. Chem. Inf. Comput. Sci.*, 1998, **38**, 379–386.
- 262 G. Maggiora, M. Vogt, D. Stumpfe and J. Bajorath, *J. Med. Chem.*, 2014, **57**, 3186–3204.
- 263 B. Lamoree and R. E. Hubbard, *Essays in Biochemistry*, 2017, **61**, 453–464.
- 264 S. A. Hudson, S. Surade, A. G. Coyne, K. J. McLean, D. Leys, A. W. Munro and C. Abell, *ChemMedChem*, 2013, **8**, 1451–1456.
- 265 G. Schneider, W. Neidhart, T. Giller and G. Schmid, 3.
- 266 Y. Hu, D. Stumpfe and J. Bajorath, *J. Med. Chem.*, 2017, **60**, 1238–1246.
- 267 F. Gonzalez-Lopez de Turiso, D. Sun, Y. Rew, M. D. Bartberger, H. P. Beck, J. Canon, A. Chen, D. Chow, T. L. Correll, X. Huang, L. D. Julian, F. Kayser, M.-C. Lo, A. M. Long, D. McMinn, J. D. Oliner, T. Osgood, J. P. Powers, A. Y. Saiki, S. Schneider, P. Shaffer, S.-H. Xiao, P. Yakowec, X. Yan, Q. Ye, D. Yu, X. Zhao, J. Zhou, J. C. Medina and S. H. Olson, *J. Med. Chem.*, 2013, **56**, 4053–4070.
- 268 P. Furet, K. Masuya, J. Kallen, T. Stachyra-Valat, S. Ruetz, V. Guagnano, P. Holzer, R. Mah, S. Stutz, A. Vaupel, P. Chène, S. Jeay and A. Schlapbach, *Bioorganic & Medicinal Chemistry Letters*, 2016, **26**, 4837–4841.
- 269 T. Cernak, K. D. Dykstra, S. Tyagarajan, P. Vachal and S. W. Krska, *Chem. Soc. Rev.*, 2016, **45**, 546–576.
- 270 J. Kallen, A. Izaac, S. Chau, E. Wirth, J. Schoepfer, R. Mah, A. Schlapbach, S. Stutz, A. Vaupel, V. Guagnano, K. Masuya, T. Stachyra, B. Salem, P. Chene, F. Gessier, P. Holzer and P. Furet, *ChemMedChem*, 2019, **14**, 1305–1314.
- 271 M. Doyle, W. Winchester, M. N. Protopopova, A. P. Kazala and L. J. Westrum, *Organic Syntheses*, 1998, **9**, 322.
- 272 H. M. L. Davies, T. Hansen and M. R. Churchill, *Journal of the American Chemical Society*, 2000, **122**, 3063–3070.
- 273 H. M. L. Davies and L. Rusiniak, *Tetrahedron Letters*, 1998, **39**, 8811–8812.
- 274 J. L. Thompson and H. M. L. Davies, *J. Am. Chem. Soc.*, 2007, **129**, 6090–6091.
- 275 Y. Liu, Z. Yu, J. Z. Zhang, L. Liu, F. Xia and J. Zhang, *Chemical Science*, 2016, **7**, 1988–1995.
- 276 P. Müller and D. Fernandez, *HCA*, 1995, **78**, 947–958.
- 277 H. Nakatsuji, J. L. Sonnenberg, T. Nakajima, F. Ogliaro, J. Normand, N. Rega, R. Gomperts, R. L. Martin, S. Dapprich and I. Gaussian, *Gaussian09, D.01*, 2009.
- 278 R. Ditchfield, W. J. Hehre and J. A. Pople, *The Journal of Chemical Physics*, 1971, **54**, 724–728.
- 279 W. J. Hehre, R. Ditchfield and J. A. Pople, *The Journal of Chemical Physics*, 1972, **56**, 2257–2261.
- 280 M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees and J. A. Pople, *The Journal of Chemical Physics*, 1982, **77**, 3654–3665.
- 281 K. Gilbert, *PCModel*, Bloomington, IN, 2004.
- 282 A. V. Brethomé, S. P. Fletcher and R. S. Paton, *ACS Catal.*, 2019, **9**, 2313–2323.

- 283 W.-W. Chan, T.-L. Kwong and W.-Y. Yu, *Org. Biomol. Chem.*, 2012, **10**, 3749.
- 284 F. Varano, D. Catarzi, V. Colotta, F. R. Calabri, O. Lenzi, G. Filacchioni, A. Galli, C. Costagli, F. Deflorian and S. Moro, *Bioorganic & Medicinal Chemistry*, 2005, **13**, 5536–5549.
- 285 T. D. Goddard and D. G. Kneller, *SPARKY 3*, University of California, 1997.
- 286 D. Baldisseri, *Practical Aspects of Fragment-Based Screening Experiments in TopSpin*, Bruker BioSpin, 2016.

Appendix A

Establishing the p53/*hDM2* Fluorescence Anisotropy Assay for Photoredox Reaction Conditions

Each component required in a photoredox Minisci reaction was tested in the established p53₁₅₋₃₁ Flu/*hDM2*₁₇₋₁₂₅ fluorescence anisotropy assay to check for assay interference. All concentrations tested for the catalysts/catalytic systems were tolerated across the entire range of the titration experiment up to 0.1-1 μ M concentrations (Figure A1), giving an estimated maximum assay window between 1-100 μ M for ADS reaction screening. Potentially higher concentrations can be used with the (Ir[dF(CF₃)ppy]₂(dtbpy))PF₆ and Acridinium photocatalysts (i.e. above 100 μ M) due to low residual emission intensities under the assay conditions (Figure A2).

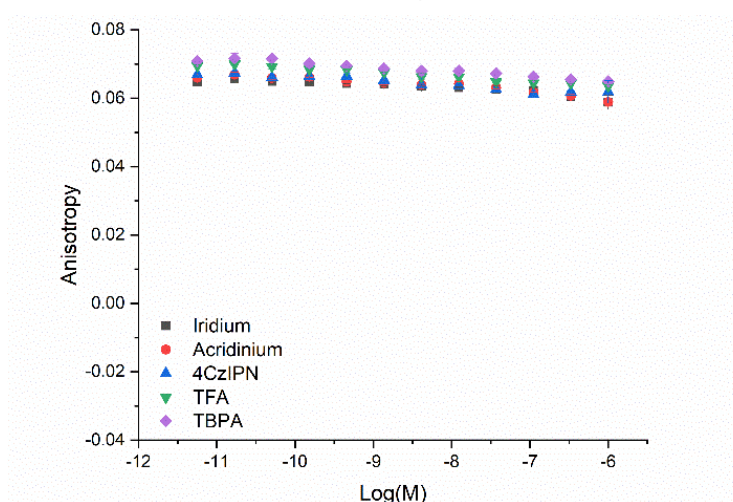


Figure A1. Dose-response of the individual photoredox Minisci reaction components, with 150 nM MDM2, 25 nM p53-tracer and pH 7.4 Tris buffer containing 50 mM Tris, 150 mM NaCl and 0.1% Triton X-100. IC₅₀ values between 45 – 140 nM and no time-dependent change in inhibition was observed.

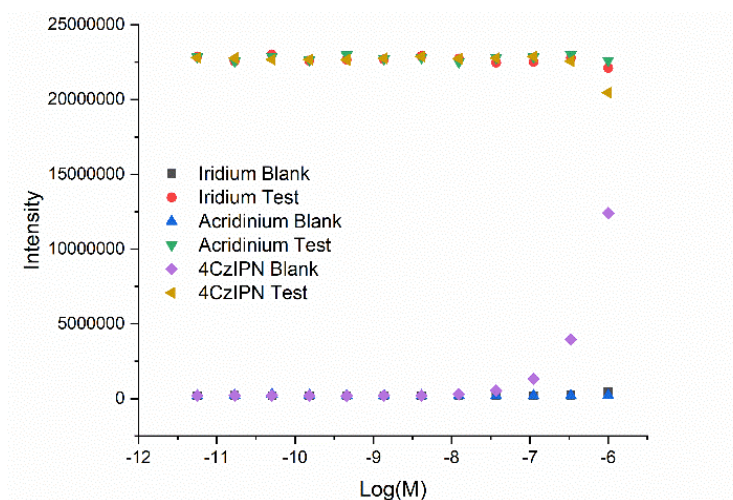


Figure A2. Measured fluorescent intensities for the photoredox catalysts in dose-response. 4CzIPN shows clear interference at 1 and 0.3 μM concentrations, but rapidly diminishes.

A dose-response titration of Nutlin-3a, identical to the procedure described in Chapter 5.3.5.2, was undertaken with 0, 1 and 2 molar equivalents of trifluoroacetic acid (TFA) to investigate if the Tris-Triton X-100 pH 7.4 buffer system could be effected by TFA. Each titration experiment gave similar results indicating that residual TFA would be tolerated when screening ADS reaction mixtures.

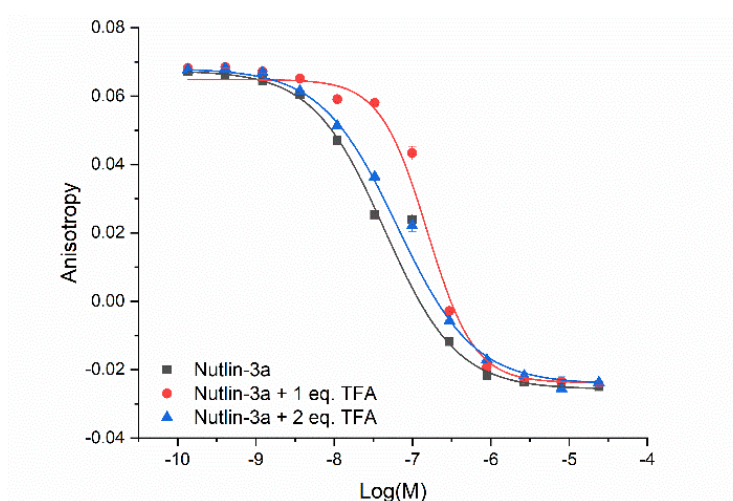


Figure 2. Dose-response of Nutlin-3a with zero, one or two equivalents of TFA, with 150 nM MDM2, 25 nM p53-tracer and pH 7.4 Tris buffer containing 50 mM Tris, 150 mM NaCl and 0.1% Triton X-100. IC_{50} values between 45 – 140 nM and no time-dependent change in inhibition was observed.