Targeting the Ras Superfamily via their Regulatory GEF Complexes

Janine Gray

St John's College

Supervisors: Professors Paul Brennan and Frank von Delft

A thesis submitted for the degree of

Doctor of Philosophy in Clinical Medicine

Nuffield Department of Medicine, University of Oxford

Diamond Light Source, UK

Hilary 2021

Declaration

The work presented in this thesis is my own except where explicitly stated otherwise in the text. No parts of this thesis have been submitted for any other degree.

Janine Gray

St John's College

Abstract

The Ras superfamily of small GTPases are guanine nucleotide-dependent molecular switches which are involved in numerous signalling pathways in the cell. Mutations or dysregulation of these proteins are associated with a variety of diseases, and so the small GTPases are important targets in drug discovery. Extensive medicinal chemistry efforts have largely been unsuccessful due to the difficulty in developing inhibitors to compete with GDP or GTP in an analogous manner to the ATP-competitive inhibitors of kinases. The activation of the small GTPases is regulated by guanine nucleotide exchange factors (GEFs). This thesis describes the efforts to identify small molecules which could be developed into novel inhibitors of the small GTPases by targeting their regulatory GEF complexes.

The second chapter details a kinetic analysis of the GEF-catalysed activation cycle to convert an inactive GDP-bound GTPase into its active GTP-bound form. The modelling of a competitive inhibitor of a GEF/GTPase complex within this cycle supported the hypothesis that a GEF/GTPase inhibitor could slow the rate of nucleotide exchange in a manner not achieved by an archetypal competitive inhibitor which binds to the GTPase alone.

The following chapter presents the cloning, protein production, and crystallisation efforts of over 50 proteins to develop a pipeline within the Structural Genetics Consortium (SGC). The chapter reports the determination of three novel crystal structures, including the GEF Kalirin and its small GTPase Rac1, which was chosen as a therapeutic target for compound screening due to its role in neurodegenerative diseases and suitability for X-ray crystallography ligand screening.

Chapter 4 describes the high-throughput X-ray crystallography fragment screening conducted on Kalirin/Rac1 and the subsequent identification of an isoxazole-urea hit fragment within the canonical binding site. Computational docking of purchased analogues and the synthesis of bioisosteres is also reported, as well as the development of a biochemical

fluorescence-based assay to detect compounds which inhibit the nucleotide exchange reaction.

Electrophilic covalent fragments were also found to bind to the guanine nucleotide pocket of Kalirin/Rac1. Determination of their co-crystal structures, and subsequent ongoing efforts to synthesise novel and merged fragments are detailed in Chapter 5, including a lead compound which demonstrated potency in the biochemical assay.

The final chapter delineates a virtual screening effort to discover new scaffolds with similar pharmacophores which bind within the GDP/GTP binding pocket of Kalirin/Rac1 and the consequent synthesis of non-covalent and covalent analogues arising from the identification of a potent non-covalent fragment.

Overall, this thesis proposes several compounds as suitable vectors for the development of probes or therapeutics for Kalirin/Rac1, as well as providing the groundwork for the expression, structural determination, and compound screening of other members of the Ras superfamily. Future work arising from this thesis, including the development of lead compounds, establishment of cellular assays, and the progression of other GEF/GTPase complexes is also considered.

Acknowledgments

First and foremost, I would like to thank both of my supervisors Paul Brennan and Frank von Delft for giving me the amazing opportunity to work on this project in their groups. I am incredibly grateful not only for the teaching and support you both gave me, but also for allowing me the freedom to explore different areas of the project I found of particular interest. I would also like to thank the Nuffield Department of Medicine and Diamond Light Source for their funding.

I would like to say a special thanks to all the other members of Team GEF: to Carmen, for teaching me how to be a better scientist and your sunny optimism; to Andrew, for bravely taking over the protein pipeline; and to Mia, for continuing with the project now that I have finished. I would also like to thank our collaborators in Germany and Israel for their hard work and assistance during this project.

They say raising a child takes a village; in this case, it took two, and so I would like to extend a massive thank you to both the past and present PX and MedChem members for your advice, encouragement and outside-the-lab excursions over the last four years. Thanks in particular must go to Mike, Romain, and Tobias for endeavouring to teach me the basics of biochemistry and crystallography, as well as the XChem team at Diamond for always managing to fit in my last minute requests for beamtime with aplomb. Many thanks to Raina for teaching me chemistry basically from scratch, as well as the rest of the MedChem group for answering my constant questions with incredible patience. Particular thanks to the brave souls who agreed to read my thesis - Carmen, Andrew, Lizbe, Mike, and Tryfon – as well as Elliot, Rachael, Raina, Hollie, Aidan, and Mary who carried out certain experiments within the project. And finally, thanks also to the rest of the SGC for their help in the lab, in particular to Jesse for his help in statistical analysis and interesting discussions.

Thanks must also go to those outside of the lab who have helped to keep me going: to Nisha, Abby and Sophie, who have been with me before I even realised I liked science; to Lara, Liz, Katie, Zahra, Emma, and Clare without whom I never would gotten past undergraduate level, never mind a DPhil; and to Rose, Melis and the rest of the St John's group for making the DPhil an incredible experience.

To Stephen, for tirelessly providing me with meals that are not on toast, for constantly supporting me during my DPhil, and for going with me on adventures.

To my sisters, Siobhan and Heather, for encouraging me to enjoy my time off with various escapades, for always being there for me and for both of your abilities to quote animated movies.

And finally, to my parents, Eddie and Kathryn, for your constant love and encouragement, from tough love - "Type, monkey, type!" - to helping me revise for exams by going to watch Star Trek at the cinema. I have no doubt that without your support – from the Gray patented method of writing presentations, to home schooling between moving countries – there is a chance I wouldn't have even studied science, let alone attempted a DPhil. This thesis is dedicated to you both.

Abbreviations

Å	Angstroms
ACE-Cl	1-Chloroethyl chloroformate
ADMET	Absorption, Distribution, Metabolism and Excretion
ADR	Adverse Drug Reaction
AKAP13	A-Kinase Anchoring Protein 13
AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
anhyd.	anhydrous
ANOVA	Analysis of Variance
AP	Alkaline phosphatase
Arf	ADP-ribosylation factor
ARHGEF	Rho Guanine Nucleotide Exchange Factor
ATP	Adenosine triphosphate
BCL-xL	B-cell lymphoma-extra large
BET	Bromo- and Extra-Terminal
BI	Boehringer Ingelheim
Bn	benzyl
Boc	*tert*-Butyloxycarbonyl
BODIPY-GTP	2'-(or-3')-O-(N-(2-Aminoethyl) Urethane guanosine triphosphate
BOP	benzotriazol-1-yloxytris(dimethylamino)phosphonium hexafluorophosphate
BRAF	B-Rapidly Accelerated Fibrosarcoma
BRD9	Bromodomain-containing protein 9
Bu	Butyl
Cdc25	Cell division cycle
CDI	Carbonyldiimidazole
CDK2	Cyclin-dependent Kinase 2
CETSA	Cellular Thermal Shift Assay
cLogP	Calculated logarithm of the partition coefficient
CMD	Centre for Medicines Discovery
CNS	Central Nervous System
COX	Cyclo-oxygenase
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats

Cryo-EM	Cryogenic Electron Microscopy
Cyth	Cytohesin
Da	Daltons
DABAL-Me3	Bis(trimethylaluminum)-1,4-diazabicyclo[2.2.2]octane adduct
Dbl	Diffuse B-cell lymphoma
Dbs	Dbl's big sister
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DCE	1,2-Dichloroethane
DCM	Dichloromethane
DEF6	Differentially expressed in FDCP 6
DENND	Differentially expressed in normal and neoplastic cells
DENND1B	DENN domain containing protein 1B
DH	Dbl-homology
DHR	DOCK-homology region
DIPEA	*N, N*-Diisopropylethylamine
DISC1	Disrupted-in-Schizophrenia 1
DLS	Diamond Light Source
DMA	Dimethylacetamide
DMAP	4-Dimethylaminopyridine
DMF	*N,N*-Dimethylformamide
DMF-DMA	*N,N*-Dimethylformamide dimethyl acetal
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DOCK	Dedicator of cytokinesis
DPPE	1,2-Bis(diphenylphosphino)ethane
DSF	Differential Scanning Fluorimetry
DSI	Diamond, SGC, iNEXT
DtBAD	Di-*tert*-butyl azodicarboxylate
EDCI	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	Ethylenediaminetetraacetic acid
EPAC2	Exchange Protein directly activated by cAMP2
eq	equivalents
EtOH	Ethanol
FBLD	Fragment-based lead discovery
FDA	Food and Drug Administration

FI	Fluorescence Intensity
FGI	Functional Group Interconversion
FKBP	FK506 binding protein
FTI	Farnesyltransferase inhibitors
G	Gibbs free energy
GAP	GTPase activating protein
GDI	Guanine nucleotide dissociation inhibitor
GDP	Guanosine diphosphate
GE	Group efficiency
GEF	Guanine nucleotide exchange factor
GGTI	Geranylgeranyltransferase inhibitor
GMP	Guanosine monophosphate
GppCHp	Guanosine-5'-[(β,γ)-methylene]triphosphate
GppNHp	Guanosine-5'-[(β,γ)-imido]triphosphate
GTP	Guanosine triphosphate
HAP	Huntingdon-associated protein
HATU	1-[Bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxid hexafluorophosphate
HDX	Hydrogen deuterium exchange
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
hERG	Human Ether á-go-go related gene
His	Histidine
HMDS	Bis(trimethylsilyl)amide
HPLC	High performance liquid chromatography
HRas	Harvey Ras
HTS	High-throughput screening
IC_{50}	Half maximal inhibitory concentration
ICEKAT	Interactive Continuous Enzyme Kinetics Analysis Tool
IDP	Intrinsically Disordered Proteins
IG	Immunoglobulin-like domain
IMAC	Immobilised Metal Affinity Chromatography
IMP	Integral membrane proteins
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IQSEC	IQ motif and Sec7 domain-containing protein
ITC	Isothermal Titration Calorimetry

ITSN1	Intersectin 1
K_D	Dissociation constant
KRas	Kirsten Rat sarcoma
L	Litre
LCMS	Liquid chromatography mass spectrometry
LE	Ligand Efficiency
LIC	Ligation Independent Cloning
LLE_{AT}	Lipophilic ligand efficiency
M	Molar
MAD	Multiple wavelength anomalous dispersion
mant-GDP	2'/3'-O-(N-Methyl-anthraniloyl)-guanosine-5'-diphosphate
MBP	Maltose Binding Protein
Me	Methyl
MeCN	Acetonitrile
MeOH	Methanol
MOA	Mechanism of action
MR	Molecular replacement
MS	Mass spectrometry
MSS4	Mammalian Suppressor of Sec4
MST	Microscale thermophoresis
MW	Microwave
M_w	Molecular weight
NET1	Neuroepithelial cell-transforming 1
NHA	Non-hydrogen atoms
NHS	*N*-Hydroxysuccinamide
nM	nanomolar
NMDA	N-methyl-D-aspartate
NMR	Nuclear magnetic resonance
NRas	Neuroblastoma Ras
NUDT	Nudix hydrolase
PAINs	Pan-Assay Interference compounds
PAK	p-21 activated kinases
PanDDA	Pan-Dataset Density Analysis
PDB	Protein Data Bank
PDE	Phosphodiesterase

PEG	Polyethylene glycol
PH	Pleckstrin Homology
Phth	Phthalimide
pIC50	$-\log_{10}(IC_{50})$
pM	picomolar
PPI	Protein-protein interactions
P-Rex1	Phosphatidylinositol-3,4,5-Trisphosphate Dependent Rac Exchange Factor 1
PROTACs	Proteolysis Targeting Chimeras
PSA	Polar Surface Area
Py	Pyridine
QTOF	Quantitative Time of Flight
Rab	Ras-related in brain
Rabex-5	Rabaptin-5-Associated exchange factor for Rab5
Rac1	Ras-related C3 botulinum toxin substrate 1
RalA	Ras-like protein A
Ran	Ras-like nuclear
Rap1B	Ras-related protein 1B
Ras	Rat sarcoma
RASGRF1	Ras-specific Guanine Nucleotide-Releasing Factor 1
RCC1	Regulator of Chromosome Condensation 1
RGL2	Ral Guanine Nucleotide Dissociation Stimulator Like 2
Rho	Ras homologous
Rit2	Ras-like without CAAX 2
RMSD	Root-mean-square-deviation
RNAi	Ribonucleic acid interference
Ro3	Rule of Three
RSCC	Real-space correlation coefficient
RSRZ	Real-Space R-value Z-score
r.t.	room temperature
SAD	Single wavelength anomalous dispersion
SAR	Structure-activity relationships
SBDD	Structure-based drug design
SCX	Strong cation exchange
SEC	Size exclusion chromatography

SGC	Structural Genetics Consortium
SH3	Src homology 3
siRNA	Small interfering Ribonucleic acid
S_NAr	Substitution nucleophilic aromatic
S_N2	Substitution nucleophilic bimolecular
Sos	Son of Sevenless
SPR	Surface Plasmon Resonance
STD	Saturation Transfer Difference
Strep	Streptavidin
t	time
tert-	tertiary
T3P®	Propylphosphonic anhydride
TB	Terrific broth
TBDMS	*tert*-Butyldimethylsilyl
TCI	Targeted covalent inhibitors
TCDI	Thiocarbonyldiimidazole
TEA	Triethylamine
TEP	Target Enabling Package
TFA	Trifluoroacetic acid
TFAA	Trifluoroacetic anhydride
THF	Tetrahydrofuran
Tiam1	T-cell lymphoma invasion and metastasis-inducing protein 1
TLC	Thin Layer Chromatography
TPP	Triphenylphosphine
TRAPP	Transport Protein Particle
Tris	Tris(hydroxymethyl)aminomethane
Trio	Triple functional domain protein
TSA	Thermal Shift Assays
WaterLOGSY	Water-Ligand Observed *via* Gradient Spectroscopy

Table of Contents

1. Introduction

1.1 Overview

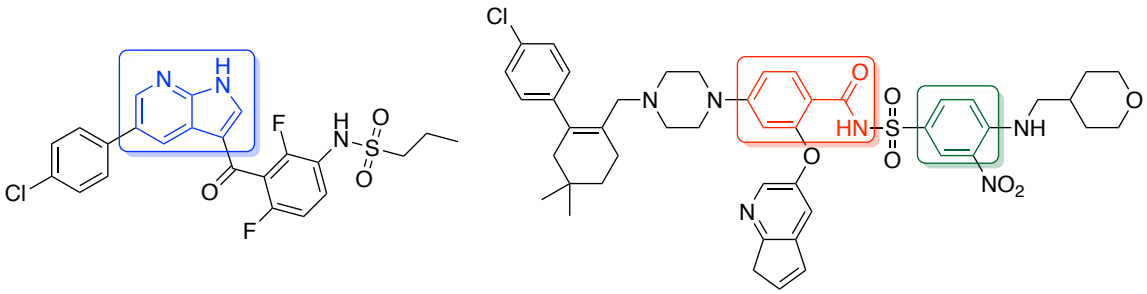
This introduction is split into three sections. The first third discusses several aspects of the drug discovery process relevant to this thesis, including fragment-based lead discovery, the development of chemical probes and their use in target validation for therapeutic development, the identification of covalent compounds to be used as probes or drugs, and the process of X-ray crystallography and its use in structure-based ligand design. The middle section details the Ras superfamily, their roles in the cell, their regulation by GEF proteins, and the attempts to date to target these proteins using small molecule inhibitors. The latter section details the specific hypothesis behind this thesis and outlines the project proposal.

1.2 Drug Discovery

1.2.1 Fragment-Based Lead Discovery

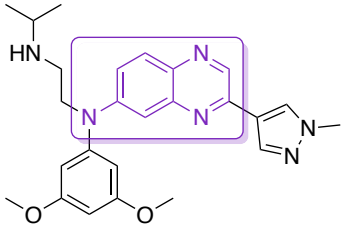
1.2.1.1 FBLD vs HTS

In the past 20 years, the low success rate (45-55%) of discovering compounds from high-throughput screening (HTS) strategies has resulted in the emergence of fragment-based lead discovery (FBLD) as an alternative for the development of pharmaceuticals.[1,2] The objective of FBLD is to synthesise potent drugs by combining or elaborating low molecular, weakly binding compounds, termed 'fragments'. This method, originating from the theoretical work of Jencks in 1981,[3] has been validated in recent years with the FDA approval of three drugs derived from FBLD screening campaigns, with at least a further 40 advancing into clinical trials (Figure 1).[4]



Vemurafenib (**1**)

Venetoclax (**2**)



Erdafitinib (**3**)

Figure 1. The FDA-approved drugs Vemurafenib (**1**), Venetoclax (**2**), and Erdafitinib (**3**) to treat BRAF-mutated metastatic melanoma, chronic lymphocytic leukaemia, and bladder cancer respectively. The lead fragment(s) identified by fragment screening are highlighted accordingly.[5–9]

FBLD, with the use of smaller compounds, yields a number of advantages over HTS. Firstly, despite being weakly potent, their small size results in fragments being able to probe macromolecular space more efficiently than the large compounds found in HTS libraries. Molecules can form both favourable and unfavourable interactions; the larger the compound, the higher the likelihood that unfavourable interactions or steric clashing arises, which can lead to lower hit rates (Figure 2). Fragments generally bind with low affinity but high-quality interactions.

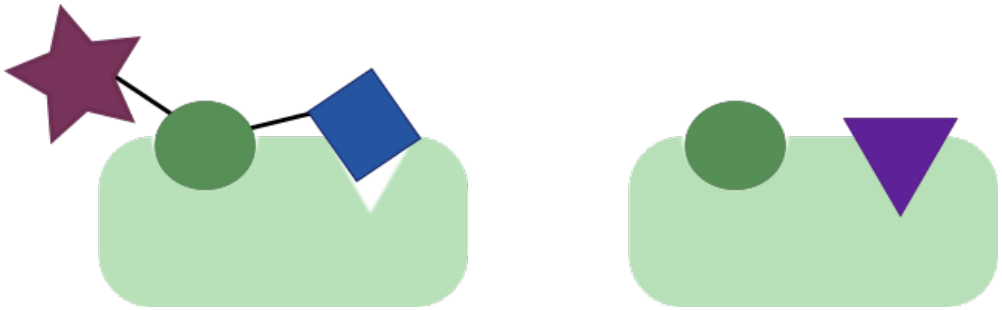


Figure 2. HTS compounds (left) are more complex than fragments (right). A risk with HTS lies with compounds not being perfectly complementary for the protein pocket they are exploring, which could prevent the compound from binding or reduce the binding affinity. They can also contain superfluous groups which do not contribute to the potency and lower ligand efficiency.

Secondly, fragments sample a greater proportion of structural space compared to HTS compounds. This is due to the astonishingly large number of potential drug molecules - estimated to be 10^{63}, which is larger than the total number of stars in the universe.[10] With fewer than 20 heavy, non-hydrogen atoms (NHA), the number of possible fragments is much smaller (estimated at ~ 10^{11} for fragments with 17 heavy atoms). This results in fragment libraries being significantly smaller (~ 10^{3} fragments) compared to HTS (normally around 10^{6}), whilst still sampling chemical space in a more efficient manner. The smaller scale and cost of fragment libraries also allows them to be accessible to academic groups as well as pharmaceutical companies.

Finally, fragments are generally easier to elaborate due to their small size and often higher solubility. There is greater flexibility in which moieties can be added and the physicochemical parameters can be monitored throughout the development process to ensure desirable pharmacokinetic properties in the final compound without loss of potency.

1.2.1.2 Fragment Library Design

The first stage of FBLD is to either obtain or design a fragment library to screen against a selected target. As indicated above, most libraries comprise of 1000-1500 fragments. Molecules in HTS libraries often follow Lipinski's Rule of Five, which provides parameters for key physicochemical properties to increase the likelihood of synthesising an orally bioavailable drug.[11] Fragments are characterised as small molecules which follow an analogous framework proposed by Congreve *et al*,[11] designated as Rule of Three (Ro3):

- Molecular weight < 300 Da
- Number of hydrogen bond donors ≤ 3
- Number of hydrogen bond acceptors ≤ 3
- cLogP ≤ 3

These rules are often extended to include number of rotatable bonds (≤ 3) and a polar surface area (PSA) (≤60) to improve oral bioavailability of eventual lead compounds derived from the fragments.[13] However, there are a number of drugs which have been successful despite violating Lipinski's rules; as a consequence, some fragment libraries have also been designed which deviate from the Ro3.[14] Solubility is an important consideration during fragment library design. Fragments are generally screened at higher concentrations due to the lower affinity of fragments compared to HTS compounds. This necessitates libraries to be comprised of fragments with high solubility in solvents such as DMSO, and to avoid known

chemical moieties or compounds which aggregate at high concentrations. Pan-assay interference compounds (PAINs), which contain reactive motifs such as Michael acceptors and often result in non-specific binding to a target, or known metal chelators, should be carefully considered before inclusion in a fragment library.[14,15]

Developments in fragment library design include attempts to address the major bottleneck in most drug discovery platforms: the speed of medicinal chemistry to synthesise a follow-up library of analogues based on the initial hit. The establishment of a 'poised' library consisting of fragments derived from a set of standard robust reactions aims to avoid extended synthetic optimisation.[16] The chemical 'handles' included in each molecule allows for simple and rapid elaboration of the fragment hit into a library of analogues using parallel chemistry – and theoretically resulting in the faster development of potent compounds.

In recent years, there has been much debate regarding the so-called 'flatness' of drugs and whether there are missed opportunities by not expanding into chiral space, especially for more difficult drug targets such as protein-protein interactions (PPIs).[17] This has resulted in the design of '3D' fragment libraries, which aim to use sp^3-containing fragments to avoid the development of planar compounds, achieving greater structural diversity and better spatial selectivity for a target protein.[18] In addition, fragments with 3D components are typically more soluble due to poor crystal packing, solving the aforementioned problem of high concentration testing. However, Hann's rule of molecular complexity predicts that increasing the complexity of fragments would reduce the probability of finding a matching binding site.[19] This could be overcome by screening more fragments, although this is dependent on the capacity of the biophysical screening method, and increases the cost of the experiment. 3D fragment libraries are often more challenging to synthesise, especially when attempting to maintain the low molecular weight threshold of fragments. The development of diversity-orientated synthesis aims to simplify the process of developing compounds or fragment libraries with greater skeletal diversity and chirality using the build/couple/pair model.[20]

Custom libraries have been designed specifically for a single target with varying success in both fragment and HTS campaigns. These focused libraries typically use structural information of an active site and known binders, and are therefore often available for so-called 'privileged' targets, which are members of families that have had proportionally significant drug discovery efforts within the genome.[21,22] For example, kinase-directed libraries containing fragments that bind to the ATP-binding region are commercially available.[23] The use of custom libraries allows a relatively small and quick screen with a large chance of identifying a binder. However, these libraries often have a higher hit rate for off-target proteins within the family and detractors argue that it is more difficult to achieve a selective novel compound for a single protein in an already congested intellectual property space.[14] Other custom libraries used in FBLD and HTS include localisation-based compounds – for example, CNS libraries containing compounds with desirable properties for brain-penetrating molecules[23] - or mechanism based, such as covalent fragment libraries or libraries designed to contain analogues that bind to the transition state of a desired enzyme.[24–26]

1.2.1.3 Hit Identification

The small number of interactions made by fragments with the target results in fragments binding with low affinity (typically μM – mM).[27] This range of potency is typically too weak to be detected by the biochemical methods used to screen HTS libraries, or by cell activity assays. Fragment screening instead uses biophysical techniques, which are often highly sensitive but are lower throughput and require large quantities of protein. FBLD was first validated upon the discovery by Fesik and colleagues of 'SAR by NMR'.[28] This seminal research identified a compound with nanomolar binding affinity to FK506 binding protein (FKBP), which was derived by the linking of two fragments. The binding mode of the fragments, and the structure-activity relationships (SARs) identifying key binding motifs were

determined by the analysis of shifts in the 2D NMR spectra of the backbone amino acids of FKBP.

Protein- and ligand- observed NMR (such as saturation transfer difference [STD] or waterLOGSY) has remained very popular for FBLD due to the immediate structural information regarding the binding site and high sensitivity for weak binders. Other biophysical methods for fragment screening include Surface Plasmon Resonance (SPR), X-Ray Crystallography, Thermal Shift Assays (TSA) such as Differential Scanning Fluorimetry (DSF), Microscale Thermophoresis (MST) and Mass spectrometry (MS). NMR, SPR and X-ray are deemed the 'classical' fragment screening methods due to their sensitivity and robustness.

Virtual screening uses *in silico* docking to screen commercially available compounds against a known target structure.[29] This can be used as a screening technique but limitations in current algorithms to accurately model the complexities of the protein-ligand interactions often results in poorly binding compounds. Meanwhile, advances in cryogenic electron microscopy (cryo-EM) may result in this becoming a leading fragment screening method in the future due to its structural information about the protein architecture without the difficult optimisation process associated with crystals. However, this will require overcoming significant technological and logistical hurdles including low resolution and throughput.[30] A summary of the main methods, with their relative advantages and disadvantages, is shown in Table 1. Further details for X-Ray crystallography as a FBLD screening method are given in section 1.2.4.6.

Table 1. The main biophysical techniques found in fragment screening campaigns.[2,10,13,31–33] Scientists will consider the advantages and disadvantages before choosing a method. Often this will rely on the capabilities of the research group and the specifics of the selected target. Fragment campaigns may use a combination of screening methods. References for potent compounds derived from fragments found in the relevant screening campaign are given in the final column.

Method	Description	Advantages	Disadvantages	Ref.
Protein-observed NMR	Changes in chemical shifts of ^{15}N-labelled protein upon ligand binding	- Binding site information - High sensitivity - Calculate K_D	- High protein consumption - Time consuming - Expensive - Proteins < 50 kDa	[34]
Ligand-observed NMR	Changes in ligand NMR spectrum upon protein binding	- Can discriminate between specific and non-specific binders - No protein size limit - Rapid	- No binding site information - Not robust - False positives	[35]
X-ray Crystallography	Fragment soaking or co-crystallisation resulting in X-ray diffraction data of ligands bound to protein	- Best method for structural binding information - Highly sensitive - Avoids false positives - Can screen large M_W targets	- Difficult to crystallise some target proteins - False negatives - High protein consumption - Often low throughput	[36]
SPR	Changes to refractive index observed upon fragment binding to immobilised protein	- Kinetic and thermodynamic information - Low protein consumption and cost-efficient - Rapid - Applicable to most target proteins	- No binding site information - Requires immobilisation of target proteins - False negatives - Artefacts - Medium sensitivity	[37,38]
DSF	Shift in thermal unfolding of protein upon ligand binding	- High throughput - Inexpensive - Rapid - Low protein consumption	- Not suitable for all target proteins - False positives and negatives - Low sensitivity	[38,39]

Method	Description	Advantages	Disadvantages	Ref.
MST	Shift in motion along a thermal gradient upon ligand binding to protein	- High throughput - High sensitivity - Amenable to most targets - Very low protein consumption	- False positives - No binding site information	[38]
MS	Enrichment of compounds following protein incubation identified on MS	- High throughput - Low protein consumption	- No binding site information - Low sensitivity	[40]
Cryo-EM	Electron microscopy to give 3D structure of ligand bound to protein	- Binding site information - No protein/complex size limit - Protein in native state - No crystallisation required	- Currently lacks the throughput and reproducibility for FBLD - Low resolution	N/A
Virtual screening	*In silico* docking of compounds in structure of target protein	- Time and cost saving - Large number of available methods	- Increased rate of false positive/negatives if using homology model - Scoring functions poor	[41]

1.2.1.4 Hit Elaboration to Potency

Fragments are typically optimised in an iterative process, guided by affinity data (generated either through the screening process itself, or by an alternative method such as isothermal titration calorimetry [ITC]) and by subsequent structural binding information. Structural data in the form of X-ray crystal structures – or 2D NMR, although this is significantly more time-consuming and less precise – is often essential for the successful development of fragments into potent lead compounds. The structural methods are used to characterise the binding mode, providing integral information as to how the fragment interacts with the target, and identifying pharmacophores that the medicinal chemist needs to maintain or pursue to gain potency. They can also show fragments binding to hotspots – defined as areas within

pockets that disproportionately contribute to the binding affinity.[42] These would be difficult to predict without structural information.

Commercially available analogues can be purchased or otherwise synthesised to elucidate SARs of the core of the fragment before further elaboration into a larger compound. Fragments are typically elaborated by three different methods: linking, merging, and growing (Figure 3).

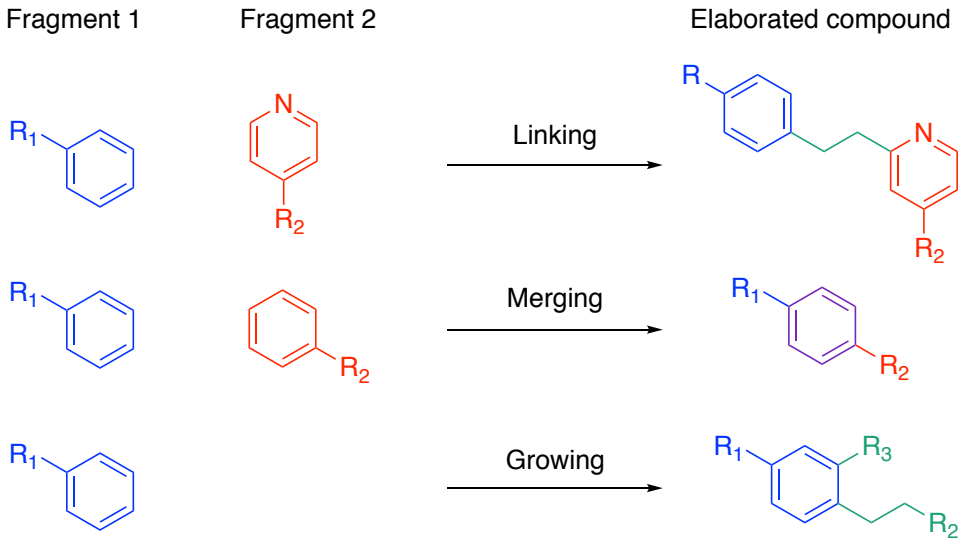
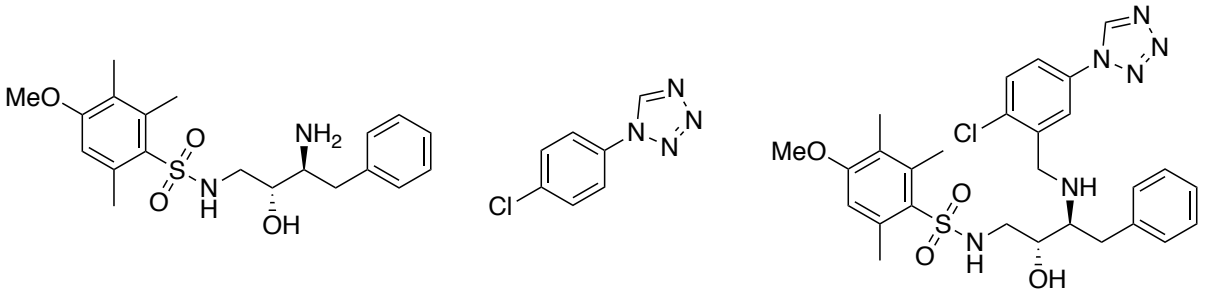


Figure 3. The three methods of fragment elaboration in FBLD.

Fragment linking involves the design of a high affinity molecule by covalently linking two fragment hits which bind to distinct neighbouring sites on a target. In theory, this would be the desirable method of fragment elaboration as it should result in 'superadditivity', whereby the binding energy of the linked compound exceeds the sum of the binding energies of the individual fragments due to a smaller entropic barrier.[43] In practice, it is often challenging to develop an ideal linker which joins the fragments whilst maintaining their optimal configurations within the binding pocket of the protein. The binding affinity of the final compound is usually lower than the theoretical value due to induced steric clashing or loss of favourable interactions resulting in a greater entropic penalty than predicted.[27] Linked compounds can sometimes exceed the molecular weight limit of 500 Da derived from

Lipinski's rules. This is dependent on the size of the original fragments and the length of the linker. The method is also reliant on the protein having two appropriate adjacent pockets which both bind to fragments.

However, several successful examples have been reported in the literature; Howard *et al*, following screening by X-ray crystallography on thrombin, found two weakly binding hits **4** and **5** in adjacent binding sites (Figure 4).[36] Linking the fragments together yielded the potent compound **6** with an IC_{50} of 1.4 nM, although the authors note the compound does not adhere to Lipinski's rules for an orally bioavailable compound. Fragment linking was also used in the identification of venetoclax.[10]



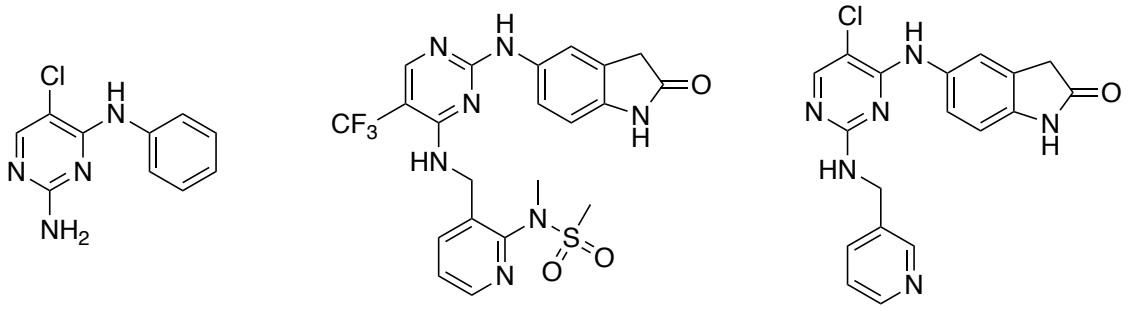
4
IC_{50} = 12 μM

5
IC_{50} = 330 μM

6
IC_{50} = 1.4 nM

Figure 4. A strategy to link fragments **4** and **5** yielded potent compound **6** with 'superadditive' IC_{50} against thrombin.

Fragment merging involves the combination of two or more fragments containing a structural component which overlaps in the binding site. This requires multiple crystal structures and careful design to ensure the binding poses of both original fragments are maintained in the final compound.[44] Fragment merging has also been extended successfully to combine screens; there are several examples in the literature where a fragment has been merged with an HTS lead to give a potent compound (Figure 5).[45–47]



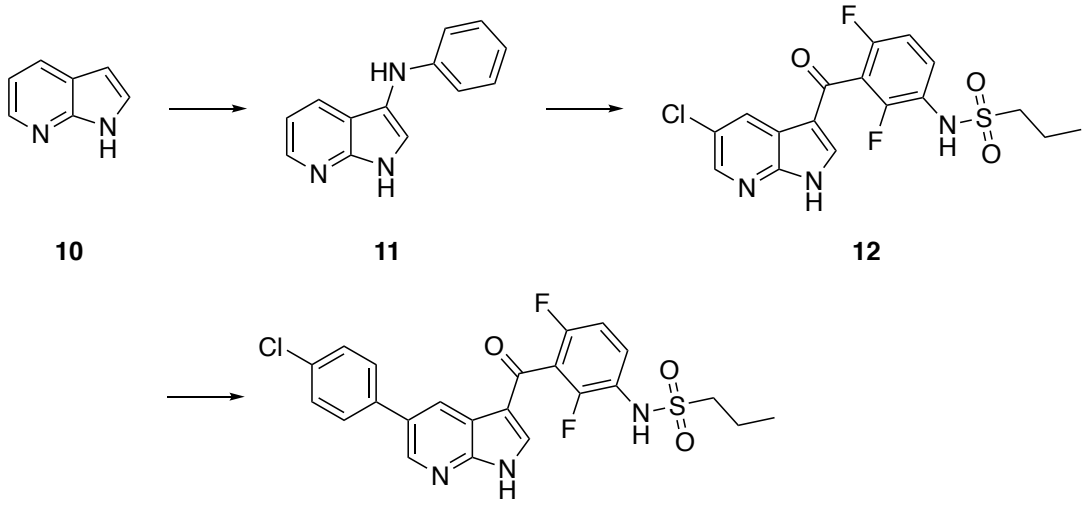
7
K_D = 68 μM

8
K_D = 88 μM

9
K_D = 78 nM

Figure 5. The fragment hit **7** and HTS compound **8** overlapped in the crystal structure of BCL6. Merging of these compounds yielded **9** with higher potency and improved ligand efficiency.[46]

Fragment growing is the most common method of fragment elaboration. It comprises of stepwise extensions of a single fragment hit by adding functional groups. This is an incremental process, aiming to explore the new interactions with neighbouring regions of the protein inferred from SARs, but without creating clashes with the protein surface. Fragment growing often involves the synthesis of a library of follow up compounds to elucidate SARs. This method was used in the discovery of vemurafenib which was developed by iterative fragment growing of the original 7-azaindole fragment **10** (Figure 6).[10]



10 **11** **12**

1

Figure 6. The fragment growing strategy from the fragment hit **10** to vemurafenib (**1**).[10]

Ligand efficiency (LE) is an important metric to evaluate during the fragment selection and optimisation process. It is defined as the binding energy of a molecule per non-hydrogen atom (LE = - ΔG / NHA or LE = pIC_{50} / NHA).[48,49] LE normalises the potency of a compound or fragment compared to its molecular weight. Selecting fragments and compounds with a LE > 0.3 kcal mol^{-1} NHA^{-1} is postulated to deliver compounds with more favourable ADMET properties with high affinity; compounds with a LE < 0.3 are less likely to become successful drugs.[50–52]

1.2.1.5 Bioaffinity Measurements

The iterative process of rational synthesis is aided by quantitative measurement of a ligand's potency. The dissociation constant (K_D) or half-maximal inhibition constant (IC_{50}) are common metrics used to compare compounds within a series. The methods used are applicable across drug discovery processes and are not specific to FBLD. General methods previously mentioned, such as ITC, NMR, or SPR can be used in binding assays, although these often require large amounts of protein and do not measure the ability of the compound to inhibit activity.[31] Activity-based assays can be developed for the target protein which measure the ability of the ligand to block a specific action, such as the synthesis of a product by an enzyme.[53] These assays are often easy to conduct, with light-based readouts such as fluorescence, and scalable to high-throughput. PAINs can cause issues at this stage, with reactive, instable or aggregatory compounds resulting in artifacts or false positives in these assays. As compounds are elaborated and increase in potency, the assay may be developed for other similar proteins to measure the selectivity of the ligand for the desired target.

Once ligands have reached the desired level of potency and selectivity, they may be progressed as either therapeutics or be used as chemical probes.

1.2.2 Chemical Probes as tools for target validation

1.2.2.1 Target Validation

Target validation is a key process in the development of successful therapeutics. Inadequate target validation prior to clinical trials has been attributed as a major factor in the high rates of drug attrition.[54] If a drug fails due to a lack of efficacy, one reason may be that the wrong target protein was chosen for that particular disease. Hence, methods to elucidate and confirm the proposed role of a target protein in a disease is of paramount importance to academic research and the pharmaceutical industry.

In recent years, and following improvements in bioinformatics and the resultant data mining of biomedical information, there has been a huge increase in the identification of targets of therapeutic interest in the literature.[55] A target must then be prioritised and thoroughly validated to ensure it is clinically relevant, tractable and non-toxic before beginning drug development efforts. Researchers often combine a variety of methods to characterise a target, from *in vivo* animal models to genetic knockdowns using CRISPR or siRNA. However, the use of biological methods alone has raised concerns because they are often associated with poor reproducibility, off-target effects, and are frequently unable to replicate the specific phenotypic effects of a small molecule for the target in question.[56] An integrated chemical biology approach is increasingly common, combining biological validation methods with small molecule tools, known as chemical probes, to thoroughly investigate a target and its role in the disease before embarking on drug development.[57]

1.2.2.2 What is a Chemical Probe?

A chemical probe is a small molecule reagent that is highly selective for a particular protein and can modulate its function so that its biological role can be investigated in *in vitro* and/or in *in vivo* studies. Their role is to validate targets for therapeutic development through

analysis of the phenotypic outcome and provide a proof of concept, thereby reducing the risk of a drug failing in the later stages of drug development due to a lack of efficacy.[58] They can also be used to analyse the safety of inhibiting a particular target in a disease, and whether reduced activity of this protein is tolerated. Ideally, they are used with complementary biological techniques to provide extra information. For example, the use of a chemical probe in parallel with RNAi knockout experiments allows a researcher to determine if the phenotypic outcome of the disease is linked to the protein's enzymatic activity inhibited by the small molecule, or related to its other potential roles in the cell, such as acting as a protein scaffold.[59]

The discovery of high-quality chemical probes has a huge impact on the research community. A good chemical probe will improve the understanding of a target and its mechanism in disease, and so the release of a probe into the literature by one group will catalyse and accelerate further research on that target. Indeed, the discovery of BET family bromodomain chemical probe JQ1 led to the implication of the BET family in several distinct diseases, the initiation of clinical programs at pharmaceutical companies for BET proteins, and was even used in early proof-of-concept experiments for the inhibition of BET proteins *via* Proteolysis Targeting Chimeras (PROTACs).[60,61] However, there is a danger with poorly characterised chemical probes being used inappropriately, leading to incorrect conclusions and drug campaigns destined to failure.[60] It is vital that going forward, care is taken to ensure that probes are selective and do not contain reactive moieties or PAINs motifs which could lead to misleading results. Concern within the chemical biology community has resulted in the peer-reviewed Chemical Probe Portal, which aims to maintain a high standard for chemical probes, and a recent surge of literature specifically defining the characteristics of a probe in order to avoid unnecessary pitfalls.[60,62]

In order to standardise the definition of a chemical probe, the Structural Genetics Consortium (SGC)[60] designated the following requirements:

- *In vitro* potency < 100 nM for the desired target
- Cellular potency < 1 μM
- > 30-fold selectivity for other proteins within the target family
- Profiled against other pharmacologically relevant off-targets (eg Eurofins panel)[63,64]
- Close inactive analogue to act as negative control (e.g. enantiomer)

It is beneficial to have more than one chemical probe for a target, with a distinct chemical scaffold to ensure related phenotypic outcomes are attributed to direct inhibition of the target and not as a result of off-target effects.[59] Other, more stringent methods of characterisation include Workman *et al*'s 'Fitness Factors'[59] and Pfizer's '4 Pillars' approach,[57] which requires researchers to confirm: 1) The probe can reach pharmacologically relevant concentrations within a cell; 2) Target engagement by the probe in the cell; 3) A developed assay to elucidate pharmacology of the probe and 4) An assay to show phenotype perturbation.

1.2.2.3 Chemical Probes versus Drugs

A chemical probe can be developed against a target using typical drug discovery methods, such as HTS or FBLD as described in section 1.1.1. Successful probes developed by FBLD methods for important cancer targets such as BCL-xL and BRD9 are well reviewed in the literature.[58] However, the criteria for drugs and chemical probes are different (Figure 7). It is important to optimise certain pharmacokinetic metrics, such as oral bioavailability and half-life, during drug development to maximise target engagement within the body. These

properties are closely monitored during the FBLD process by characterising compound properties such as LE or clogP in early-stage development. This is not a requirement for a probe compound which will be used only in biochemical or cell-based assays.

Chemical probes require exquisite selectivity to ensure that any effects seen in assays are directly attributable to inhibition of the desired target.[60] The body can tolerate therapeutics with some promiscuity and off-target effects; certain drugs exhibit 'polypharmacological' phenomena, where the drug acts therapeutically *via* the interaction with multiple targets.[65] However, drugs are required to be selective against protein families known to cause adverse risk if inhibited, such as the hERG channels and cytochrome P450 famillies.[62] Chemical probes also need a well-characterised mechanism of action in order to ensure their utility in assays. On the other hand, there have been many circumstances of approved drugs with unknown mechanisms of action, especially in the early days of drug discovery.[66]

Drugs are also subject to intellectual property law, whereas the use of chemical probes is free and widely advertised within the chemical biology community, with the recent development of the Chemical Probes Portal advocating for open access to well-detailed probes throughout the chemical biology community, and allowing for peer review to ensure the optimal probe can be used by a researcher (Figure 7).[67]

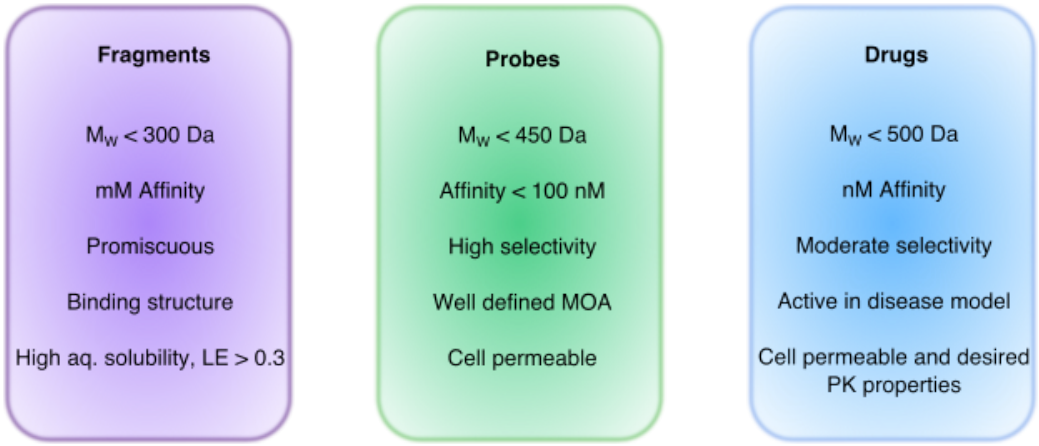


Figure 7. The major differences between fragments, probes, and drugs. Fragments can be developed into either small molecules probes or drugs depending on which criteria are followed during the fragment elaboration process.

The examples of small molecule drugs and chemical probes given thus far are reversible binders. However, a subset of inhibitors contain a warhead within their scaffold which is capable of reacting with desired target, causing irreversible inhibition. These covalent compounds are considered below.

1.2.3 Covalent Therapeutics and Probes

1.2.3.1 What is a Covalent Therapeutic/Probe?

The majority of small molecule therapeutics and probes are reversible binders. However, around 30% of approved drugs are covalent compounds.[68] These are small molecules which contain an electrophilic handle as part of their substructure that is capable of reacting to form a covalent linkage with their target, which can be irreversible. Covalent drugs will typically inhibit an enzyme, although other classes of covalent drugs exist, such as DNA alkylating agents.[68] The binding of the compound to the target protein is generally a two-step process; the unreactive part of the compound binds to the target through intermolecular interactions before the handle reacts to form a covalent bond with a suitable residue, often a cysteine or serine, within the binding pocket.[69] The intermolecular interactions must be very tight and specific and the warhead generally unreactive in order to retain specificity and prevent off-target effects. The use of covalent inhibitors has existed since the 19th century, with the use of aspirin as a painkiller, which was later discovered to irreversibly inhibit cyclo-oxygenase (COX) enzymes 1 and 2.[69]

1.2.3.2 Advantages and Disadvantages of Covalent Compounds

Covalent compounds have several advantages over their non-covalent counterparts. Firstly, the (often) irreversible nature of binding results in a higher affinity in relation to their size, making it easier to maintain desirable pharmacokinetic properties during the design

process when increasing the potency in an efficient manner. There are also often promising pharmacodynamic properties. A reversible small molecule inhibitor binds under equilibrium conditions; when not bound, it can be cleared from the body whilst the protein remains fully functional. As the covalent bond is durable, and often irreversible, the body must fully resynthesise the protein to have functional protein, resulting in a prolonged duration of action for the drug. This translates into lower doses and a less frequent dosing regime, which is often preferable for the patient. However, this effect may not be seen in targets with high levels of protein turnover in the cell.[70] The non-equilibrium binding kinetics and extra binding energy of a covalent bond also give covalent compounds an advantage over high concentrations of any potential endogenous substrates, whereas a reversible binder would be in competition with such molecules.[71]

Covalent compounds can provide the opportunity to inhibit intractable targets where the development of a non-covalent binder would be hindered due to shallow binding pockets where interactions are generally weaker. If a suitable residue is present, the introduction of a strong covalent linkage may result in marked inhibition of the target despite few non-covalent interactions.[68] Covalent compounds may also represent a chance to overcome drug resistance mechanisms. An important mechanism in drug resistance occurs due to a mutation of one or several key residues responsible for compound binding. Irreversible covalent drugs would be susceptible to this resistance mechanism if the point mutation happens to be the reactive residue, resulting in complete resistance to the drug; otherwise some inhibition of the protein may still be observed.[72]

However, there has been some reluctance until relatively recently to pursue covalent inhibitors from the pharmaceutical industry. Covalent drugs gained a poor reputation as some of the original therapeutics produced extremely toxic metabolites *in vivo*. A famous example is paracetamol, which when metabolised by the liver produces reactive quinone intermediates which are hepatotoxic, leading to adverse drug reactions (ADRs).[71] The electrophilic handles

are particularly susceptible to metabolism, hence the increased risk of toxic metabolites associated with covalent inhibitors. The covalent warheads also pose a risk of off-target effects from non-specific binding due their high reactivity. Recent developments have offset this problem using less reactive covalent handles and optimising the non-covalent interactions for greater affinity. It is thought that this off-target binding in some patients can provoke an immune response, leading to idiosyncratic toxicity effects which are difficult to predict prior to later stage clinical trials. This has resulted in a bias against the development of covalent drugs due to safety concerns from the unpredictability of ADRs and failures in clinical trials.[72,73]

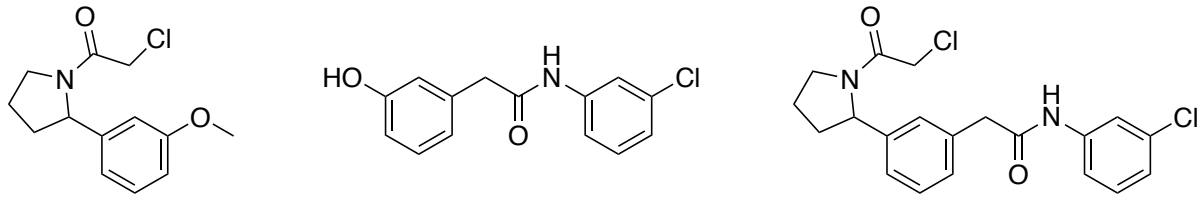
The durability of the covalent bond can also raise safety concerns if the targeted protein is essential to the body, and turnover of the protein is very slow, meaning covalent inhibitors may not be suitable for every target. The long binding duration also makes it difficult to measure and compare the potency of covalent inhibitors, as IC_{50} values are normally calculated assuming equilibrium dynamics, whereas the potency values of covalent inhibitors are time dependent. These measurements are often used as parameters for determining SARs, and so the rational design of a series of compounds becomes more arduous in the pursuit of a drug or probe.[70]

1.2.3.3 Recent Developments

A problem associated with many early covalent therapeutics was that they were designed to react with a catalytic residue essential for the protein's function; such a residue is often conserved across families, leading to off-target effects and non-selective inhibition. A more recent approach involves the development of 'targeted covalent inhibitors' (TCIs).[70,74] TCIs combine a non-covalent molecule which has high specificity for the desired target in combination with the covalent handle, which has low reactivity. They target non-catalytic residues which are poorly conserved, thereby reducing off-target side effects. A variety of covalent warheads have now been developed for covalent drug and probe development.[74]

These warheads are designed to be only reactive enough to form a bond with a residue when proximally orientated, following the reversible binding of the non-covalent section of the molecule within the binding pocket. The handles will also vary depending on the amino acid residue being targeted, with the development of warheads capable of targeting other residues apart from cysteine and serine, such as tyrosine, lysine or histidine.[74,75]

The discovery of most of the original covalent therapeutics was serendipitous; rational discovery has remained more difficult. Typically, the non-covalent section and covalent handle of TCIs are optimised separately to ensure each part has high affinity for the desired protein.[70] The reversible section can be found by traditional methods, such as HTS or FBLD. However, it can be difficult to efficiently add a covalent handle whilst maintaining the potency of the reversible ligand. A way to overcome this pitfall is to screen a library of fragments with covalent handles, and then develop them into potent inhibitors by classic fragment elaboration techniques. Covalent fragment screening is robust *via* mass spectrometry, as the covalent labelling can easily identified by an increase in the mass of the protein. Recent developments by the London group at the Weizmann Institute of Science have optimised the high throughput mass spectrometry screening of a large library of mild electrophilic fragments, combined with a thiol reactivity assay and counter-screening platform to avoid prioritisation of promiscuous binders.[40] Merging of a covalent fragment (**13**) found by this approach with a non-covalent fragment (**14**) of NUDT7 discovered by X-ray crystallography fragment screening resulted in the development of a promising covalent lead compound **15** for a probe which is able to engage NUDT7 in a CETSA assay (Figure 8).



13 **14** **15**

Figure 8. The development of a lead covalent compound **15** for NUDT7 by the merging of a covalent fragment **13** and non-covalent fragment **14**.

The development of both reversible and irreversible inhibitors, as indicated by the example given above, is aided through detailed structural information that can be ascertained using X-ray crystallography.

1.2.4 X-ray crystallography

1.2.4.1 Overview of X-ray Crystallography

Protein X-ray crystallography is the determination of the 3D structure of a protein to atomic resolution, generated from the diffraction pattern from electrons when a protein crystal is subject to X-ray radiation. The exquisite detail of the 3D structure as the result of an X-ray crystallography experiment has resulted in it often being considered the 'gold standard' for structural information in drug development when compared to NMR spectroscopy or cryo-EM. It was known since the 19th century that proteins could be crystallised, with the first crystal structures discovered in the 20th century, pioneered by Perutz, Kendrew and Hodgkin.[76]

The use of X-ray crystallography in drug discovery is a multi-step process, from construct design and protein production to structure determination. An overview of the process is as thus: the gene is first analysed, and suitable protein constructs are designed and cloned into a vector. The protein is expressed in a suitable organism and purified for crystallisation. After protein crystals are obtained, they are then mounted and rotated through a high intensity

X-ray beam, producing diffraction patterns resulting from the scattering of X-rays by the electrons in the structure. These diffraction images can be converted into an electron density map using a Fourier transform equation. A 3D atomic model of the protein can then be built into the electron maps. If the protein has a ligand bound to it, the ligand will also be visualised in the electron density maps, which provides the foundation for structure-based drug design (SBDD). Once the model of the macromolecule is of suitably high quality, it can be deposited into the Protein Data Bank (PDB). The PDB is a worldwide accessible repository for non-proprietary macromolecular structures, with around 90% of structure models being determined by X-ray crystallography.[77]

1.2.4.2 The Requirement for Protein Crystals

Protein crystals are regular arrays of identical macromolecules that form a 3D crystal lattice that is held together by weak interactions. X-rays behave as electromagnetic waves; a diffraction pattern can be detected from the scattering of photons by the electrons in an atom. The diffraction patterns from a single molecule are very small and hard to detect. However, if lots of molecules are present in a periodic lattice, then the corresponding diffraction is very strong and can be detected – hence the requirement for protein crystals. A space group is used to describe the symmetry in which the protein is packed in the crystal lattice. The crystal lattice can be broken up into unit cells, which is the basic building block of the crystal; the entire crystal may be generated by repeating the unit cell in three dimensions. A single unit cell may contain one or more macromolecules.[53] The asymmetric unit is a small part of a unit cell that can generate the complete unit cell following the symmetry operations of the space group.

1.2.4.3 Protein Production

The production of high-quality, well-diffracting protein crystals is a dominant bottleneck of the X-ray crystallography process. To generate protein crystals, one first requires protein. Construct design and protein engineering is often vitally important for crystallisation success.[78] In order to form a well-ordered crystal, the protein must be able to form favourable intermolecular interactions. It may be necessary when designing the protein construct to remove or change parts of the protein which could prevent favourable packing. For example, bioinformatic programs can predict disordered regions, which, if possible, should be excluded from the protein constructs, as flexible regions may hamper the ability of the protein to form an ordered lattice. It may also be necessary to mutate certain surface residues which could cause aggregation of proteins during the crystallisation process. Protein production and purification is also an important consideration as crystallisation often requires a vast amount of protein, and proteins are more likely to crystallise if they are pure, homogenous, and monophasic.

1.2.4.4 Protein Crystallisation

Proteins in their native structure are in aqueous solution and are not ordered. To form a crystal lattice, a concentrated protein solution is slowly forced into being supersaturated. Given the right nucleation conditions, the protein will separate from solution and form an organised crystal lattice (Figure 9). There is often a fine line between supersaturated and unstable protein, leading to denaturation. Crystallisation trials normally involve the testing of hundreds of conditions at once, especially as there is no definitive process to predict the conditions required for crystallisation for a particular protein. pH and temperature often have a significant effect on the protein solubility, as does the presence of salts and precipitants (such as PEG).[76,79] It is also possible to control the nucleation event by the process of seeding, which is the addition of microcrystals to act as nucleation sites, encouraging the formation of

crystals instead of precipitated protein. The success rate of producing protein crystals which suitably diffract is estimated as only 10%, meaning the use of multiple constructs and crystallisation conditions is essential to improve the odds of structure determination. For more difficult targets, such as integral membrane (IMPs) or intrinsically disordered (IDPs) proteins, the success rate is even lower.

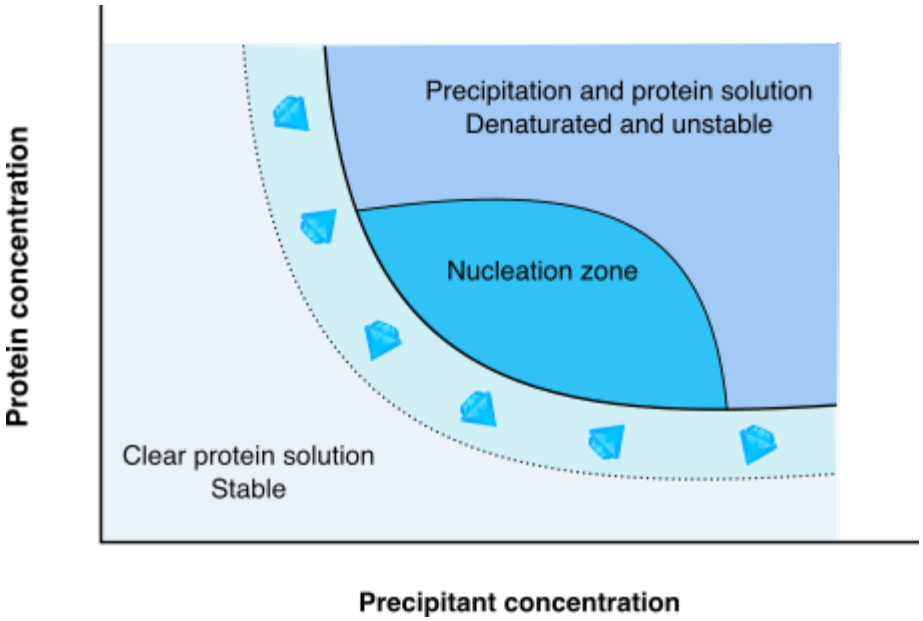


Figure 9. Solubility phase diagram for protein crystallisation. Proteins can be forced into a metastable supersaturated solution by either increasing the protein or precipitant concentration. If the protein mixture is too dilute, only clear solution will be seen. Too concentrated, and the protein will denature. In the metastable solution, a protein will separate into either amorphous protein or crystals in a protein rich phase and a different saturated solution. Inspired by Spiliopoulou *et al* [80] and Rupp[76].

Typical experiments at the SGC use the sitting-drop vapour diffusion method, where a protein solution is mixed with the crystallisation conditions in a single drop, with a neighbouring separate reservoir containing mother liquor. Over time, water moves from the sitting drop through to the mother liquor *via* vapour diffusion, resulting in the protein solution becoming more concentrated and hopefully resulting in the development of crystals. The mother liquor will vary from well to well, with a single plate containing 96 conditions varying the buffer pH

and precipitants, with three different protein concentrations per well. The plate can be incubated at different temperatures to maximise the possibility of growing a crystal.

1.2.4.5 X-ray Diffraction and Model Refinement

Upon the production of crystals, they are mounted onto small pins and frozen in liquid nitrogen. This is to minimise the radiation damage to the crystal from the X-ray beam and requires the use of a cryoprotectant such as ethylene glycol to prevent the formation of ice. The crystals are then rotated through an X-ray beam and the diffraction patterns recorded. The diffraction data can be converted into electron density maps of the protein structure by a Fourier transformation. However, the diffraction spots collected only provide the amplitude of the diffracted X-rays – which is determined by the size of the atoms causing the X-ray scattering – and the phase information is lost. The phases for each reflection are dictated by the position of the atoms in the structure and are therefore essential for building the electron density maps. This discrepancy is known as the 'phase problem', as the electron density cannot be reconstructed using Fourier transforms of the amplitude data alone.[81] Hence, the phases must be determined experimentally using heavy atom methods such as multiple wavelength anomalous dispersion (MAD) or single wavelength anomalous dispersion (SAD). Alternatively, they can be estimated by using a known, structurally similar model to calculate the initial phases for the first reconstruction of electron density in a process known as molecular replacement (MR). The phases can be improved by refinement against the actual protein model to fit the experimental data.

The quality of the model is measured by its R values, R_{work} and R_{free}.[53,76] The R_{work} calculates how well the model fits the experimental data. Depending on the resolution, a well-refined structure will have a R_{work} value of 0.2-0.25. The R_{free} value is the R_{work} value derived from experimental data (~ 5% of reflections) that were not used in the construction of the electron density map. This value is used to ensure the model is not over-fitted to the electron

density compared to the experimental data. If the model is overfitted in a round of refinement, the R_{work} will drop but the R_{free} will either remain constant or increase. This value is typically below 0.3 and within 0.05 of the R_{work} value.

The fitting of individual atoms can be analysed by their B-factor, otherwise called the isotropic displacement factor. This value measures the thermal movement of a particular atom in a structure - the mean movement from its coordinates. A large B-factor means that either the atoms moved significantly between individual macromolecules in different unit cells, or, in the case of a ligand, it is not present in high occupancy. Large B-factors in a region can indicate disordered or flexible chains in a protein, and these regions should be modelled with caution or emitted from the structure. The resolution (in angstroms, Å) refers to the shortest spacing between atoms whose reflections have been used in the construction of the electron density map. When the resolution value in angstroms is lower, the resulting electron density maps are sharper and more detailed.

1.2.4.6 Ligand Co-crystallisation

Approximately 75% of the crystal structures of the PDB contain ligands bound to the macromolecule, either unintentionally bound from the purification or crystallisation conditions, or intentionally added in order to decipher the binding mode of the compound.[82] It is possible to intentionally obtain protein-ligand complexes by two different methods; soaking or co-crystallisation.

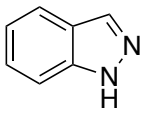
Protein crystals are by nature typically 50% solvent. This allows for the diffusion of small ligands through the solvent channels to bind to the macromolecule following the 'soaking' of a ligand solution onto protein crystals. Heavy atoms for experimental phasing experiments can also be soaked into crystals. Soaking is a quick and easy method and is preferred if a large number of molecules are to be tested. However, some ligands can cause conformational changes within the protein, which can either destroy the crystals or cause the

diffraction quality to suffer. Compounds may also take too long to diffuse into the crystal structure, especially if the crystal is tightly packed, or may not be soluble in the conditions of the crystallisation drop. Hence, for some compounds it is more suitable to try co-crystallisation, where the ligand is first pre-incubated with the protein in solution before attempting protein crystallisation. This method is time consuming, requires larger amounts of protein, and may be difficult to optimise, as the protein-ligand complex may not crystallise in the same conditions as the protein by itself. With both procedures, care must be taken when fitting the ligand in the crystal, especially if the ligand does not bind with 100% occupancy. It is easy to register false positives by fitting a ligand to poor electron density, which could be attributable to another molecule such as a component from the crystallisation conditions.

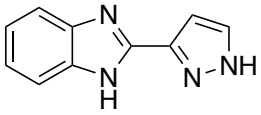
Protein-ligand crystallography has long been used in the drug discovery process as part of structure-based drug design.[83] Following the identification of a hit during either a HTS, FBLD or natural product screening process, it is typical to obtain binding information of the ligand to the structure. X-ray crystal structures are preferred over other methods such as NMR due to the exquisite information and resolution they can provide. These crystal structures can form the basis for iterative rational *de novo* design by medicinal chemists or used by computational chemists for *in silico* screening of follow-up libraries. They can also confirm that the hit obtained from screening acts on the protein of interest and is not a false positive, which is especially common in HTS screens. However, even though the crystal structures can enhance the efficiency of lead optimisation, it is worth taking into consideration that crystal packing may produce artificial interactions or conformations. The static nature of X-ray crystal structures also does not represent the dynamic nature of a protein which may be accessed by a small molecule, nor does it provide information regarding thermodynamic parameters of binding that might be obtained from a different assay.

Whilst X-ray crystallography is frequently used, as detailed above, as a SBDD tool,[84] until recently it was not often used an initial screening method due to being highly resource

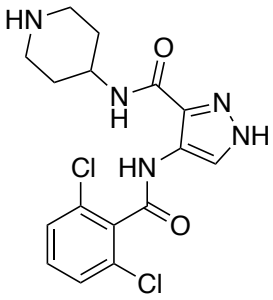
intensive and relatively low throughput. However, improvements in soaking methodologies, automated data collection, and high-throughput crystal production and harvesting have resulted in the use of X-ray crystallography as a FBLD biophysical method, where the libraries are sufficiently small to be able to screen.[85] In the Practical Fragments blog, X-ray crystallography was identified as the most popular screening method by readers in 2019.[86] There are now several compounds in the literature[87], including Astex's CDK2 inhibitor AT7519 (**18**) and Aurora Kinase inhibitor AT9283 (**19**), which were discovered by screening a cocktail of fragments *via* X-ray crystallography and subsequently developed using SBDD (Figure 10). These compounds are currently in clinical trials for the treatment of cancer.[88,89]



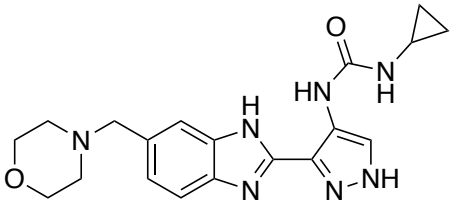
16



17



AT7519 (**18**)



AT9283 (**19**)

Figure 10. The compounds **18** and **19** are currently in clinical trials as CDK2 and pan-Aurora inhibitors respectively. They were developed by fragment growing from **16** and **17** identified in an X-ray crystallography fragment screen on CDK2 using cocktails of compounds. Subsequent testing found **17** had affinity for Aurora Kinase and was subsequently optimised to develop **19**.

XChem[90], at Diamond Light Source (DLS), UK, is an example of an X-ray crystallography screening platform. It is automated and accessible, even for novice crystallographers, and is available for both academic and industrial users. The development of high-throughput methods, such as acoustic dispensing[91], rapid mounting using Shifter

technology[92] and robotic automation on the beamline allows the screening of over 1000 fragment-soaked crystals per week. The acoustic dispensing also means that individual compounds per crystal can be screened instead of a cocktail. This allows higher concentrations to be soaked, reduces the number of false negatives, and is more successful in preventing the destruction of crystals during the soaking process.[91] Upon data collection, XChemExplorer[93] is designed to manage the processing and analysis of thousands of datasets and the PanDDA[94] software allows for the easy identification of even weakly binding ligands by increasing the signal to noise ratio through a background reduction density analysis algorithm. The XChem platform has been involved in the successful generation of chemical probes, such as the identification of a leading fragment in the probe development for the Wnt signalling pathway.[95] The high-throughput fragment screening platform using X-ray crystallography at DLS results in a theoretically quicker and more streamlined process from fragment to drug or probe, as shown in Figure 11.

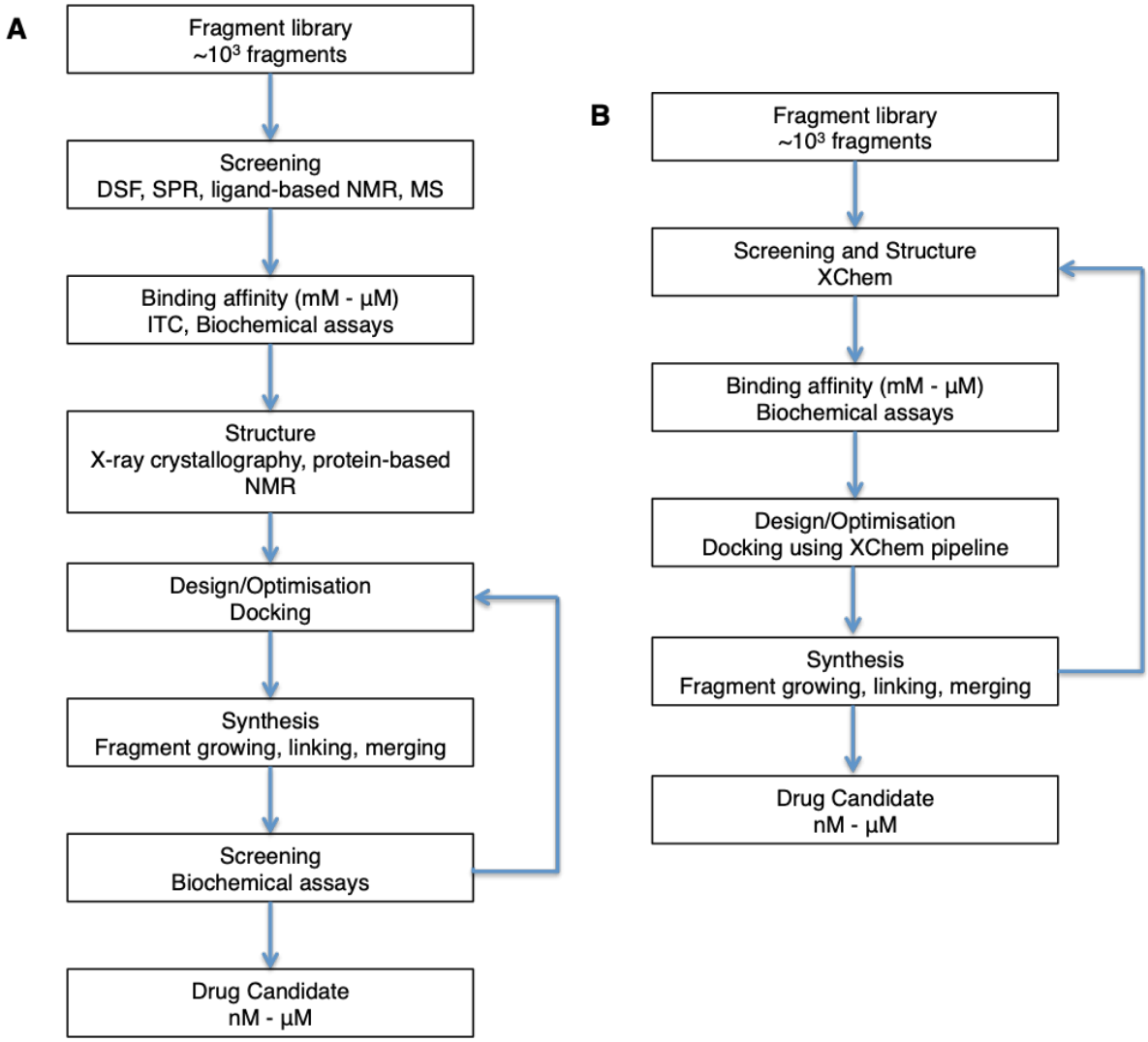


Figure 11. **A**. The traditional pipeline for FBLD. **B**. The development of XChem allows several of the earlier stages to be amalgamated, as screening and structural information are obtained at the same time, allowing for more efficient rational design.

1.3 GTPases/GEFs

1.3.1 Ras Superfamily of Small GTPases

The Ras superfamily of small GTPases are guanine-nucleotide dependent molecular switches essential in the regulation of numerous cell processes. Consisting of over 150 members, the superfamily can be subdivided into five evolutionarily conserved families - Ras, Rho, Rab, Ran and Arf – based on their sequence, structural similarity, and function in the cell (Figure 12A).[96]

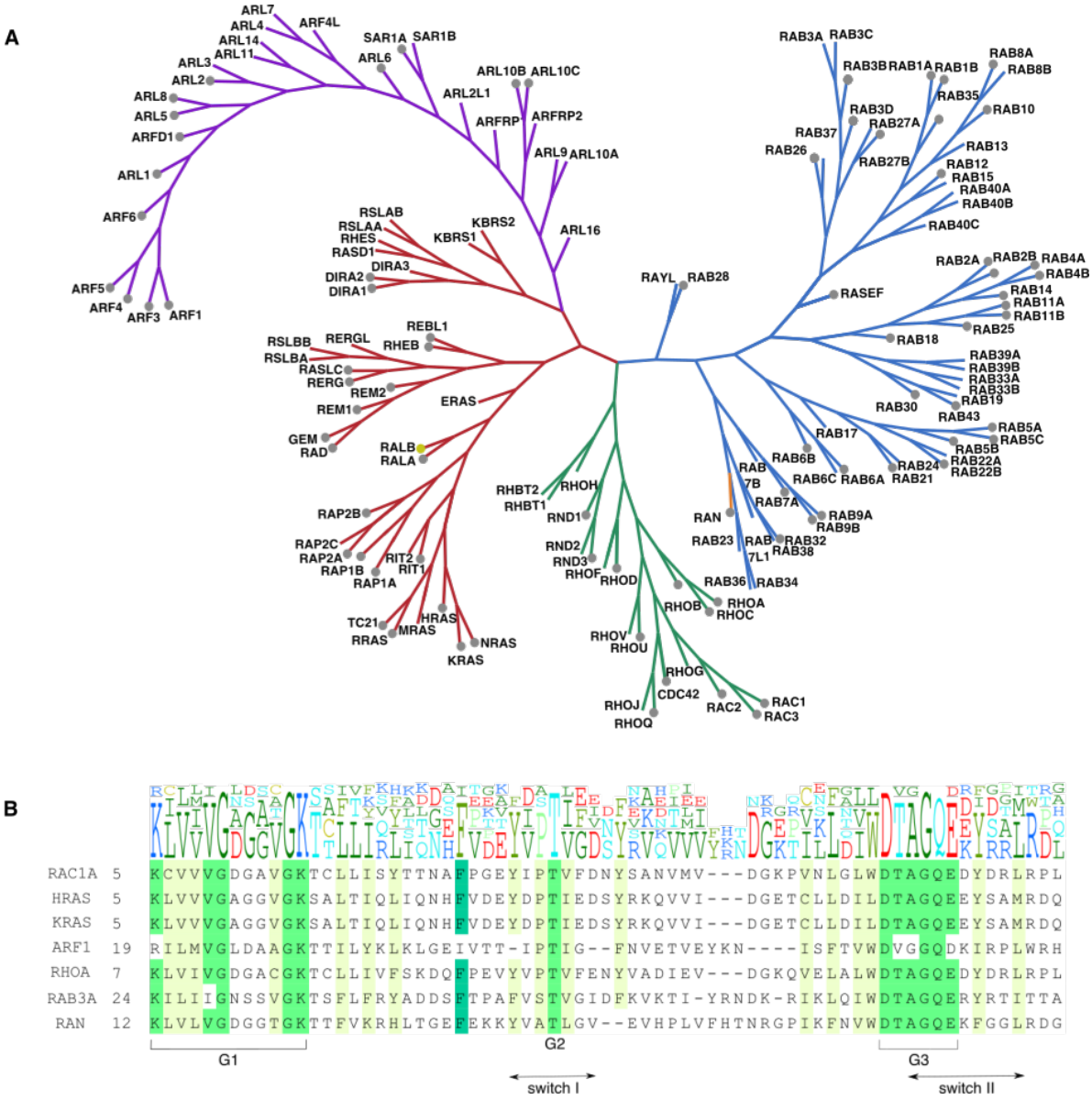


Figure 12. **A**. Phylogenetic tree of the Ras superfamily. The five subfamilies are shown as follows: Ras in red; Rho in green; Rab in blue; Ran in orange and Arf in purple. Grey dots indicate solved X-ray crystal structures documented in the PDB. Yellow dots indicate an NMR structure. **B**. Comparison of sequences between subfamilies. Characteristic 'G-boxes', essential for guanine nucleotide and magnesium ion binding are largely conserved across the superfamily. Reproduced from Gray *et al*.[96]

The Ras subfamily, consisting of 36 members, is primarily responsible for the regulation of cell proliferation, differentiation and apoptosis signalling networks. Comprised of 20 members, the Rho subfamily is key for regulation of gene expression, actin organisation and cell cycle progression. Rab is the largest subfamily, with over 60 members, and regulates

intracellular trafficking. Similarly, Arf proteins, with 30 members, are also involved in vesicle transport. The Ran subfamily is the smallest, consisting of only a single protein which is involved in nuclear transport.[97,98]

The members of the Ras superfamily cycle in the cell from an inactive, GDP-bound state to an active GTP-bound form. This exchange reaction causes conformational changes in the proteins which results in high affinity for their effectors and allows the GTPases to regulate cellular processes. GTPases modulate the function of their effectors *via* a variety of processes, including removal of autoinhibitory intramolecular interactions, catalysing translocation to membranes or inducing conformational changes affecting substrate binding.[99–101]

There is structural similarity throughout the superfamily, and the conserved regions are extremely important for GTPase activity. Unique to the superfamily is the 'G-domain', a region of the protein with five largely conserved fingerprint motifs, known as 'G-boxes', which are responsible for binding to the guanine nucleotides GDP and GTP (Figure 12B). Closest to the N-terminus is the P-loop (also known as G1 or Walker A motif), consisting of residues GXXXXGKS/T, which is essential for nucleotide binding as it is interacts with the β- and γ-phosphates of the guanine nucleotides. The G2 region, also known as 'switch I', contains a single conserved residue across the superfamily, T35 (number in Rac1), which is interacts with both the guanine nucleotide and a magnesium ion that is crucial for binding. The switch I loop is flexible and flips between an 'open' confirmation when GDP is bound, to a 'closed' formation with GTP. This 'closed' formation of switch I is recognised by effectors and allows for the activation of signalling pathways.[102] The G3 (or Walker B) motif DXXGQ is close to the switch II region, which is also involved in the interaction with effectors and regulators. The backbone NH groups of the glycine of switch II and threonine of switch I are responsible for hydrogen bonding with the oxygen atoms of the γ-phosphate of GTP. This helps to maintain the conformational changes seen in the switch I and II loops when GTP rather than GDP is

bound. The G4 region N/TKXD contains an aspartate which forms key hydrogen bonds with the guanine base moiety of the nucleotides. The final G-domain, G5, is weakly conserved as SAK/L, and the backbone amine of the alanine interacts with the oxygen of the guanine base. This generates specificity for the guanine nucleotides, as the adenine amino group cannot be tolerated due to steric reasons.[97,102,103] The G-boxes are highlighted in Figure 13.

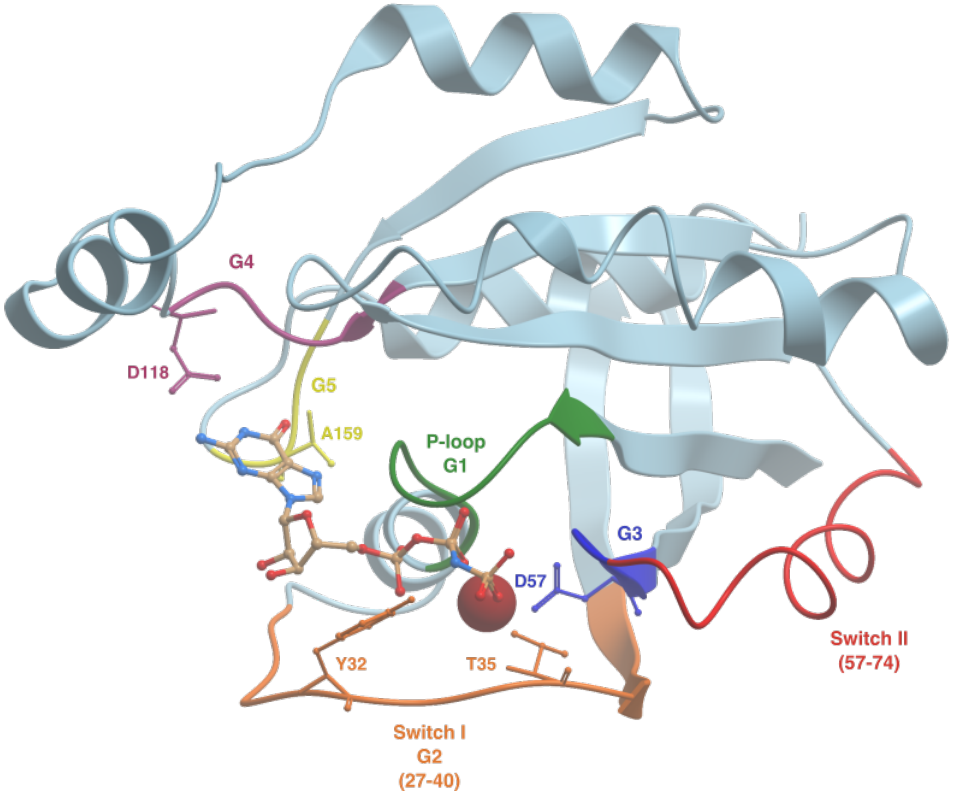


Figure 13. The conserved G-domains of the Ras superfamily, as shown on Rac1 bound to non-hydrolysable GTP analogue GNP (brown) (PDB code: 3TH5). The numbering for switches I and II refers to the residue numbers of Rac1. Key residues are shown as sticks, and the magnesium ion as a red sphere. Y32 and T35 of the switch I loop and D57 of the G3 motif are essential for their interactions with the γ-phosphate of GNP or GTP. Their orientation differs when GDP is bound.[102]

The GDP and GTP bound forms of the GTPases are conformationally distinct in the switch I and II regions.[103] Whilst the GDP-bound form varies between the subfamilies, the GTP-bound state is remarkably similar (Figure 14). It is postulated that in the GTP-bound state, the switch loops are held in a specific conformation due to the interactions of T35 and

G60 with the γ-phosphate of the GTP, which is recognised by effectors. However, even in the GTP-bound form, there are two 'states' of the switch conformations. 'State 2' refers to a closed, ordered conformation favourable for effector binding, whilst 'state 1' refers to a GTP-bound conformation where there is some flexibility in the switch I loop. The 'state 1' conformation of switch I is not conducive to effector binding and more similarly resembles the GDP-bound orientation than GTP-bound 'state 2' (Figure 14). Upon GTP hydrolysis to GDP, these interactions no longer occur; the switch regions are far more dynamic and flexible, leading to variations in their conformations.

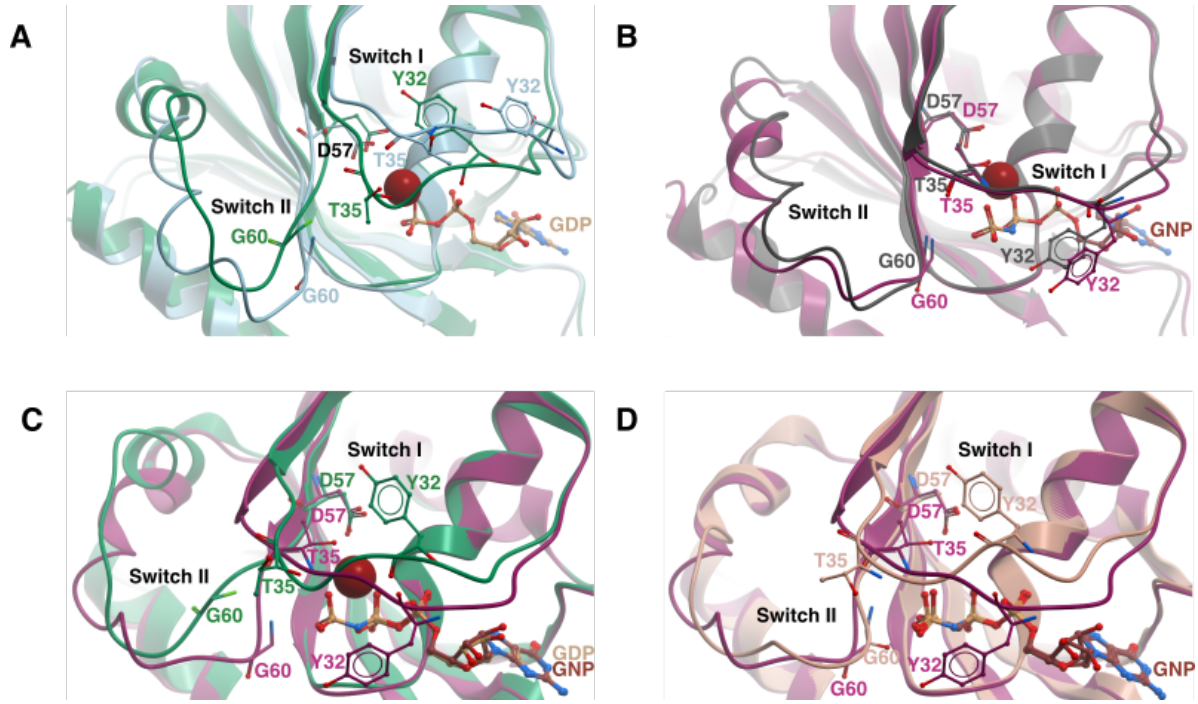


Figure 14. **A**. Comparison of Rho GTPase Rac1 (light blue, PDB code: 5N6O) and Ras GTPase HRas (green, PDB code: 4Q21) bound to GDP. The switch regions are significantly different between the subfamilies in the GDP-bound state. **B**. Comparison of Rac1 (grey, PDB code: 3TH5) and HRas (purple, PDB code: 4Q21) bound to GNP, a non-hydrolysable GTP analogue. The switch II region in both proteins is relatively ordered compared to the GDP bound state. The key residues in the switch I region that interact with the γ-phosphate of GTP, such as T35 and D57, are in the same orientation. Overall, the GTP-bound conformations of the switch regions are more similar than their GDP-bound counterparts when comparing between subfamilies. **C**. Comparison of GDP-bound HRas (green) and GTP-bound HRas (purple). Both switch I and II loops are more disordered in the GDP-bound state

compared to the GTP-bound form. **D**. Comparison of GTP-bound HRas in the ordered 'state 2' conformation (purple) versus GTP-bound HRas in the 'state 1' conformation (light pink, PDB code: 6Q21, chain D). The 'state 1' conformation can be seen due an artefact resulting from the crystal space group. Whilst the switch II region remains relatively ordered in both conformations, key residues in switch I such as Y32 are orientated differently, more similar to the GDP-bound orientation, despite being bound to GTP.

1.3.2 Guanine Nucleotide Exchange Factors (GEFs)

The small GTPases bind to the guanine nucleotides with picomolar binding affinity.[104,105] The intrinsic exchange from the inactive GDP-bound form to the active GTP-bound form of the GTPase, allowing for effector binding, is too slow for rapid control of cellular processes. Conversely, the rate of inherent hydrolysis of GTP to GDP is too slow and could result in constitutively active GTPase without any external input. Hence, the activation cycle of GTPases is tightly regulated by guanine nucleotide exchange factors (GEFs), GTPase activating proteins (GAPs) and, for the Rho and Rab subfamilies, guanine nucleotide dissociation inhibitors (GDIs) (Figure 15).

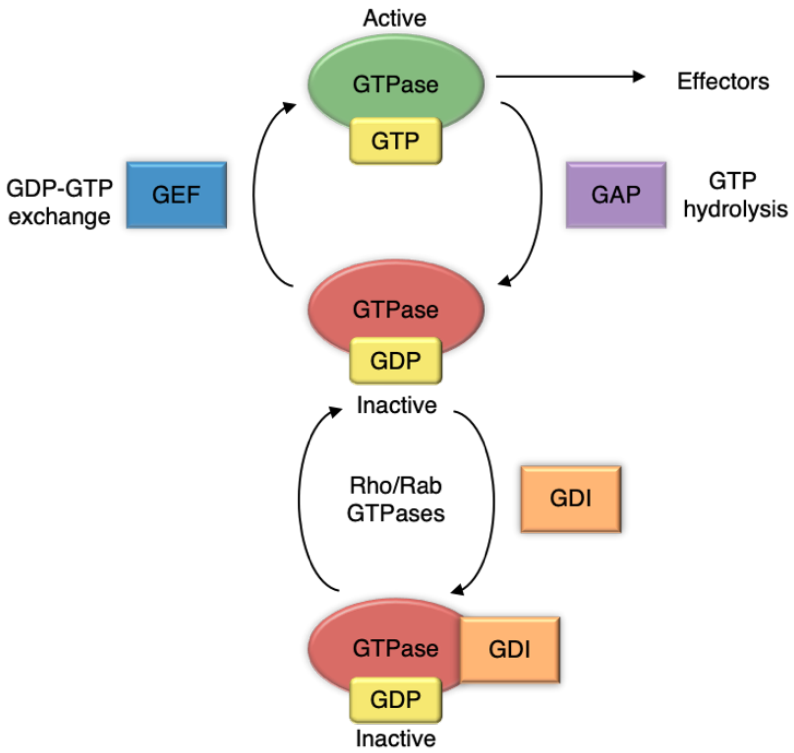


Figure 15. Activation cycle of the small GTPases is controlled by GEFs, GAPs and, in certain subfamilies, GDIs. Reproduced from Gray *et al.*[96]

GEFs bind to and catalyse the dissociation of GDP from the GTPase, allowing for GTP binding and therefore activation of the small GTPase. Conversely, GAPs act as negative regulators of the cycle and promote the hydrolysis of GTP to form GDP and a free phosphate group, leading to the inactive conformation of the GTPase. GDIs can bind to the inactive form of Rho or Rab proteins and sequester them from cellular membranes, preventing interaction with effector molecules or allowing the dissociation of GDP, thereby providing an extra level of control over the activities of these subfamilies.

Each GTPase subfamily is associated with its own subset of GEFs (Table 2). The number of GEFs for each subfamily differs; the Ras subfamily (36 members) is larger than its GEF family (27 members), which are characterised by their Cdc25 domain, whereas the Rho GTPases are outnumbered by their GEFs by a factor of three, allowing for a rigorous spatial control over their activity. The Rho and Arf families are both associated with two distinct sets

of GEFs characterised by different domains. The Dbl-homology (DH) and DOCK-homology region (DHR) GEF families are associated with the Rho subfamily of GTPases; the Sec7 and Sec12 GEFs bind to the Arf GTPases. Ran has a single GEF, RCC1, whereas the Rab subfamily has been linked to as many as 4 different subsets of GEFs, with DENN and Vsp9 proteins binding to a variety of Rab GTPases, whilst the Sec2 and TRAPP proteins are more selective with their Rab binding partners.[106] Interestingly, there are examples of some GEFs which contain two different subfamily domains, such as Sos, which contains a Cdc25 and DH domain.

Table 2. The types of GEFs associated with each subfamily in the human genome. Details from references.[106–109] *Varying numbers reported/not yet identified.

GTPase subfamily	GEF domain	Number of GEFs	GEF/GTPase in PDB
Ras	Cdc25	27	8
Rho	Dbl homology (DH)	> 70	18
	DOCK homology (DHR)	11	5
Rab	Vsp9	10	2
	DENN	18	2
	Various (unrelated)	*	3
Ran	RCC1	1	1
Arf	Sec7	16	5

Whilst the precise mechanism of activation of the small GTPases by GEFs varies between subfamilies, some general similarities can be identified (Figure 16). The GEF will bind to the GDP-bound GTPase, creating a low-affinity complex. GEF binding induces conformational changes in the switch I loop through steric hindrance. A relatively large interface between the GEF and GTPase switch II region stabilises the GTPase and prevents

it from unfolding when guanine nucleotides are not bound. Specific residue interactions within the interface allow for recognition between GEF/GTPase binding partners for formation of the correct complexes. The GEF reduces the affinity of the GTPase for GDP by inserting residues that interfere with the interactions of the GDP phosphate groups with the binding site and/or the magnesium ion bound to the GTPase, or by rearranging the switch regions to destabilise nucleotide binding. Upon dissociation of the nucleotide, the complex exists as a nucleotide-free, high-affinity GEF/GTPase complex. GTP is present in ten times the concentration in the cell as so is more likely than GDP to re-associate with the complex. When GTP binds, the complex rearranges to a low-affinity state between the GEF and GTPase. Eventually the GEF will be released, resulting in the active GTP-bound form of the GTPase. Compared to the intrinsic rate of exchange, GEFs can catalyse GDP/GTP exchange by several orders of magnitude.[96,106,109]

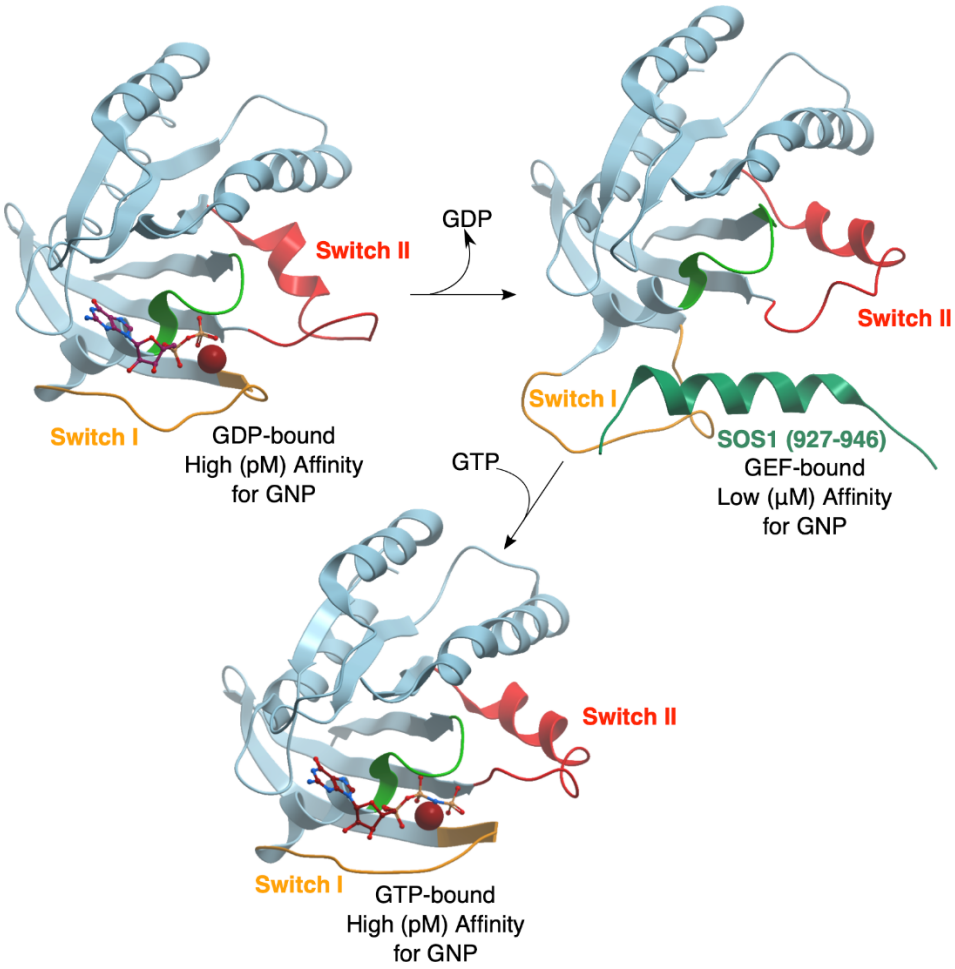


Figure 16. General mechanism and conformational changes during the exchange of GDP for GTP catalysed by the GEFs. Switch I is shown in orange, switch II in red, P loop in green, GDP/GTP as purple sticks. GNP = GDP or GTP. In the GDP bound form of Ras (PDB code: 4Q21), the switch II region is relatively disordered. Upon Sos1 binding to form the nucleotide free GEF/GTPase complex (PDB code: 1NVU), the switch I loop moves significantly and the switch II loop is more ordered. When GTP binds (PDB code: 3L8Z), the switch I loop moves back but the switch II loop remains helical. Replicated from Gray *et al.*[96]

1.3.3 Small Molecule Drug Discovery Efforts for the Ras Superfamily

1.3.3.1 Historical Drug Discovery Efforts

Due to their vital roles in regulating essential cellular processes, dysregulation of GTPase activities has been linked to many forms of disease. The most famous example of this are the Ras proto-oncogenes KRas, HRas and NRas, which are mutated in 20-30% of cancers.[110,111] A common mutation in cancer involves the mutation of a key residue (G12) used by the GAP protein to catalyse hydrolysis of the GTP to GDP; without the correct amino acid, the GTPase remains constitutively active, driving the formation of cancer cells.[112] Whilst the isoforms of Ras are the best represented in the literature, other members of the superfamily have also been linked to disease including cancer, neurodegenerative disorders and autoimmune conditions.[113–116] Additionally, abnormal regulation of the activation of Ras proteins controlled by GEFs, GAPs and GDIs can result in disease; overexpression or mutations in the regulatory proteins have been linked to many types of neurodegenerative diseases and cancer.[117–121] Ras superfamily proteins have also been identified as important biomarkers in disease types, but research has been hampered by a lack of suitable chemical probes.[122–124] Their essential roles in the cell and ubiquitous presence has resulted in the Ras superfamily being extremely important targets in both drug discovery and the design of chemical probes.

Despite this longstanding interest of the Ras superfamily, little chemical matter of note has been forthcoming until recently; indeed, for a long time the Ras proteins were considered ‘undruggable’. Medicinal chemists originally intended to synthesise competitive inhibitors for GDP in a manner similar to the ATP-competitive inhibitors successfully produced for kinases.[125] However, this method of targeting the small GTPases was unsuccessful.[126] The GTPases have extremely high affinity for their nucleotides, in the picomolar range, in comparison to the kinases’ millimolar affinity for ATP, which in combination with the high

concentrations of GDP (> 30 μM) and GTP (> 300 μM)[127] in the cell rendered attempts at developing potent or selective inhibitors at the nucleotide binding site to be fruitless.[105,128]

Other attempts to target the superfamily were undertaken in a less direct manner. Farnesyltransferase inhibitors (FTIs) and geranylgeranyltransferase inhibitors (GGTIs) were developed to inhibit the association of Ras GTPases with the plasma membrane. The farnesyltransferases and geranylgeranyltransferases are responsible for the post-translational modification of the C-terminus, which allows for the association with the plasma membrane and is essential for Ras activity. Despite promising success *in vitro* and *in vivo* for FTI inhibitors, these compounds were ultimately unsuccessful in clinical trials due to the inaccuracy of these preclinical models, off-target inhibition and toxicity effects.[129–132]

It was also hoped that the inhibition of proteins downstream of the small GTPases may be of success. Indeed, some kinase inhibitors in the Ras signalling pathway, such as sorafenib, which inhibits c-Raf – an effector of Ras - have been approved for clinical use. However, these compounds are associated with side effects due to a lack of selectivity, the development of drug resistance and the issues associated with complex feedback mechanisms.[105,129,133] Certainly, whilst there is some evidence that inhibition of c-Raf provides clinical benefit in hepatocellular carcinoma, in all other approved uses of sorafenib it appears that inhibition of other kinases is responsible for its therapeutic effect.[105] There are also reports of severe side and adverse effects and the development of resistance in patients.[134]

1.3.3.2 Targeting the G12C Mutation of Ras

However, the recent development of covalent inhibitors which target a specific glycine to cysteine mutation in the guanine nucleotide binding site in KRas G12C has seen success. The KRas G12C mutation is found in ~7% of lung adenocarcinomas and 3-5% of colorectal cancers.[135] The glycine at position 12 is essential for the hydrolysis of GTP to GDP by the necessary GAP protein; the larger cysteine blocks the insertion of a key arginine finger by the

GAP protein.[118,136] This mutated KRas form was identified as theoretically tractable for drug development, with the cysteine residue as a handle for a targeted covalent inhibitor and preventing the formation of the active form of the GTPase.

Shokat and co-workers[137] discovered a series of inhibitors from a fragment screen which bound to C12 in an allosteric pocket underneath the switch II loop and forced a conformational change in the switch I region (Figure 17). Elaboration using structure-based drug design yielded ARS-853 (**20**), which had a cellular engagement IC_{50} of 1.6 μM and could attenuate Sos-catalysed nucleotide exchange. Unsuitable pharmacokinetic properties resulted in the development of ARS-1620 (**21**). This had an IC_{50} of 120 nM and was the first molecule to demonstrate *in vivo* efficacy in a mouse model.[138] Based on the success of **21**, the company Wellspring Biosciences was started with the aim of developing a targeted covalent therapeutic and a compound, ARS-3248, was recently entered for Phase I clinical trials.[139] Zeng *et al* also used the quinazoline series as a basis to develop compounds which could bind both to switch II and the GDP binding site, and managed to synthesise potent inhibitors capable of causing a novel conformation of the switch II loop.[140]

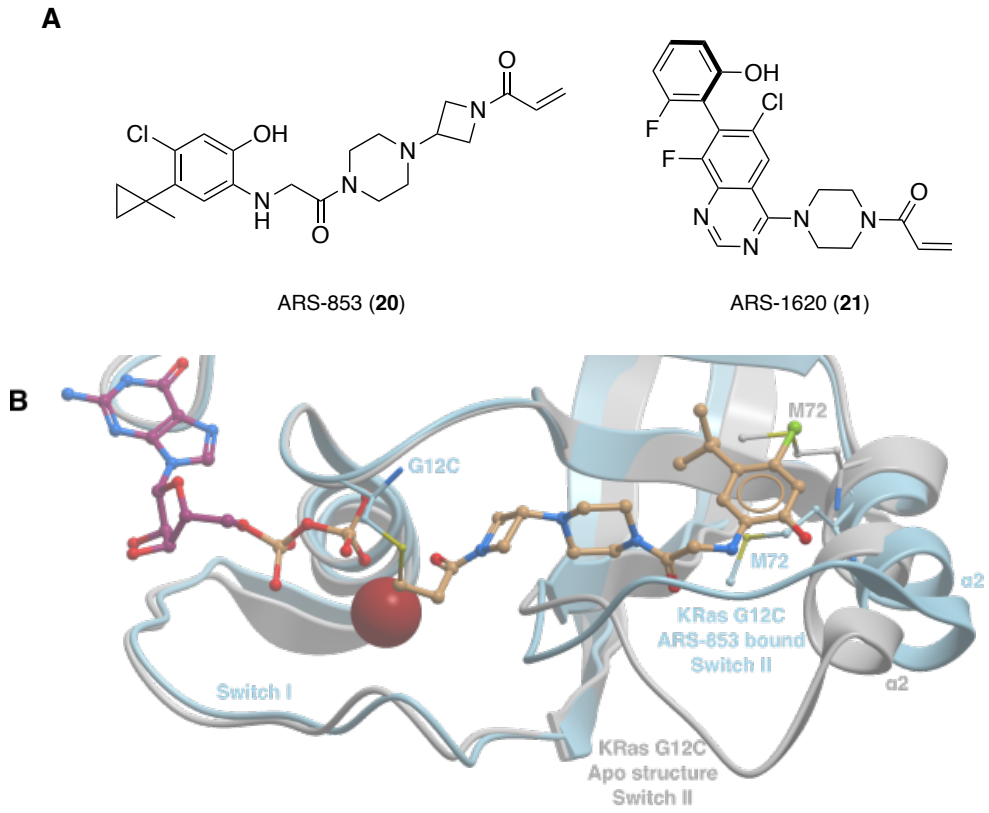


Figure 17. **A**. The first KRas G12C covalent inhibitors. **B**. The binding of **9** to KRas causes M72 and the α2 helix of switch II to move. The acrylamide warhead clashes with the γ-phosphate of GTP, forcing KRas G12C to favour the GDP-bound form and preventing nucleotide exchange. Replicated from Gray *et al*.[96]

At a similar time, Amgen collaborated with Carmot Therapeutics in the development of the acrylamide AMG-510 (sotorasib, **22**), which is currently in Phase II clinical trials.[141] They used the Chemotype Evolution platform developed by Carmot, which identifies a reactive 'bait' molecule, which interacts with the desired target but also contains a reactive functionality and combines this 'bait' with fragments to create a biased screen for the target.[142] Subsequent rounds of optimisation using SBDD eventually yielded a lead indole compound, **23**, which bound to a novel cryptic pocket on KRas G12C, and inhibited KRas in biochemical and cellular assays but was found to be metabolically unstable. Comparison of the binding modes of **21** and **23** led them to hypothesise that the quinazoline core of **21** could be modified to access the cryptic pocket identified by **23** (Figure 18A).[143] Several rounds of improvements involved

extending into the cryptic pocket, improving pharmacokinetic properties such as high oral bioavailability and cell permeability, and preventing atropisomer interconversion within the molecule by restricting rotation. These combined developments eventually lead to the discovery of **22** (IC_{50} = 68 nM, Figure 18B).

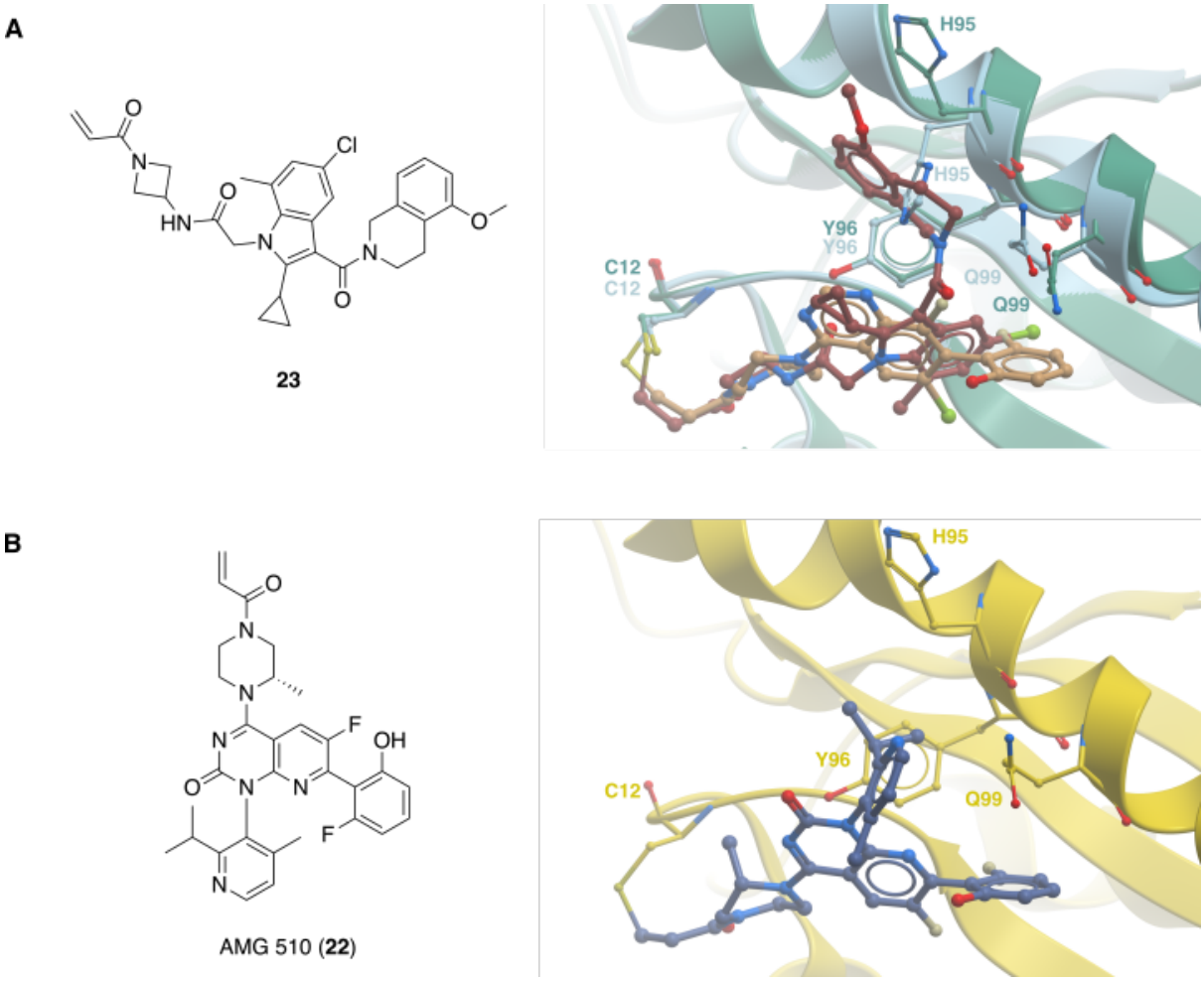
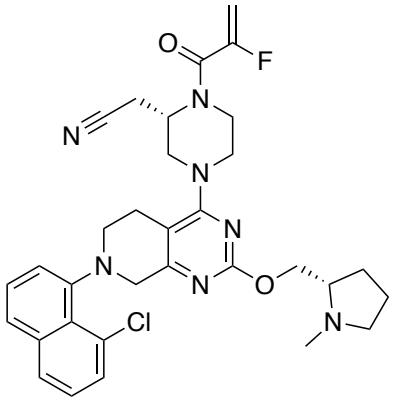


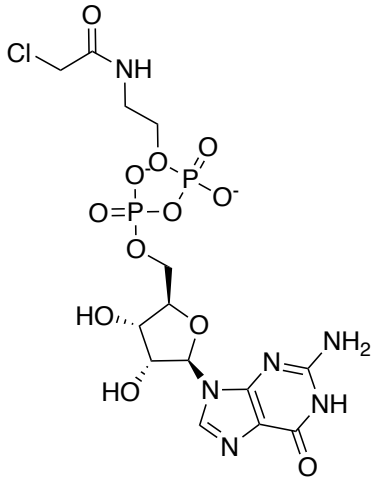
Figure 18. **A**. (L) The lead compound derived from the Chemotype Evolution Platform. (R) The overlay of ARS-1620 (light brown) bound to KRas G12C (light blue, PDB code 5V9U) and lead **23** (red) bound (green, PDB code 6P8Z). The key residues forming the cryptic pocket, H95, Y96 and Q99 are shown as sticks. H95 'flips up' in the **23** structure to allow for binding, thereby creating the cryptic pocket. **23** overlaps with the original position of H95 in the ARS-1620 bound structure. **B**. (L) Clinical candidate AMG 510. (R) **22** (dark blue) bound to KRas G12C (yellow, PDB code 6OIM). **22** remains more similar in scaffold to **21** but exploits the cryptic pocket configuration of KRas with the H95 residue shifted into the same position seen for **23**.

A third compound, MRTX849 (adagrasib, **24**), is also in clinical trials following a SBDD campaign by Mirati Therapeutics and is reported to covalently bind to the same cysteine residue and stabilise the inactive GDP bound state (Figure 19).[144,145] **24** has demonstrated antitumour activity in 26 KRas G12C cancer models, with significant regression of tumour size in 65% of the models, and shown promising clinical activity in Phase I/II trials.[146] **24** has also been used as a warhead in the first active PROTAC for KRas.[147,148] However, resistant mutants of KRas for this compound have already been identified in patients.[149]

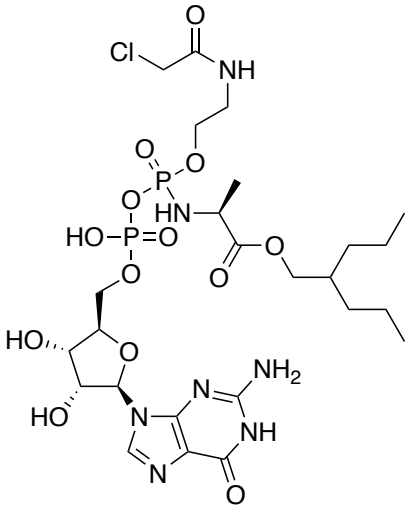
The Gray lab at the Dana-Farber Cancer Institute also targeted the G12C mutation, although their compound, an electrophilic analogue of GDP, binds within the GDP binding site rather than in an allosteric pocket. They hoped that a covalent nucleotide analogue might have an advantage over a non-covalent compound with irreversible or long-acting inhibition overcoming the issues associated with the high affinity of GTPases for their endogenous substrates. Their optimised compound, SML-8-73-1 (**25**) labelled 95% of KRas G12C at cellular levels of guanine nucleotides and stabilised the inactive form of Ras, with crystal structures showing a binding mode comparative to that of GDP (Figure 19).[150,151] However, the negatively charged phosphate groups meant the compound was unable to enter cells, so a 'prodrug' caged version, SML-10-70-1 (**26**, Figure 19), was developed where the phosphate groups are masked by a caged moiety to allow for cellular uptake. The caged moiety is typically cleaved enzymatically in the cell. **26** labelled KRas in the cell and showed anti-proliferative effects in three cancer cell lines with a mid-micromolar EC_{50}. However, no further updates since 2014 have been published, perhaps showing the difficulty in improving the potency or selectivity of direct GDP analogues.



MRTX849 (**24**)



SML-8-73-1 (**25**)



SML-10-70-1 (**26**)

Figure 19. Allosteric covalent inhibitor **24**, the GDP covalent analogue **25** and its caged derivative **26**.

1.3.3.3 Targeting Ras Protein-Protein Interactions

The method detailed above of targeting Ras exploits a cysteine mutation specific to KRas and so is not applicable to every member of the superfamily. Hence, attentions have turned to the development of inhibitors outside of the nucleotide binding pocket. Many attempts in the literature report the identification of inhibitors for the Ras GTPases with their GEFs and effectors. Several of these efforts have been hampered by improper chemical consideration, and little research has been devoted to the development of molecules targeting GAPs and GDIs.[96,128] However, the field has improved as of late, with recent successful

development of chemical probes by Bayer and Boehringer Ingelheim, giving traction to the idea of developing GEF/GTPase PPIs as drugs .

Hillig *et al*[152] produced a potent inhibitor with an IC_{50} of 21 nM for the KRasG12C-Sos interaction by combining hits found in a fragment and HTS screen (Figure 20). **27** stabilised the KRas-Sos complex in NMR and SPR. **28** was identified in a high-throughput fluorescent enzymatic assay of over three million compounds to inhibit the Sos-induced KRas activation with sub-micromolar affinity, with selectivity over other members of the Ras superfamily. Merging of these compounds and further elaboration eventually yielded BAY-293 (**29**) with the improved IC_{50} of 21 nM for the Sos-medicated activation of KRas. Whilst only demonstrated as a probe for *in vitro* experiments, promising *in vivo* antiproliferative data means the authors are hopeful for its use in animal models following further bioavailability data.

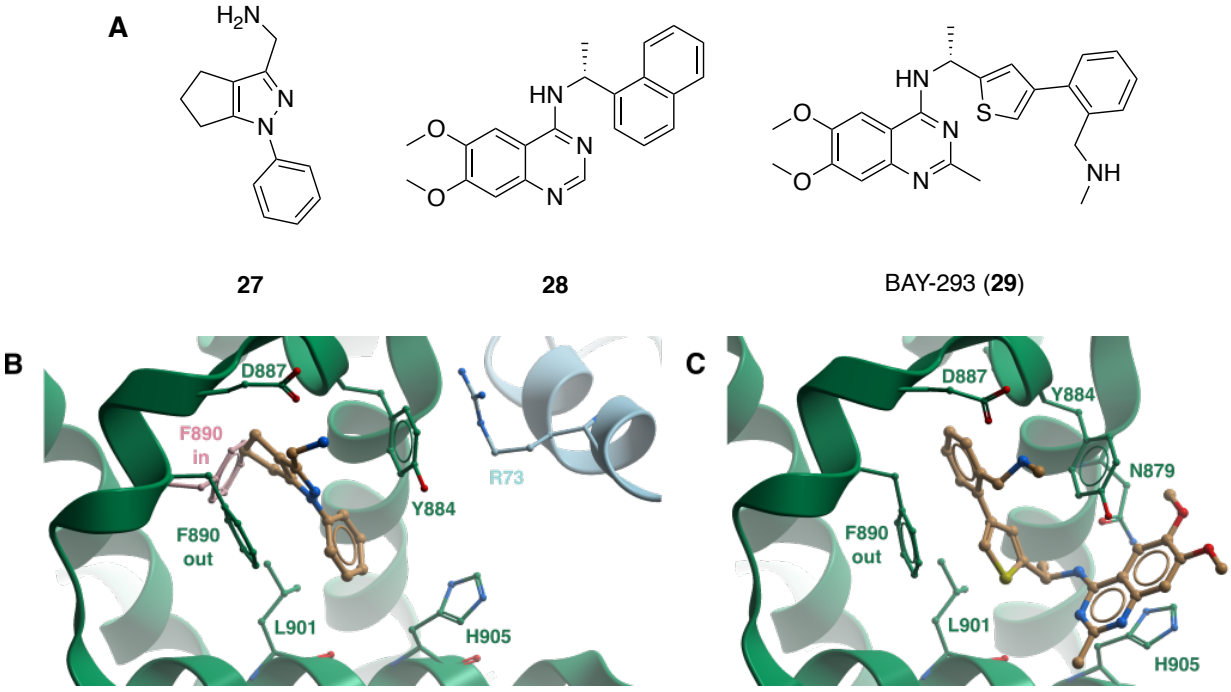
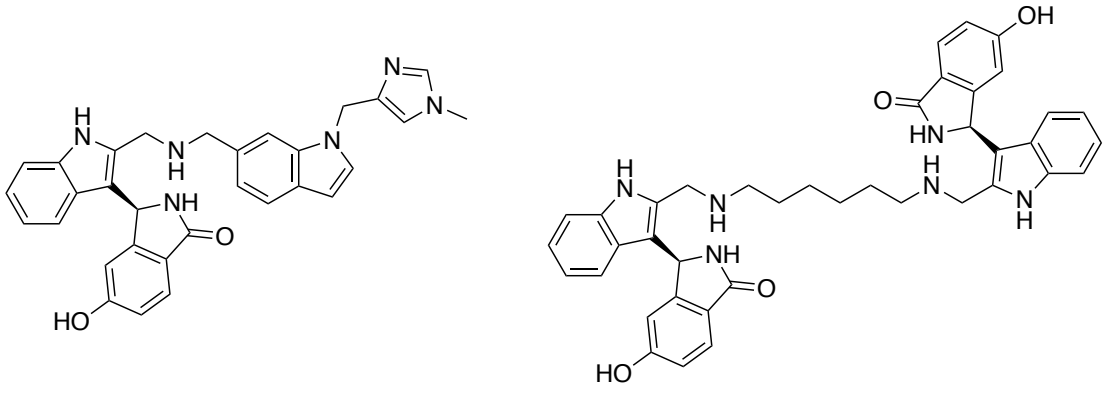


Figure 20. **A**. Compounds found to bind to Sos1. **B**. **27** bound to a pocket of Sos1 (green) adjacent to KRas (blue) (PDB code: 6EPM). **27** forms hydrogen bonds with Y884 and D887. **27** forces F890 into an 'out' conformation in comparison to the 'in' position of the apo Sos protein (pink, PDB code 6EPL). **C**. **29** binds to the same pocket as **27**, maintaining the F890 'out' conformation, bind expands into the neighbouring pocket occupied by **28**. Reproduced from Gray *et al*.[96]

Discovered by a 2D NMR fragment screening campaign and SBDD, BI-2852 (**30**) from Boehringer Ingelheim (BI) binds to GTP-bound KRas with nanomolar affinity (K_D = 750 nM, Figure 21). It also inhibits the interaction of KRas with its GEFs and effectors in a dose dependent manner at nanomolar concentrations.[153] **30** showed dose-dependent antiproliferative effects in cellular assays with an EC_{50} of 7 μM. It is hypothesised by the authors that this is due to a key hydrogen bond interaction between **30** and E37, a key residue for the interaction of Ras with its GEFs, GAPs and effectors. It was proposed separately that **30** may exert its effects by inducing a KRas dimer[154], although this was disproved by the original authors as an artefact of the latter experiment.[155] However, it was realised that the development of a compound which could stabilise a non-functional dimer of KRas would be an interesting therapeutic approach. Using the linker approach used in the development of PROTACs[156], they developed **31**, which is capable of dimerising KRas with a K_D of 3.8 μM. A therapeutic candidate, BI-1701963, has also been submitted by BI for clinical trials[157], and is proposed to be a pan-KRas inhibitor by selectively binding to Sos1 and preventing KRas activation by disrupting the PPIs. BI have published the results of a closely related analogue, which is also similar to Bayer compound **28**.[158]



BI-2852 (**30**) **31**

Figure 21. Compounds developed by Boehringer Ingelheim.

1.3.3.4 Future Developments in Ras Drug Discovery

Whilst this discussion has mainly examined the small molecule strategies targeting the Ras superfamily, there are examples of biologics designed to inhibit Ras action, with several examples of peptides designed to mimic the GEF reviewed in the literature.[96] However, for the most part, whilst they achieve higher potencies than their small molecule counterparts, achieving selectivity and favourable pharmacokinetic properties is more problematic. An anti-KRas antisense oligonucleotide clinical candidate from AstraZeneca, AZD4785, has been discontinued following Phase I clinical trials where it failed to show sufficient engagement of KRas and lowering of activity in cells.[159,160]

Whilst there is optimism for the success of compounds in clinical trials, there is still no universal mechanism for the development of therapeutics targeting the Ras superfamily. Many of the most promising candidates target cysteine mutations, which is not generally applicable to the entire superfamily. There has also been a bias in the literature for the three isoforms of Ras due to their importance in cancer, with comparatively little chemical matter targeting the other four subfamilies, or even other members of the Ras subfamily. With the development of FBLD and its application to PPIs and more challenging targets, it is hopeful that a direct inhibitor of a Ras protein, potentially by targeting the complex of the GTPase with a regulator or effector, is a promising possibility for future drug discovery efforts.

1.4 Project Outline

The majority of compounds developed to act as Ras GTPase inhibitors attempt to inhibit the GTPase alone, inhibit the effector, or act as PPIs of the complex. In general, due to the picomolar binding affinities of the guanine nucleotide and their high micromolar concentrations in the cell, the development of guanine nucleotide competitive inhibitors has been unsuccessful. However, we hypothesise that it may be possible to generate a competitive inhibitor of the GTPase in complex with its GEF.

The GEF catalyses the exchange of GDP to GTP by reducing the affinity of GTPase to the guanine nucleotides from picomolar to micromolar. The pathway from GDP to GTP occurs *via* a tightly binding guanine nucleotide free GEF/GTPase complex. This complex is very stable, with the GEF and GTPase binding with picomolar binding affinity, but with low micromolar affinity for the guanine nucleotides. We propose that it may therefore be possible to develop a competitive inhibitor for GDP and GTP targeting this nucleotide-free GEF/GTPase complex, as the development of a competitive inhibitor should in theory be easier due to the lower affinity for the endogenous substrates (Figure 22).

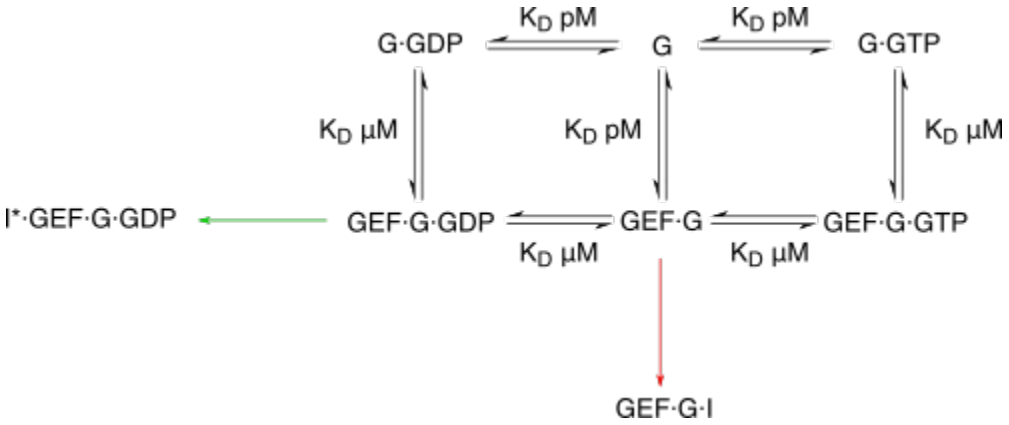


Figure 22. An overview of the GDP/GTP exchange reaction catalysed by GEFs. 'G' denotes GTPase. The GEF binding reduces the affinity of the GTPase to the GDP or GTP from picomolar to micromolar, thereby catalysing the exchange reaction. An inhibitor that attempts to compete with the GDP/GTP binding site for the GTPase alone would have to compete with picomolar affinities and high cellular

concentrations. An inhibitor, I (red arrow), which could bind and sequester the GEF/GTPase complex, only has to compete with micromolar affinity for the guanine nucleotides. An alternative would be a molecule, I* (green arrow) which modulates the GEF-GTPase protein-protein interactions. Adapted from Randazzo *et al.*[161]

This method would also have the added advantage of superior selectivity than an inhibitor targeting a GTPase alone. GTPases often have many distinct roles in the cells which are regulated spatially and temporally by GEFs. By targeting a single GEF/GTPase complex, an inhibitor would be able to isolate a small subset of GTPase activity, delivering greater effect and, in theory, fewer side-effects. This method of inhibition would also be applicable to the entire superfamily.

The proposal for this DPhil project was to solve the crystal structure of at least one GEF/GTPase complex, with the aim to use this complex for X-ray crystallography fragment screening and subsequent structure-based design. It was hoped that novel chemical matter binding to the GEF/GTPase complex within the canonical binding site would be identified, so that this may in future be optimised into a competitive inhibitor that could be used to explore the hypothesis that inhibition of the complex may be more successful than inhibition of a GTPase alone.

The GEF/GTPase complexes, with examples in the literature documenting straightforward purification in *Escherichia coli*, successful crystallisation and fragment screening campaigns, are ideal candidates for the XChem pipeline. The use of XChem would provide exquisite structural information to aid the difficult structure-based design and improvement in potency of the compounds targeting the complex. X-ray crystallography would also be particularly useful for GEF/GTPase complex screening as it would ‘trap’ the complex in the desired target state that may not be easily observed by other structural methods due to the fast nature of the GEF/GTPase activation cycle. The compounds identified could be activators or inhibitors, either binding in the guanine nucleotide binding pocket as detailed

above, or by binding to the interface of the complex and acting as PPI inhibitors. The potential of identifying PPI inhibitors is increased by using FBLD, which historically has had better success than HTS.[10]

This thesis details the evaluation of the kinetic parameters of the GEF-catalysed GTPase activation cycle to test the reasonability of the GEF/GTPase complex inhibitor hypothesis in Chapter 2. Chapter 3 describes the construct design, molecular cloning, protein expression and purification of complexes and crystallisation trials in the SGC. The chapter also discusses the crystal structures of three complexes and comparison to previously solved structures within the PDB. Chapters 4, 5 and 6 subsequently concentrate on the ongoing efforts to target a single GEF/GTPase complex, Kalirin/Rac1. Chapter 4 describes an XChem fragment screening campaign, with successive computational and chemical synthesis to generate follow-up compounds and bioisosteres. This chapter also depicts the development of a nucleotide exchange assay to measure compound potency. Chapter 5 records two subsets of covalent binders identified to target the nucleotide binding site of Kalirin/Rac1. Co-crystallisation of these compounds and the subsequent structure-based design of a lead compound is also shown. Finally, Chapter 6 documents the efforts to identify fragments binding to Kalirin/Rac1 with novel vectors for expansion into the nucleotide binding pocket using *in silico* and XChem screening, and the rational structure-based design which is in progress within the CMD (formerly SGC). Future work arising from this project is also considered in Chapter 7.

2. Kinetic Analysis of the GEF/GTPase System

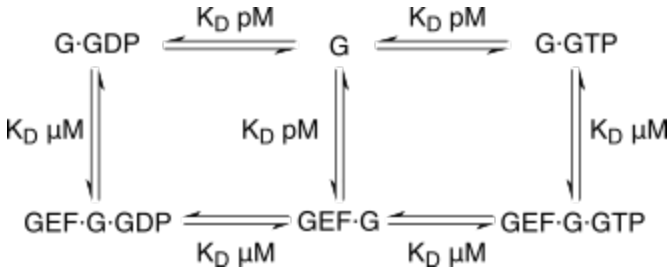
2.1 Overview

Before significant resources were committed to the design of novel competitive inhibitors targeting GEF/GTPase complexes, it was important to establish the validity of the proposed hypothesis. This chapter details the kinetic modelling used to predict whether an inhibitor with nanomolar affinity for a GEF/GTPase complex can inhibit the GEF-catalysed nucleotide exchange reaction using cellular data for Ras. Three models of reversible inhibitors targeting GTPases and their regulatory GEF complexes are presented. These models are then adapted to represent an irreversible, covalent inhibitor and its effect on the GEF-catalysed activation of a GTPase. The model predicted that a competitive inhibitor for a GTPase alone would not inhibit nucleotide exchange, in agreement with literature reports.[162] Conversely, both reversible and irreversible GEF/GTPase inhibitors were predicted to inhibit GDP/GTP exchange. The validity and key assumptions of the kinetic models are discussed.

2.2 Kinetic Modelling using KinTek Explorer

2.2.1 GEF-catalysed Nucleotide Exchange Model

A collaboration was instigated with Professor Roger Goody of the Max Planck Institute, Dortmund, Germany, an expert in GEF kinetics.[163–165] A thermodynamics diagram of the GEF/GTPase system was modelled in KinTek Explorer, an enzyme kinetics software designed to simulate complex kinetic models (Figure 23).[166] The model was designed to be physiologically relevant, with appropriate k_{on} and k_{off} values for each reaction derived from literature values (Figure 23).[163] The GTPase was modelled to have picomolar binding affinity for both GDP and GTP. The GEF was represented as exhibiting picomolar binding affinity in the absence of guanine nucleotides, and micromolar when they are bound, again reflective of the potency observed *in vivo*. The guanine nucleotides were depicted to have reduced affinity for the GTPase when it is in the GEF complex. For ease, the binding of effectors or GAP-stimulated hydrolysis of GTP that can occur following GEF-catalysed nucleotide exchange was not modelled as it was not expected to substantially affect the kinetic parameters of the exchange reaction.



Reaction	K_D (nM)	k_{on} ($\mu M^{-1}s^{-1}$)	k_{off} (s^{-1})
G + GDP ⇔ G·GDP	0.01	1	1×10^{-5}
G + GTP ⇔ G·GTP	0.01	1	1×10^{-5}
G·GDP + GEF ⇔ GEF·G·GDP	1000	1	1
GEF·G + GDP ⇔ GEF·G·GDP	10000	1	10
GEF·G + GTP ⇔ GEF·G·GTP	10000	1	10
G·GTP + GEF ⇔ GEF·G·GTP	1000	1	1
GEF + G ⇔ GEF·G	0.001	1	1×10^{-6}

Figure 23. GEF/GTPase activation cycle modelled in KinTek Explorer. Values of K_D are representative of literature values for the Ras superfamily. 'G' denotes GTPase.[161,163]

Relative concentration values of the GEF and GTPase were consistent with quantitative proteomic data for Ras (Table 3).[167] The model assumes that at t = 0, the GTPase is only in the GDP-bound state as no GEF-catalysed exchange has occurred. The rate of intrinsic exchange of the GTPase is very slow, so it was assumed that the concentration of the apo-GTPase or GTP-bound GTPase was 0 at t = 0. GTP was modelled at a ten-fold higher concentration to GDP.[127] As GDP and GTP both bind with picomolar binding affinity, the exchange for GDP to GTP is driven by the higher concentration of GTP in the cell, resulting in the active form of the GTPase.[104]

Table 3. Concentration of variables used in the kinetic model of the GEF-catalysed guanine nucleotide exchange reaction. 'G' denotes GTPase.

Variable	Concentration at t = 0 (μM)
G	0
GDP	30
GTP	300
GEF	0.01
G·GDP	2
G·GTP	0
GEF·G·GDP	0
GEF·G·GTP	0
GEF·G	0

The model was performed to predict the relative concentrations of GDP and GTP-bound GTPase over an hour. The model showed the exchange reaction to occur rapidly, with the level of activated GTP-bound GTPase to reach a maximum in around 30 minutes (Figure 24). This is consistent with the rapid exchange kinetics observed in cellular contexts and *in vitro* assays.[168–170]

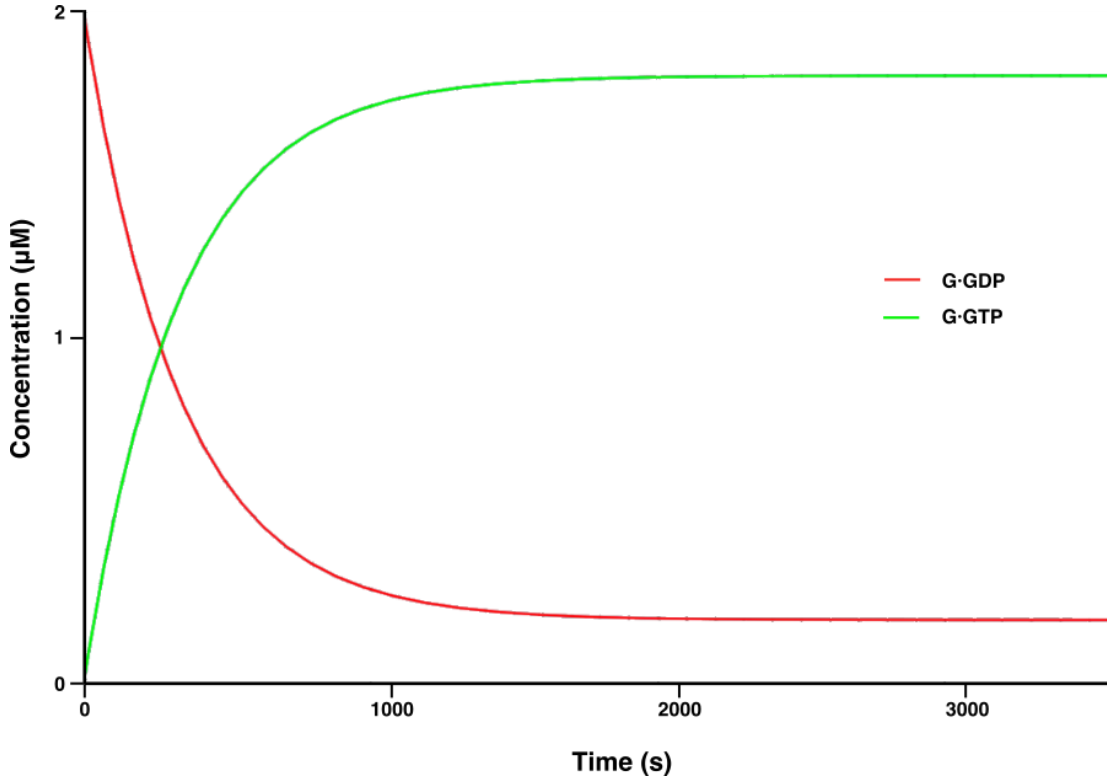


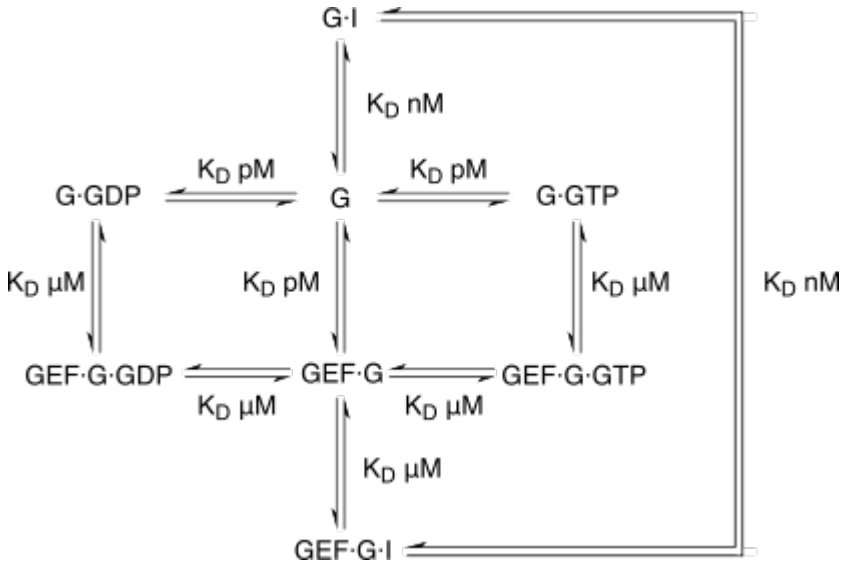
Figure 24. Outcome from KinTek Explorer of the GEF-catalysed activation of a GTPase. The concentrations of GDP-bound GTPase (red) and GTP-bound GTPase (green) are shown over 1 hour (3600 seconds). The model predicts rapid exchange of the guanine nucleotides.

With a representation of the GEF/GTPase cycle incorporated within KinTek Explorer, the model was then adapted to predict the effects of different classes of inhibitors.

2.2.2 'Classic' Competitive Reversible Inhibitors

The first type of inhibitor considered was the 'classical' reversible competitive inhibitor of the canonical binding site of GTPases, analogous to ATP-competitive inhibitors of kinases. Historically the development of these inhibitors were unsuccessful due to the picomolar affinity of GTPases for their guanine nucleotides.[96] Thus, it was predicted that these inhibitors would not induce a significant change in the rate of nucleotide exchange in the KinTek model.

The kinetic model was hence adapted to incorporate such a ligand (Figure 25). The ligand was modelled to inhibit the GTPase alone with a K_D of 10 nM and to be incapable of binding when either guanine nucleotide is present in the binding pocket. The inhibitor was depicted to minimally bind to the GEF/GTPase nucleotide-free complex, analogously to the guanine nucleotides. This was due to a predicted loss of affinity arising from GEF-catalysed movement of residues within the orthosteric pocket. Hence, it could also be assumed that the GEF would have lower affinity for the inhibitor-bound GTPase, comparable to its reduced affinity for guanine-nucleotide bound GTPase.



Reaction	K_D (nM)	k_{on} ($\mu M^{-1}s^{-1}$)	k_{off} (s^{-1})
G + I ⇔ G·I	10	1	0.01
GEF·G + I ⇔ GEF·G·I	10000	1	10
GEF + G·I ⇔ GEF·G·I	10	1	0.01

Figure 25. Modelling of a competitive inhibitor for a GTPase. The inhibitor is modelled with nanomolar affinity for the GTPase and is selective for the GTPase over the GEF/GTPase complex. 'G' denotes GTPase. 'I' denotes inhibitor.

As anticipated, the model predicted the inhibitor having minimal effect on the rate of nucleotide exchange at 1 μM or 1 mM, with the activated GTPase reaching a maximal concentration within an hour (Figure 26). Very high millimolar concentrations of a ligand are unlikely to be tolerated *in vivo*, indicating a nanomolar-affinity competitive inhibitor would be ineffective at targeting GTPases in the cell. The hypothetical inhibitor still had nominal effect if its affinity was increased to from nanomolar to picomolar, further demonstrating the futility of developing a guanine nucleotide competitive inhibitor of the apo-GTPase as a therapeutic strategy.

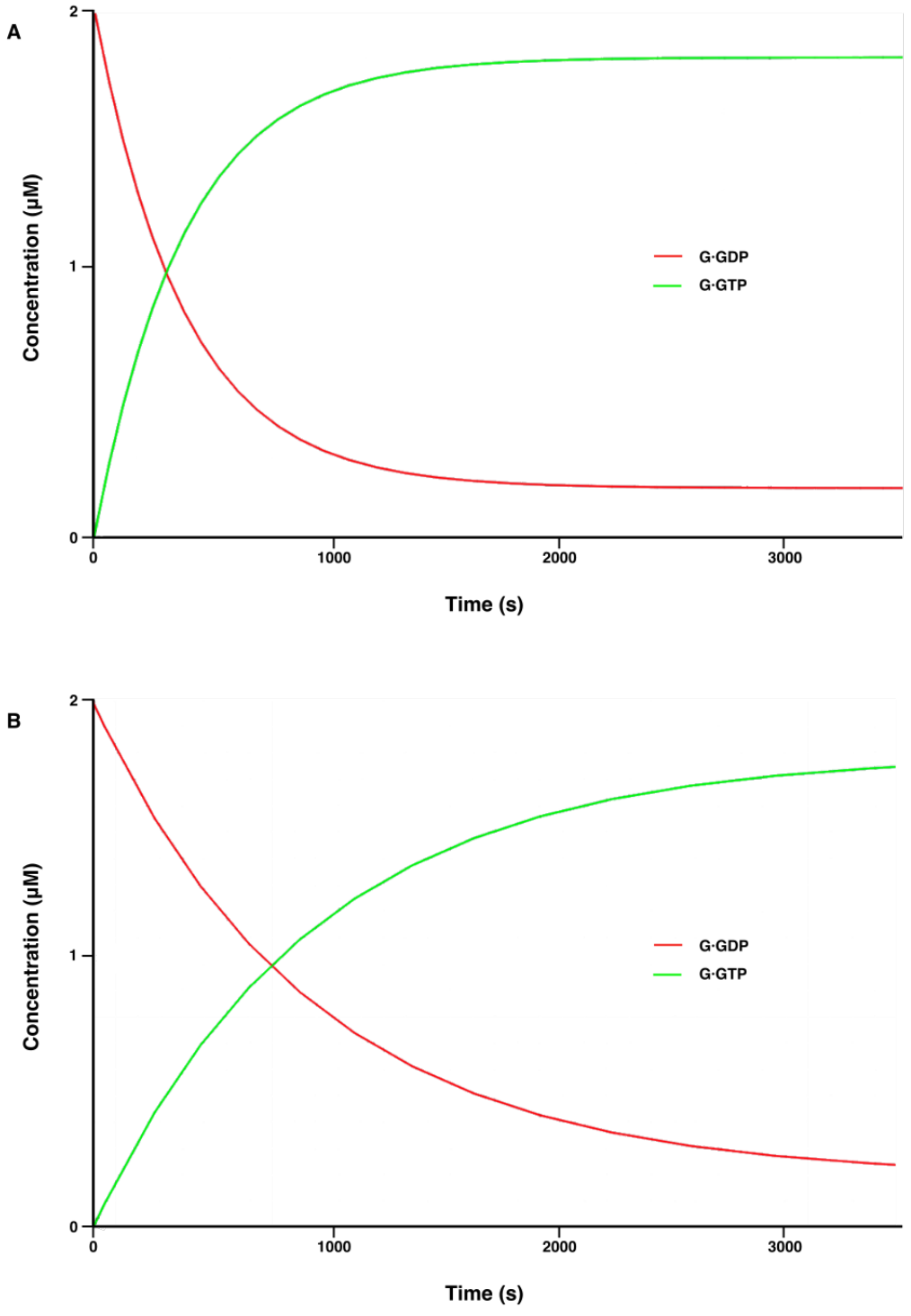
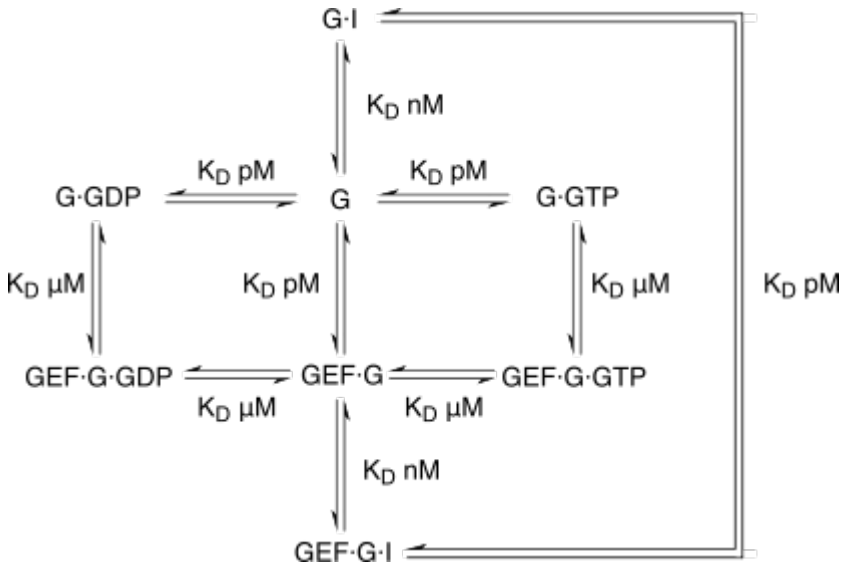


Figure 26. Outcome from KinTek Explorer of a nanomolar GTPase-selective competitive inhibitor. The concentrations of GDP-bound GTPase (red) and GTP-bound (green) are shown over 1 hour (3600 seconds. **A**. Concentration of inhibitor set as 1 μM at t = 0. No distinct change is seen compared to the activation cycle with no inhibitor modelled. **B**. Concentration of inhibitor set as 1 mM at t = 0. There is a small retardation of the exchange reaction.

2.2.3 GEF/GTPase Complex Reversible Inhibitors

2.2.3.1 Non-selective Inhibitor for GTPase and GEF/GTPase Complex

With the model validated against known ineffectual inhibitors, competitive inhibitors capable of targeting the GEF/GTPase complex were next considered. First modelled was a non-selective nanomolar inhibitor for the GTPase and the GEF/GTPase complex (Figure 27).



Reaction	K_D (nM)	k_{on} ($\mu M^{-1}s^{-1}$)	k_{off} (s^{-1})
G + I ⇔ G·I	1	1	0.001
GEF·G + I ⇔ GEF·G·I	1	1	0.001
GEF + G·I ⇔ GEF·G·I	1×10^{-3}	1	1×10^{-6}

Figure 27. Modelling a non-selective competitive inhibitor of the GTPase and GEF/GTPase complex. 'G' denotes GTPase. 'I' denotes inhibitor.

The model assumed that the inhibitor does not adversely affect the binding of the GEF to the GTPase. This was a reasonable assumption based on the equal affinity of the inhibitor for both the GTPase alone and the GEF/GTPase complex; if it reduced the affinity of the GEF for its GTPase, it would presumably bind to the complex with lower affinity than to the GTPase.

The graph generated indicated that the dual affinity of an inhibitor had a significant impact on the predicted rate of nucleotide change. A concentration of 1 μM of a nanomolar affinity inhibitor significantly slowed the rate of exchange, and 10 μM was associated with minimal GTPase activation (Figure 28). Monitoring the GEF concentration during this reaction showed it gets trapped in the inhibitor bound GTPase complex. The lower concentration of the GEF compared to GTPase means once trapped in the GTPase complex, there is insufficient free GEF remaining to catalyse the exchange reaction, so the GTPase remains in the GDP-bound inactive state. This is promising data that the development of a GEF/GTPase complex inhibitor could impact the exchange reaction in a cellular context.

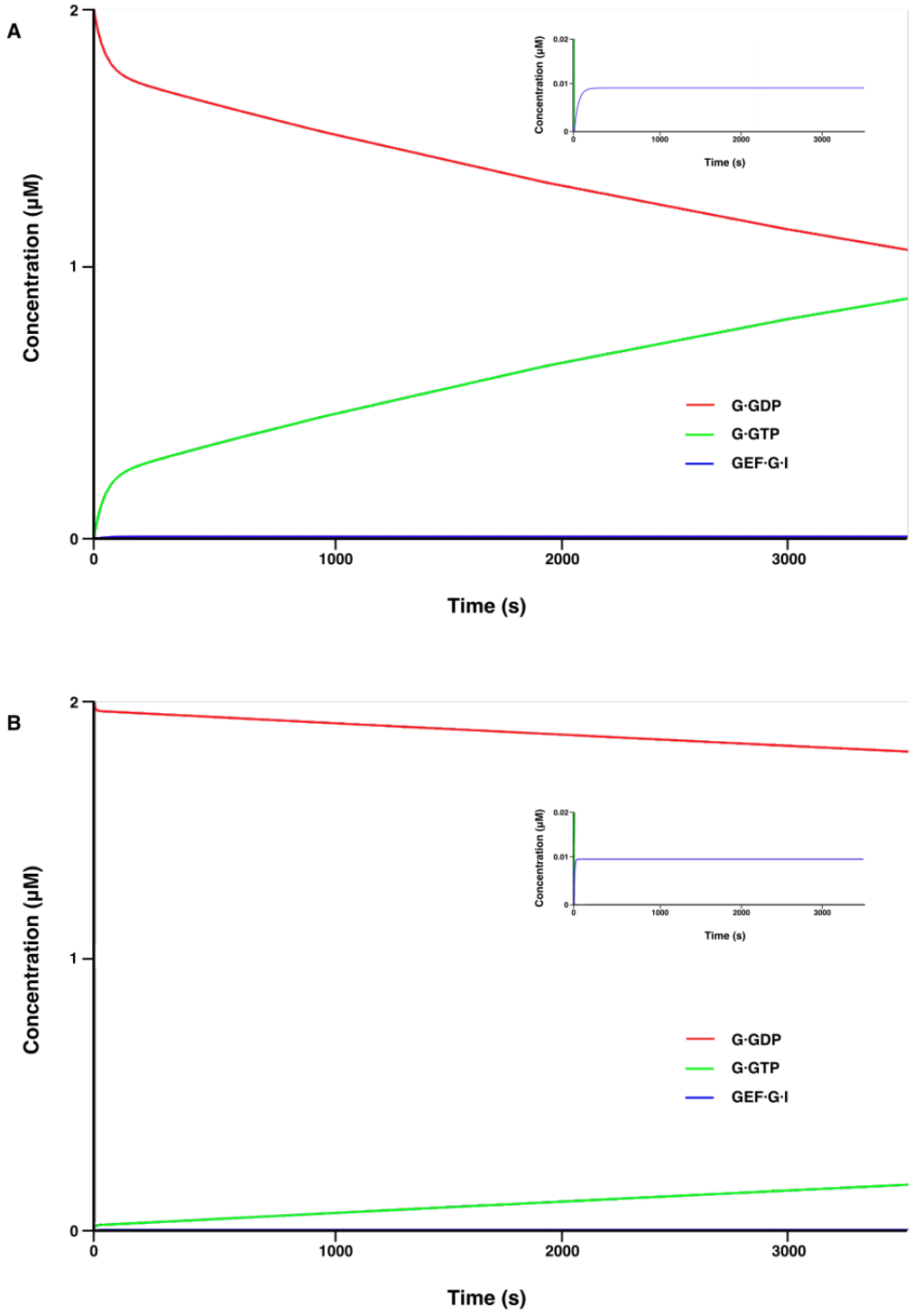
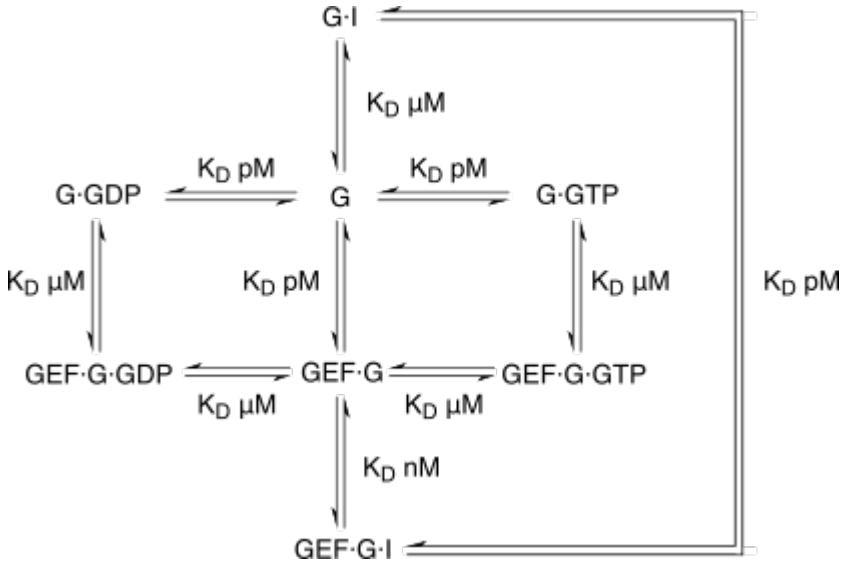


Figure 28. Outcome in KinTek Explorer of a non-selective competitive inhibitor for the GTPase alone and in the GEF/GTPase complex. The concentrations of GDP-bound GTPase (red) and GTP-bound (green) and the inhibitor-bound GEF/GTPase complex (blue) are shown over 1 hour (3600 seconds). **A**. Concentration of inhibitor is set at 1 μM at t = 0. Rate of GDP/GTP exchange is slowed. **B**. Concentration of inhibitor is set at 10 μM at t = 0. Minimal exchange occurs over the course of an hour.

2.2.3.2 Selective Inhibitor for a GEF/GTPase Complex

GEFs catalyse the dissociation of the guanine nucleotides in a subfamily-dependent mechanism. Generally, this involves the insertion of residues to displace the Mg^{2+} ion, reducing interactions with the phosphate groups of guanine nucleotides, and rearrangement of the switch I and II loops.[106] If an inhibitor does not extend into the area which binds to the phosphate groups and/or interacts with the rearranged residues within the canonical binding pocket of the GEF/GTPase complex, then it in theory would have greater affinity for the complex over the GTPase alone. Depending on the orientation of the GEF with respect to the guanine nucleotide binding pocket, an inhibitor might also be able to interact with residues of the GEF, again introducing selectivity for the complex. Thus, it is entirely conceivable that an inhibitor could be designed with higher affinity for the GEF/GTPase complex.

An inhibitor that was selective for the GEF/GTPase complex with nanomolar affinity was thus modelled (Figure 29).



Reaction	K_D (nM)	k_{on} ($\mu M^{-1}s^{-1}$)	k_{off} (s^{-1})
G + I ⇔ G·I	1000	1	1
GEF·G + I ⇔ GEF·G·I	1	1	0.001
GEF + G·I ⇔ GEF·G·I	1×10^{-3}	1	1×10^{-6}

Figure 29. Modelling of a selective inhibitor for the GEF/GTPase complex. 'G' denotes GTPase. 'I' denotes inhibitor

A micromolar dissociation constant was included for the GTPase, as it was presumed that even with selectivity for the GEF complex, the ligand would still be able to bind to the endogenous pocket of the free GTPase. The effect was predicted to be similar to the non-selective inhibitor in KinTek Explorer, with the GEF isolated in the GEF·G·I complex and therefore not available to catalyse nucleotide exchange (Figure 30).

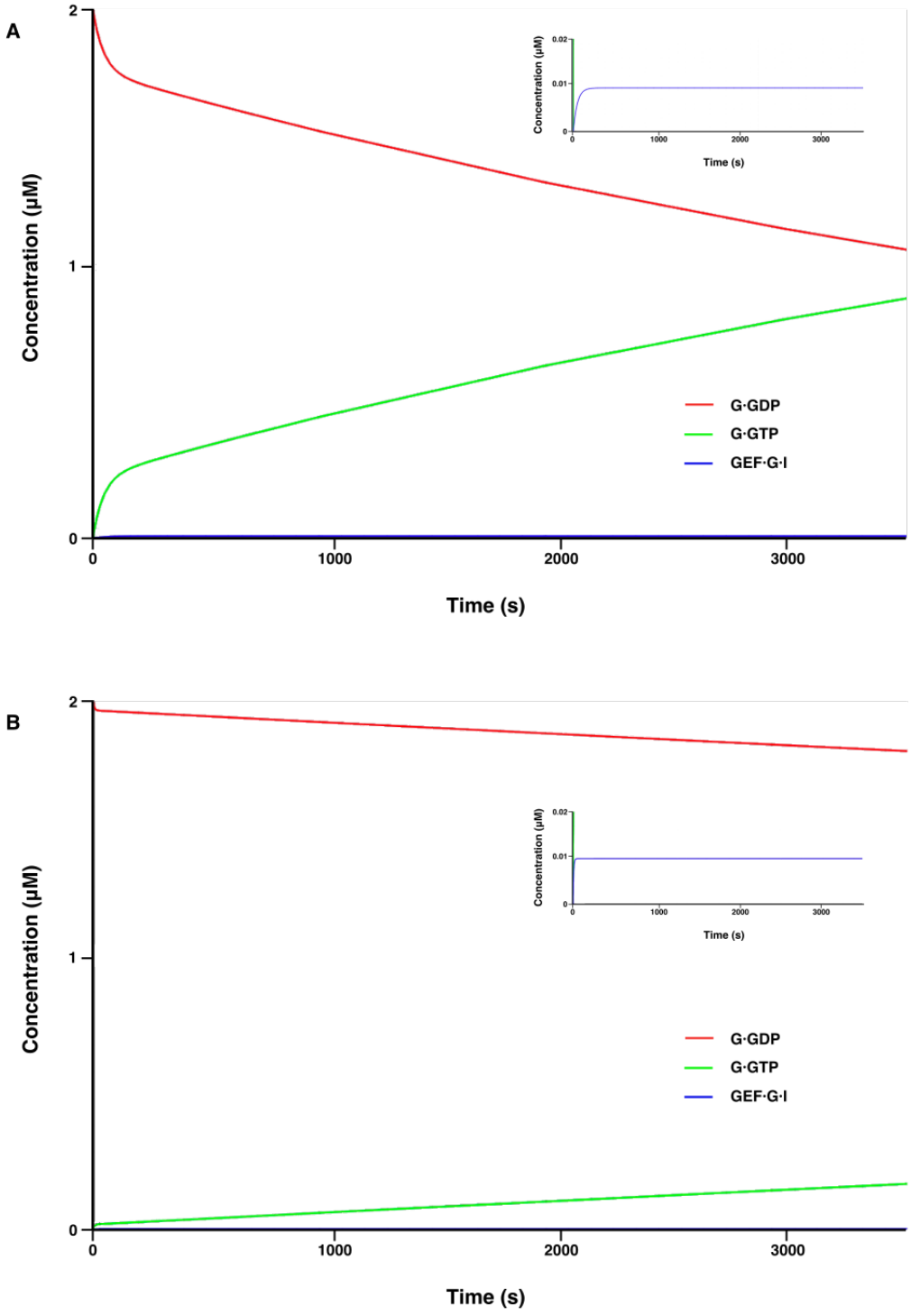


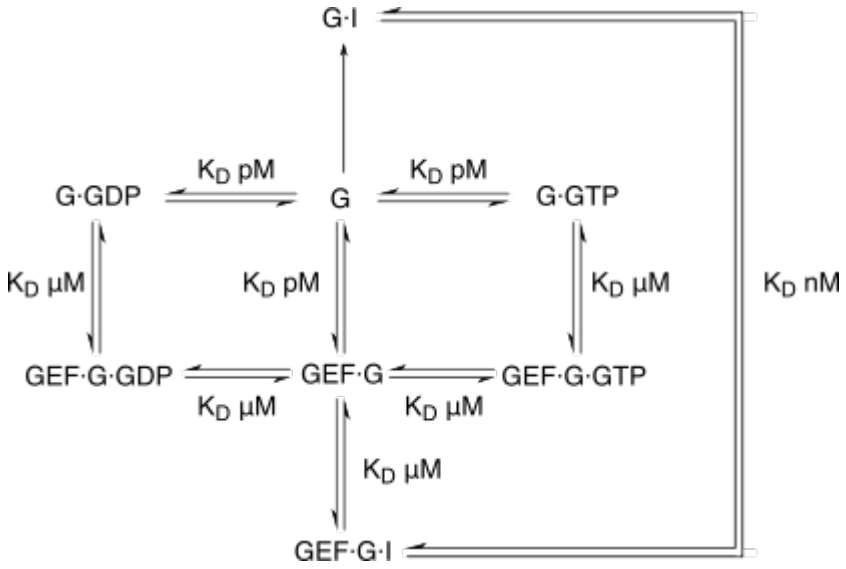
Figure 30. Outcome in KinTek Explorer of a selective nanomolar competitive inhibitor for the GEF/GTPase complex. The concentrations of GDP-bound GTPase (red) and GTP-bound (green) and the inhibitor-bound GEF/GTPase complex (blue) are shown over 1 hour (3600 seconds). **A**. Concentration of inhibitor is set at 1 μM at t = 0. Rate of GDP-GTP exchange is slowed. **B.** Concentration of inhibitor is set at 10 μM at t = 0. Minimal exchange occurs over the course of an hour.

2.2.4 Competitive Covalent Inhibitors

It was anticipated that the generation of competitive covalent inhibitors selective for GEF/GTPase complexes could be achieved if suitable residues – such as cysteine or serine – were present in the nucleotide binding site. The successful covalent inhibitors for Ras/Sos1 detailed in section 1.2.3.2 inhibit GEF/GTPase complex formation rather than competing directly with the guanine nucleotides. It was therefore necessary to model the proposed inhibitors under physiological conditions to establish whether pursual of a covalent inhibitor was worthwhile. The experiments in KinTek Explorer were established analogously to the reversible inhibitors. The covalent bond formation was modelled using a very slow k_{off} rate to represent an irreversible inhibitor.

2.2.4.1 'Classic' Competitive Covalent Inhibitor

A covalent inhibitor selective for the GTPase over the GEF/GTPase complex was modelled analogously to the 'classic' reversible inhibitor modelled in section 2.2.2 (Figure 31). The inhibitor was modelled with micromolar affinity for the GEF/GTPase complex, with the hypothesis that GEF binding would be unfavourable for the correct orientation of the compound within the binding site, and the subsequent covalent bond formation.



Reaction	K_D (nM)	k_{on} ($\mu M^{-1}s^{-1}$)	k_{off} (s^{-1})
G + I ⇔ G·I	-	1	1×10^{-20}
GEF·G + I ⇔ GEF·G·I	10000	1	10
GEF + G·I ⇔ GEF·G·I	10	1	0.01

Figure 31. Modelling of a competitive covalent inhibitor for a GTPase. The inhibitor is modelled to irreversibly bind to the GTPase and is selective for the GTPase over the GEF/GTPase complex. 'G' denotes GTPase. 'I' denotes inhibitor.

Whilst there is a very subtle, minimal change in the maxima and minima being approached, the results shown in Figure 32 are very similar to the classic reversible inhibitor modelled in Figure 26. The modelling therefore again supports the hypothesis that the affinity of an inhibitor for the GEF complex, whether it be reversible or irreversible, is of paramount importance for potency under physiological conditions.

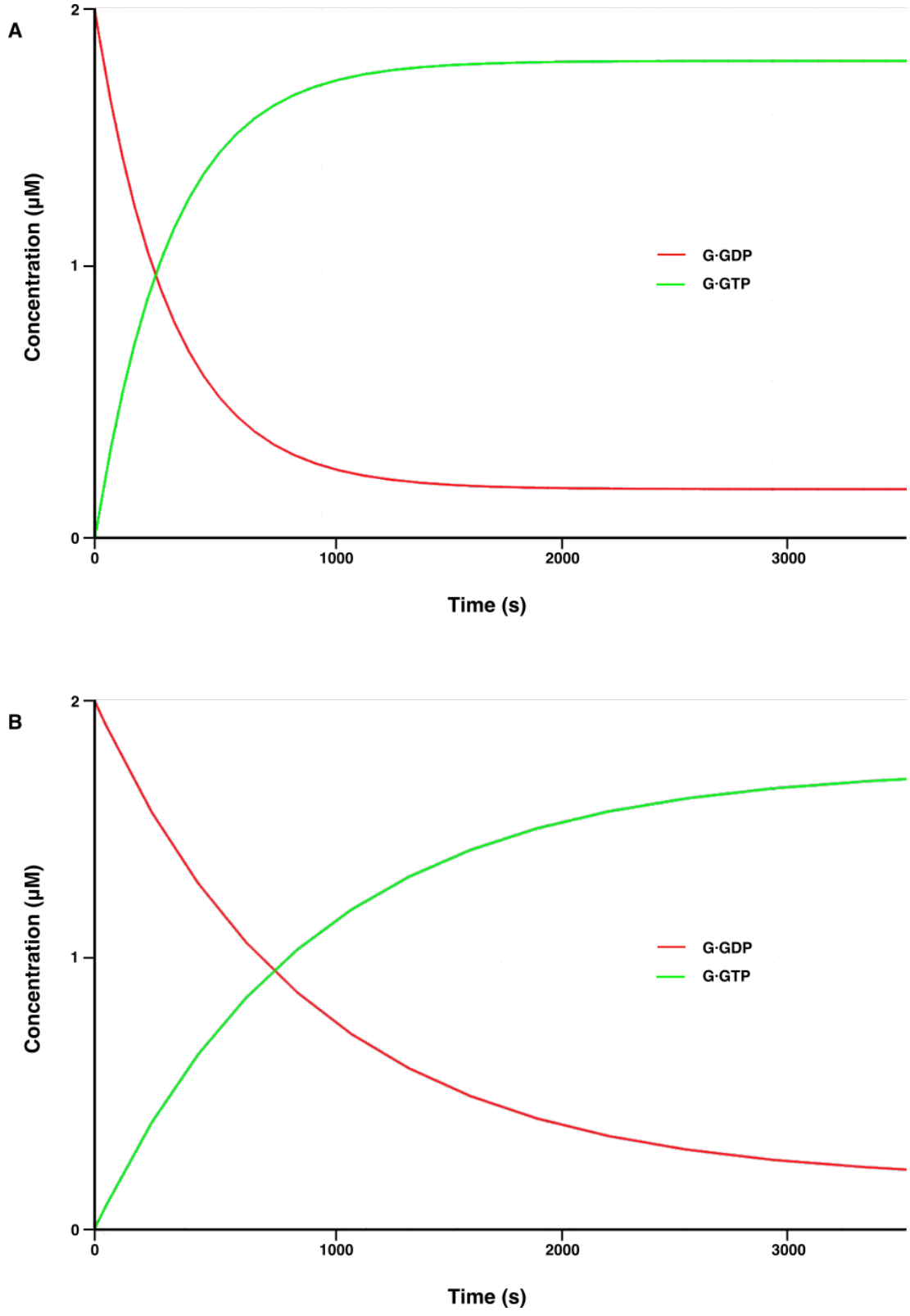
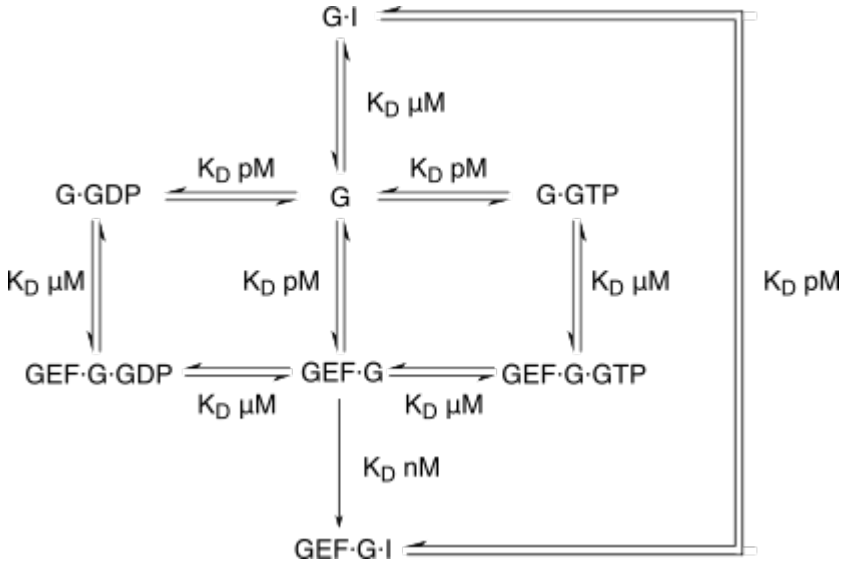


Figure 32. Outcome from KinTek Explorer of an irreversible GTPase-selective competitive inhibitor. The concentrations of GDP-bound GTPase (red) and GTP-bound (green) are shown over 1 hour (3600 seconds. **A**. Concentration of inhibitor set as 1 μM at t = 0. No distinct change is seen compared to the

activation cycle with no inhibitor modelled **B**. Concentration of inhibitor set as 1 mM at t = 0. There is a small retardation of exchange reaction.

2.2.4.2 Selective Covalent GEF/GTPase Inhibitor

The model shown in Figure 33, analogous to the selective model shown in for a selective reversible inhibitor of a GEF/GTPase complex, was submitted into KinTek Explorer.



Reaction	K_D (nM)	k_{on} ($\mu M^{-1}s^{-1}$)	k_{off} (s^{-1})
G + I ⇔ G·I	1000	1	1
GEF·G + I ⇔ GEF·G·I	-	1	1×10^{-20}
GEF + G·I ⇔ GEF·G·I	1×10^{-3}	1	1×10^{-6}

Figure 33. Modelling of a selective covalent inhibitor for the GEF/GTPase complex. 'G' denotes GTPase. 'I' denotes inhibitor

The formation of the covalent bond was represented by a very slow k_{off} rate, as for the classic covalent inhibitor. It should be noted that the same rates of exchange were observed when a non-selective inhibitor with irreversible binding to GTPase alone as well as the GEF/GTPase complex (not shown).

The difference between this model (Figure 34) and that of the analogous reversible inhibitor shown in Figure 30 is more marked than that of the classic covalent inhibitor. Even at 1 μM, the covalent compound significantly inhibits the rate of nucleotide exchange, with the curves for G:GDP and G:GTP plateauing more quickly than for the equivalent reversible inhibitor. These results are very promising for the development of an irreversible competitive compound for the GEF/GTPase complex, as the predicted rate of nucleotide exchange is extremely slow, with minimal exchange over 1 hour. If covalent and non-covalent analogues of the same inhibitor could be generated, it would be interesting to test whether this prolongation of nucleotide exchange can be observed in cellular conditions.

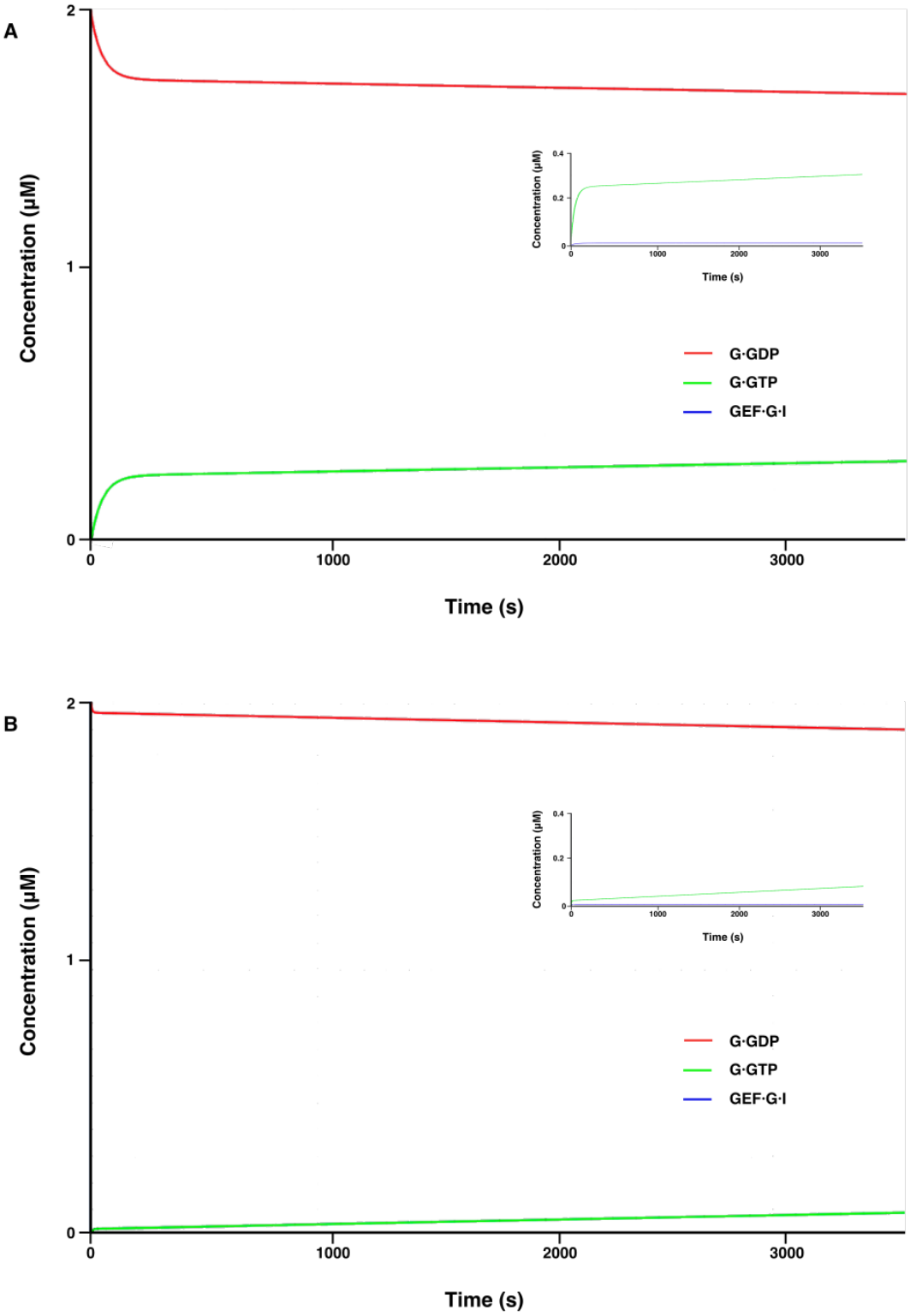


Figure 34. Outcome in KinTek Explorer of a selective irreversible competitive inhibitor for the GEF/GTPase complex. The concentrations of GDP-bound GTPase (red) and GTP-bound (green) and the inhibitor-bound GEF/GTPase complex (blue) are shown over 1 hour (3600 seconds). **A**.

Concentration of inhibitor is set at 1 μM at t = 0. Rate of GDP-GTP exchange is slowed. **B**. Concentration of inhibitor is set at 10 μM at t = 0. Minimal exchange occurs over the course of an hour.

2.3 Discussion

These models predicted that a compound with affinity for the GEF/GTPase complex rather than for the GTPase alone would inhibit GDP/GTP exchange in physiological conditions. What was perhaps surprising is that the affinity of the inhibitor for the GTPase alone did not impact the rate of the nucleotide exchange. A crucial aspect of these inhibitors is that their binding to the GTPase does not adversely affect the affinity of the GEF for the GTPase. For example, for a selective inhibitor of the GEF/GTPase complex, if the affinity of the GEF to its GTPase binding partner is reduced from picomolar to micromolar, the effect of the inhibitor is essentially null (Figure 35). With reduced affinity, the GEFs are not trapped in the GEF·G·I complex and so are available to catalyse nucleotide exchange.

This indicates that it is essential during the design of inhibitors that the compound is not a direct analogue of the guanine nucleotides; once bound, these reduce the affinity of the GEF to its GTPase from picomolar to micromolar. The inhibitor instead should bind within the orthosteric pocket in a manner which is not disruptive to GEF binding. Careful analysis of the relevant GEF/GTPase structure to identify key residues should be conducted to avoid clashing interactions which would reduce the affinity for the GEF and therefore the efficacy of the compound.

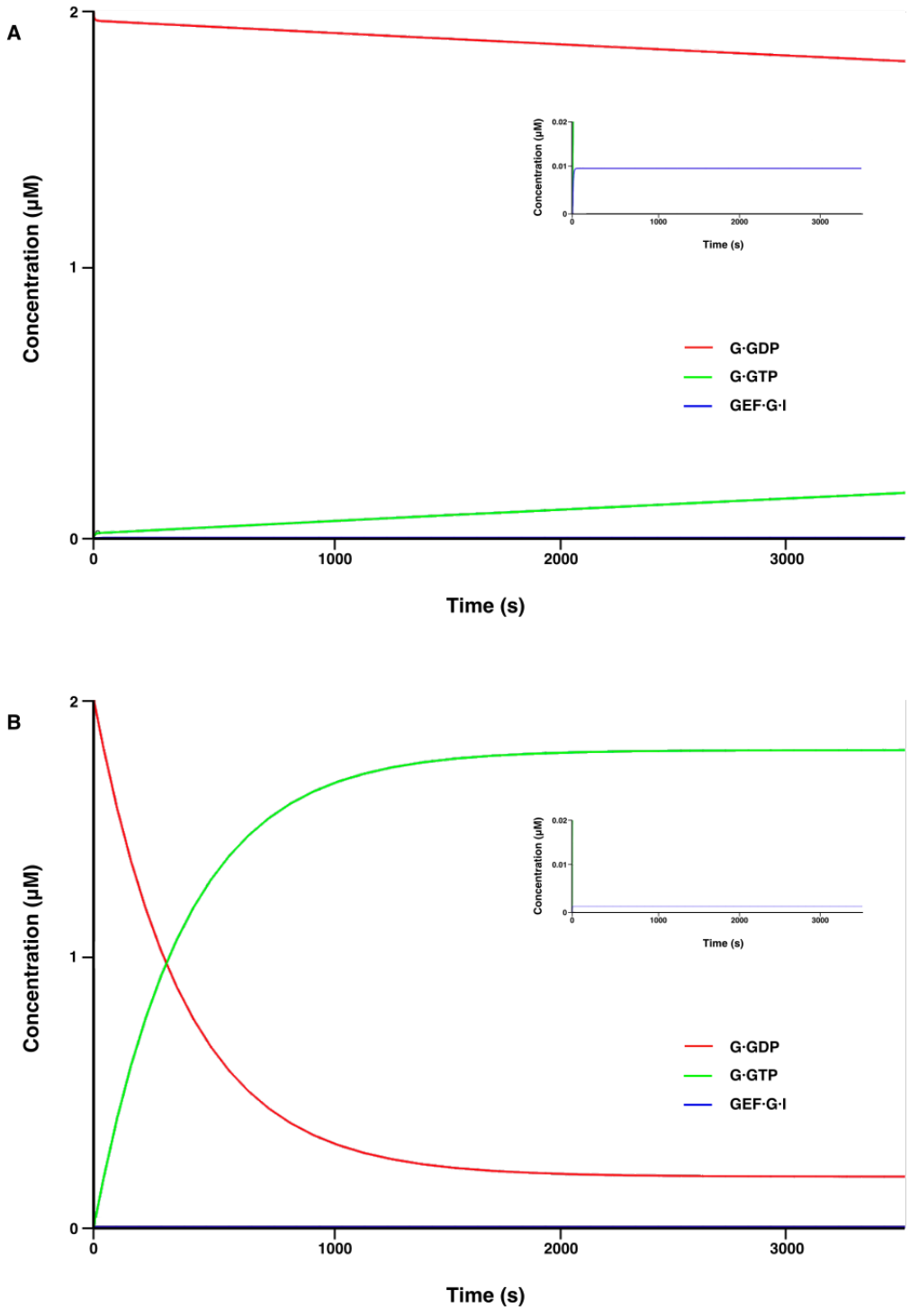


Figure 35. Outcome in KinTek Explorer comparing the change in the rate of exchange if the affinity of the GEF for its GTPase is reduced when the GTPase is bound to an inhibitor. The concentrations of GDP-bound GTPase (red) and GTP-bound (green) and the inhibitor-bound GEF/GTPase complex (blue) are shown over 1 hour (3600 seconds). **A.** GEF has high picomolar affinity for the G·I complex. Selective nanomolar inhibitor set at concentration of 10 μM at t = 0. Exchange is very slow. **B**. GEF has

micromolar affinity for the G·I complex. Selective nanomolar inhibitor set at concentration of 10 μM at t = 0. Rate of nucleotide exchange is rapid.

A second important factor that can be observed is the relevant concentration of the GEF in comparison to the GTPase. Whilst the concentrations of the GTPase and GEF were modelled using literature values, these can vary in cells and in disease, where GEFs can be over- or under-expressed.[117,118,120,121] If the concentration of GEF increases, the efficacy of the compound in preventing nucleotide exchange is reduced for a given concentration. This is observed in Figure 36, where a nanomolar inhibitor at 1 μM does not affect the rate of exchange when the GEF concentration is increased tenfold. This concentration was sufficient to demonstrably reduce the rate of formation of the activated GTPase species when the GEF concentration was lower (Figure 30). The concentration of the inhibitor needs to be significantly larger for a similar effect on nucleotide exchange to be observed. The higher concentration of GEF means less is trapped in the GEF·G·I complex, resulting in more free protein available to catalyse nucleotide exchange. It was observed during the KinTek experiments that if the potency of the inhibitor was increased, the rate of exchange remained the same, with only the maximal concentration of the activated GTPase being reduced. In diseases where GEFs are overexpressed, higher concentrations of inhibitor would therefore be required to generate the desired outcome in the cell.

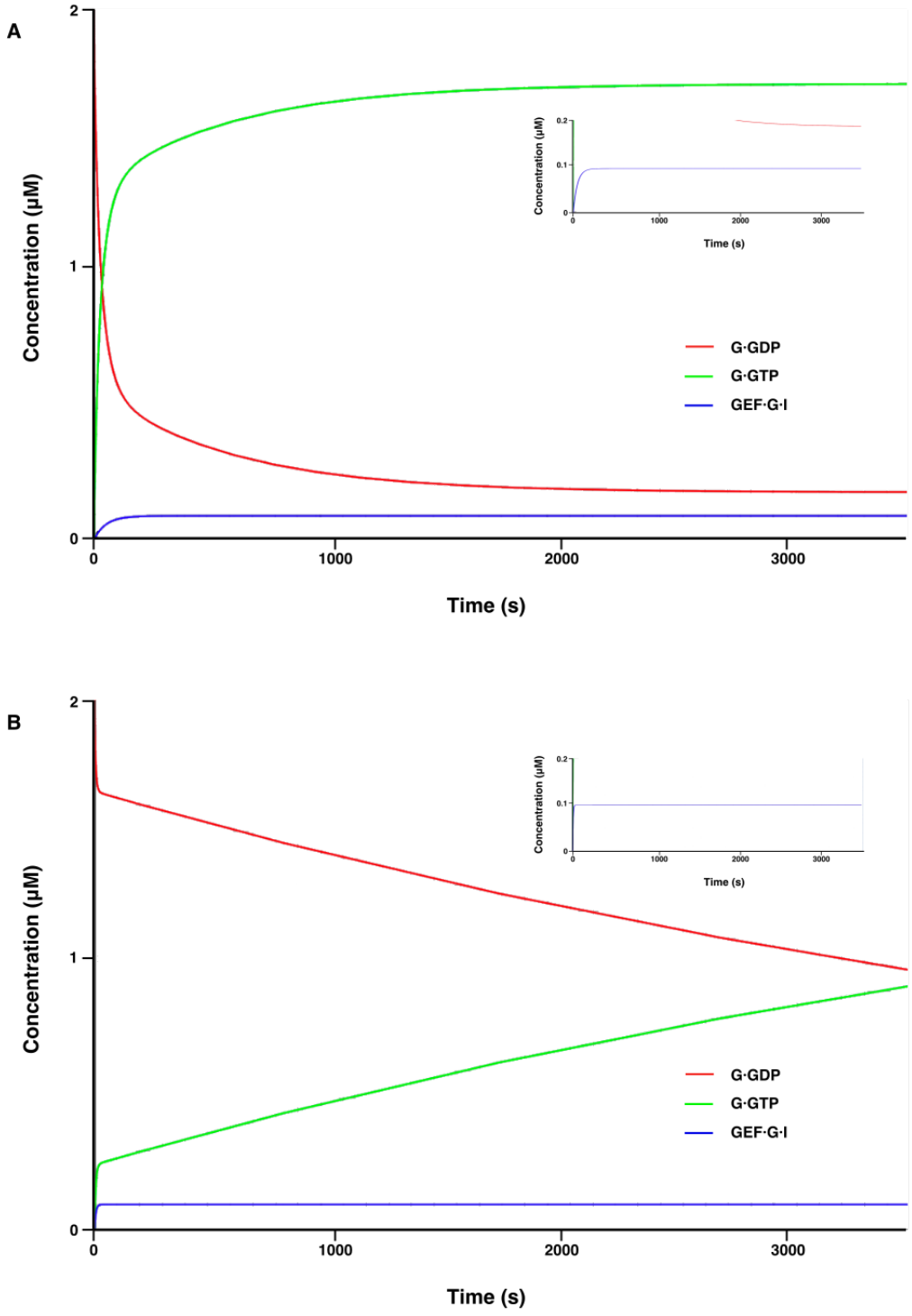


Figure 36. Outcome from KinTek Explorer modelling the effect when the concentration of GEF is increased from 10 nM to 100 nM. The concentrations of GDP-bound GTPase (red) and GTP-bound (green) and the inhibitor-bound GEF/GTPase complex (blue) are shown over 1 hour (3600 seconds). **A**. A selective nanomolar inhibitor with concentration set as 1 μM at t = 0. The rate of exchange is

extremely rapid. **B**. A selective nanomolar inhibitor with concentration set as 10 μM at t = 0. Some inhibition of nucleotide exchange is observed.

However, it should be noted that if the GEF is significantly overexpressed compared to the concentration of GTPase, it acts as a competitive inhibitor in its own right, as the GTPase is trapped in complex with the GEF within the activation cycle. Hence, the GEF expression levels could have a significant impact on the success of a competitive GEF/GTPase complex inhibitor. This mechanism of inhibition should be carefully considered before pursued in diseases associated with GEF overexpression. It is also important to note that high cellular turnover of the GTPases and their GEFs would significantly decrease the impact of a proposed inhibitor, especially for the covalent compounds. This could not be modelled in KinTek Explorer and so would have to be measured and observed in cellular experiments.

Finally, all of these scenarios modelled an inhibitor which had nanomolar affinity for the pertinent species. Whilst generating high levels of specificity can be difficult for certain targets, drugs and probes commonly have nanomolar potency, and so this should be achievable with significant medicinal chemistry effort. When the affinity within the model was lowered to high nanomolar, this significantly reduced the impact of the inhibitor (Figure 37). If a nanomolar compound cannot be synthesised, higher cellular concentrations would be required inhibit GTPase activation in cells.

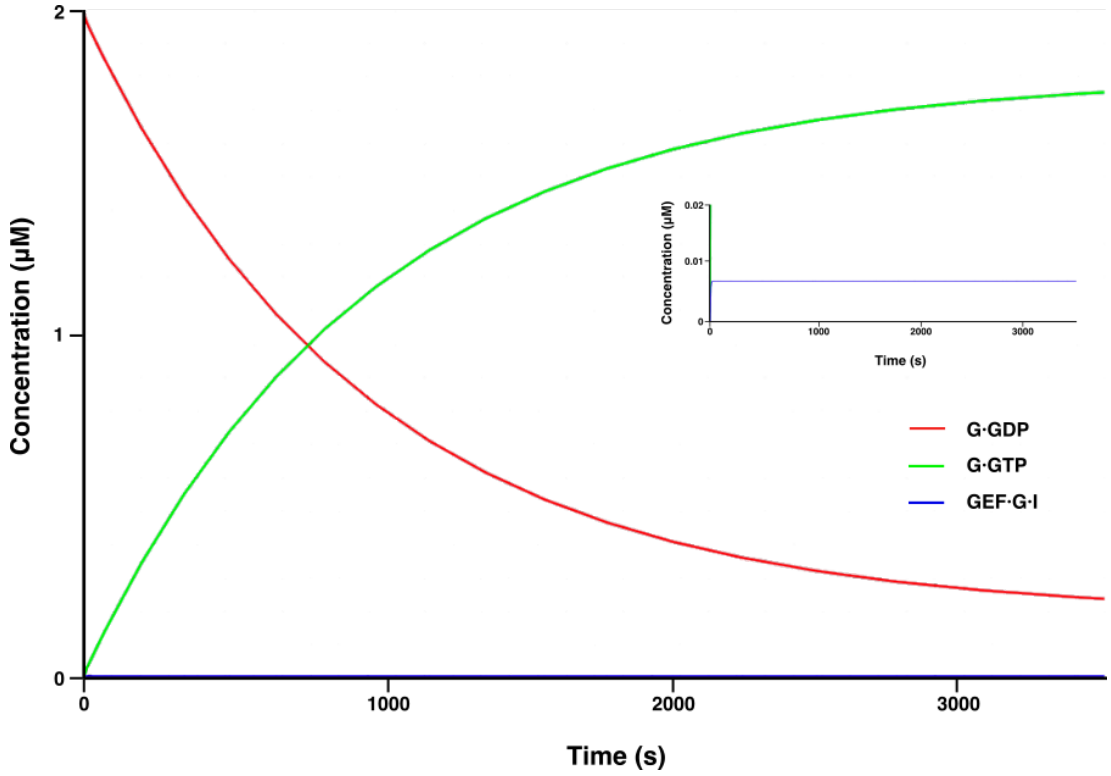


Figure 37. Outcome from KinTek Explorer modelling a 100 nM selective competitive inhibitor for the GEF/GTPase complex at a concentration set at 10 μM at t = 0. The concentrations of GDP-bound GTPase (red) and GTP-bound (green) and the inhibitor-bound GEF/GTPase complex (blue) are shown over 1 hour (3600 seconds).

2.4 Conclusions

Using a kinetic simulation of the GTPase activation cycle in physiological conditions, it has been demonstrated that a covalent or reversible inhibitor with a nanomolar K_D for the GEF/GTPase complex could reduce the rate of nucleotide exchange and thereby prevent the rapid activation of the small GTPase. The model makes several assumptions already discussed: the capacity to generate a low nanomolar affinity inhibitor; that the relative GEF/GTPase concentrations remain constant; and that the GEF maintains a high affinity for the GTPase once the inhibitor is bound.

The kinetic models also raise certain questions to be investigated once a sufficiently potent inhibitor has been identified in cellular contexts. The model is an isolated system and does not address that many GTPases are activated by multiple GEFs for temporal and spatial regulation. It is unknown whether an inhibitor can be specific for a single GEF/GTPase complex, although careful structure-based drug design and subsequent targeting of key residues should aid in this endeavour. It is also unclear whether non-specific inhibition would be beneficial.

An important conclusion of the simulation was the prediction that the activation cycle could be significantly slowed by the presence of the inhibitor. It is unknown based on these experiments and existing compound data whether retardation of the activation cycle through sequestration of the complex will be sufficient to produce a phenotypic response. It could be that complete inhibition is required to cause physiological changes; this seems more achievable with an irreversible covalent inhibitor but could also have damaging effects in the cell if the activation cycles are completely inhibited.

Whilst the above queries could not be addressed with these kinetic simulations, the results were promising enough to continue with the pursuit of a GEF/GTPase inhibitor within the SGC. At the very least, any compounds produced could be used as probes to elucidate

the roles of these complexes *in vivo*, even if the hypothesised mechanism of inhibition is ultimately unsuccessful as a therapeutic strategy.

3. Development of a GEF/GTPase Pipeline for the SGC

3.1 Overview

The difficulties associated with the heterologous expression of pure proteins in high yield, coupled with the low success rate of crystallisation means the chances of progressing a protein from construct to crystal are slim. Hence, in pursuing this project, a high-throughput pipeline was developed for the cloning, production, and crystallisation of GEFs and GTPases to maximise the likelihood of success in achieving a high-quality well-diffracting crystal of a complex suitable for ligand experiments. This chapter details the development of this pipeline, including considerations for construct design, the use of ligation independent cloning (LIC) optimised by the SGC, the high-throughput PrepX method for easy expression and production of multiple protein constructs simultaneously, and the crystallisation trials. Three novel crystal structures were obtained and are discussed, with one (Kalirin/Rac1) proving to be suitable for XChem and co-crystallisation with ligands, which is detailed in chapters 4, 5 and 6. The status of the pipeline, with optimisation of other crystal systems, is also discussed. A summary for the status of all the constructs designed during this project is included in Appendix 8.1.

3.2 Pipeline Round 1

3.2.1 Construct Design

To maximise the chance of obtaining high quality crystals, the initial subset of GEF and GTPase proteins included in the pipeline were chosen based on their previous structural determination having been documented in the PDB (Table 4). Additional criteria included the use of an *E. coli* expression and purification system. Members from each subfamily were represented, with the hope that the method of targeting GEF/GTPase complexes would translate across the Ras superfamily.

In general, a single construct for each protein was selected using the end points recorded in the PDB. For Rac1, a Rho GTPase, multiple constructs were present in the PDB that varied the length of the flexible C-terminus. Three constructs of Rac1 were included in the pipeline, as disordered regions can prevent crystallisation or cause poorly diffracting crystals. Other exceptions include the Rho GEF Kalirin, as the NMR structure of the DH domain of Kalirin alone had been previously determined, and members of the DOCK and IQSEC GEF subfamilies (Rho and Arf respectively), which did not have their structure solved by either X-ray, NMR or cryo-EM. These proteins were deemed of particular interest in neurodegenerative diseases.[171,172] Additional constructs of Kalirin were also devised to include the DH GEF domain and the neighbouring PH domain, which was thought to be necessary for full GEF activity based on evidence for the Rho GEF Trio.[173] Several constructs for these proteins were designed based on the cut-off points used in the crystallisation of homologous proteins. The lengths of the N- and C- termini were varied to avoid predicted disordered regions or hydrophobic residues which could cause aggregation during the purification or crystallisation processes.

Table 4. Subset of GEF/GTPase complexes chosen from the PDB organised by their subfamily. The relevant PDB codes for all constructs are listed in Appendix 8.16.1.1. Proteins that are underlined had not been previously crystallised as part of a GEF/GTPase complex. Proteins in italics had no X-ray structure in the PDB at the beginning of the project.

GTPase Subfamily	GTPase	GEF
Ras	HRas	Sos1
	RalA	RGL2
	Rap1B	EPAC2
Rho	RhoA	NET1, ARHGEF11, ARHGEF12, AKAP13, Dbs
	Cdc42	ITSN1, DOCK9
	Rac1	P-Rex1, Trio, *Kalirin*, Vav1, Tiam1, DOCK2, *DOCK3*
Rab	Rab35	DENND1B
	Rab21	Rabex-5
	Rab8	MSS4
Ran	Ran	RCC1
Arf	Arf1	IQSEC1, *IQSEC2, IQSEC3*, Cyth2, Cyth1, ArfGEF1, ArfGEF2
	Arf6	Cyth3, IQSEC2, IQSEC3

3.2.2 Molecular Cloning

The initial strategy was to clone the GEF and GTPase constructs into a sole pNic28 bicistronic expression vector with a single promoter using LIC optimised at the SGC (Appendix 8.2).[78] This strategy allows simultaneous expression and immediate formation of the complex in one *E. coli* system. Unfortunately, this method was largely unsuccessful, with only three clones being obtained and one (IQSEC2/Arf6) resulting in protein expression. The reasons behind this failure are unknown, although it can be hypothesised that, in most cases, the combination of the GEF and GTPase in the vector were lethal to the bacterial cell. This method was not optimised or explored further, as the expression of IQSEC2/Arf6 encoded within the bicistronic vector was also problematic. The first complement in the vector was significantly overexpressed in comparison to the second complement, which would have created issues with complex purity for crystallisation. Hence, this strategy of cloning was abandoned in favour of cloning the GEF and GTPase genes into vectors separately and combining them during either the expression or purification stages of the pipeline.

Using LIC, GEFs were encoded into pNIC28 His-tagged vectors and GTPases into pNIC28 Strep-tagged vectors. The different tags were utilised so that co-purification using immobilised metal affinity chromatography (IMAC) and StrepTactin columns could be attempted during the purification stage of the pipeline. It was noticed that either the ligation into the vector or transformation into bacteria was often inefficient and had to be attempted several times. It is unknown as to why this was the case, although it was noted that the transformation was generally more unsuccessful for GEFs rather than the GTPases. The inefficiency could thus either be attributed to the relative larger size of the GEFs, or that they were more toxic to the *E. coli* competent cells during transformation. However, the overall process was generally successful, with all proteins being cloned into *E. coli* BL21 DE3 cells

for expression apart from the Rho GEFs AKAP13, NET1, Dbs and ITSN1, and some constructs of IQSEC2, IQSEC3 and DOCK3 (Table 5).

3.2.3 Protein Production

Successfully cloned constructs were trialled for expression using the high throughput PrepX protocol developed by the Mike Fairhead of Frank von Delft's group at the SGC. PrepX streamlines protein expression and purification to allow the production of proteins in batches of 24 within the time frame of one week. Generally, the GTPases were expressed and soluble with the exception of HRas, RhoA and Ran, and were easy to purify, whilst the GEFs were either not expressed or not soluble, with the exception of most of the Kalirin constructs and some of the Arf GEFs (Table 5). The initial trial expression was completed using 100 mL of cell culture; scaling up of the procedure to 1 L resulted in enough protein for purification for a subsection of the GEFs, but this method was not generally applicable.

Table 5. A summary of targets which failed at the cloning or expression stage of the pipeline. GTPases written in italics also failed to clone/express. Starred (*) proteins were later successfully cloned into MBP2 vectors.

		Target failing during stage	
GTPase Subfamily	**GTPase**	**Cloning**	**Expression/Purification**
Ras	*HRas*		Sos1
	RalA	RGL2*	
	Rap1B		Epac2
Rho	*RhoA**	NET1, Dbs, AKAP13*	AKAP13, ARHGEF11
	Cdc42	ITSN1	DOCK9
	Rac1	TIAM1*, Trio*	PRex-1, Trio, TIAM1, DOCK2, DOCK3, Vav1
Rab	Rab35		DENND1B
	Rab8		MSS4
	Rab21		Rabex-5
Ran	*Ran**	RCC1*	
Arf	Arf6	ARFGEF1*, CYTH2*, IQSEC3*	Cyth3

Several strategies were then employed in an attempt to improve the expression and purification of the unsuccessful proteins (Appendix 8.3). First, the remaining constructs were cloned into a pNIC28-MBP2 vector to create a maltose binding partner (MBP)-fusion protein. The cloning was inefficient and only successful for ten proteins, but the MBP tag significantly improved the solubility and yield, resulting in sufficient amounts of eight proteins for

crystallisation trials, albeit with a significant amount of protein lost during purification following the removal of the MBP tag.

Further trials on problematic proteins were conducted to see if a general method could be applied to optimise production. Methods included extension of the expression time or reduction in temperature, use of benzyl alcohol to promote endogenous chaperone production during expression to prevent aggregation[174], transformation into T7 competent cells designed to facilitate the expression of potentially toxic genes[175] and expression using a range of IPTG concentrations, and co-lysis of GEF and GTPase partners. All of these methods were generally unsuccessful in improving the yield of the soluble fraction of the proteins (Appendix 8.3). As Frank von Delft's group had commonly only observed improvements in yields following MBP incorporation or change in construct design, it was decided to deprioritise the unsuccessful GEFs.

For proteins that were successfully expressed and purified, methods to form the complex were investigated. It was attempted to form the GEF/GTPase complex during purification by mixing the proteins prior to application to an IMAC column, with the expectation that the proteins remain bound as a complex to the nickel resin *via* the GEF His tag. Subsequent elution and addition to a Strep-Tactin column would ensure complex purity. However, this was unsuccessful as only GEFs were obtained from the elution of the IMAC column. Hence, the complexes were formed following the purification of GEFs and GTPases separately by mixing the proteins in an approximately 2 to 1 ratio. The complexes were subsequently purified using an additional round of size-exclusion chromatography (SEC).

3.2.4 Crystallisation and Structure Determination

The outcomes of the proteins submitted into crystallisation trials are shown in Table 6.

Table 6. The status of crystallisation for each GEF/GTPase complex in the first pipeline. The starred GEFs refer to crystal trials of the GEF alone rather than in a complex. Kalirin-c001 refers to a construct containing only the DH domain of Kalirin. Kalirin-c002 refers to a construct containing both the DH and PH domains of Kalirin.

GTPase Subfamily	Crystal trials	Crystal Optimisation	Structure Determination	FBLD screening
Ras	RGL2 / RalA EPAC2 / Rap1B			
Rho		Kalirin-c001*	Kalirin-c002/Rac1	Kalirin-c001/Rac1
Ran	RCC1 / Ran			
Arf	Cyth2 / Arf1 ArfGEF1 / Arf1 Cyth1 / Arf1 ArfGEF2 / Arf1 IQSEC2/Arf6	IQSEC3* IQSEC2* IQSEC3 / Arf1	IQSEC2 / Arf1	

Each protein complex was tested in 1-6 coarse screens at 4 and 20 °C using an initial concentration between 10-20 mg/mL, depending on the quantity of protein available. Microcrystals were obtained for the GEFs Kalirin, IQSEC2 and IQSEC3 alone but follow-up fine screens and seeding approaches were ineffective at producing larger crystals. Other complexes were ultimately unsuccessful, with at most spherulites being observed in the

coarse screens. The SGC has generally observed the largest improvement in crystal quality upon the use of seeding or varying the construct boundaries. As these proteins were based on a single PDB structure, they were only represented by one construct in the pipeline. It was thus decided to concentrate on progressing the crystallised complexes, IQSEC2/Arf1 and two complexes of Kalirin/Rac1, rather than protracted optimisation of microcrystals and spherulites. Data collection and refinement statistics of these complexes are presented in Appendix 8.4.

3.2.5 Kalirin-c001/Rac1

Crystallisation trials were attempted for complexes containing the Kalirin DH GEF domain (Kalirin-c001) bound to either full length Rac1 or Rac1 with the flexible C-terminus truncated. Cube shaped crystals of Kalirin-c001 bound to truncated Rac1 were obtained in the JCSG coarse screen from Molecular Dimensions (Figure 38).[176] Fluorescent labelling using different coloured dyes of the GEF and GTPase components showed that the crystals were of the complex rather than the GEF or GTPase individually. Further optimisation yielded crystals which were solved to high resolution (1.64 Å) at the DLS using the homologous model of Trio/Rac1 (PDB code: 2NZ8) for phasing.

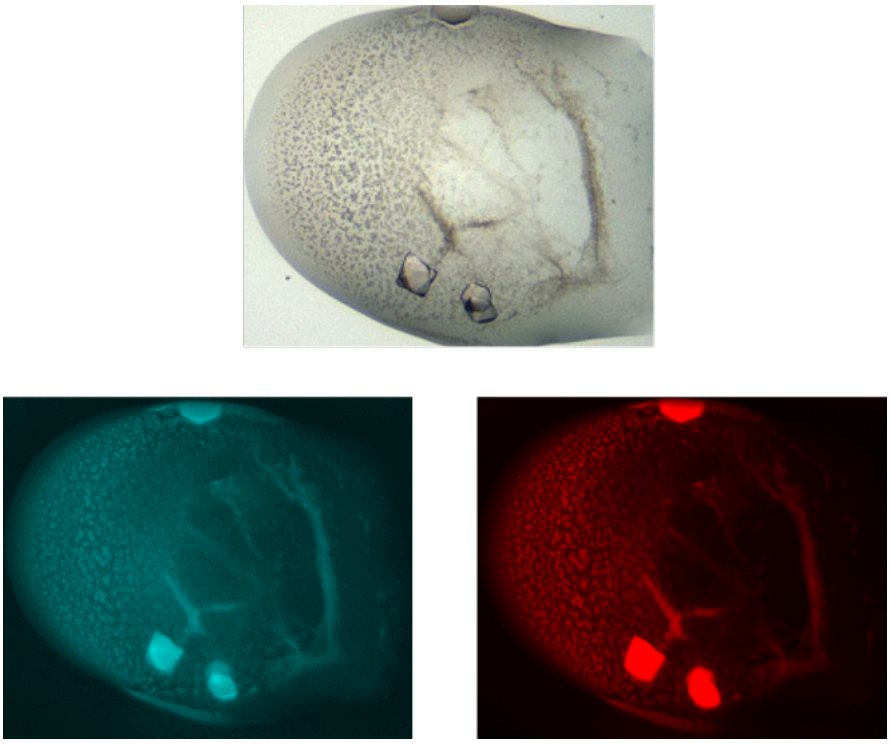


Figure 38. Crystals of the Kalirin/Rac1 complex. Top: visible wavelength image of the crystals. Bottom left: Rac1 was labelled with Alexa-488 dye, which fluoresces green at 488 nm. Bottom right: Kalirin was labelled with Alexa-555 dye, which fluoresces red at 555 nm. Both colours being present in fluorescent images indicate that the crystals are of the complex rather than the individual components. The fluorescent dye was also useful for identifying crystals buried under denatured protein which are hard to identify. This method had not previously been used before at the SGC.

The crystal structure of GDP-bound Kalirin/Rac1 (PDB code: 5O33), shown in Figure 39A, was novel, with no crystal structure of Kalirin previously available in the PDB. The Kalirin DH domain consists of a six α-helical bundle characteristic of the GEF domains of the Dbl Rho GEFs. Superposition of the complex structure with the homologous Trio/Rac1 shows that the orientations are similar, which is of little surprise considering the 92% homology of the DH domains of Trio and Kalirin.[177]

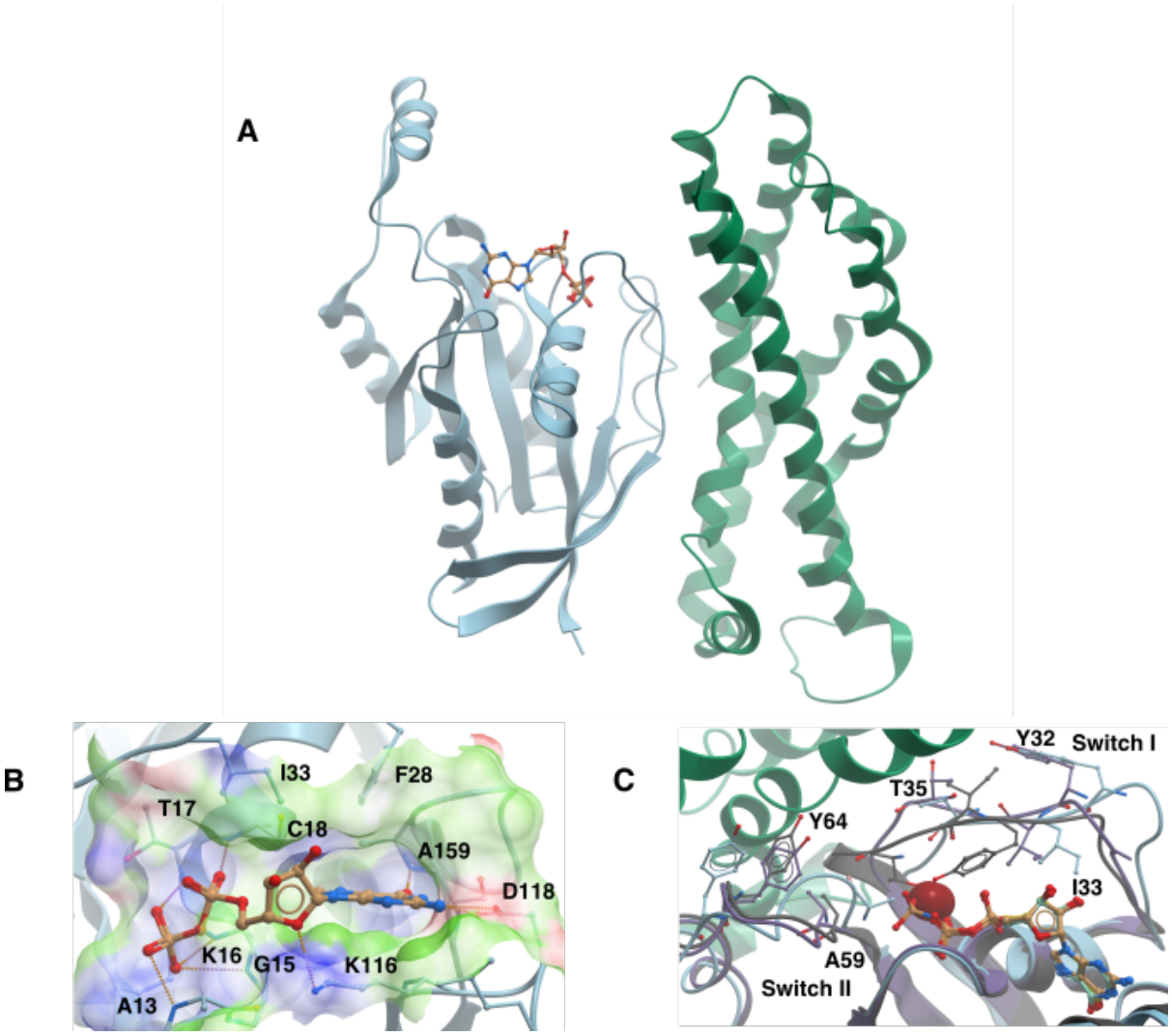


Figure 39. **A**. The novel structure of the Kalirin/Rac1 complex with GDP bound (PDB code: 5O33). Kalirin is shown as dark green, Rac1 as light blue and GDP as brown sticks. **B**. A magnified representation of the GDP binding pocket of Rac1 in the Kalirin/Rac1 complex. Key residues forming hydrogen bond interactions (such as D118 or the backbone amines of K16 and T17) are visualised as sticks with the hydrogen bonds as small orange spheres. The position of C18, which could be targeted with a covalent small molecule, is also shown. **C**. Overlay of Kalirin/Rac1 with Rac1 alone bound to GDP (PDB code: 5N6O, purple) and GTP (PDB code: 3TH5, dark grey). Some movement in the switch I and II loops can be observed, with some key residues depicted as sticks. Mg^{2+} are shown as red spheres.

GDP is present in high occupancy within the switch I loop of the Kalirin/Rac1 complex, with the guanine base forming hydrogen bonds with aspartate residue 118 and the phosphate

groups interacting with the NH group of amine backbones (Figure 39B). GDP was not added during the purification process and must originate from the bacterial cell. No evidence for Mg^{2+} was observed, but there is literature precedent for Rac1 being capable of binding to guanine nucleotides in the absence of Mg^{2+}. It suggests that the conformation of Rac1 is unfavourable for Mg^{2+} binding; an analogous summation was made in the structure of Tiam1/Rac1 due to positioning of key residue T35.[178] The structures of Rac1 in both of these complexes is extremely similar (not shown).

Comparison of the switch I and II regions of Rac1 bound in the complex compared to Rac1 alone bound to GDP or GNP show movement in the loops which could be exploited for selectivity for the complex. (Figure 39C). The switch I loop in the complex moves out to allow more space within the pocket for easier GDP/GTP exchange, which could improve accessibility for a small molecule binder. Key residues in the switch I loop which could be targeted for interactions by a small molecule binding within the GDP pocket include Y32 and I33, which are significantly shifted in the activated GTP-bound form of Rac1. A59 and Y64 of the switch II loop are also markedly different in the GEF-bound complex. It is interesting to note the orientation of A59 towards the position of Mg^{2+} in the crystal structure of Rac1 alone, indicating it may also be involved in the displacement of Mg^{2+} in the complex.

The presence of a cysteine residue (C18) in the pocket is of particular interest, as it could be targeted with a small molecule covalent inhibitor. Alignment analyses indicate that this cysteine is present in around 20% of the Ras superfamily, but mainly in the Rab subfamily, where other residues in the vicinity vary (Appendix 8.5).

Comparison of the structure of Kalirin with its free solution NMR structure (PDB code: 2KR9, Figure 40A) shows them to be relatively similar, perhaps indicating that the conformation of Kalirin seen in the crystal structure is energetically favourable rather than an artefact of the crystal space group. The N-terminus of the NMR solution of Kalirin alone overlaps with the switch I loop of Rac1, indicating that this disordered region must move in the

complex. This region was removed prior to designing the construct as it was predicted that this disordered region may hamper with crystallisation, a clear indication of the importance of construct design to the success of a crystallisation project.

The main movements occur in residues involved in PPIs at the interface of the complex. Previous experiments have shown that T1244 and N1406 are required for full GEF activity of Trio on Rac1.[179] T1244 forms van de Waals interactions with V36 of the Rac1 switch I region which are required for the binding efficacy of Trio to Rac1, and N1406 forms a hydrogen bond with D65 of the G3 box which is required for GDP/GTP exchange.[179] These amino acids and their neighbouring residues in the helices shift slightly in the complex structure in order to form these interactions.

Attempts to remove GDP from the system using EDTA appeared to be successful following the crystallisation of the complex which had little to no evidence of occupancy by GDP in the nucleotide binding site. This was conducted to ensure the binding site was available for fragment screening and to simplify the background analysis for fragment identification over a large subset of structures. Optimisation of the crystal system conditions, and the development of a stock seed solution resulted in reproducible diamond-shaped crystals which consistently diffracted to better than 2.0 Å. They also tolerated high concentrations of DMSO (> 20%). Hence, this complex was determined as suitable as an XChem target and was taken forward for fragment screening (Chapter 4 onwards).

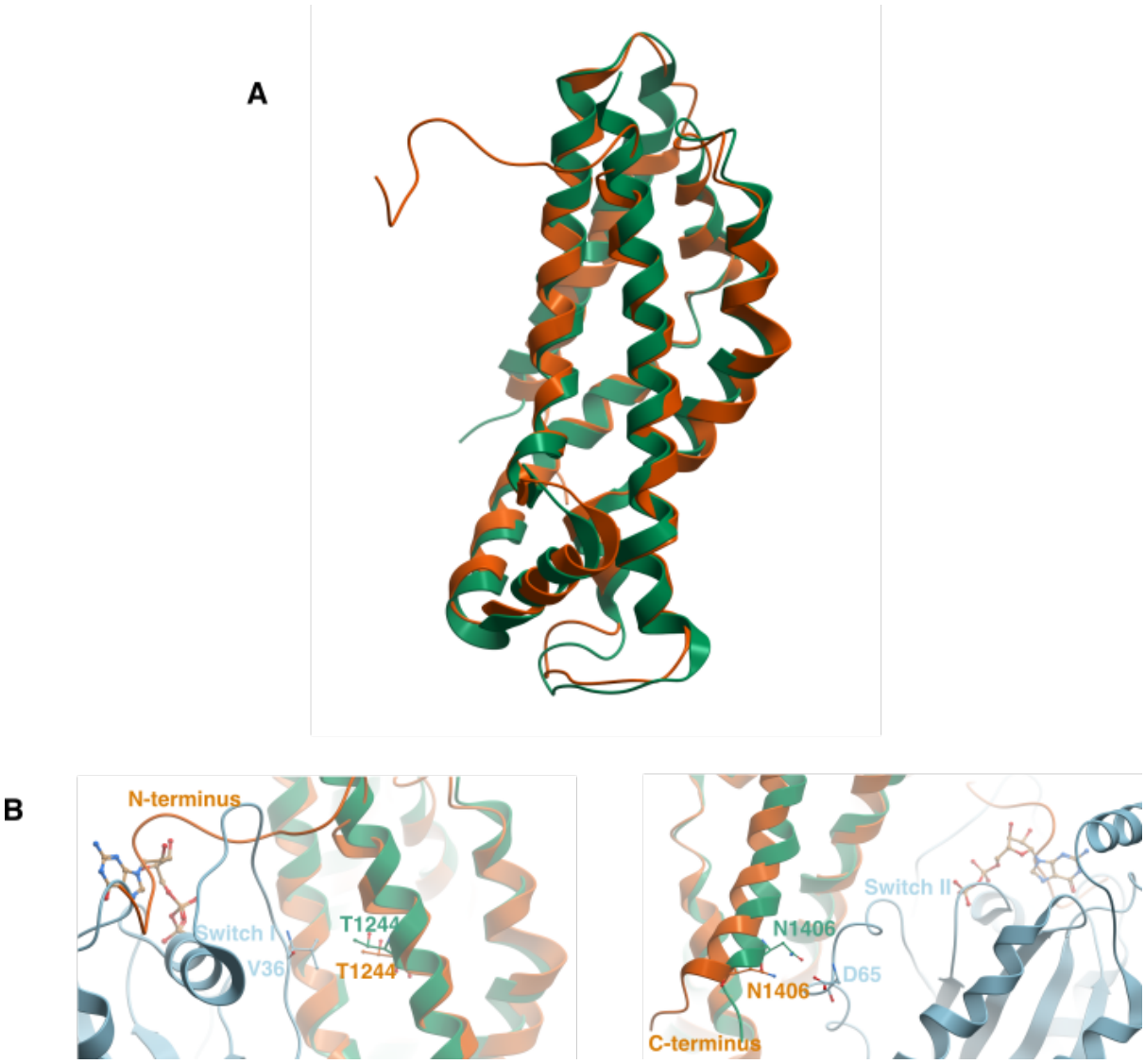


Figure 40. **A**. Overlay of the structure of Kalirin in Kalirin/Rac1 complex (green) compared to the solution NMR structure determination (orange, PDB code: 2KR9). The NMR structure was used to design the construct of Kalirin present in the complex. The DH domains in the 2 structures generally look similar, although some of the residues in loops that interact with Rac1 have shifted. The similarities suggest that the structure observed is favoured in solution and is not a crystallographic artefact. **B**. Some important differences between the two structures. Left: The key residue T1244 has shifted slightly to interact with V36. The disordered N-terminus of the NMR structure is also overlapping with the switch I of Rac1, suggesting this section moves when Kalirin is bound to Rac1. Right: N1406 moves significantly to interact with D65 of Rac1.

3.2.6 Kalirin-c002/Rac1

Crystallisation trials were also attempted for a construct of Kalirin, c002, which contained both the DH and PH domains in complex with the truncated Rac1 protein. The crystals of the Kalirin-c002/Rac1 complex were poorly diffracting in comparison to the Kalirin-c001/Rac1 complex. Optimisation of conditions using a series of fine screens and different seed stocks resulted in crystals which diffracted to 2.85 Å on the I04 beamline at DLS. The structure was solved using the homologous model of Trio/Rac1 (PDB code: 2NZ8) for phasing (Figure 41A). Data collection and refinement statistics are presented in Appendix 8.4.

There are two copies of the complex within a single unit cell of the $P2_12_12$ space group. GDP is present in the binding site, but with much lower occupancy than observed in the Kalirin-c001/Rac1 complex. The PH domain of Kalirin has a similar seven-stranded β-sandwich architecture seen in the structure of Trio/Rac1. The majority of the interactions between Kalirin and Rac1 remain between the DH domain and Rac1, with only a few residues of the PH domain involved at the interface (Figure 41).

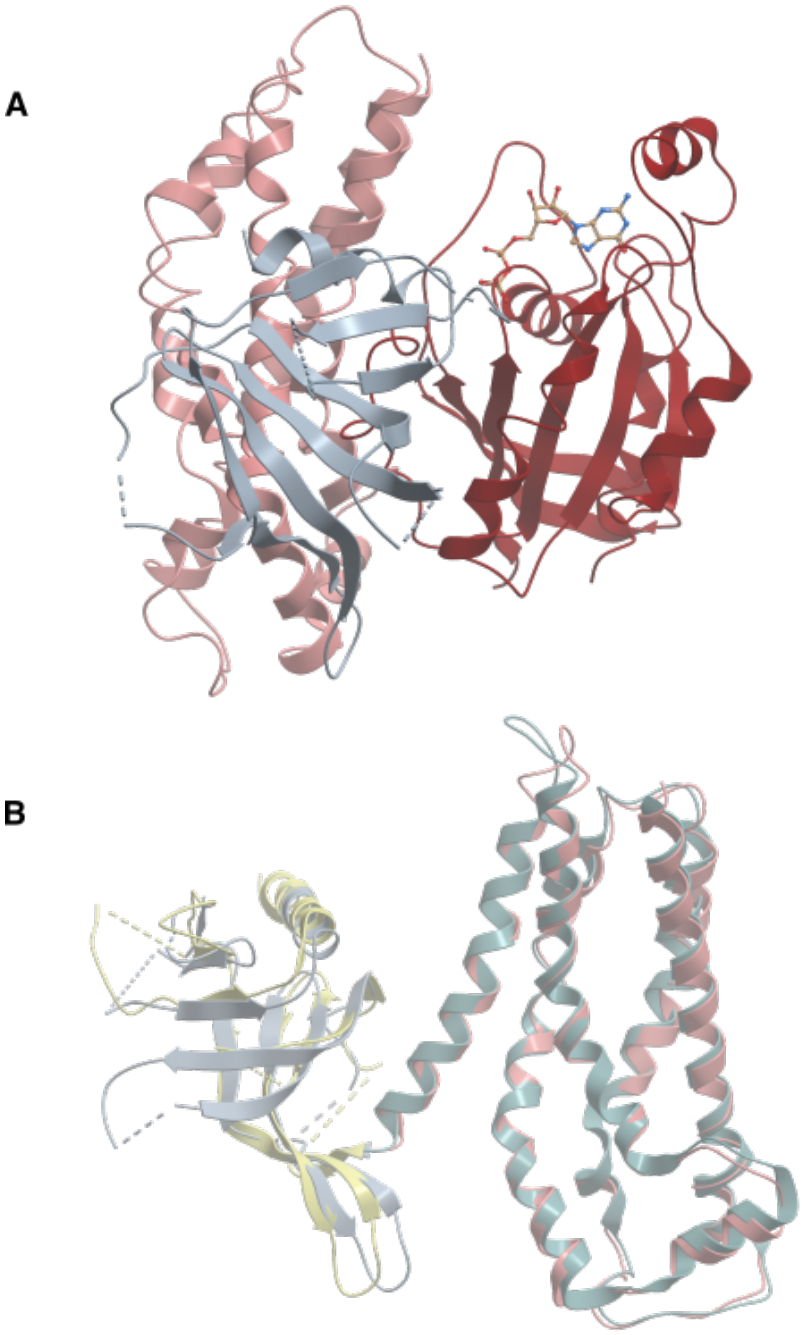


Figure 41. **A.** The novel structure of Kalirin-c002/Rac1 complex with GDP bound. Rac1 is shown as dark red, Kalirin DH domain in light pink, Kalirin PH domain as blue-grey and GDP as brown sticks. The structure of Rac1 is very similar to that of the Kalirin (DH)/Rac1 structure. **B**. Comparison of the Kalirin (DH-PH) structure to that of Trio (PDB code: 2NZ8). The structure is rotated 180° compared to **A**. Trio DH domain is shown as light green, Trio PH domain as light yellow. The PH domain of Kalirin forms a similar β-sandwich motif. Both structures contain unmodelled regions (dashes) within the relatively disordered PH domain.

Whilst the electron density for Rac1 and the DH domain of Kalirin were reasonably well-defined, sections of the PH domain had poorly characterised electron density and were difficult to model. This resulted in segments of the PH domain either being removed from the model or were modelled with high B factors. The flexibility or relative disorder of the PH domain has been documented previously for Trio and may account for the lower quality diffracting crystals (Figure 41B).[173] Chhatriwala *et al* state that the flexible PH regions of Trio that are not in contact with Rac1 have extremely high B-factors, with the average anisotropic B-factor of the PH region of Trio documented as 143.6, which is consistent with the high B-factors in the Kalirin PH domain.

Whilst it is known that the PH domain is required for full GEF activity in Trio, with the DH domain being four times less active *in vitro*, few changes can be observed between the two Kalirin/Rac1 structures. A major difference is the orientation of the key residue H1410 in Kalirin, which is needed for GEF recognition and activity. R66 in the switch II region of Rac1 significantly moves to allow space for H1410 (Figure 42). Whilst R66 is extended in the DH structure, it reorientates back towards Rac1 to allow key hydrogen bonding interactions between D65 and N1406, H1410 and Y1472 of Kalirin. Analogous interactions have been observed in Trio and Dbs and are vital in cellular models for the full activity of these GEFs. The side chain of H1410 in Kalirin-c001/Rac1 could not be modelled due to poor electron density resulting from it being the end residue of the construct. Fewer interactions between the PH domain and the relevant GEF has been observed in the crystal structures of less related Rho GEFs such as Tiam1 or LARG. This is suggestive that if a ligand was found to bind near the switch II region or the interacting residues of Kalirin, then selectivity over the larger Rho subfamily could be achieved.

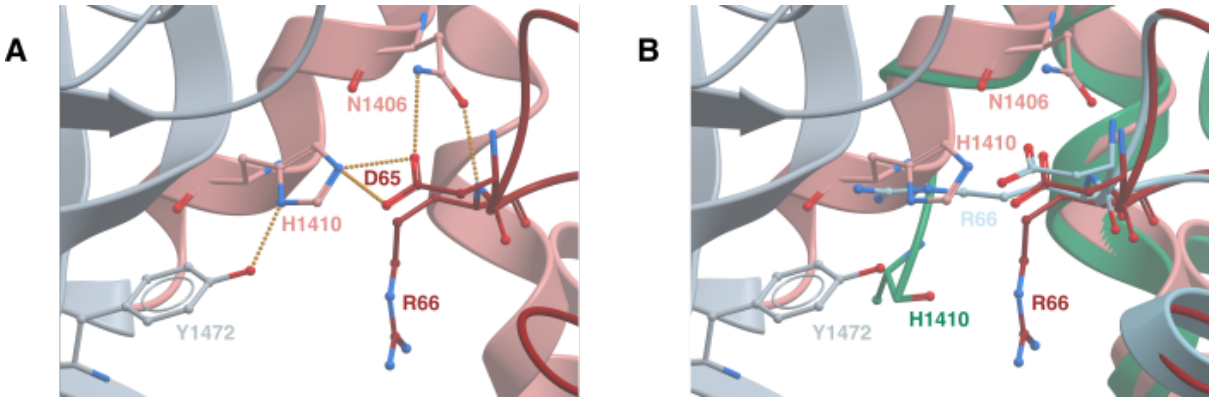


Figure 42. **A**. H1410 of Kalirin-c002 is key for GEF recognition and activity. It forms a hydrogen bond network with Y1472 of the PH domain (blue-grey), D65 of Rac1 (dark red) and N1406 of the DH domain (pink). **B**. Comparison of Kalirin-c002/Rac1 with Kalirin-c001/Rac1 (PDB code: 5O33, Kalirin-c001 in green, Rac1 in blue). R66 moves significantly in the Kalirin-c002/Rac1 structure to allow H1410 to be orientated correctly.

Apart from these residues mentioned, regions of the Rac1 switch I and II loops remain remarkably similar, and the presence of the PH domain does not seem to significantly change the orientation of Rac1. It has been suggested for Trio that the crystallographic conformation may not be representative of the true orientation of the PH domain in solution due to the marked lack of interactions and the high flexibility observed in this region.[173]

Efforts to optimise the crystal quality or obtain a different crystal space group reaching higher resolution using coarse screen, cross seeding and additives were unsuccessful. Hence, it was decided to continue with the Kalirin-c001/Rac1 crystals for XChem and structural determination (from now on, referred to as just Kalirin/Rac1 and Kalirin-c002 will be specified). The structures of the DH domains were not significantly different, and the guanine nucleotide binding site appeared to be representative and so could be used in structure-based drug design.

3.2.7 IQSEC2/Arf1

IQ motif and Sec7 domain-containing protein 2 (IQSEC2) is a member of the Sec7 Arf subfamily of GEFs, and representations of its structure have not previously been deposited in the PDB.[180] Brush-shaped crystals of the IQSEC2/Arf1 complex were originally obtained in the HCS coarse screen from Molecular Dimensions.[176] Optimisation and seeding led to rod-shaped crystals that diffracted to 2.7 Å at the DLS. Molecular replacement (MR) using IQSEC1/Arf1 as a homologous model (PDB code: 4C0A) was used to solve the structure but phase bias resulted in a model of insufficient quality. Attempts to spike the crystal with Br- or I-labelled GDP were unsuccessful. Soaking with mercury salts produced crystals that diffracted to 2.3 Å and was used to determine the phases of the electron density map. Further exploration of additives to the crystallisation conditions yielded a rod-shaped crystal that diffracted to 2.35 Å. The electron density maps from this crystal were used to refine the model originally built in the heavy atoms experiment, as the diffraction data was of better quality due to less radiation damage. The structure was deposited in the PDB with code 6FAE (Figure 43A).

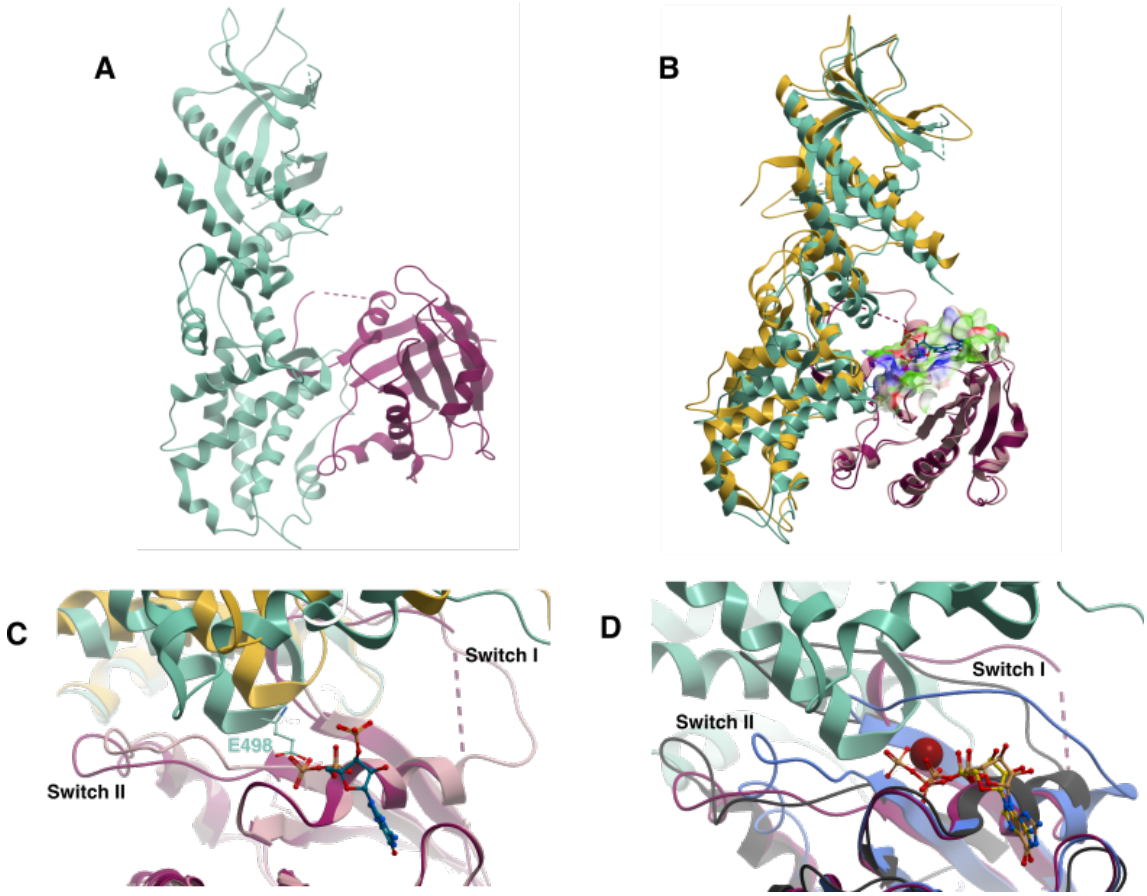


Figure 43. **A**. The novel structure of the IQSEC2/Arf1 complex (PDB code: 6FAE). IQSEC2 is shown in light green, Arf1 in maroon. Disordered, non-modelled regions are represented as dashed lines. **B**. Overlay of IQSEC2/Arf1 with the homologous IQSEC1 (yellow) bound to Arf1 (light pink) structure (PDB code: 4C0A). A large solvent accessible nucleotide binding site is shown as a surface. Using this homologous model in a MR strategy resulted in phase bias. **C**. The position of the glutamic finger, E498, in the IQSEC2/Arf1 complex in comparison to the GDP-analogue bound IQSEC1/Arf1. The glutamic finger in IQSEC1/Arf1 was mutated to lysine and so was not modelled. E498 is positioned in the same vicinity as the β-phosphate of the guanine nucleotide, competing for interaction leading to a reduction in the binding affinity. **D**. Comparison of the IQSEC2/Arf1 structure with that of Arf1 alone bound to GDP (black, PDB code: 1U81) and GTP (blue, PDB code: 1O3Y) with guanine nucleotides shown as sticks. Significant differences are observed in the switch regions.

The structure is nucleotide free, and when overlaid with the homologous GDP-bound IQSEC1/Arf1 structure shows an extremely large and accessible guanine-nucleotide binding pocket (Figure 43A). Comparing the complexes shows them in a similar conformation, albeit with IQSEC2 shifted slightly towards Arf1 in comparison to IQSEC1; this might result from IQSEC2/Arf1 being nucleotide free and forming a more tightly bound complex. The conformations of the switch I and II regions of Arf1 bound to IQSEC1 are different in the IQSEC2 complex. These conformational changes may explain the phase bias conundrum when using IQSEC1/Arf1 as a homologous model.

The IQSEC2 Sec7 domain consists of ten alpha-helices typical of the Arf GEFs.[181] The characteristic 'glutamic finger', the key residue of Arf GEFs which catalyses GDP/GTP exchange can be observed in the IQSEC2/Arf1 structure in the expected orientation (Figure 43C).[182] The Arf GEFs catalyse exchange by insertion of the glutamic finger into the nucleotide binding site, where it competes with the β-phosphate group of the guanine nucleotides for key residues in the Arf1 binding pocket, thereby leading to a reduction in affinity of GDP or GTP for Arf1.

Large differences in the Arf1 structure, especially in the switch I region, are observed between the GEF complex and free Arf1 bound to guanine nucleotides (Figure 43D). In the GTP- and especially GDP-bound form, the switch I loop must re-orientate significantly to avoid clashing with IQSEC2. In the complex, this region is relatively disordered and could not be modelled accurately. The switch II loop of Arf1 in the complex is orientated more similarly to GDP-bound Arf1, whereas the GTP-bound switch II loop is in a conformation closer to the nucleotide binding site to stabilise the binding of GTP.

It is interesting to compare the structures of Arf1 and Rac1 in their different GEF complexes (Figure 44). Whilst in general, the conformational arrangement of various secondary elements was the same in Arf1 and Rac1, no real similarities can be observed between the two GEFs. The greatest difference in the small GTPases can be perceived in

the disordered switch I and II regions, which are orientated outwards away from the rest of the GTPase in Arf1 compared to the analogous position in Rac1. Whilst this may be due to the presence of GDP in the Kalirin/Rac1 complex resulting in a 'closed' instead of 'open' formation of the switch I loop, it may also be attributed due to the difference in switch arrangement between the two Ras subfamilies. These differences could be exploited for compound selectivity. It is also possible to identify the α-helical insertion characteristic of the Rho subfamily in the Rac1 structure.

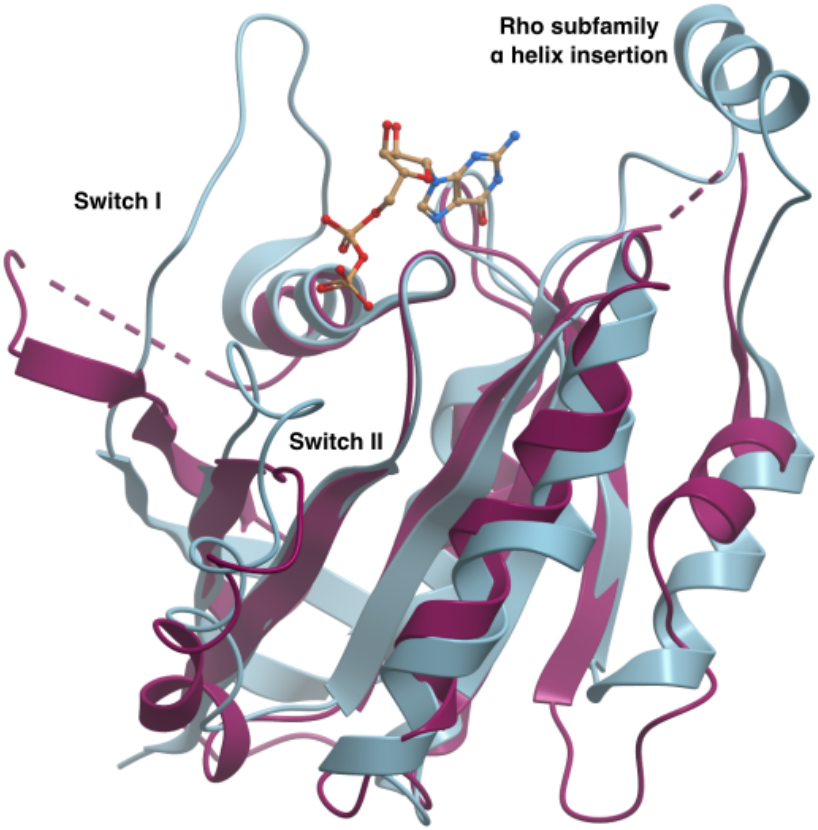


Figure 44. Comparison of Arf1 (maroon) and Rac1 (blue) in their GEF complexes. There are significant differences in the switch I and II regions which could be exploited for selectivity. The α-helix insertion characteristic of the Rho subfamily is also highlighted.[107] The 3D structures are remarkably similar considering the proteins share approximately 30% sequence similarity.[177]

However, attempts to optimise the IQSEC2/Arf1 structure for an XChem fragment screen were ultimately unsuccessful. The effects of the additive which produced the 2.35 Å diffracting crystal, 0.1 M KCl, could not be replicated in follow-up screens. Multiple strategies were used in an attempt to improve the consistent diffraction quality from ~4 Å to below 2.5 Å, including other additive screens, new seeds, cross-seeding from a different crystal system and the use of artificial 'glues' instead of seeds, but the crystal quality did not improve. Attempts to crystallise other constructs of IQSEC2 were also unsuccessful, as were efforts to obtain crystals of a different morphology in a variety of coarse screens.

A DSF assay using different buffer conditions was conducted to analyse the stability of the complex in the standard PrepX buffer containing HEPES pH 7.5.[183] Shown in Figure 45, a biphasic curve was observed, indicating that IQSEC2 and Arf1 were not stable as a complex in the PrepX buffer and were unfolding separately or aggregating. A different buffer system with slightly higher pH, such as bicine pH 8, resulted in a monophasic curve and so was predicted to be a more suitable buffer for the stabilisation of the IQSEC2/Arf1 complex. Literature reports of DSF assays found that proteins with a monophasic curve are more likely to be monodisperse and thus produce better quality crystals.[184] However, crystallisation trials of IQSEC2/Arf1 in coarse screens following dialysis to alter the buffer only yielded brush-like crystals with similar diffraction quality as observed previously. Hence, it was decided that this crystal system was not suitable for XChem without extensive optimisation in the future.

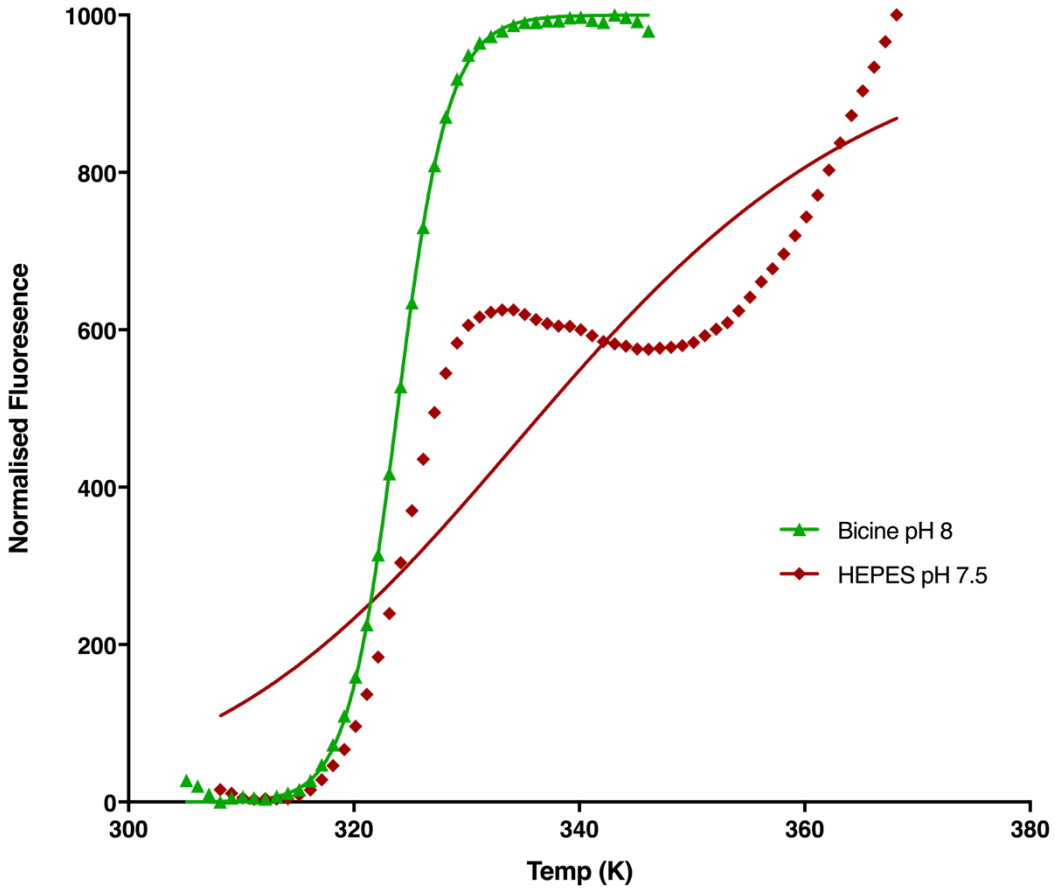


Figure 45. DSF screening assay for the optimal buffer system for IQSEC2/Arf1 (n=3). 0.1 M HEPES pH 7.5 (red, $R^2 = 0.748$), was not favourable for the complex as the proteins unfolded separately at high temperatures, resulting in a biphasic curve. 0.1 M Bicine pH 8 (green, $R^2 = 0.999$) resulted in a monophasic curve, indicating the proteins were more stable as a complex in this buffer system. Least squares regression analysis was completed in GraphPad Prism.[185]

3.3 Pipeline Round 2

3.3.1 Construct Design

Following the successful introduction of the first subset of proteins into the pipeline, a second group of GEFs and GTPases were chosen based on their predicted roles in neurological disorders (Table 7).[171] These GEFs and GTPases had not been crystallised before, and so several novel constructs were designed for each protein to optimise the chance of successful cloning, expression, purification and crystallisation. The proteins chosen were generally members of the Rho subfamily, with the exception of the Ras GTPase Rit2, which is of particular interest in Parkinson's disease.[186] Multiple constructs were also designed for DOCK2 and 3, which were poorly expressed in the first round of the pipeline but were deemed as targets of particular interest.

Instead of using homologous models as in the first subset, constructs were designed using bioinformatic servers Crysalis[187], DisMeta[188] and ProteinCCD[189] to predict the secondary structure of the relevant domains. Several constructs for each target were chosen by systematically varying the ends of the N- and C-termini, being careful not to cut into predicted secondary structures. The end termini were controlled to exclude low complexity regions, which may hamper protein crystallisation, and terminal hydrophobic residues were removed to avoid aggregation during protein production or crystallisation.

Table 7. The second subset of proteins entered into the GEF/GTPase pipeline. Fewer proteins were chosen but more constructs were designed per target to increase the success rate of crystals suitable for fragment screening.

GTPase Subfamily	GTPase	GEF
Ras	Rit2	RASGRF1, Sos2, Sos1
Rho	RhoA	ARHGEF15
	Cdc42	DOCK6, DOCK8, DOCK9, DOCK11
	Rac1	DOCK2, DOCK3, DOCK6, DOCK7, DEF6

3.3.2 Molecular Cloning

Observations from the first round of the pipeline indicated that greater purity was achieved when a Strep tag was utilised instead of a His tag. Hence, constructs of Rit2, Sos1, DOCK2, 3, 7, 9 and 11 were attempted to be cloned into a pNIC-NStIIt vector. However, whilst the PCRs were successful, the transformations failed multiple times. Hence, a wider range of constructs were designed, in combination with the remaining proteins detailed in Table 7 (a total of 190 constructs) and were cloned using a LIC procedure by the Biotechnology group within the SGC into both bacterial and insect cells. It was observed that the transformations of these proteins were especially difficult in bacteria, requiring multiple rounds of optimisation, indicating some level of incompatibility between these proteins and either the vector or the bacterial cell line. These constructs are summarised in Appendix 8.1.2.

3.3.3 Protein Production

Optimisation of the expression and purification of the second subset of proteins was difficult. The standard PrepX protocol was generally unsuccessful, producing sloppy, foul smelling pellets which could not be centrifuged, suggesting that toxic levels of the protein had been produced and were causing cell lysis during protein expression. Experiments using the original subset of proteins yielded similar results, and it was eventually discovered that the composition of the autoinduction media had changed and was not longer suitable for expressing the GEFs and GTPases. The expression of proteins was adapted to use terrific broth (TB) media and IPTG, and careful management of growth times produced usable cell pellets. Exceptions included the DEF6 constructs, which were relatively easy to express in either media.

Test expressions conducted indicated no usable expression of most of the DOCK2, DOCK3 or RIT2 constructs in bacteria, whereas DOCKs 6, 7, 8 and 11 all had constructs with suitably high yields and purity to progress into crystal trials. Scaling up of these proteins is currently being pursued by Andrew Thompson of Frank von Delft's group, as are some promising constructs of DOCK3 and RIT2 in insect cells.

3.3.4 Crystallisation Trials

Certain constructs had sufficient yields from 100 mL test expressions for progression into crystallisation trials, as shown in Table 8. In general, the drops were mainly clear, indicating that for these proteins, higher concentrations than 10 mg/mL – which was suitable for the crystallisation of Kalirin-c001/Rac1 - were required to reach supersaturation. This is perhaps not unexpected due to the larger size of the DOCK proteins resulting in a lower molar

concentration at 10 mg/mL. The exceptions were the complexes of DOCK6/Rac1 and DEF6/Rac1. The optimisation of these crystals is being pursued by Andrew Thompson.

Table 8. Current status of the crystallisation trials of the second subset in the GEF/GTPase pipeline. Constructs for Rit2, DOCKs 6, 7, 8 and 11 are currently being optimised for crystallisation trials by Andrew Thompson. To date, no system suitable for XChem screening has been achieved.

GTPase Subfamily	Crystal trials	Crystal Optimisation	Structure Determination	FBLD screening
Rho	DOCK2/Rac1 DOCK7/Rac1	DEF6/Rac1 DOCK6/Rac1		

3.4 Conclusions

Despite encountering challenges during cloning and protein production, the GEFs and GTPases were tractable for large scale production within the SGC. The purification and crystallisation trials of several complexes of interest in neurodegenerative diseases are currently being pursued within Frank von Delft's group. It is interesting to note that none of the complexes that crystallised were previously represented in the PDB, with no documented complexes yielding promising crystal-like conditions. It may be that the general PrepX method requires optimisation on a case-by-case basis; certainly, performing the DSF buffer assay for other complexes may increase the rate of crystallisation success.

The structures of three novel structures were solved. The IQSEC2/Arf1 complex produced needle-like crystals which were difficult to optimise. Subsequent heavy atom soaking, and fine screening of crystallisation conditions yielded a high-resolution crystal, but this system was ultimately not suitable for fragment screening due to its low reproducibility. Two structures of Kalirin/Rac1 were solved, with one representing the GEF DH domain alone and the other showing the DH and PH domains of Kalirin in tandem. The crystal of Kalirin-c002 in complex with Rac1 was solved in low resolution due to the high disorder in areas of the PH domain, consistent with literature reports of homologous proteins. However, the Kalirin-c001/Rac1 complex was extremely robust and suitable for ligand screening, and so was used to test the hypothesis that the generation of a novel competitive inhibitor of the GEF/GTPase complexes was a suitable method for targeting the Ras superfamily.

4 Kalirin/Rac1 XChem Fragment Screening

4.1 Overview

This chapter begins with an introduction to the Rho subfamily and their GEFs, with a focus on Kalirin, its role in disease and its potential as a therapeutic/probe target. It then details the XChem fragment screens which were conducted on Kalirin/Rac1, which included the development of an assay to remove GDP from Rac1. The identification of fragment hits that bound within the Rac1 guanine nucleotide binding site formed the basis for a computational docking protocol, which led to the purchase of commercial analogues and the chemical synthesis of bioisosteres. This chapter also describes the development of a fluorescence-based assay to identify which fragments inhibited the loading of fluorescent GTP analogues into the Kalirin/Rac1 complex. This thereby allowed ranking and prioritisation of chemical synthesis for scaffold optimisation. These findings formed the basis of a proposal which led to Kalirin/Rac1 being accepted as a Target Enabling Package (TEP) for the SGC.[190,191]

4.2 The Rho Subfamily and their GEFs

The Rho subfamily of the Ras GTPases consists of 20 members and are essential regulatory proteins involved in actin organisation and cytoskeletal dynamics.[97] The proteins Rac1, RhoA, and Cdc42 are the most extensively studied and well-known members of the subfamily. Dysregulation of Rac1 and its downstream effectors including p-21 activated kinases (PAK), have been linked to a number of neurodegenerative diseases. Inhibition of PAKs and their aberrant activation by Rac1 has been suggested as a possible therapy for Alzheimer's disease.[192]

The activity of the Rho GTPases is regulated by their GEFs, of which there are two distinct sub-groups, Dbl and DOCK.[193,194] The Dbl family of GEFs contain a DH domain which is responsible for their GEF activity. The DH domain is commonly associated with a Pleckstrin homology (PH) domain in tandem which is responsible for membrane localisation and required for the full catalytic GEF activity. With over 70 GEFs in the Dbl subfamily, each Rho GTPase can bind to several Dbl Rho GEFs, allowing spatial control over their activity (Figure 46). There are also two GEFs, Sos1 and RASGRF, which contain a DH domain as well as their Ras Cdc25 GEF domain, and so are capable of activating both Ras and Rho GTPases. Certain Dbl Rho GEFs contain other domains important for non-GEF activities in the cell, such as serine-threonine kinase domains for Rho GTPase-independent phosphorylation events.[193] More recently discovered, the DOCK subfamily instead contain two DHR domains, DHR1 and DHR2. Only the DHR2 domain is responsible for GEF activity, with the DHR1 being a C2-like domain responsible for membrane localisation.[195] With only 11 members, the DOCK families are cited in the literature to activate Rac1 and Cdc42. The Dbl and DOCK families do not have any sequence or structural similarity.[194]

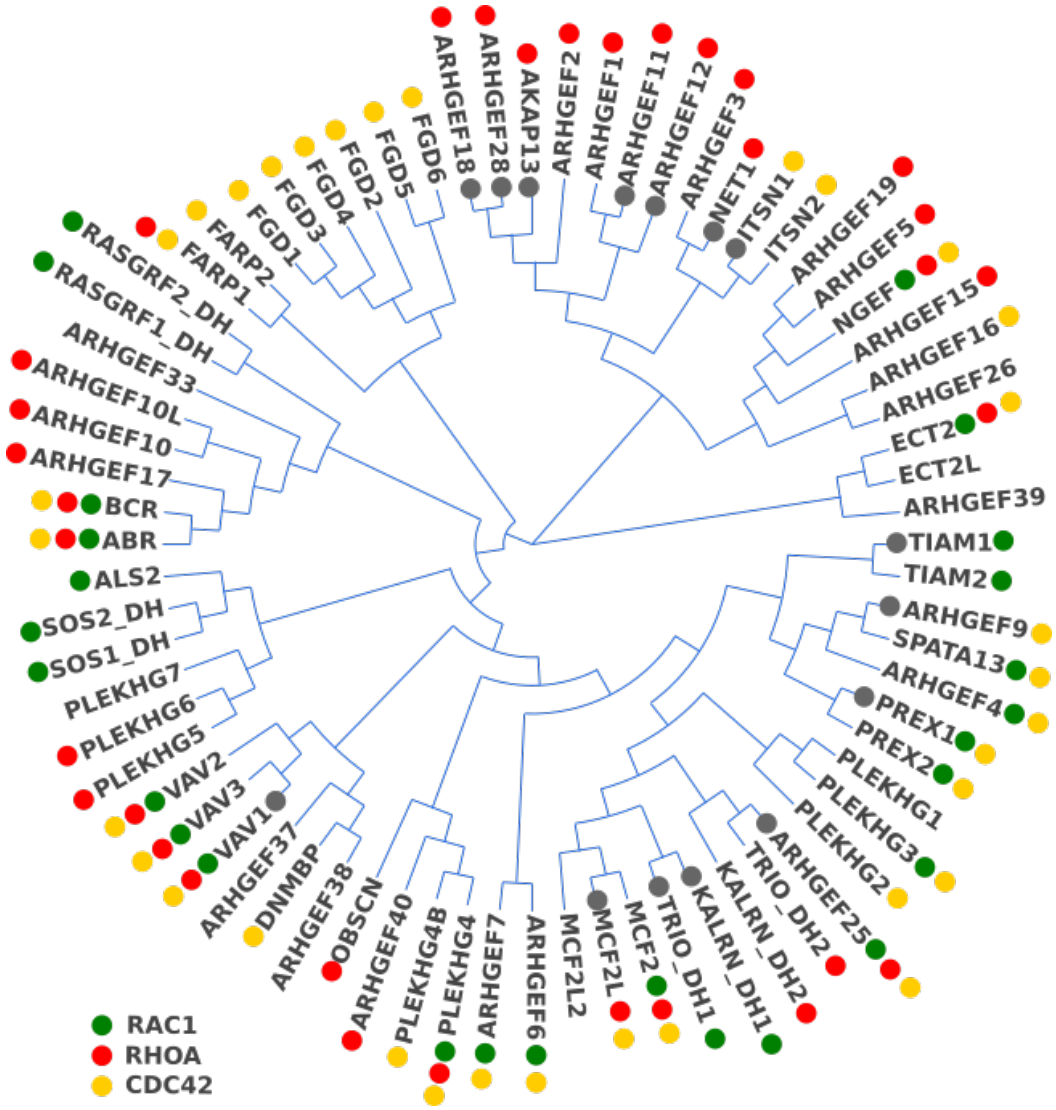


Figure 46. Phylogenetic tree for the Rho Dbl family of GEFs. Grey dots refer to proteins that have structural information available in the PDB. The green, red and yellow dots indicate if there is evidence in the literature for the GEFs targeting one of the three main Rho GTPases; Rac1, RhoA or Cdc42. Inspired by Cook *et al.*[193]

Kalirin is a Dbl Rho GEF named after the multi-armed Hindu goddess Kali, as it contains multiple domains which enable it to bind to several proteins.[196] It was first identified as binding to Huntington-associated protein (HAP) before its GEF activity was established. Kalirin is an unusual exchange factor in that it contains two GEF DH domains; the Rac1 activating DH1 domain, and the RhoA activating DH2 domain (Figure 47). These domains

have distinct roles in the cell.[197] Kalirin shares ~60% homology with its paralog Trio, although this increases to ~90% conservation when comparing only the DH1 domain; however, whilst Trio is ubiquitously expressed, Kalirin is most enriched in the brain.[198,199]

Figure 47. Diagram of the domain sequence of the full length Kalirin gene. The domains that form part of the truncated Kalirin-7 isoform have been indicated. Full length Kalirin contains a Sec14p-like ("S") domain, multiple spectrin-like domains, a DH domain that binds to Rac1 (DH1) and RhoA (DH2), PH domains associated with each DH domain, 2 SH3 domains, an IG-like C2-type domain ("I"), a Fibronectin type-III ("F") and a kinase ("Kin") domain. The DH and PH Kalirin-7 domains that were crystallised as part of the Rac1 complex are drawn with an aura. The multiple other non-GEF related domains in the protein, such as the C-terminus kinase, indicate other non-GEF roles in the cell. Reproduced from Gray *et al.*[191]

The most abundant isoform is Kalirin-7, which is truncated such that the only GEF domain present is the Rac1-specific DH1 domain (Figure 47).[200] Kalirin-7 mediates dendritic spine formation and actin remodelling through its activation of Rac1 and its downstream effectors such as PAKs.[200] Hence, Kalirin has been linked to a number of neurodegenerative disorders associated with dendritic spine pathology.[200,201] Kalirin is also at the epicentre of signalling pathways with other key targets in neurodegenerative disorders, such as regulating the function of neurotransmitters NMDA and AMPA receptors.[202,203] Kalirin-7 knockout mice exhibit decreased spine density leading to anxiety-like behaviour, but they lack the age-dependent characteristics such as hyperactivated locomotion or reduced spatial memory found in KALRN gene knockout mice.[204] This indicates that Kalirin-7 has separate functions to the other isoforms, which may be attributed to the RhoA GEF DH domain in Kalirin-8, -9 and -12 isoforms. Gene expression of KALRN is reduced in Alzheimer's disease and

schizophrenia, and Kalirin interacts with several proteins associated with schizophrenia; Disrupted-in-Schizophrenia 1 (DISC1) anchors to Kalirin-7, regulating its access to Rac1. These studies found that constitutive Rac1 activation leads to a decrease in spine size.[205]

However, the roles of Kalirin, both related to its GEF and non-GEF activity, are relatively unknown with experiments currently limited to indeterminate studies such as knockout mice and RNA interference models. Hence, a compound which can specifically inhibit Kalirin GEF activity could be used to probe the various roles of Kalirin in psychiatric disorders and indicate how inhibition of the GTPase signalling networks impacts normal brain function.

4.3 XChem Fragment Screening and Hit Identification

Due to its proposed involvement in neurodegenerative disease, and its reproducible and high-quality crystals, the complex of the Kalirin DH domain bound to Rac1 identified in section 3.2.5 was deemed suitable for XChem screening.

4.3.1 First XChem Screen

An initial screen of 450 crystals were soaked using subsets of the DSI-Poised[16], OxXChem[190] and covalent libraries[40] at 100 mM concentration for one hour at DLS. The success rate for obtaining soaked crystals diffracting better than 3 Å – following low quality crystals, mounting failures and crystals melting – was 72%. Selected datasets were submitted through the XChem Explorer pipeline for molecular replacement and refinement[93] using the GDP-bound Kalirin/Rac1 structure (PDB code: 5O33) as a reference dataset. It was occasionally necessary to reprocess using iMosflm[206] due to an autoprocessing failure resulting from the crystal being mounted in an orientation where the beam was shot down the long axis of the unit cell, which led to an overlapping of spots which were difficult to differentiate.

The Pan-Dataset Density Analysis (PanDDA) algorithm was then used to identify ligands.[94] PanDDA takes a random subsection of maps from the screen and uses them to construct a 'background' ground state electron density map. This map is then subtracted from the electron density map of every individually soaked crystal. If no ligand is bound, then the difference between the average map and the soaked map is zero. If a ligand is present, then the difference between the average map and the electron density from the relevant crystal should be flagged by the program. This enables the identification and fitting of even weakly bound ligands into the event map created by PanDDA.

However, in the case of Kalirin/Rac1, the PanDDA program was largely unsuccessful. Despite the crystal structure obtained prior to the XChem screen showing no evidence of GDP, the occupancy of GDP varied over hundreds of data sets and differing batches of protein. This resulted in GDP often being identified by PanDDA as a ligand in crystals with higher-than-average occupancy of GDP. The difference in electron density between the maps due to the varying GDP concentration made analysis challenging, with changes in GDP density obscuring electron density from ligand binding events and creating noise (Figure 48A). There was concern that the presence of GDP could also result in false negatives; fragments are often weakly binding and so could be unable to compete with GDP, despite being capable of binding within the orthosteric site, resulting in their being missed during the XChem screen.

There were also several cases where potential binding events were identified in the switch I region. However, it became apparent that this density was misidentified due to a rearrangement of the switch I loop. The presence of different loop arrangements superimposed in a single crystal meant the electron density maps were ambiguous. The background density analysis was unable to differentiate between a loop movement and ligand binding densities, even at low contours of the difference maps (Figure 48B). This was further complicated by the proximity of the switch I loop to its symmetry copy, which led to overlapping electron density maps so that the residues of the loops could not be fitted with any certainty.

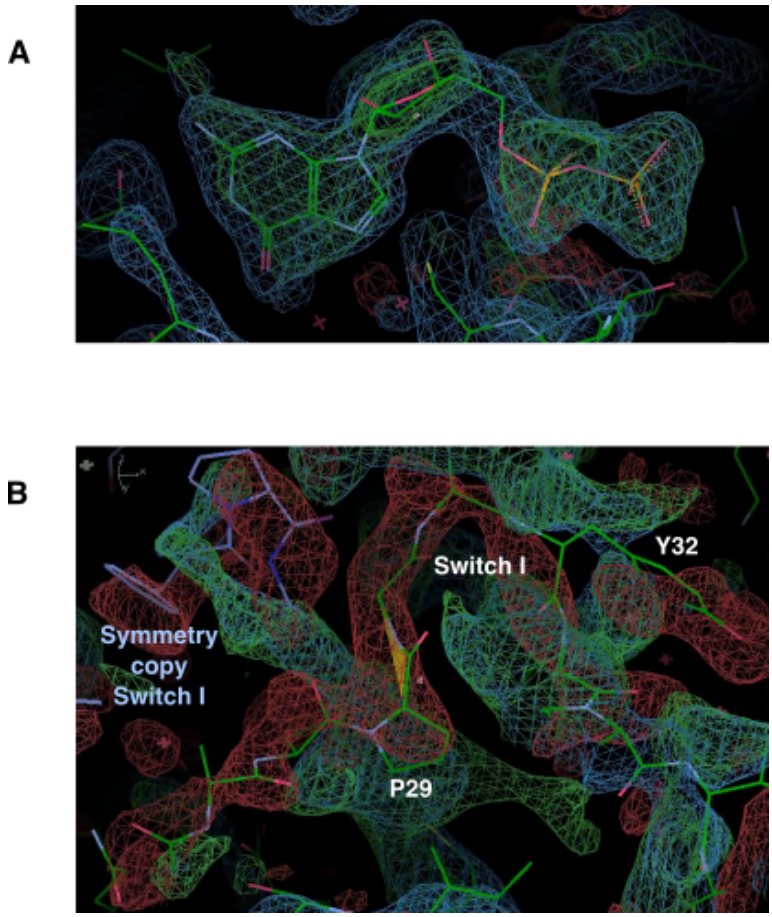


Figure 48. The PanDDA analysis from the first XChem was generally problematic. The PanDDA event map is shown as a blue mesh (contoured at 2σ). The Fo-Fc map is shown as a green (positive density, +3σ) and red (negative density, -3σ) mesh. **A**. GDP was present in varying amounts in the crystals. Crystals containing GDP at a higher occupancy than the ground state maps would be flagged by PANDDA as a hit. This made the maps ambiguous and potentially masked the density of weakly binding fragments. **B**. The residues in the switch I loop did not always adopt the same conformation as the background density model. This made the maps in this area difficult to model and resulted in the Fo-Fc map covering any ligand binding events. The proximity of the loop to the symmetry copy (light blue sticks) also contributed to the challenge in modelling the conformation correctly.

4.3.2 Fragment Identification from the First XChem Screen

Despite these challenges, careful evaluation of the maps led to the identification of two fragment hits binding to Rac1 in the Kalirin/Rac1 complex. The first fragment hit, **32**, bound within the guanine nucleotide binding pocket with high occupancy, suggesting it displaced GDP from the nucleotide binding site (Figure 49A). The urea moiety forms two hydrogen bonds with D118 and the isoxazole nitrogen interacts with the peptide bond of A159, generating a similar hydrogen bond donor/acceptor pattern to GDP.

There were also movements within the switch I loop and active site near where **32** binds that could be observed in the Kalirin/Rac1 complex that were hypothesised could be exploited for selectivity over Rac1 alone. Some of these movements are shown in Figure 49B. F28, whilst not changing position significantly, is orientated closely to where the urea binds and so could be targeted to increase the number of interactions between the ligand and the binding pocket. Significant shifts are observed in the switch I loop between the urea-bound complex and Rac1 alone, with E31, Y32 and I33 shown as examples. Interactions with these residues could be targeted, or the ligand built into the space vacated by them in the complex structure. This ligand would then clash with the favoured conformations of the Rac1 when bound to guanine nucleotides.

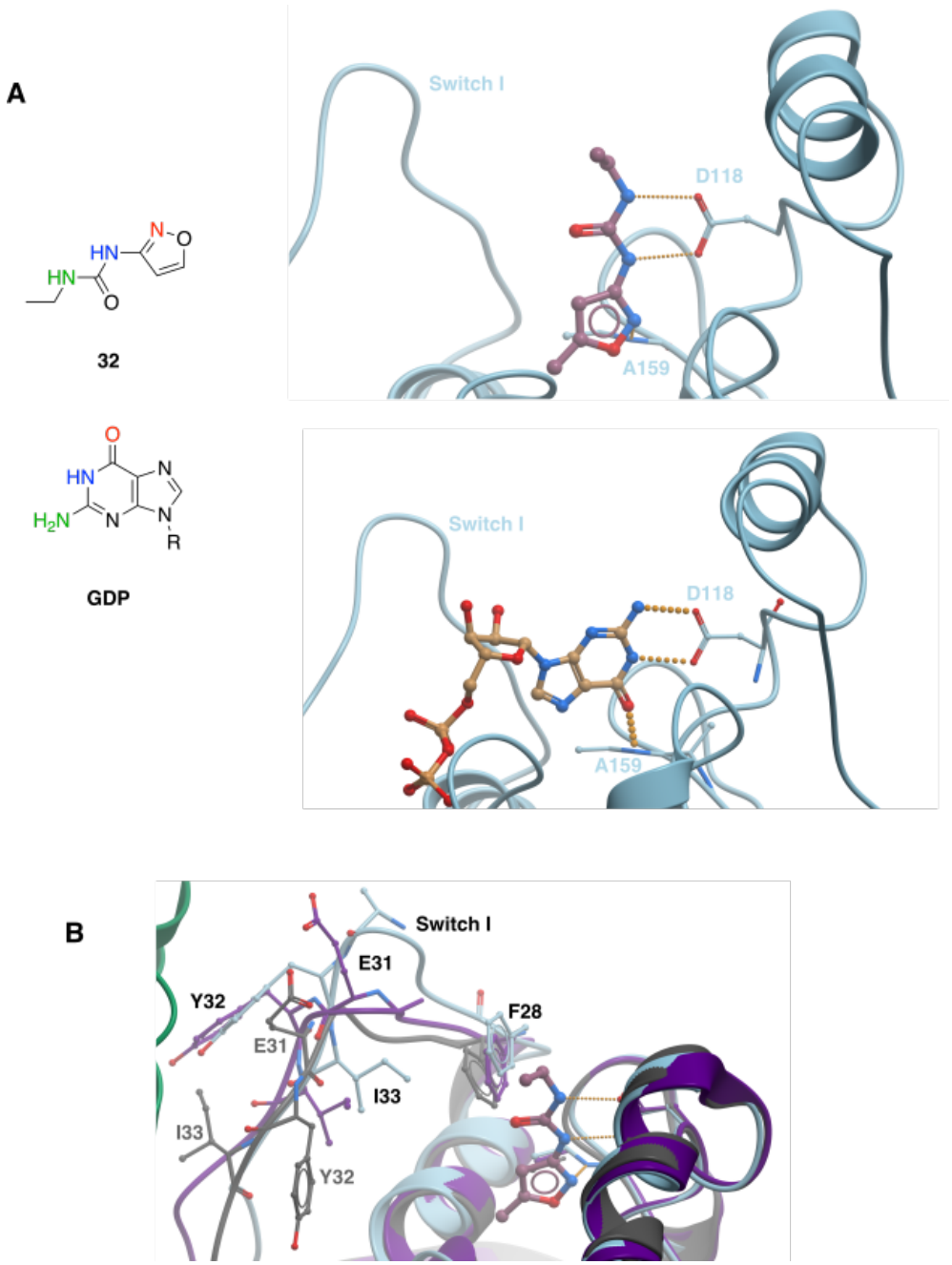


Figure 49. **A**. The urea-isoxazole fragment **32** (purple sticks) binds to Rac1 (light blue) within the guanine nucleotide binding pocket (PDB code: 5QQD). It mimics the key hydrogen donor-acceptor pattern seen in the guanine ring of GDP, with the residues involved highlighted in each structure. The nitrogen atoms from the urea motif acts as donors to form the same key hydrogen bonds (orange spheres) with the oxygens of D118 of the G4 box as GDP. The isoxazole ring forms an additional hydrogen bond with the backbone amide of A159 of the G5 box. **B**. Switch I loop movements between the urea-bound Kalirin/Rac1 complex and GDP-bound Rac1 (purple, PDB code: 5N6O) and GNP-bound Rac1 (grey, PDB code: 3TH5). These could be exploited for selectivity or potency.

The second fragment hit **33** bound in a 'putasteric' – defined as a putative, non-orthosteric site identified in a structure but with no activity assay data to support a definition of allostery[207] - polar pocket behind the switch I loop and the C-terminal α-helix (Figure 50). The pocket was tight, and the fragment only formed a single hydrogen bond with the protein, creating concerns regarding potency and potential for optimisation. However, the fragment was observed bound in two independently soaked crystals, indicating that this was a genuine hit.

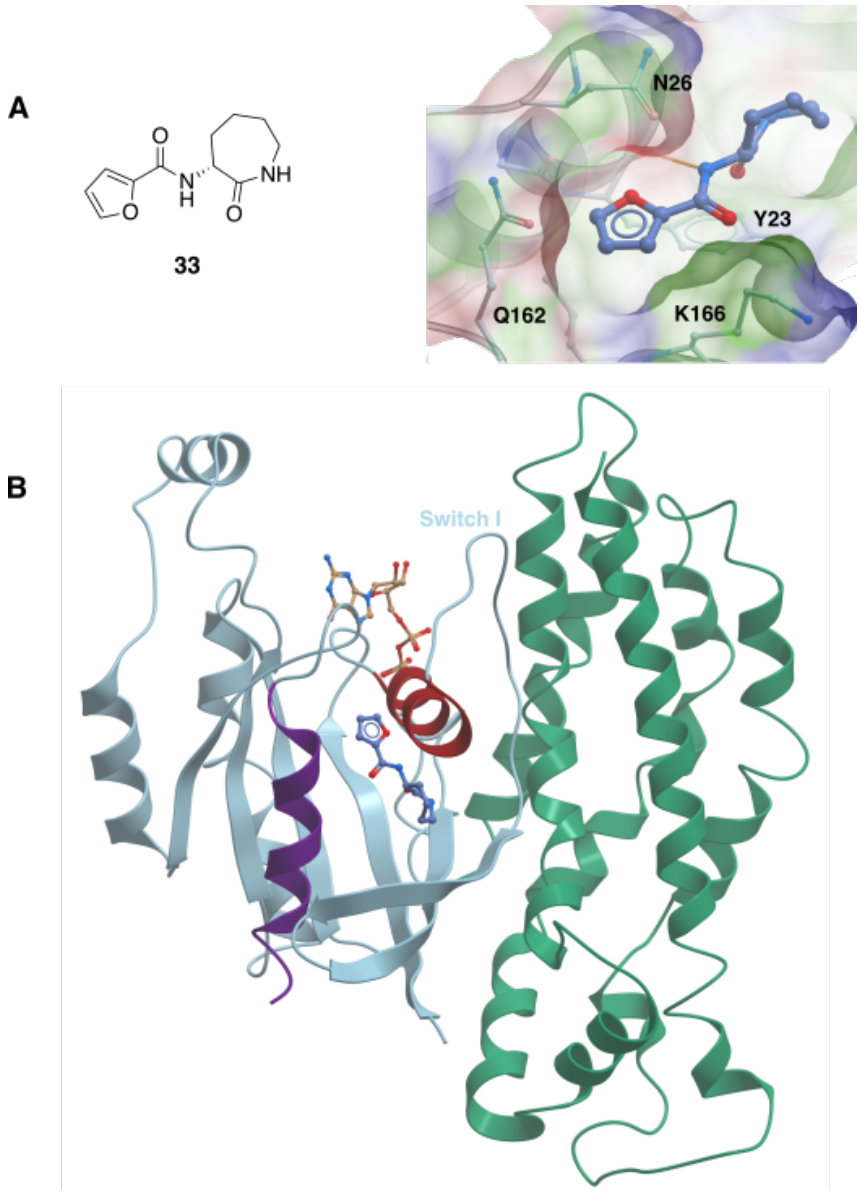


***Figure 50*. A**. The second fragment, **33**, found in the XChem screen. It forms a single hydrogen bond (orange spheres) with the C=O backbone of residue Y23. **B**. The binding pocket is situated adjacent to

the C-terminal α-helix 5 (purple) and behind the switch I loop and α-helix 1 (dark red) of Rac1 (light blue) bound to Kalirin (green).

The success rate of identifying hits (0.4%) was considered very low. It was thus decided that explorations of the SARs of the two fragment hits **32** and **33** would be pursued following the development of a reliable GDP removal protocol. It was hoped that a subsequent nucleotide-free XChem campaign would yield more fragment leads, which could be elaborated with **32** and **33** concurrently (section 4.3.5).

4.3.3 Strategies to Remove GDP from Kalirin/Rac1 Complex

4.3.3.1 Attempted Optimisation of the EDTA Method

Multiple reports in the literature detailed the addition of EDTA to purification buffers in order to obtain nucleotide-free Rac1.[123,173,208] The GDP removal method initially involved dialysis of the complex into a buffer containing 10 mM EDTA prior to purification by SEC to chelate Mg^{2+} and stimulate GDP dissociation. As this was insufficient to remove all GDP from crystals, the concentration of EDTA was thus increased to 20 mM.[208] Native mass spectroscopy performed by the Biotechnology group at the SGC indicated that only 5% of the complex was bound to GDP. This was presumed to be unlikely to complicate the PanDDA analysis or mask a ligand with higher occupancy.

A further 523 crystals were soaked using subsets of the DSI-Poised[16], OxXChem[190], Covalent[40], Leeds3D[209] and Maybridge[210] libraries. However, disappointingly, there was again variation in the GDP density between crystals which made analysis challenging. It was postulated that GDP bound in the complex was stabilising for crystal formation and so the bound complex was preferably crystallised over the unbound complex. The same protein batch was subjected to a small range of crystallisation conditions and the subsequent crystals were mounted and analysed. Although electron density for GDP could not be observed at a

1σ contour level (2Fo-Fc), lower levels showed that the phosphate groups of GDP were present in low occupancy. This would cause problems with the electron density PanDDA analysis and suggested sufficient GDP was present in crystals to compete with the soaked ligands.

It was suggested that the relatively high salt concentration of the PrepX buffer (500 mM NaCl) may have been affecting the rate of exchange of the GDP following EDTA addition. Hence, a different dialysis protocol was attempted using EDTA in a salt-free buffer. However, native mass spectrometry performed by the Biotechnology group at the SGC showed that this was unsuccessful in completely removing GDP from the complex. Further experiments implied the EDTA method was unreliable, despite publications reporting the use of EDTA to generate nucleotide-free Ras proteins.[173] This was probably due to the lack of Mg^{2+} observed in the crystal structure. Native mass spectrometry comparing different purification batches of protein identified a range of 5-20% GDP bound to the complex. Subsequent experiments testing other literature protocols used activated charcoal as an alternative to EDTA in the dialysis buffer, but this was completely unsuccessful in removing GDP.[211,212] Attempts at removing the GDP from Rac1 alone using EDTA before formation of the complex were also ineffective.

In hindsight, it was perhaps not surprising that it was difficult to remove the GDP from Rac1 using EDTA, as there is evidence that Rac1 binds strongly to GDP even in the absence of magnesium.[208] Whilst the EDTA protocols may have been sufficient for assays or the production of a single nucleotide-free crystal structure, the analysis of the average electron density over 1000 crystals means complete removal of GDP in a robust manner was required. An alternative to native mass spectrometry to monitor the GDP removal was also desired, as it was time consuming, required large amounts of protein and required an expert to run the experiment.

4.3.3.2 Development of an Alkaline Phosphatase GDP-removal Method

Several reports in the literature detail the degradation of nucleotides by alkaline phosphatase to prepare nucleotide-free complexes.[164,213] A non-hydrolysable GTP analogue is added in excess to the desired protein with a buffer that catalyses the exchange of the bound natural nucleotide with the non-hydrolysable analogue. Alkaline phosphatase (AP) is added to degrade free GDP into GMP and guanosine, leaving only the non-hydrolysable analogue bound to the GTPase. Following the complete degradation of GDP, snake venom phosphodiesterase (PDE) is added to degrade the non-hydrolysable analogue. This results in only guanosine or GMP bound to the GTPase (Figure 51). Guanosine and GMP have extremely low affinity for the GTPases in comparison to GDP (mM instead of pM) and can be easily removed by gel filtration. The degradation of GDP or the non-hydrolysable analogue into guanosine or GMP can be monitored using HPLC, as each compound has a different retention time (Appendix 8.6).

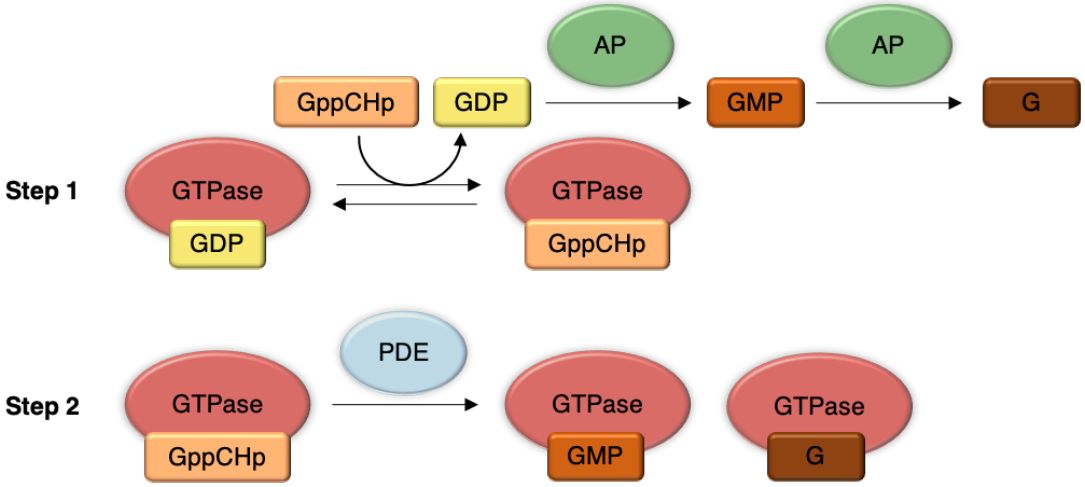


Figure 51. Outline of the alkaline phosphatase method used to remove GDP from Kalirin/Rac1. In the first step, alkaline phosphatase (AP) is added to the mixture along with GppCHp to degrade the GDP bound to the GTPase into GMP or guanosine (G). In the second step, phosphodiesterase (PDE) is added to digest the alkaline phosphatase resistant GppCHp into GMP or G. B. This reaction can be monitored by reverse phase HPLC. As GDP and GppCHp are degraded by AP or PDE respectively,

the relative peak size of guanosine increases. Over several hours, the peaks for GDP and GppCHp disappear entirely.

The published method[213] was optimised at the SGC in collaboration with Carmen Jimenez Antunez. The method was successful in removing GDP for Rac1 alone and the Kalirin/Rac1 complex. It was observed that the exchange of GDP for the non-hydrolysable analogue was faster for the complex than for Rac1 alone, as expected from the GEF Kalirin catalysing the exchange reaction. The method was also successful for other members of the Ras superfamily in the pipeline such as IQSEC2/Arf1 and DENND1A/Rab35. The monitoring by HPLC was quicker and simpler than that of native mass spectrometry and required little protein.

Once it was confirmed by HPLC that Kalirin/Rac1 was bound to guanine or GMP, it was purified by gel purification. Subsequent analysis by HPLC indicated that the complex was nucleotide-free. Crystal trials of the nucleotide-free protein complex resulted in crystals in the same space group but with a change of morphology, from diamonds to sword or rocket ship shaped. The nucleotide-free crystals were on average lower quality than GDP-bound (diffracting ~2.2 Å rather than ~1.7 Å) and could only tolerate up to 10% DMSO when soaked for one hour. Analysis of the electron density maps at < 1 σ showed no evidence of GDP, GMP or guanosine.

4.3.4 Nucleotide-free XChem Screen

With the successful generation of nucleotide-free crystals, an XChem screen of 269 fragments from the DSI-Poised[16] library was conducted. The campaign was less successful than before, with more crystals dissolving upon ligand soaking or not diffracting to high resolution. Analysis of successful crystals by PanDDA only identified the urea **32** as a fragment hit apart from crystal contact binders. Further soaking of 107 Fraglites[214] and MiniFrags[215] to

identify small starting points did not yield any interesting binding events. It had been hoped that these would bind near the binding sites of **32** and **33** to aid easy elaboration. Hence, it was decided to continue with **32** and putasteric binder **33** instead of pursuing fragment screening any further.

From the XChem screens, two important observations can be considered. First, GDP is important for Kalirin/Rac1 crystal formation and stability. The removal of GDP resulted in a change of morphology and comparatively poorly diffracting crystals. When GDP is bound, the switch I and II loops are held in specific orientations. In the nucleotide free complex, the conformations of these loops are the same as in the GDP-bound structure. However, the drop in resolution suggests that these loops have greater flexibility in the nucleotide-free complex, making their subsequent crystallisation more difficult. Second, the urea isoxazole **32** binding to both the GDP-bound and nucleotide-free crystals indicates that is it well suited to the binding site, and that despite its small size, its key interactions are sufficient to recognise the binding site in multiple soaking experiments.

4.3.5 Docking of Follow-up Compounds from Molport

SARs of the lead compounds **32** and **33** were investigated by computational docking and the subsequent purchase of commercially available analogues. Compounds were prioritised based on their docking pose, cost, and availability from the commercial vendor Molport. The docking procedure used in the software Molsoft ICM-Pro is detailed below.

*4.3.5.1 Urea Hit **32***

Substructure searches for the urea hit were divided into three groups: variations of the ethyl chain, bioisosteres of the urea group and alternatives to the isoxazole ring (Figure 52). Compounds were docked using the Kalirin/Rac1 structure bound to **32** (PDB code: 5QQD).

The receptor pocket was generated in ICM and **32** was redocked to ensure that the docking software accurately predicted the correct binding pose. The docked pose of **32** overlapped exactly with its crystal structure and so ICM-Pro was deemed a suitable docking software for this ligand.

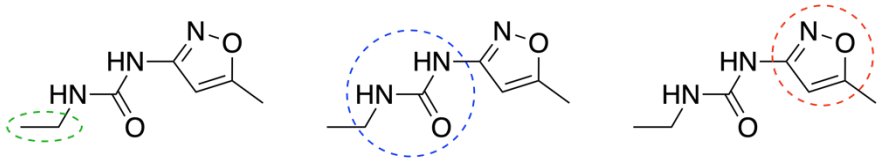


Figure 52. Docking strategy for compound **32**. The Molport database was systematically searched for compounds varying the ethyl chain (green), the urea motif (blue) or the isoxazole ring (red).

A substructure search of commercial vendor Molport varying the ethyl group resulted in 1000 potential analogues. The molecules were clustered into sets to avoid docking numerous iterations of similar compounds. Stereocentres were generated to allow for racemic mixtures of molecules. 269 compounds were docked, allowing for three poses each. Analogues were ranked based on their root-mean-square-deviation (RMSD) calculated on their docking coordinates from the position of **32** rather than the docking score predicted by ICM due to the limitations associated with docking scores.[216,217] The compounds were then analysed manually and chosen for their predicted overlap with **32** and maintenance of the hydrogen bonding pharmacophore. A total of 105 analogues varying the ethyl group were purchased based on this docking method.

For bioisosteres of the urea moiety, the structures in Figure 53 were submitted to the SwissBioisostere Database[218], with 500 compounds chosen for each scaffold.

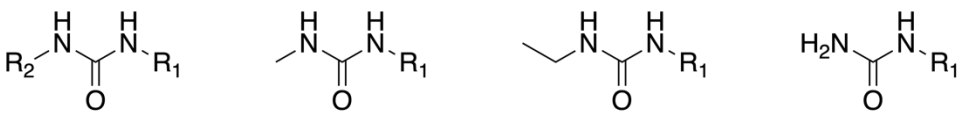
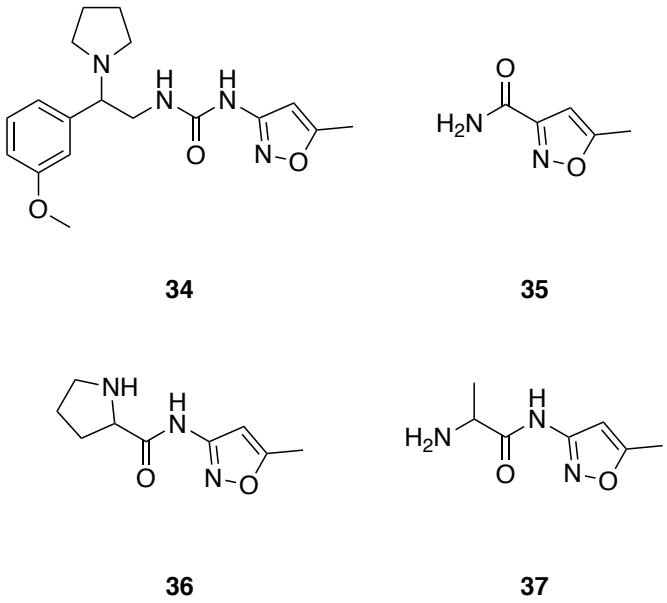


Figure 53. Bioisostere search of the SwissBioisostere Database.

In ICM, a 3-methyl isoxazole ring was added to each R1 group before the compounds were ranked based on properties such as number of hydrogen bond donors, with the top 85 compounds taken forward. The compounds were then searched in the Molport database and the 14 commercially available were docked and ranked according to the procedure detailed above for the ethyl analogues. Four bioisosteres were subsequently ordered (Figure 54).



34 **35**

36 **37**

Figure 54. Commercially available bioisosteres identified using SwissBioisostere.

Other compounds were designed with the isoxazole ring and urea fused, maintaining the hydrogen bonding pattern but adapting the core scaffold or generating isosteres of the isoxazole ring, with examples given in Figure 55. Compounds from the Molport database that were too large or lost the hydrogen bond donor/acceptor pattern were removed before docking. The analogues were docked according to the procedure detailed above. Only seven compounds were ordered based on the RMSD and docking pose. In general, modifications that enlarged the isoxazole ring were not predicted to fit in the binding pocket. Further searches and docking of compounds changing the isoxazole to an oxazole yielded a further three compounds predicted to bind in a similar manner to **32**.

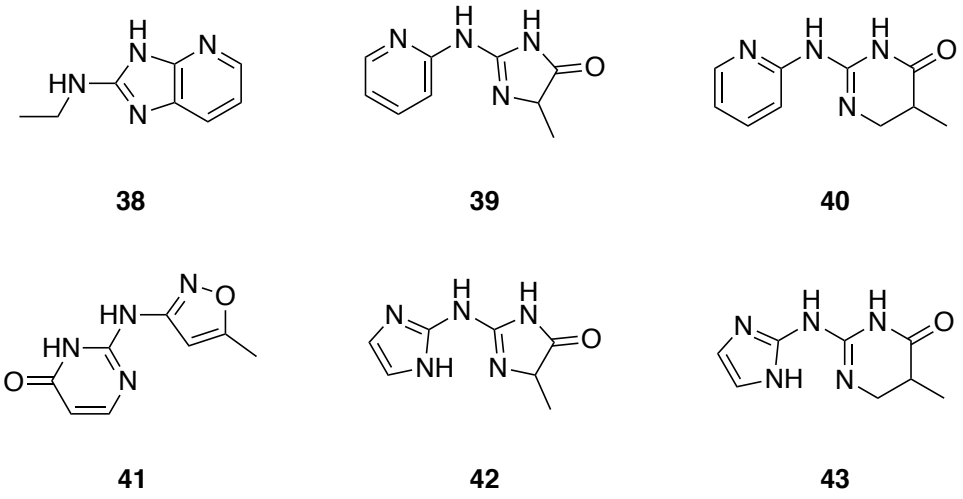


Figure 55. Examples of isoxazole replacements and fused ring motif of fragment hit **32** in searches of the Molport Database.

Overall, 123 analogues of **32** were purchased from Molport. Overlay of all the docked compounds in the guanine nucleotide binding site showed that the compounds purchased were predicted to explore the entire pocket whilst maintaining the desired pharmacophore (Figure 56).

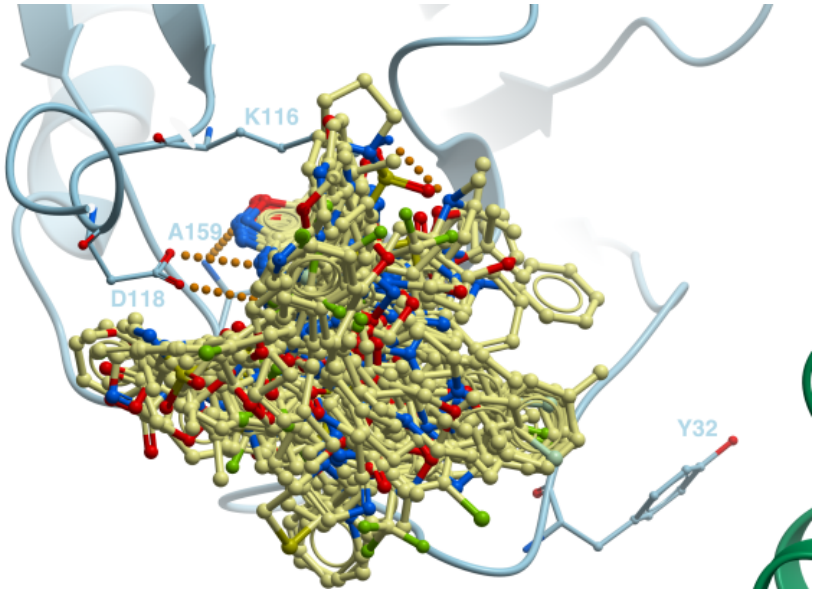


Figure 56. Overlay of the docked structures of compounds (yellow sticks) ordered from Molport derived from fragment **32**. All the compounds were predicted to form hydrogen bonds (one representation,

orange spheres) with D118 of Rac1 (light blue) in an analogous manner to **32**. Other compounds are predicted to form hydrogen bonds with K116. The ordered compounds represented a diverse mix and were predicted to thoroughly explore the pocket, with the distant switch I Y32 residue shown to indicate the end of the pocket.

*4.3.5.2 Putasteric Hit **33***

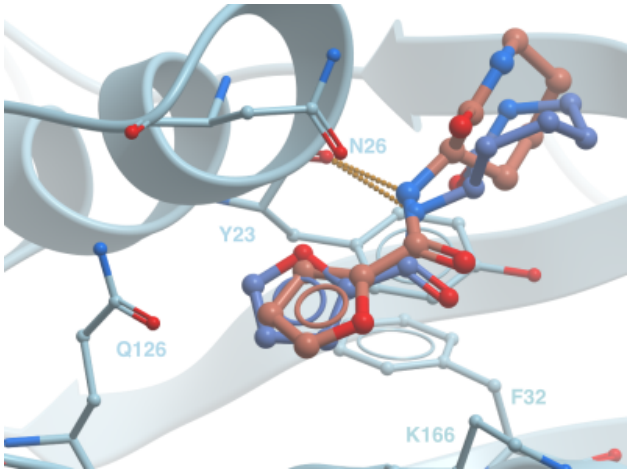


Figure 57. Docking of putasteric hit **33** within a pocket of Rac1 (light blue). The docked (peach sticks) and crystallised (violet sticks) poses do not directly overlap, especially for the 7-membered ring.

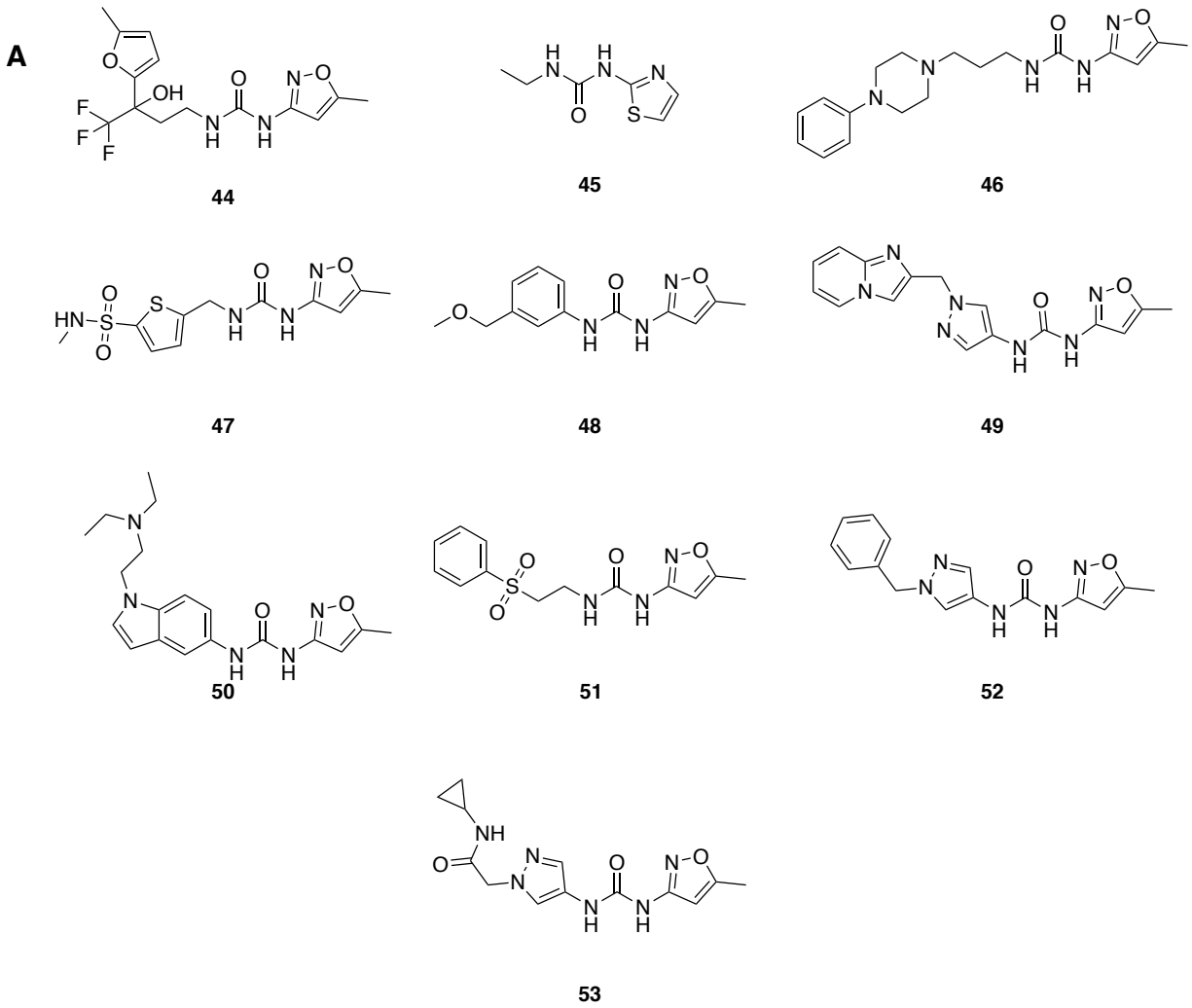
The same docking procedure in ICM was carried out for putasteric hit **33**. The docking software predicted the compound to bind analogously but slightly shifted, with a different orientation of the 7-membered ring (Figure 57). However, it was deemed suitable for a basic docking of follow-up compounds. Two searches yielded compounds from Molport; the entire structure was used as a substructure to obtain screening compounds which either grew the original hit or varied the furan motif for another heterocycle whilst maintaining the 7-membered ring. In total, 142 compounds were docked and a total of 19 ordered for a follow-up screen. In general, the compounds which elaborated on the original structure of **33** were deemed too large for the pocket.

4.3.6 XChem Screening of Follow-up Compounds

In total, 142 follow-up analogues were ordered from Molport at 100 mM concentration in a format suitable for acoustic dispensing. Of these, 20 were unable to be formatted by Molport, likely due to low solubility in DMSO arising from the urea component. The remaining compounds were screened in duplicate at 20 mM concentration in both GDP-bound and nucleotide-free crystals. In the nucleotide-free crystals, ten structures with ligands were identified, with the compounds shown in Figure 58A. Their data collection and refinement statistics are documented in Appendix 8.7. It was observed that a significant number of crystals did not diffract, or the compound precipitated upon addition to the crystal, leading to no data from several compounds. These compounds were retested at 10 mM, but no further binding events were observed. The crashing out of compounds could have been due to poor solubility in the low pH buffer (citric acid pH 5.5) used in the crystallisation conditions of Kalirin/Rac1. Attempts to develop another crystal system at higher pH, using different Kalirin constructs in several coarse screens, were ultimately unsuccessful.

Of the ten follow-up structures, nine contained modifications of the ethyl chain. Analysis of their structures showed that the elaborated chains extended out of the nucleotide binding pocket into the solvent channel of the crystal with no new interactions (Figure 58B). They were therefore predicted to have poor LE with little improvement in potency over **32**. Only one compound, **45**, had the isoxazole ring replaced with an alternative heterocycle. However, the electron density showed that **45** had shifted compared to the original hit **32**, with both **45** and D118 located slightly further into the binding pocket. No other residues changed their position.

No analogues of **33** were identified as hits. This was not entirely unexpected due to the small size of the pocket and its position near the solvent channel, and that **33** only had a single weak interaction with Kalirin/Rac1.



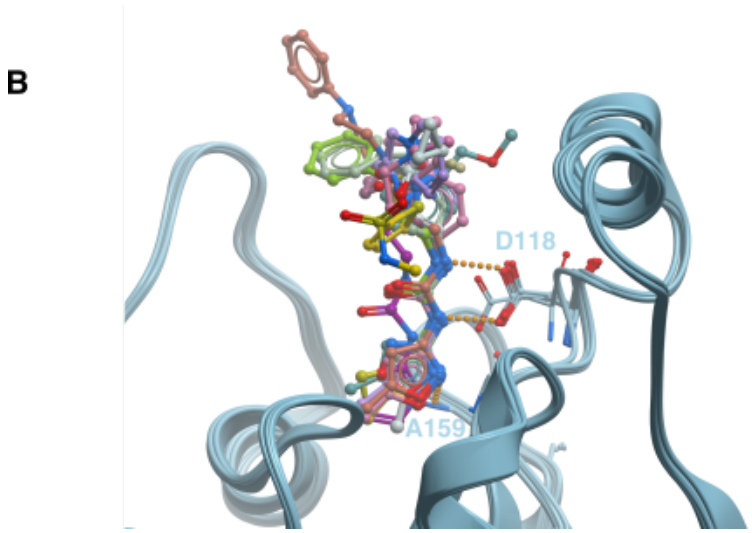


Figure 58. **A**. Follow up compounds identified in an XChem screen. All apart from **45** varied the ethyl chain; **45** contains a thiazole ring instead of an isoxazole. **B**. Overlay of compounds in the guanine

nucleotide binding site of Rac1. A representation of the hydrogen bond interactions (orange spheres) with D118 and the peptide backbone is shown for one compound.

Overall, the results of the follow-up screening indicated that elaboration of this core scaffold might be more difficult than originally anticipated. The majority of the compounds that bound did not form new interactions or explore into the nucleotide binding pocket. However, **32** displaced GDP from the canonical binding pocket, indicating appreciable affinity that should be capitalised upon.

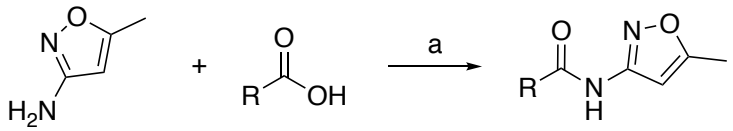
4.4 Urea Bioisosteres

Urea motifs are commonly associated with poor solubility.[219] Several bioisosteres of the urea moiety were identified in the SwissBioisostere search (section 4.3.5.1) but were not commercially available. It was hoped they would have greater potency or improved physiochemical properties than **32**, and so were synthesised.

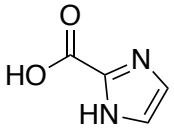
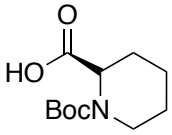
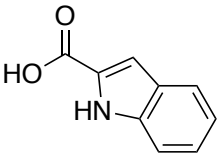
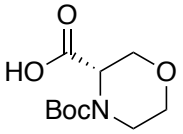
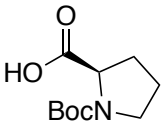
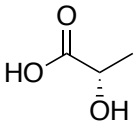
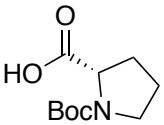
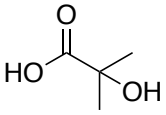
4.4.1 Synthetic Procedures

4.4.1.1 Amide Bond Formation

The synthesis of amide bond-containing analogues was first attempted using amide coupling reagents to activate the requisite carboxylic acids for reaction with amine **54**. Originally the coupling agent HATU was used, but this did not generate the amide bond with any of the substituents. It was predicted that the activated carboxylic acid-HATU intermediates were too reactive for the poorly nucleophilic **54**, and so the activated ester was degraded in the longer reaction time required for amide coupling. T3P® was tried instead; this generated amides **55**-**60** (Scheme 1). The lower yields for **55** and **56** were attributed to the free amine group interfering in the coupling reaction, resulting in side-products. Attempts to improve the yield by increasing the temperature or equivalents of T3P® were unsuccessful. The use of T3P® did not generate **61** and **62**, likely due to the presence of the unprotected alcohol.

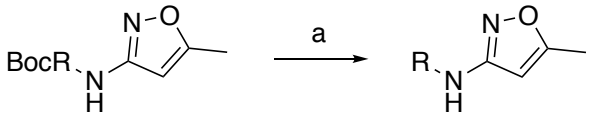


54 **55-62**

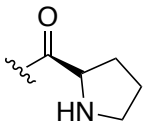
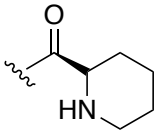
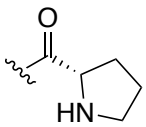
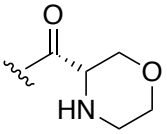
Compound	R-COOH	Yield (%)	Compound	R-COOH	Yield (%)
55		7	**59**		10
56		2	**60**		37
57		60	**61**		0
58		49	**62**		0

Scheme 1. Synthesis of amide bioisosteres using T3P®. Reagents and conditions: (a) T3P® 50 wt% in EtOAc, DIPEA, DCM, r.t., 16 h.

For amides **57**-**60**, the Boc protecting group was removed under acidic conditions to generate final products **63**-**66** in high yields apart from **65**, which required an additional purification step (Scheme 2).

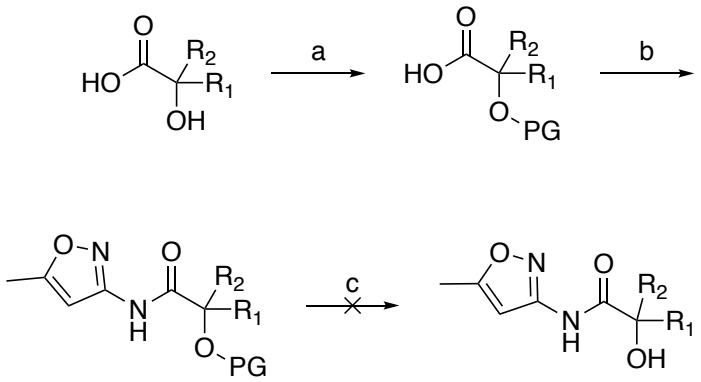


57-60 **63-66**

Compound	R	Yield (%)	Compound	R	Yield (%)
63		93	**65**		91
64		75	**66**		17

Scheme 2. Boc deprotection following amide synthesis. Reagents and conditions: (a) TFA, DCM, r.t., 16 h.

The silyl ether TBDMS was trialled as a protecting group for the free alcohol of **61** and **62**; the reaction failed for the carboxylic acid of **62,** potentially due to steric hindrance of the 3° alcohol. Whilst the protected acid for **61** could be observed by TLC, the final product could not be isolated. It was hypothesised that the protected product was volatile as only starting material could be observed following removal of solvent *in vacuo*. The use of a benzyl (Bn) group was then chosen to simplify purification by introducing a chromophore. Despite following literature protocols for its synthesis, the benzyl-protected alcohol of **62** could not be synthesised. The benzyl-protected starting material for **61** was commercially available; reaction with amine **54** using T3P® with the above conditions was successful to form **67**, but removal of the benzyl protecting group using mild hydrogenation conditions caused to the isoxazole ring to open (Scheme 3) and so final product **61** was not accessible by this route.



67 R_1 = Me, R_2 = H

61 R_1 = Me, R_2 = H

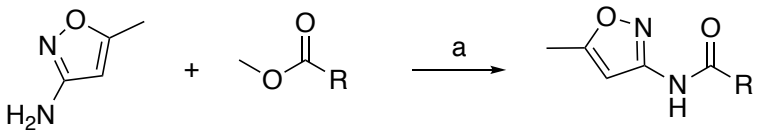
62 R_1 = Me, R_2 = Me

Reaction	R_1	R_2	PG	Conditions	Result
a	-Me	-H	TBDMS	TBDMSCl, Imidazole, r.t., 16 h	0 %[a]
a	-Me	-Me	TBDMS	TBDMSCl, Imidazole, r.t., 16 h	NR[b]
a	-Me	-Me	Bn	BnBr, NaH, DMF, 0 °C-r.t., 16 h	NR[b]
b	-Me	-H	Bn	T3P®, DIPEA, DCM, r.t 16 h	**67**[b]
c	-Me	-H	Bn	Pd on C, H_2, r.t., 16 h	Ring opened[c]

PG = Protecting Group NR = No reaction [a]Compound decomposed during purification [b]Determined by LCMS analysis [c]Determined by LCMS and NMR

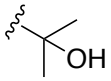
Scheme 3. Unsuccessful attempts to synthesise amides **61** and **62** by protecting the free hydroxyl group.

It has been noted in the literature that amides can be formed from the methyl ester directly using $AlMe_3$; DABAL-Me3 is a safer alternative and is tolerant of a variety of substituent groups.[220] Use of DABAL-Me3 with the relevant methylated ester yielded **61** and **62** in reasonable yields (Scheme 4). The *R*-isomer of **61**, **68**, was also synthesised.



54 **R**

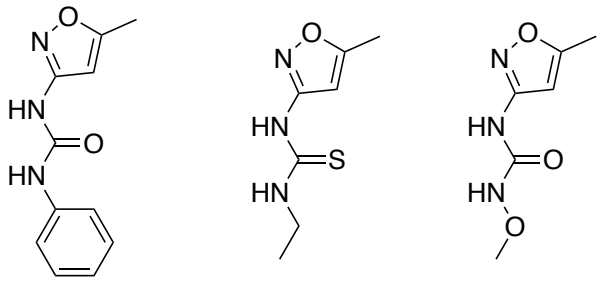
61 R_1 = Me, R_2 = H
62 R_1 = Me, R_2 = Me
68 R_1 = Me, R_2 = H

Compound	R	Yield (%)
61		30
62		21
68		35

Scheme 4. Alternative synthesis of alcohol-containing amides from the methyl ester. Reagents and conditions: (a) DABAL-Me3, THF, 60 °C, 16 h.

4.4.1.2 Urea-like Bond Analogues

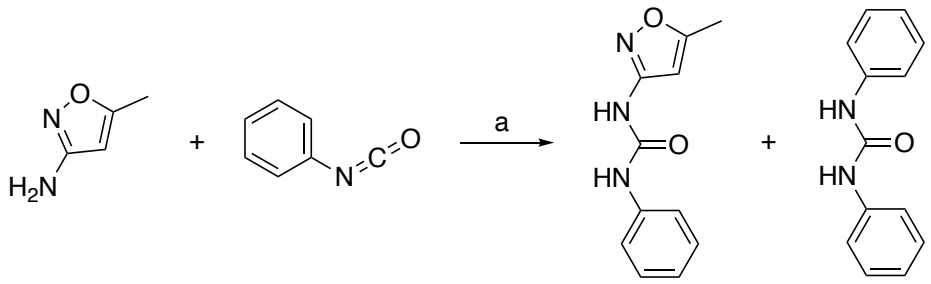
Compounds **69-71** all contained a urea-esque motif where there are two NH groups on either side of a carbon atom double bonded to a heteroatom (Figure 59). Hence, it was originally conceived that the synthesis of the analogues may proceed in a similar manner.



69 **70** **71**

Figure 59. Bioisosteres that all contained a urea-esque motif

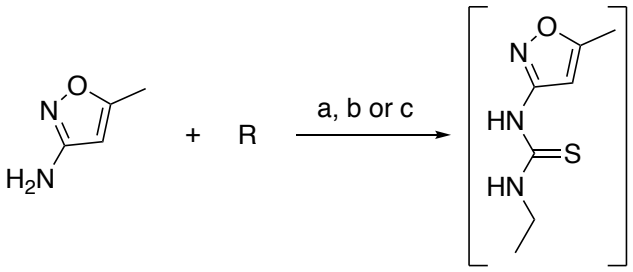
The urea motif of **69** was synthesised from the reaction of phenyl isocyanate with **54** (Scheme 5).[221] The reaction had to be repeated carefully in anhydrous conditions to prevent the formation of 1,3-diphenylurea (**72**). In wet conditions, the starting isocyanate was hydrolysed to aniline, which then reacted with non-hydrolysed isocyanate to form **72** instead of the desired product **69**. Despite the use of anhydrous conditions to reduce this side reaction, the yield of **69** was still low.



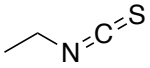
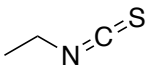
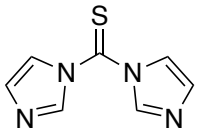
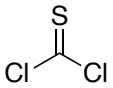
54 **69** **72**

Scheme 5. Synthesis of **69**. Reagents and conditions: (a) DIPEA, MeCN, 75°C, 4 h, 18%.

A similar method was attempted for the synthesis of **70** with the use of ethyl isothiocyanate (Scheme 6). However, the ^{1}H NMR of the product showed that the isoxazole ring had opened during the reaction, suggesting **70** was not stable upon heating. To improve the reaction rate at lower temperatures and avoid self-condensation, ethyl isothiocyanate was switched for TCDI and ethylamine.[222] TCDI was mixed with **54** to produce an activated electrophilic centre. Ethylamine is a better nucleophile than **54** and so was used in the second stage of the reaction to prevent it from reacting with TCDI twice. Despite this, LCMS and NMR indicated that the major product was diethylurea suggesting that the **54** was not nucleophilic enough to react with TCDI before the addition of ethylamine. Hence, the more reactive thiophosgene was used to create the electrophilic intermediate before the addition of ethylamine.[223] Whilst this was successful in generating **70** in low yield, the product ring-opened and decomposed during the NMR analysis in both chloroform and anhydrous DMSO. Due to the instability of **70**, it was deprioritised for assay screening.



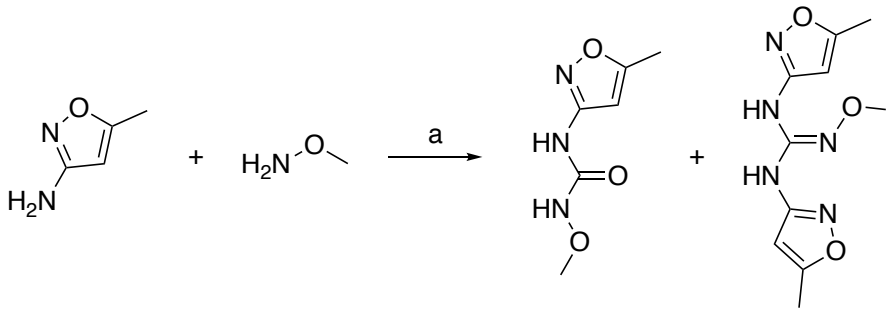
54 **70**

Reaction	R	Conditions	Result
a		MeCN, 75 °C, 4 h.	Isoxazole ring opened[a]
a		Toluene, reflux, 16 h	NR[b]
b		Ethylamine, DCM, 0 °C- r.t., 16 h	Isoxazole ring opened and diethylurea formed[a]
c		Ethylamine, TEA, THF, 0 °C- r.t., 16 h	**70**[a, c]

NR = No reaction [a]Determined by LCMS and NMR [b] Determined by LCMS [c] Compound decomposed during characterisation

Scheme 6. Different methodologies to synthesise **70**.

In the synthesis of **71**, the necessary isocyanate was not commercially available. Due to the lack of success of generating **70** using TCDI, the more reactive triphosgene was used to activate **54** in the synthesis of **71** instead of CDI.[223] Methoxyamine was added following the generation of the active intermediate. The reaction was successful in generating **71**. However, due to the excess of **54** compared to methoxyamine, the disubstituted **73** was produced as a major side product (Scheme 7). Both **71** and **73** were purified for screening.

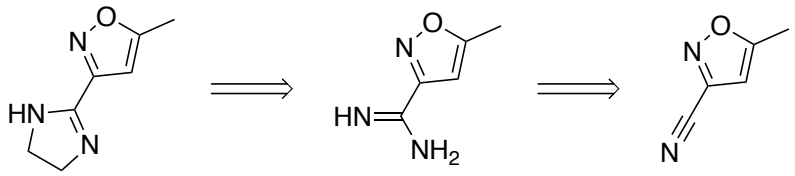


54 **71** **73**

Scheme 7. Synthesis of **71** and the side product **73**. Reagents and conditions: (a) Triphosgene, TEA, THF, DMSO, 0 °C-r.t., 16 h, 9% (**71**), 8% (**73**).

4.4.1.3 Heterocycle Ring Formation

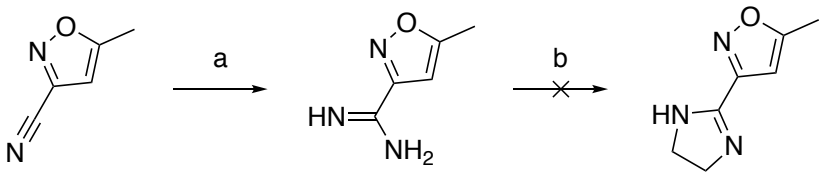
It was realised that the amidine bioisosteres **75** could be generated as an intermediate in the synthesis of the 2-imidazoline ring of **74** (Figure 60).



74 **75** **76**

Figure 60. Retrosynthetic strategy to generate **74** and **75** from common starting material **76**.

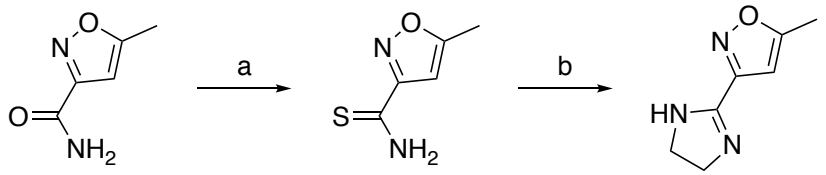
The nitrile **76** was converted to an imidate intermediate using sodium methoxide in a Pinner reaction, before being converted into **75** following the addition of ammonium chloride (Scheme 8).[224] The yield of **75** was low due to its polarity; extraction from the aqueous phase was difficult, even using 3:1 chloroform:isopropanol. The pure product was eventually generated in better yield by triturating in methanol before application to a SCX column. However, the envisaged conversion of **75** into **74** using 1,2-dichloroethane (DCE) was unsuccessful; the amidine did not react, even with the addition of further equivalents of DCE or heating the reaction, probably due to the poor nucleophilicity of **75** (Scheme 8).



76 **75** **74**

Scheme 8. The successful synthesis of **75** from **76**. The conversion of **75** into **74** was unsuccessful. Reagents and conditions: (a) i) Sodium methoxide, MeOH, r.t., 2 h; ii) Ammonium chloride, r.t. 16 h, 33%. (b) DCE, MeOH, 45 °C, 72 h.

An alternative route was pursued for **74** through the generation of an electrophilic isoxazole intermediate. Based on published protocols for generating the 2-imidazoline ring *in situ* using a thioamide, the synthetic route in Scheme 9 was designed.[225] The commercially available amide **77** was converted into thioamide **78** using Lawesson's reagent.[226] **78** was taken on crude into the cyclisation reaction with ethylenediamine and refluxed overnight; product **74** was successfully synthesised in reasonable yield.

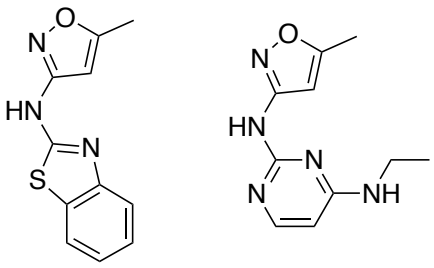


77 **78** **74**

Scheme 9. Alternative synthetic route for generating **74**. Reagents and conditions: (a) Lawesson's reagent, toluene, reflux, 16 h. (b) Ethylenediamine, THF, reflux, 16 h, 51%.

4.3.1.4 Nucleophilic Substitutions

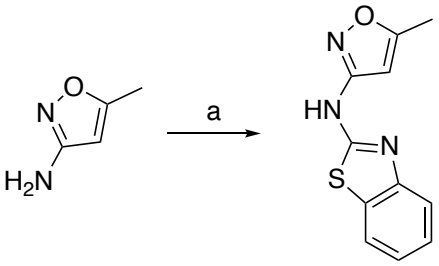
The compounds **79** and **80** (Figure 61) were envisaged to be synthesised in analogous nucleophilic aromatic substitutions (S_NAr).



79 **80**

Figure 61. Bioisosteres **79** and **80**.

The brominated version of benzothiazole was commercially available; the corresponding S_NAr reaction was successful (Scheme 10). However, during purification using flash column chromatography, the residual amine starting material eluted at the same retention time as **79**. Exploration using thin-layer chromatography (TLC) showed **54** and **79** had similar retention times in a number of solvent systems. Further purification using preparative HPLC resulted in pure product in low yield.



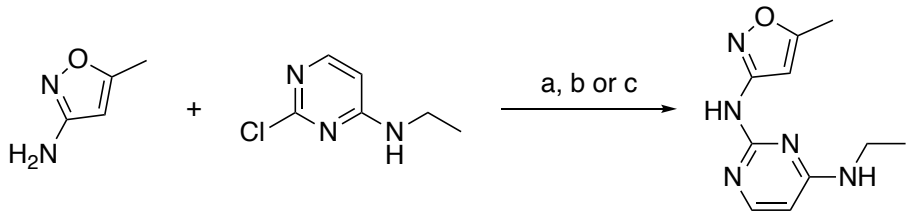
54 **79**

Scheme 10. S_NAr reaction to form **79**. Reagents and conditions. (a) i) NaH, DMF, 0 °C, 1 h; ii) 2-bromobenzothiazole, r.t., 16 h, 15%.

Based on the success of **79**, the same method was used to synthesise **80**, replacing 2-bromobenzothiazole with 2-chloro-4-ethylaminopyrimidine (Scheme 11). However, this reaction failed to generate the desired product, even with heating. A Buchwald-Hartwig cross-coupling was then attempted as an alternative. Whilst Buchwald-Hartwig reactions can be successful when S_NAr fails due to low reactivity of the reagents, the wide variety of palladium catalysts, ligands, and bases means they can be difficult to optimise. In the case of **80**, there

were selectivity concerns due to the presence of the free amine on the aryl coupling partner. In order to avoid self-coupling, BrettPhos Pd G3 was chosen as the palladium catalyst and BrettPhos as the ligand due to its general selectivity for the arylation of primary amines.[227] LiHMDS was chosen as a strong base; it is documented in the literature as tolerant of protic groups on the aryl, with LiHMDS used in the Pd-catalysed amination of heterocycles containing a free NH group.[227] Refluxing only generated a trace amount of **80** identified by LCMS. Attempts to increase conversion by using microwave irradiation to achieve higher temperatures and pressures was unsuccessful. Increasing the equivalents of LiHDMS or replacing with sodium *tert*-butoxide, a more typical Buchwald-Hartwig coupling base and often giving higher reaction rates,[227] also did not improve product yield.

It was predicted that the lack of success achieved by basic changes to the reaction conditions meant that the optimisation of this reaction would be protracted, and so the Buchwald-Hartwig method was discarded. Before embarking on lengthier routes to install the two amines on the aryl separately, a last attempt was made at a one-step S_NAr using acid catalysis. As the aryl and amine were not particularly reactive in base, it was hypothesised based on literature precedent that the use of strong acid may result in protonation of a nitrogen in the pyrimidine ring.[228,229] The reaction proceeded smoothly using 12 M HCl, circumventing a longer, less efficient pathway to generate **80** (Scheme 11).



54 **80**

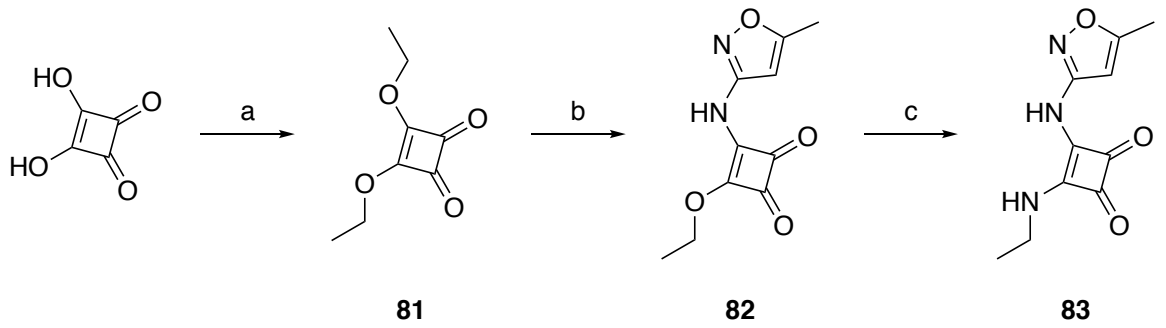
Pd catalyst	Ligand	Base	Conditions (b)	Result
BrettPhos Pd G3	BrettPhos	LiHDMS	1,4-Dioxane, reflux, 16 h	Minimal conversion[c]
BrettPhos Pd G3	BrettPhos	LiHDMS	MW, 120 °C, 4 h 2.2 eq of base	Minimal conversion[a]
BrettPhos Pd G3	BrettPhos	NaO*t*-Bu	MW, 100 °C, 16 h. 2.2 eq of base.	NR[a]

NR = No Reaction [a]Determined by LCMS

Scheme 11. Attempted reaction for the synthesis of **80**. Reagents and conditions: (a) i) NaH, DMF, r.t., 16 h; ii) 80 °C, 48 h. (b) See table. (c) 12M HCl, 2-propanol, MW 100 °C, 2 h, 10%.

*4.4.1.5 Squaramide **83***

Synthetic routes for similar compounds to the squaramide **83** were available in the literature.[230] Squaric acid was converted into squaric acid diethyl ester (**81**) using ethanol. The ester groups were then displaced, first by the poorer nucleophile **54,** then ethylamine, to prevent ethylamine reacting twice with **81** to form the disubstituted side product.[230] It was predicted that **54** would not be reactive enough to displace both ethoxy groups, with the risk further reduced by using dilute conditions. Whilst both substitutions occurred in reasonable yield to successfully generate **83** *via* the intermediate **82**, they were very slow and did not reach completion (Scheme 12). Increasing the temperature from 45 to 55 °C resulted in isoxazole ring opening, and so the reaction was stopped to prevent further loss of yield.

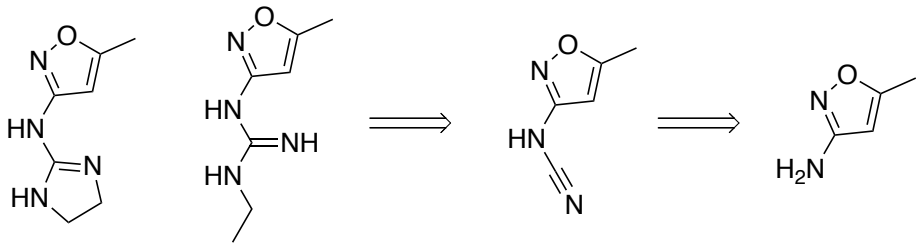


81 **82** **83**

Scheme 12. Synthetic route for the squaramide **83**. Regents and conditions: (a) EtOH, reflux, 8 h, 92%. (b) **54**, DIPEA, EtOH, 45 °C, 120 h, 42%. (c) Ethylamine, EtOH, r.t., 16 h, 83%.

4.4.1.6 Guanidines

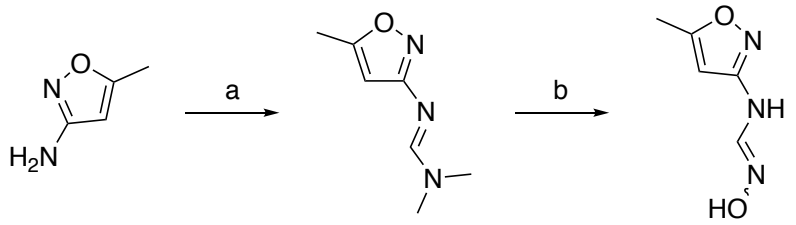
It was hypothesised that **84** and **85** could be synthesised from the same common cyanamide intermediate **86**, which in turn could be generated from **54** (Figure 62).



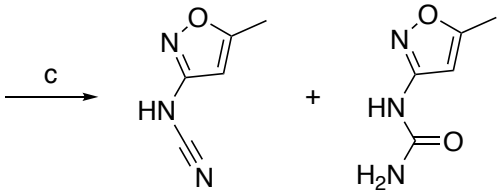
84 **85** **86** **54**

Figure 62. Retrosynthetic analysis for the derivation of **85** and **84** from a common cyanamide intermediate **86.**

Conversion of **54** into oxime **88** *via* the intermediary tri-substituted amidine **87** occurred in high yields following literature protocols for amines with different heterocycles.[231] Unfortunately, the subsequent conversion of the oxime to the cyanamide was less successful. Several different conditions were attempted using dehydrating agents as detailed in Scheme 13. The reactions gave an intractable mixture of products or yielded the hydrolysed **89**, even under strictly anhydrous conditions.



54 **87** **88**



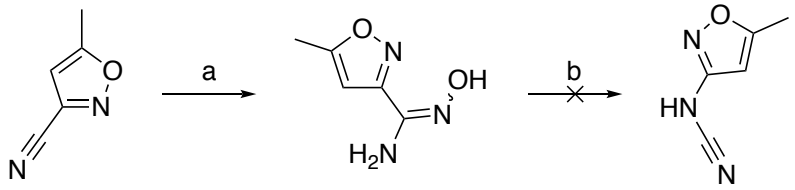
86 **89**

Conditions (c)	Result
$POCl_3$, TEA, 2-MeTHF	Mixture[a]
$POCl_3$, Py, DME	Mixture[a]
MsCl, TEA, DME	**89** formed[b]
MsCl, DIPEA, DME	**89** formed[b]
MsCl, Py, DME	**89** formed[b]
BOP, DBU, DCM	Mixture[a]

[a]Determined by LCMS [b] Determined by LCMS and NMR

Scheme 13. Strategies to convert **88** into **86**. Reagents and conditions: (a) DMF-DMA, toluene, reflux, 2 h. (b) Hydroxylamine hydrochloride, MeOH, r.t., 2 h, 74% (over two steps).

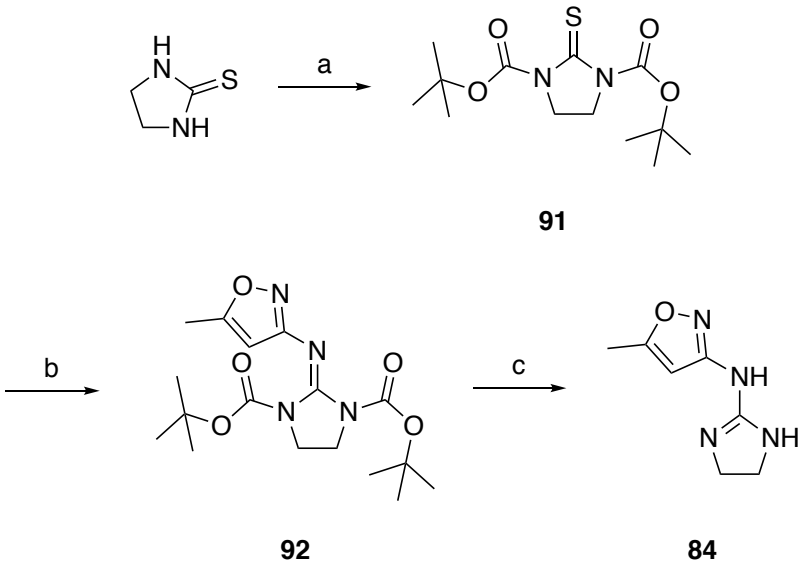
In a similar strategy, the aldoxime isomer of **88**, **90**, was synthesised from **76** with the intention of forming the cyanamide *via* a Tiemann rearrangement of the aldoxime (Scheme 14).[232] The rearrangement either did not occur or again generated **89**. Hence this strategy was deprioritised in favour of different route.



76 **90** **86**

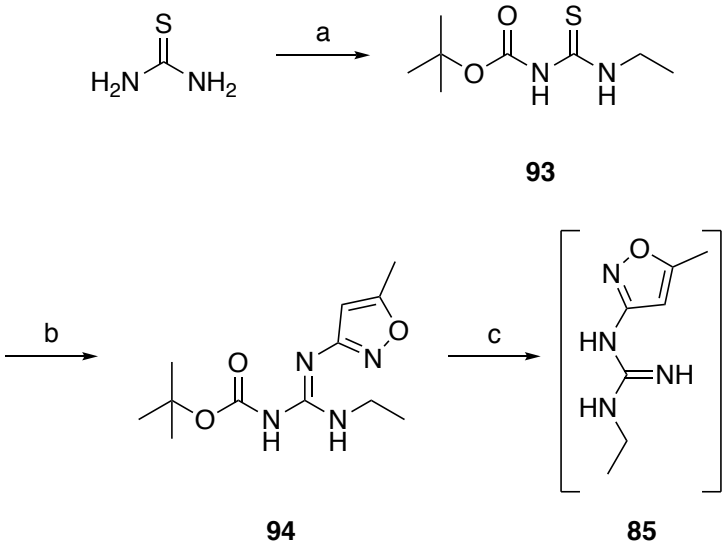
Scheme 14. Attempt to form common intermediate **86** from the aldoxime **90**. Reagents and conditions: (a) NH_2OH, EtOH, reflux, 3 h, 20%. (b) TsCl or 2-nitrobenzenesulfonyl chloride, DIPEA, DCM, reflux, 1 h.

Protocols in the literature were found detailing the generation of guanidines from Boc-protected thioureas.[233] The synthesis of **84** was relatively facile using this route due to the readily available starting material imidazoline-2-thione (Scheme 15). The Boc-protection of imidazoline-2-thione followed literature procedures using Boc anhydride in high yield.[234] The substitution of sulphur within a thiourea with a weakly nucleophilic amine to produce a guanidine was reported using $HgCl_2$ as a catalyst.[233] Copper chloride was also documented as a safer alternative to catalyse this reaction but with lower yields than mercuric chloride. Nonetheless, the reaction using $CuCl_2$ proceeded in reasonable yield to give **92**, and a simple Boc-deprotection generated **84**.



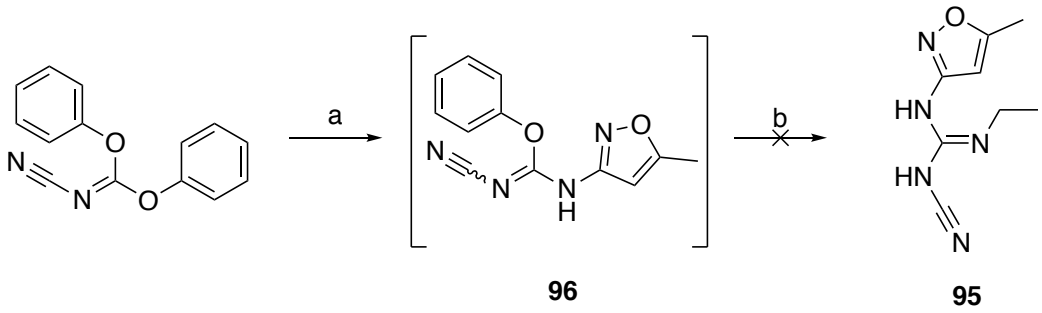
Scheme 15. Synthesis of **84** through the Boc-protected thiourea **91.** Reagents and conditions: (a) Boc anhydride, NaH, THF, 0 °C-r.t., 16 h, 88%. (b) **54**, $CuCl_2$, TEA, DCM, 0 °C-r.t., 16 h, 41%. (c) TFA, DCM, r.t., 16 h, 68%.

The synthesis of **85** was more difficult due to the necessity of making a non-symmetrical Boc-protected thiourea. A one-pot procedure using Boc anhydride and TFAA generated **93** from thiourea in low yields (Scheme 16). Subsequent substitution using $CuCl_2$ yielded **94** but in substantially lower conversion than for **92**. Whilst the Boc-deprotection was successful, the isoxazole ring of **85** was unstable upon addition of NMR solvent in a similar manner to **70**. The compound was therefore discarded as it would not be stable for testing.



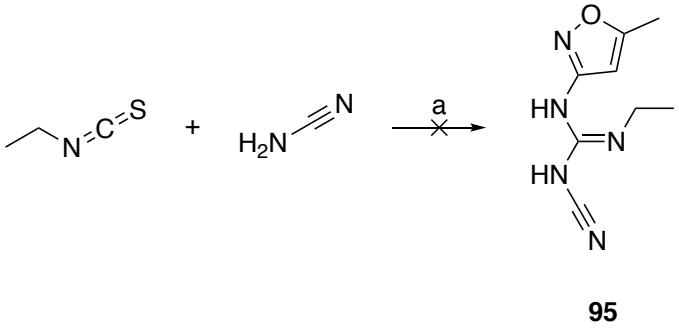
Scheme 16. Attempted synthesis of **85**. Reagents and conditions: (a) i) NaH, Boc anhydride, 0 °C-r.t., 16 h, ii) NaH, TFAA, Ethylamine, r.t. 16 h, 11%. (b) **54**, $CuCl_2$, TEA, DCM, 0 °C-r.t., 16 h. (c) TFA, DCM, r.t., 16 h. Product decomposed during purification.

Access to trisubstituted **95** was attempted by a different method as it could not be generated *via* a Boc-protected thioamide intermediate. There is literature precedent for the synthesis of cyanoguanidines by substituting the phenol groups of diphenyl *N*-cyanocarbamimidate with the desired amine side chains which avoided the formation of a complicated Boc-protected intermediate (Scheme 17).[235] This method was relatively unsuccessful using **54** and ethylamine. **54** was substituted first due to its lower reactivity. The intermediate **96** could be observed by LCMS, but in low yield and extended reaction times. **96** was taken on crude to the next step; no formation of **95** was observed. Heating to 50 °C resulted in the isoxazole ring of **96** opening rather than the formation of **95**.



Scheme 17. Unsuccessful route to generate **95** through nucleophilic substitution. Regents and conditions: (a) 2-propanol, r.t., 16 h. (b) Ethylamine, THF, 16 h.

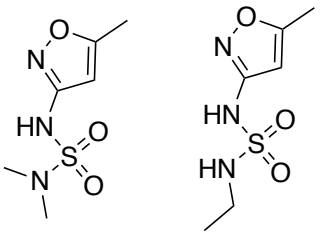
A final attempt was made to synthesise **95** in a one-pot procedure *via* a thiourea using ethylisothiocyanate and cyanamide, before converting to the cyanoguanidine using EDCI and **54**.[236] Unfortunately, even whilst refluxing, cyanamide and ethylisothiocyanate did not react. Synthesis of **95** was then abandoned due to time constraints and the probability that the compound was unstable.



Scheme 18. One-pot procedure to synthesise cyanoguanidine **95**. Reagents and conditions: (a) **54**, EtOH, r.t., 16 h.

4.4.1.7 Sulfamides

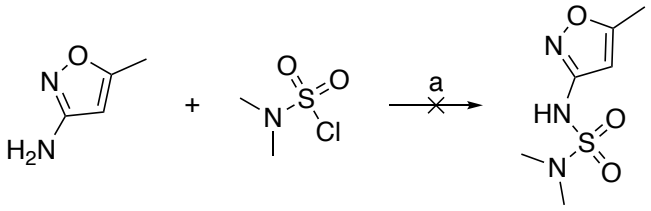
It was conceived that **97** and **98** could be synthesised in a similar manner using **54** and the relevant sulfamoyl chloride (Figure 63).[237,238]



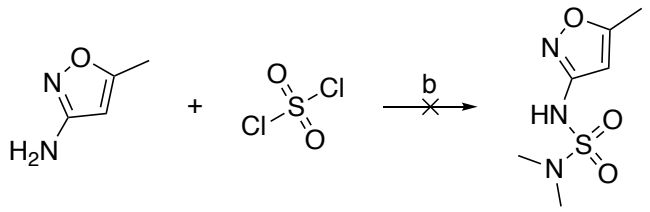
97 **98**

Figure 63. Bioisosteres **97** and **98** with a sulfamide motif.

The dimethylsulfamoyl chloride of **97** was commercially available. A simple nucleophilic substitution using TEA gave no reaction, even under microwave conditions (Scheme 19). Use of the stronger base NaHMDS again resulted in unreacted starting material. It was considered that **54** was insufficiently nucleophilic to react with the sulfamoyl chloride. Hence, **54** was mixed with sulfuryl chloride to form the relevant sulfamoyl chloride before subsequently adding with dimethylamine, a more reactive aliphatic nucleophile (Scheme 19).[239] LCMS indicated a product with the correct mass; NMR analysis post purification identified that the isoxazole ring had opened. The synthesis of these analogues was consequently ceased due to the instability of the isoxazole ring.



54 **97**



54 **97**

Scheme 19. Attempts to synthesise **97.** Reagents and conditions: (a) TEA, DCM, 0 °C-r.t., 8 h. (b) Dimethylamine, imidazole, DCM, MW, 2 h.

4.5 Design of a Fluorescence-based Assay

4.5.1 Overview

Once most of the urea bioisosteres were synthesised, they were soaked into Kalirin/Rac1 crystals, but no electron density was observed for the fragments. This may be attributable to the insolubility of the fragments at high concentrations in the acidic crystal conditions, as crashing out was commonly observed in the XChem screen.

Fragments are soaked at millimolar concentrations in XChem screens to ensure high occupancy, even for ligands with low affinity. As such, fragments identified can be extremely weak and difficult to optimise due to forming few interactions with the protein of interest. It is also not possible to determine quantitative values of potency (such as K_D or IC_{50}) from crystal structures, nor can the relative potency of different compounds be assessed. It was therefore necessary to develop a sensitive biochemical assay in parallel to monitor the ability of compounds to inhibit nucleotide exchange of the GEF/GTPase complex.

Fluorescence-based nucleotide exchange assays using either fluorescent GDP or GTP analogues have been detailed extensively in the literature for Ras/Sos1 surface binders and covalent compounds targeting KRas G12C. [137,240–242] An analogous assay suitable for our laboratory was developed with Carmen Jimenez Antunez of Paul Brennan's group, who optimised the parameters of the assay.

The assay measures the exchange rate of the guanine nucleotide bound within the GEF/GTPase complex with a fluorescent nucleotide analogue in solution (Figure 64). The GDP or GTP analogue contains a fluorophore which has low quantum yield in solution that increases upon binding to the GTPase. The assay can be designed to measure either the dissociation or association of the fluorescent analogue with the GTPase, catalysed by a specific GEF. For a dissociation assay, the fluorescent analogue is incubated with the

GEF/GTPase complex. Either GDP or a compound capable of displacing the fluorescent nucleotide analogue is added and the decrease in fluorescence intensity (FI) is used to measure the rate of exchange. An association assay has the GEF/GTPase complex bound to a non-fluorescent nucleotide and is pre-incubated with a compound of interest. Ligands which inhibit the exchange reaction cause a slower increase of FI upon addition of the fluorescent analogue.

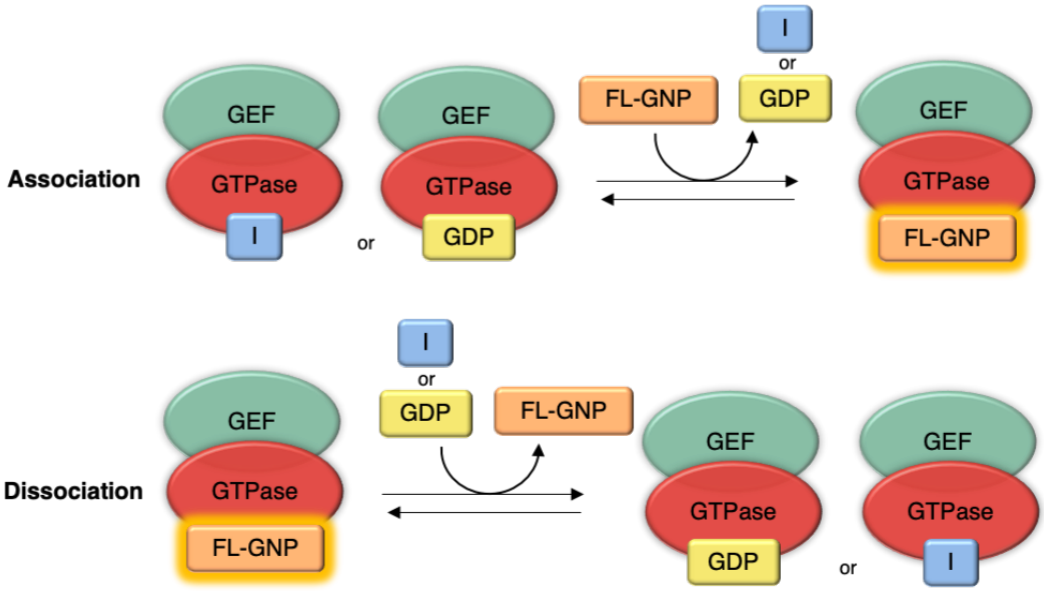
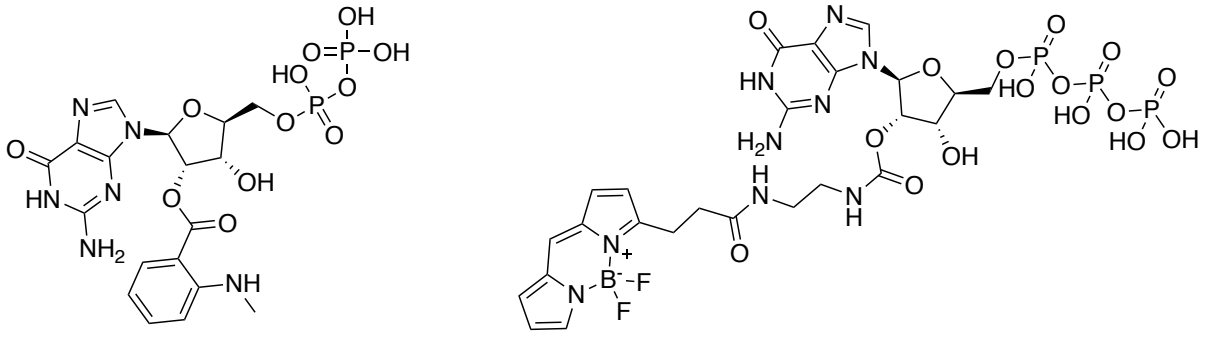


Figure 64. The two types of nucleotide exchange assay in the literature. The assay measures the change in fluorescence intensity induced by either the association or dissociation of a fluorescent guanine nucleotide analogue (FL-GNP). Compounds which inhibit this association can be identified by incubating with the GEF/GTPase complex and monitoring the change in the initial rate. 'I' denotes inhibitor.

4.5.2. Optimisation of the Assay

The protocol for the assay was originally taken from Eberth and Ahmadian[213], which uses a GTPase preincubated with the GDP analogue 2'/3'-O-(N-Methyl-anthraniloyl)-guanosine-5'-diphosphate (mant-GDP, ex/em 355/448 nm) (Figure 65). Mant-GDP is commonly used in GDP/GTP exchange assays, as the fluorophore modification is relatively small, minimising disruption of the interactions between the nucleotide and GTPase. However, early experiments were unsuccessful, with no change in FI observed over several hours. No difference was seen in different purification samples of either Rac1 or Kalirin. It was thought that the construct of Kalirin may be inactive as a GEF, especially as it has been reported that the PH domain adjacent to the DH GEF domain is required for the full activity of Trio.[173] Hence, the assay was instead optimised using a commercially available Rac1 Dbl GEF, Tiam1, to isolate whether the Kalirin DH construct was the source of the problem.



2'-Mant-GDP **2'-BODIPY-GTP**

Figure 65. Common fluorescent guanine nucleotide analogues.

Nevertheless, even using commercial Tiam1, no significant change in FI was detected during the reaction. It was observed that the background fluorescence was extremely high, resulting in a large signal to noise ratio. The buffer used in Eberth and Ahmadian contained 10 mM potassium phosphate. Comparison to a standard Tris buffer without potassium

phosphate showed a significant difference in background fluorescence, indicating that either the buffer was contaminated, or that potassium phosphate interfered with the assay.

Further analysis of the literature highlighted that mant-GDP has a small quantum yield and is sensitive to its environment. Hence, an alternative fluorescent analogue was required; BODIPY-GTP (ex/em 485/510 nm) contains a boron-dipyrromethene moiety attached to either the 2' or 3' position of the ribose ring and is cited for having a greater difference in quantum yield between being free in solution and when bound to the GTPase (Figure 65)[243]. This immediately showed an improvement in the assay, where a significant change in FI associated with nucleotide exchange could be observed.

The assay was then attempted for two Kalirin constructs; Kalirin-c001, containing only the DH GEF domain, and Kalirin-c002, comprising of both the DH and PH domain. Gratifyingly, both constructs showed greater GEF activity than commercial Tiam1, alleviating fears that the DH domain present in the Kalirin/Rac1 crystal structure was catalytically inactive and thus an inaccurate representation of the GEF (Figure 66). However, consistent with data published for Trio, the DH-PH construct was more active than the DH domain alone, implying the PH domain is required for full activity.[173] Thus, Kalirin-c002 was used in this assay going forward.

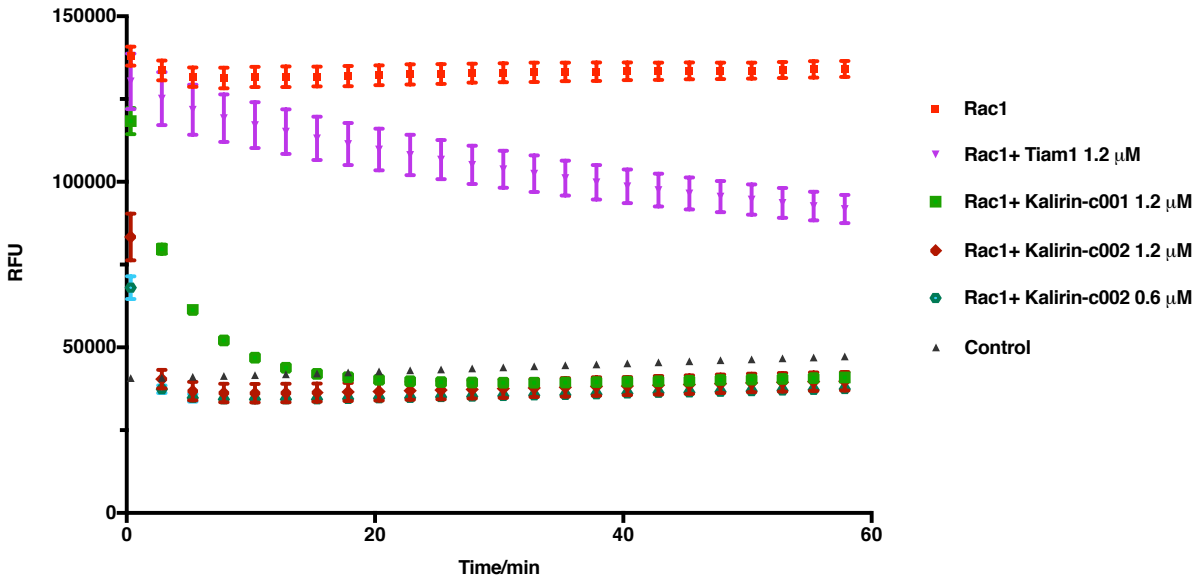


Figure 66. Comparison of different GEFs in the nucleotide dissociation assay. Both constructs of Kalirin rapidly catalyse the exchange reaction in comparison to Tiam1. This experiment was conducted by Carmen Jimenez Antunez. The concentrations in the legend refer to the concentration of GEF used in this assay. Control refers to a buffer control.

It was decided due to the problems encountered with the Eberth and Ahmadian method that a different protocol should be followed. The conditions were thus optimised using the assay conditions provided by Cytoskeleton using BODIPY-GTP and a phosphate-free buffer.[244] This protocol instead measured the change in rate of association of BODIPY-GTP with GDP-bound GEF/GTPase. This assay better represents cellular conditions, as compounds must displace GDP and compete with BODIPY-GTP, rather than in the dissociation assay where competition with free guanine nucleotides is not considered.

Positive and negative controls were next established. Non-hydrolysable nucleotide analogues GppCHp and GppNHp bind with high potency to GTPases and can displace fluorescent analogues. Initial experiments using GppCHp were inconsistent, so GppNHp was adopted as a positive control. The tolerance of Kalirin/Rac1 was tested for increasing concentrations of DMSO. The complex was stable up to 2% DMSO, which was consistent with the stability of the complex observed in XChem and co-crystallisation experiments. DMSO

was also not observed to cause a change in FI at the wavelengths measured. DMSO was used as a negative control, with all wells containing the same concentration. Finally, a Z' test was performed to ensure that the DMSO and GppNHp controls were valid, and that the measurement of the assay across a 384 plate was robust. With a value of 0.84, the assay was considered acceptable and could be used for the testing of compounds identified during screening.

The analysis of compounds in the assay was completed as thus. The derivatives of the slopes were calculated using ICEKAT[245] and normalised using the rate value of the postitive control. Each compound was tested three times with triplicate technical repeats in each plate. The rates per well, as well as the mean and standard deviation, were plotted using GraphPad Prism in SuperPlots format.[185,246] The means of the technical repeats were compared to the negative control using a one-way ANOVA and subsequent Dunnett's test to identify compounds with statistically different initial rates to the DMSO control. Only results with a p value lower than 0.05 were considered of interest for further elaboration. The results of compound screening are detailed with their syntheses.

4.5.3 Potency Determination for Urea Follow-ups and Bioisosteres

Unfortunately, no significant change in the initial rate was observed at either 250 or 500 μM concentration for any of the compounds tested (Table 9, Appendix 8.8). This was despite clear evidence of electron density in the crystal structure for initial hit **32** and follow-up compounds **44-53**. This suggests that while **32** formed hydrogen bonds with D118, these interactions are insufficient to inhibit the exchange of GDP for GTP under the assay conditions. The bioisosteres are still fragments and are therefore likely to exhibit weak potency for the complex. As for the purchased analogues, it is not surprising that they did not show significant

activity over **32** based on their crystal structures. These compounds generally varied at the ethyl chain, extending out into the crystal channel without the formation of novel interactions.

Table 9. The initial rate of nucleotide exchange for urea follow-ups and bioisosteres (n=3). SuperPlots are depicted in Appendix 8.8. Fragments were tested at 500 μM. Compounds were incubated for two hours. All results were normalised using the positive control (GppNHp). No compounds exhibited significant inhibition compared to DMSO control (defined as $p < 0.05$, one-way ANOVA followed by Dunnett's multiple comparisons test). * n=2 for compound **46**.

Compound	Mean ΔRFU/min n=1	Mean ΔRFU/min n=2	Mean ΔRFU/min n=3	p Value
DMSO	6006 ± 1546	6326 ± 2323	3583 ± 922	N/A
32	5883 ± 417	8990 ± 352	4373 ± 156	>0.99
44	9050 ± 3959	10483 ± 1156	1980 ± 1191	>0.99
45	7533 ± 444	8960 ± 1381	2550 ± 375	>0.99
46	7397 ± 1721	10205 ± 559	*	0.55
47	7110 ± 1754	9840 ± 781	4610 ± 1487	>0.99
48	5833 ± 783	10763 ± 1335	5813 ± 398	0.98
49	6873 ± 1057	7497 ± 2767	4453 ± 746	>0.99
50	7407 ± 1404	7037 ± 776	5490 ± 1125	>0.99
51	6553 ± 313	10623 ± 1344	5187 ± 280	>0.99
52	4547 ± 553	7780 ± 1067	4687 ± 630	>0.99
53	7753 ± 196	7283 ± 2330	4063 ± 992	>0.99
55	4533 ± 657	8100 ± 460	4530 ± 1491	>0.99
56	1737 ± 119	5543 ± 751	3123 ± 412	0.97
61	8683 ± 1425	8993 ± 310	4213 ± 107	>0.99
62	6387 ± 2350	6647 ± 1120	7013 ± 747	>0.99
63	5117 ± 897	6300 ± 1154	5930 ± 809	>0.99

Compound	Mean ΔRFU/min n=1	Mean ΔRFU/min n=2	Mean ΔRFU/min n=3	p Value
64	6370 ± 1065	7567 ± 1457	4973 ± 361	>0.99
65	3677 ± 145	7667 ± 744	4937 ± 444	>0.99
66	5720 ± 1851	7870 ± 646	4837 ± 1155	>0.99
68	5290 ± 1670	6750 ± 1038	4463 ± 263	>0.99
69	3873 ± 444	6830 ± 1711	5163 ± 309	>0.99
71	5587 ± 2945	6937 ± 80	3190 ± 372	>0.99
73	5607 ± 3739	5050 ± 584	4330 ± 462	>0.99
74	7093 ± 656	7507 ± 839	4177 ± 330	>0.99
75	4593 ± 189	5903 ± 934	6083 ± 624	>0.99
79	3933 ± 1598	7893 ± 1903	5840 ± 475	>0.99
80	7367 ± 1701	6710 ± 1653	4767 ± 327	>0.99
82	7550 ± 1761	5717 ± 1114	4977 ± 840	>0.99
83	3263 ± 1162	7173 ± 826	4053 ± 106	>0.99
84	4173 ± 2586	5737 ± 350	4733 ± 509	>0.99
88	3917 ± 3203	6280 ± 363	4057 ± 299	>0.99
89	4487 ± 2422	6807 ± 1131	3763 ± 788	>0.99
90	3567 ± 1210	7660 ± 754	5083 ± 870	>0.99
93	4080 ± 481	7967 ± 690	4343 ± 1047	>0.99

4.6 Conclusions

Although not surprising that **32** and the urea bioisosteres were not active in the nucleotide exchange assay due to their small size, it was disappointing that no compounds showed significant activity at the concentrations measured. It had been hoped the **32** would demonstrate activity as it was capable of displacing GDP within Kalirin/Rac1 crystals. The assay results suggest that its association with Rac1 is not potent enough to compete with high physiological concentrations of guanine nucleotides, despite their lower affinity for the GEF/GTPase complex.

A potential caveat could be that these compounds were relatively insoluble in the assay conditions. Further solubility experiments should be conducted to ensure this is not a factor influencing the outcome of the nucleotide exchange reaction. However, these results suggest that any future modifications of **32** should concentrate on extending into the pocket rather than changing or extending the urea side chain. This would involve the development of analogues with growth vectors on the 4-position of the isoxazole or isosteres of urea which can extend from the carbonyl position. Further exploration of non-covalent analogues to **32** was suspended to pursue the more promising leads detailed in Chapters 5 and 6.

5. Kalirin/Rac1 Electrophilic Fragment Screening

5.1 Overview

As detailed in section 1.2.3, covalent compounds can often be advantageous as drugs or probes, with exquisite selectivity and high potency. This chapter details the identification of two sets of electrophilic fragments that formed a covalent linkage to a cysteine present in the nucleotide binding site of Kalirin/Rac1. Section 5.2 describes the identification of different conformations of the switch I loop accessed by small molecules, both in the XChem fragment screen and by covalent compounds. It also highlights current attempts to elucidate the precise orientation of this loop in Frank von Delft's group. The latter half of the chapter concentrates on the SBDD of electrophilic fragments with similar pharmacophores to GDP and **32**, with the subsequent identification of a covalent inhibitor that is active in the nucleotide exchange assay. The ongoing medicinal chemistry strategy to extend these compounds into the nucleotide binding pocket is also discussed.

5.2 Electrophilic Mass Spectrometry Fragment Screening

5.2.1 Co-crystallisation Experiments

Initial analysis of the Kalirin/Rac1 structure indicated two solvent accessible cysteines; C18 in the guanine nucleotide binding site and C105 at the bottom of the α3 helix, adjacent to the switch II region and closer to the Kalirin/Rac1 interface (Figure 67). A collaboration with the London group at the Weizmann Institute of Science, Israel was initiated for the mass spectrometry screening of Kalirin/Rac1 with an electrophilic fragment library.[40] It was hoped that the library would target C18 in the nucleotide pocket, with the intention that any fragments identified could be co-crystallised for structure-based drug design.

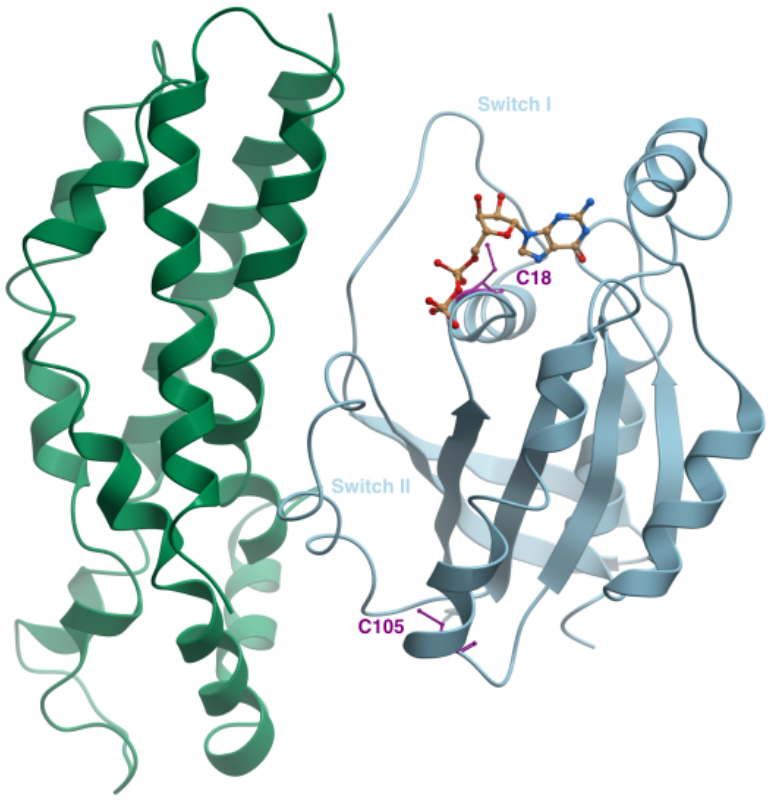
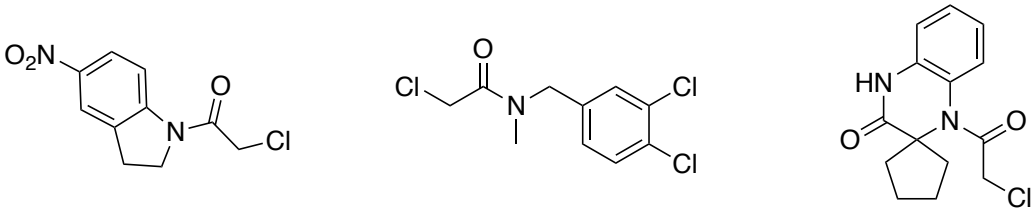


Figure 67. The solvent-accessible cysteines (purple) of Rac1 (light blue) bound to Kalirin (green, PDB code: 5O33). GDP is shown as light brown sticks.

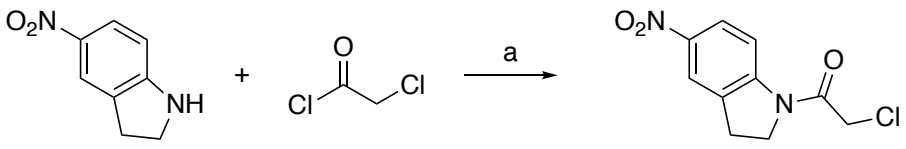
Their initial screen at 200 μM found 40 hits (Appendix 8.9.1). 15 of these compounds were removed as they were identified as promiscuous binders based on previous experience in the London group. Of the compounds remaining, these could be organised into three subcategories; those containing an indoline motif, chlorobenzylamines, and 'miscellaneous', which were 14 compounds not significantly related to the others (Figure 68).



99 **100** **101**

Figure 68. Examples of compounds identified in the mass spectrometry screen. Compound **99** contains an indoline motif, **100** is an example of a chlorobenzylamine, and **101** represents a 'miscellaneous' compound.

Eight indolines bound to Kalirin/Rac1, which constituted the majority of the indolines represented in the first electrophilic library. The indolines were therefore advanced as chemical leads. Additional experiments comparing the labelling of Rac1 monomer to the complex with additional indolines from the London lab showed that labelling was quicker and larger for Kalirin/Rac1 (Appendix 8.9.2). This provided evidence for the hypothesis that the selective targeting of the GEF/GTPase complex may have more amenable than targeting the GTPase alone. **99** was chosen for crystallisation due to the substructural amine **102** being commercially available, with the simple acylation conducted in good yield (Scheme 20).



102 **99**

Scheme 20. Synthesis of covalent indoline fragment hit **99**. Reagents and conditions: a) TEA, DCM, -20 °C, 43%.

Several attempts were made to crystallise the compound by soaking into Kalirin/Rac1 crystals, as this is simpler and requires less protein than co-crystallisation. The compound was first soaked into crystals overnight, but this destroyed the crystal at higher concentrations of ligand. Lower concentrations did not destroy the crystals but evidence for the compounds could not be seen in the solved crystal structure. Next efforts dispensed 5-10 nL droplets of compounds with the protein on the crystal plate to allow **99** to bind to the complex before the crystallisation buffer was added. However, this was also unsuccessful, with most crystallisation wells containing precipitated protein. The few crystals that formed did not show evidence of compound binding, indicating that the compound-free complex may be selectively crystallising. The difficulties of obtaining crystal structures of covalent compounds using soaking methods had been previously experienced by both Frank von Delft's group and the London group on other crystal systems, and so the lack of success in Kalirin/Rac1 was not unanticipated. Hence, this method was deprioritised for co-crystallisation.

99 was combined with Kalirin/Rac1 at four times the protein concentration and incubated overnight at 4 °C. Sufficient labelling (> 80%) was confirmed by QTOF intact mass spectrometry. Crystallisation trials using a small screen of conditions around the optimal conditions for the complex yielded crystals which diffracted to 1.96 Å. Data collection and refinement statistics are given in Appendix 8.10.

The electron density of **99** bound to the complex was elucidated using the GDP-bound Kalirin/Rac1 structure for phasing (PDB code: 5O33). It was clear from the electron density maps that the indoline covalent handle was bound to C18 rather than C105, with no unfitted electron density near any other cysteine residue. However, the electron density maps surrounding C18 were ambiguous. The same rearrangement of the switch I loop identified in the non-covalent XChem fragment screen could be observed in this structure (Figure 69A).

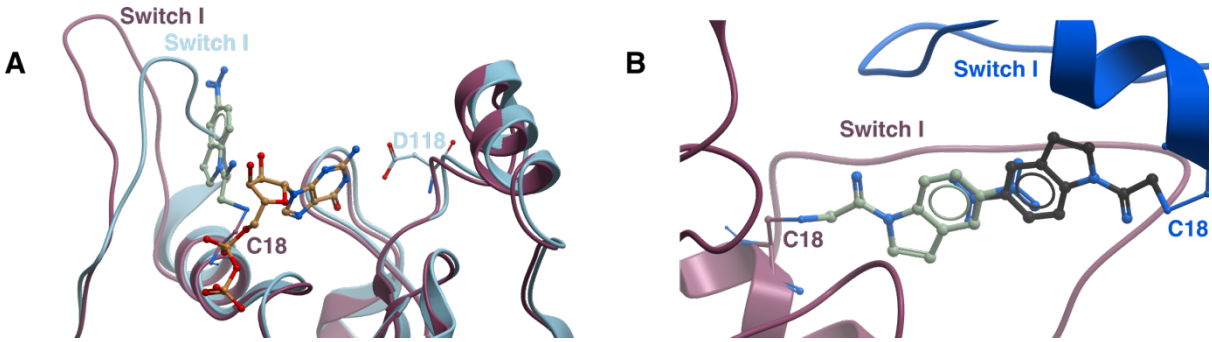


Figure 69. **A**. Comparison on the indoline **99** (green sticks) bound to the C18 residue of Rac1 (purple) in the Kalirin/Rac1 complex compared to the GDP-bound Rac1 (light blue) in complex with Kalirin (PDB code: 5O33). A large rearrangement of the switch I region can be observed due to the orientation of the indoline compound overlapping with the normal orientation of the switch I loop in the GDP-bound conformation. The indoline does not interact with any residues within the switch I region. **B**. **99** overlaps with its symmetry copy (black sticks). The symmetry copy of Rac1 is shown in dark blue.

Due to the relatively high occupancy of **99**, the loop was mainly in its reorganised state. It thus could be fitted using the Fo-Fc map into its correct orientation and the model subsequently refined. This would in theory mean that the electron density of the compound could be identified as it would no longer be camouflaged by the Fo-Fc maps arising from the incorrect modelling of the switch I region. This could not be achieved for the XChem fragments due to their small size and low occupancy.

However, upon refinement, the density for **99** overlapped with that of its symmetry copy, with no predicted interactions with the residues of Rac1 (Figure 69B). The guanine nucleotide binding pocket is near its symmetry copy within the crystal lattice. It was unknown whether the binding pose adopted by **99** was an artefact due to its proximity to the symmetry copy, or an indication that the indolines were non-specifically reacting with the cysteine residue in the binding pocket. Further crystallisation trials using other coarse screens or different constructs of Kalirin were attempted to discover an alternative space group of Kalirin/Rac1 where the guanine nucleotide binding pocket was not situated near its symmetry copy. Crystals were obtained in higher pH and PEG with a hexagonal morphology but had the

same space group of $P6_5$ 2 2 and so no difference was observed in the position of the switch I loop in relation to the symmetry copy.

The London group subsequently conducted further screening of the indolines on other non-related proteins, with the results indicating that the indolines were fairly reactive and not selective for Rac1. This suggested that the lack of interactions observed in the crystal structure of **99** could arise from non-specificity. Hence, this meant that the indolines were no longer attractive leads. However, further analysis of the loop arrangement was still desired, and so other compounds in the electrophilic fragment screen which were not identified as promiscuous were prioritised.

Consequently, co-crystals of the chlorobenzylamines and 'miscellaneous' compounds were obtained to determine whether the same movement in the switch I region occurred. Four compounds were commercially available and co-crystallisation was attempted as described above, successfully generating three ligand-bound crystals (Figure 70, Appendix 8.10). All of the compounds bound to C18 and caused an identical loop movement in the switch I region. Interestingly, not all of the compounds overlapped with their symmetry copy, indicating that the switch I loop rearrangement was induced by compound binding rather than being an artefact of the crystallisation process. This had been a theory following the XChem screen due to the lack of electron density that could be elucidated for the weakly-binding fragments, and the close proximity of the symmetry copy.

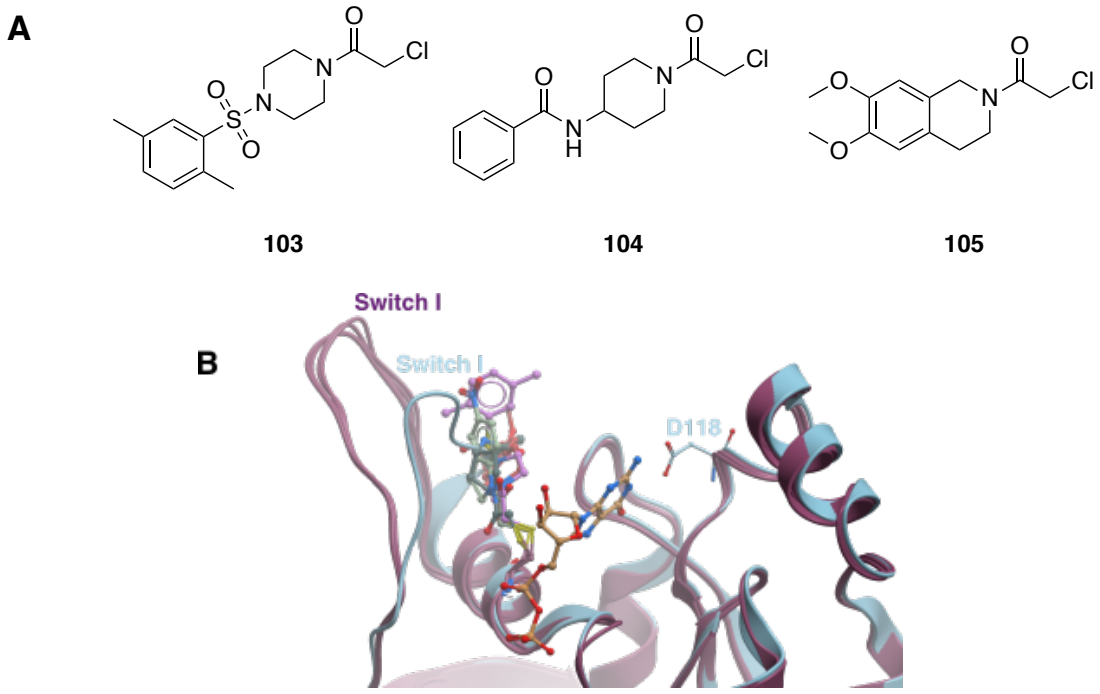


Figure 70. **A**. Commercially available electrophilic fragments **103-105**. **B**. Structures of **103-105** (various coloured sticks) overlaid with **99** (light green sticks) bound to Rac1 (all purple) and GDP (brown sticks) bound to Rac1 (light blue, PDB code: 5O33). All covalent compounds caused the same loop rearrangement, despite the binding poses of the compounds not completely overlapping. None of the compounds overlap where GDP binds within the guanine nucleotide binding pocket.

With this in mind, the PanDDA structures from the original XChem screen were closely re-analysed. Subsequent repeat soaks of the non-covalent compounds causing the loop rearrangement yielded structures with the same alternative conformation of the switch I region, again indicating that the switch I conformational change is compound dependent. Comparison of the real-space correlation coefficient (RSCC) of each structure was undertaken by Elliot Nelson of Frank von Delft's group to examine the differences in the positions of the switch I residues (Appendix 8.11). This allowed categorisation of the structures into three different sub-groups depending of the occupancy of the switch I loop: first, the complex was almost 100% in the original loop position of Kalirin/Rac1, suggesting the presence of non-covalent compounds in low occupancy; second, the loop was nearly 100% rearranged, as was the case for high occupancy covalent ligands; or third, the largest sub-set, where the structures

contained a mixture of the original loop and the rearranged loop within the crystals. These results explained why it was extremely difficult to identify ligands using PanDDA (section 4.3.1), as neither the loop representation of the GDP-bound structure or the rearranged loop would represent a reasonable model for the superposition of both loops in the background electron density algorithm used by PanDDA. Attempts to run PanDDA on the different subsets was ultimately not successful due to the small number of datasets.

It is not without precedent that the binding of a compound causes alternative conformations of the switch regions; indeed, the Gray lab has published a novel arrangement of the switch II region of Ras induced by a subset of their covalent inhibitors.[140] The specific arrangement of the loop may not be the 'correct' orientation occupied in solution, perhaps necessitated by the space available in the crystal or the crystal space group. However, from the data it can be ascertained that the same compounds are causing a movement, rather than this rearrangement being a random orientation in a subset of crystals only identified due to the large number of crystals analysed during the XChem process.

The considerable effort into establishing the orientation of the switch I loop was undertaken due to its key role in the activation cycle of the GTPase. This region is essential for binding to effectors and to the guanine nucleotides, and thus its rearrangement could have physiological relevance. Displacement of this loop by compounds could inhibit nucleotide exchange or effector activation, thereby acting as a novel mechanism to inhibit GTPase activity. The compounds bind in the presence of GDP. This would suggest these compounds, if they inhibit GTPase activation, would circumvent the issue of competing with the highly potent guanine nucleotides. Therefore, efforts to elucidate the loop structure and compound binding poses are essential for the development of further compounds which may act as inhibitors. Attempts in Frank von Delft's group are currently ongoing using the PDB REDO programme to provide a better solution of the electron density of the loops to allow correct modelling of XChem fragment binding poses.

5.2.2 Potency determination of the Initial Electrophilic Fragment Screen

The same nucleotide exchange assay as described in section 4.5 was used to test the covalent compounds detailed above. Unfortunately, none of the ligands which caused the loop rearrangement inhibited the association of BODIPY-GTP with Rac1, indicating that despite causing movement within the switch I region, it was either insufficient to inhibit GDP/GTP exchange or the compounds bind too weakly (Figure 71). No significant inhibition was seen when the experiment was repeated over 24 hours (Appendix 8.13). This does not necessarily mean that these compounds are inactive; this assay does not measure the effect of compounds on the formation of the GTP-bound GTPase, only on the GEF/GTPase complex. It is unknown based on the current experiments whether the GEF can dissociate or, if this occurs, whether the GTP-bound GTPase is able to access the activated 'state 2' orientation for effector binding when the covalent compounds are present near the switch I loop.

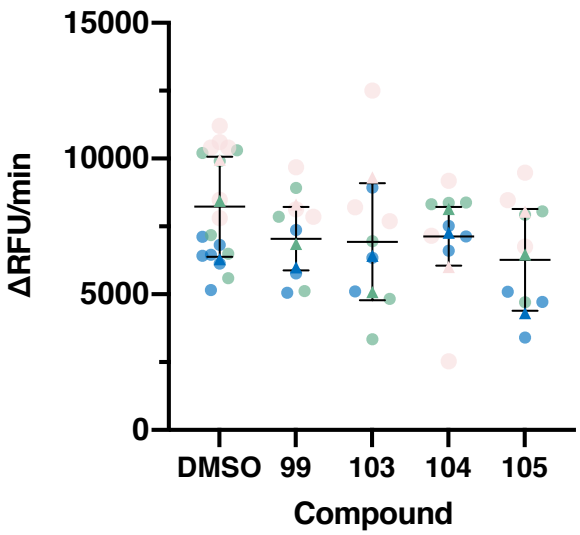


Figure 71. Nucleotide exchange assay for the initial electrophilic fragment screening tested at 500 μM (n = 3, *** $p < 0.001$, one-way ANOVA followed by Dunnett's multiple comparisons test). Compounds were incubated for two hours. Blue, green and pink circles represent three technical replicates within each repeat. Triangles represent the mean for each repeat. Black bars represent the overall mean ±

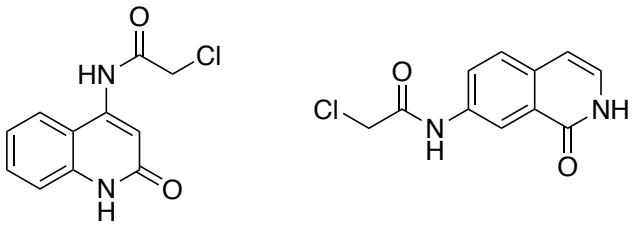
SD. All results were normalised using the positive control (GppNHp). No compounds showed significant inhibition compared to the negative control.

The development of an assay to measure GTPase activation and effector activity, and the PDB REDO of structures is considered of academic interest, but pursuit of these compounds as therapeutics was deprioritised based on the difficulties in structure-based design, their lack of activity in the nucleotide exchange assay and the identification of a more promising set of covalent compounds detailed in the next section.

5.3 Covalent Pharmacophore Docking Hits

5.3.1 Co-crystallisation of Pharmacophore Docking Leads

Following the decision to deprioritise the indolines due to promiscuity and the conformational changes in the crystal, the London group conducted an *in silico* covalent screen of 67,127 chloroacetamides against the GDP pocket of Rac1 in its GEF complex using CovDock.[247] Compounds were preferentially sorted by the predicted ability to mimic the hydrogen bonding pharmacophore of GDP. Of these, eight compounds were chosen for follow-up screening. Dose-response labelling and reactivity experiments compared to the indolines were conducted on the Kalirin/Rac1 complex in the London group (Appendix 8.12). Two compounds, **106** and **107**, were chosen as leads due to their ability to label Kalirin/Rac1 in the dose-response experiment but were significantly less reactive in comparison to the indolines (Figure 72).



106 **107**

Figure 72. Covalent hits **106** and **107** identified from the CovDock virtual screen.

Co-crystallisation experiments were conducted in a similar manner to those described above (section 5.2, Appendix 8.10). The structures of **106** and **107** were solved to 1.35 Å and 1.65 Å respectively (Figure 73). The morphology of the covalent co-crystals was more similar to the GDP-bound Kalirin/Rac1 crystals than the nucleotide-free complex.

It is interesting to note that the binding poses of both compounds accurately align with the covalent docking prediction. Both compounds bind to C18 of the guanine nucleotide binding pocket and their heterocycles orientate in the same position as the guanine ring of GDP. No rearrangement of the switch I loop was observed. Both heterocyclic rings in **106** and **107** form a single hydrogen bond with D118. Due to lower reactivity in the selectivity assay, but similar potency in the dose-labelling response assay, **106** was chosen as a lead over **107** for further optimisation.

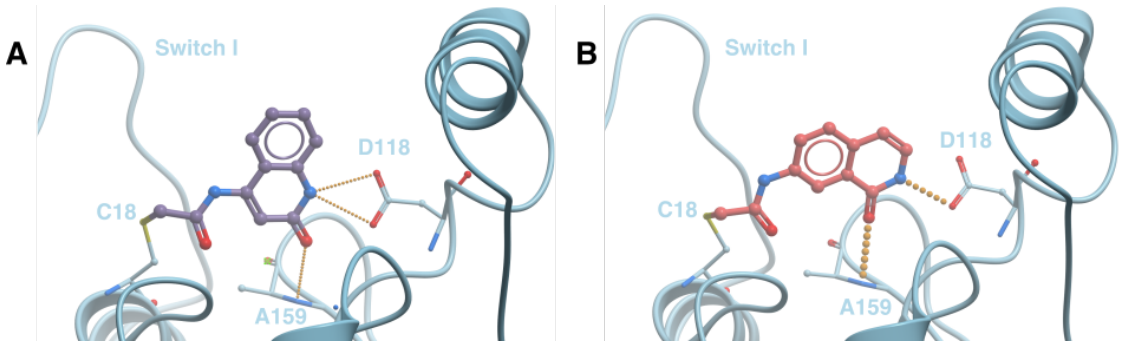


Figure 73. **A**. Structure of **106** bound to C18 of the Rac1 guanine nucleotide binding site. It can form a hydrogen bond with either oxygen of the key D118 residue of Rac1, as predicted in the docking software.. **B**. Structure of **107** bound to C18 of the Rac1 guanine nucleotide binding site.

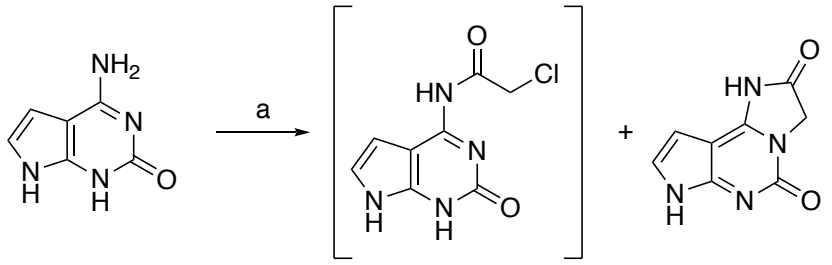
5.3.2 Fragment Screening of Kalirin/Rac1 Bound to 106

The consistently high resolution of **106**-bound co-crystals meant they were suitable for XChem fragment screening. The space in the guanine nucleotide binding pocket between the switch I loop and C18 could accommodate another fragment. It was hypothesised that fragments occupying the pocket at the same time as **106** could be used in a fragment linking or merging strategy to improve the potency and number of interactions with Kalirin/Rac1. 176 fragments from the DSI-Poised Library were soaked into **106** co-crystals. Unfortunately, only fragments binding at the crystal contact areas were identified. It was not known whether this meant the nucleotide pocket was too small to accommodate another full-sized fragment, or if the fragment library tested was simply unfavourable for the desired area of the binding pocket. Due to the difficulties in obtaining consistent numbers of good quality crystals due to presence of DMSO, the fragment screening campaign of **106** co-crystals was deprioritised in favour of fragment growing and structure-based design of the original compound **106**.

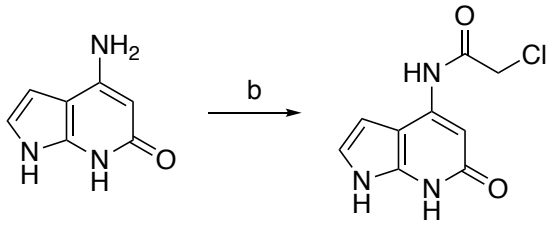
5.3.3 Optimisation of 106

GDP forms two hydrogen bonds with D118 as it contains two hydrogen bond donors in its structure. From analysis of the co-crystal structure, it was anticipated that the replacement of a carbon in the ring of 106 with an NH group would introduce a second hydrogen bond interaction, thereby improving the potency. Mary Wilson, a rotation student in Paul Brennan's group, tried to synthesise the relevant compound **108**, with the pyrrole predicted to be simpler to make than the pyrimidine (Scheme 21). Whilst they were successful in synthesising the basic amine scaffold **108**, the final acylation was unsuccessful. The nitrogen adjacent to the covalent handle displaced the chlorine in the final step to produce a linked heterocycle **110** instead of the desired chloroacetamide **108**. It was hypothesised that

112 would form identical hydrogen bonding interactions but would not have the same issues with cyclisation due to the removal of the nitrogen near to the warhead. The core scaffold **111** was synthesised by a commercial CRO Wuxi, and Raina Seupal from Paul Brennan's group conducted the acylation to generate **112** in reasonable yield (Scheme 21).



108 **110**



111 **112**

Scheme 21. The synthesis of **109** from **108** by Mary Wilson was unsuccessful and **110** was generated instead. The synthesis of **111** was conducted by Wuxi and **112** was successfully synthesised by Raina Teupal. Reagents and conditions: (a) 2-chloroacetyl chloride, DIPEA, DCM, r.t., 1 h, 14% (**110**) (b) TEA, DCM, DMF, r.t. 2h, 10%.

Co-crystallisation of **112** was conducted as described previously and solved to 1.68 Å. Overlay of the structure of **112** bound to C18 and GDP in the same pocket shows them to overlap nearly perfectly, with **112** forming the additional predicted hydrogen bond with D118 (Figure 74). Comparison of **106** and **112** shows a slight shift by **112** into the pocket to allow the formation of the second hydrogen bond.

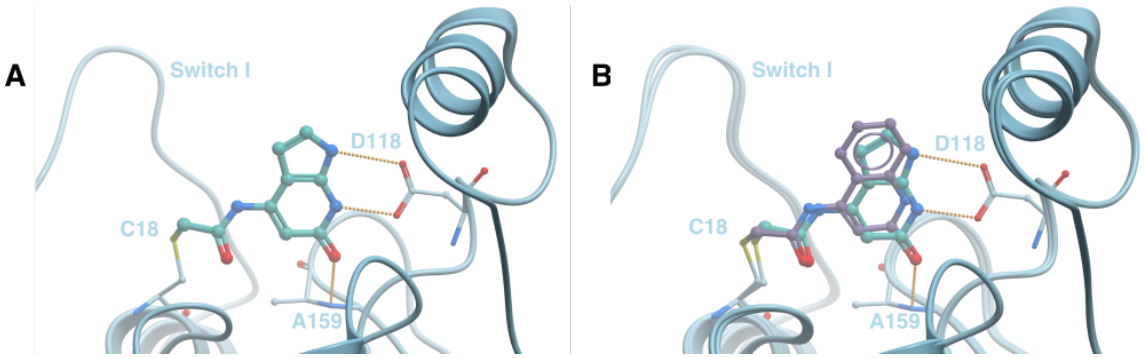
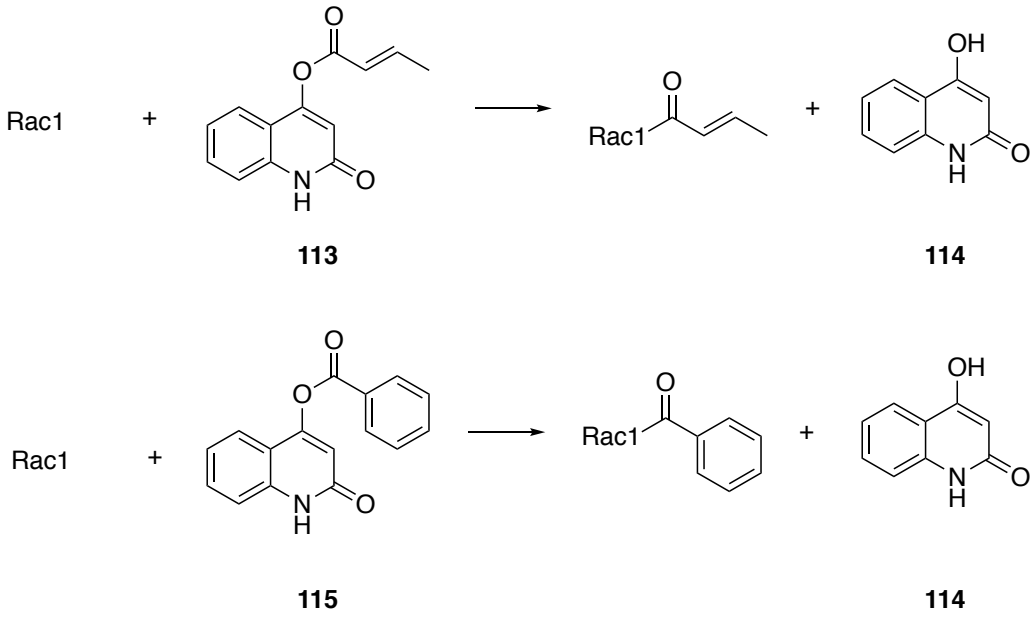


Figure 74. **A**. Crystal structure of **112** (green sticks) bound to Rac1 (light blue) in the Kalirin/Rac1 complex. It can form two hydrogen bonding interactions with D118. **B**. Comparison of the **106** and **112** binding poses. The heterocyclic rings of **112** are shifted in comparison to **106** to form the second hydrogen bond with D118.

5.3.4 Adapting the Covalent Warhead

The London group synthesised **113** to explore different covalent handles of **106**, as chloroacetamides are often too reactive and non-selective to be used as probes or therapeutics. Substituting the handle could also be a method to extend further into the guanine nucleotide binding site. The compound labelled around 20% of Rac1 in their mass spectrometry assay (200 μM concentration). Instead of reacting in a similar manner to **106**, with the heterocyclic ring forming an adduct with Rac1, **113** formed an adduct with a mass attributable to the electrophilic handle. This suggested that the heterocycle **114** was instead acting as a leaving group (Scheme 22). To test this theory, **115** was synthesised, and the adduct observed matched with benzoate addition. It was hypothesised that **113** and **115** actually bound to K116 in the guanine nucleotide pocket, with precedence of esters in the literature reacting with lysine residues.[248]



Scheme 22. The unexpected labelling of **113** and **115** with Rac1. The residue of Rac1 reacted in 1,2 direct addition instead of the expected 1,4 conjugation addition, resulting in **114** acting as a leaving group and the electrophilic handle present in the Rac1 adduct. This was further supported by the synthesis of **115**.

115 was co-crystallised to 2.15 Å according to the method previously described. Not only did **115** bind to C18 instead of K116 as predicted, but it caused the same loop rearrangement as seen before for **99, 103-105,** and non-covalent XChem fragments (Figure 75, Appendix 8.10).

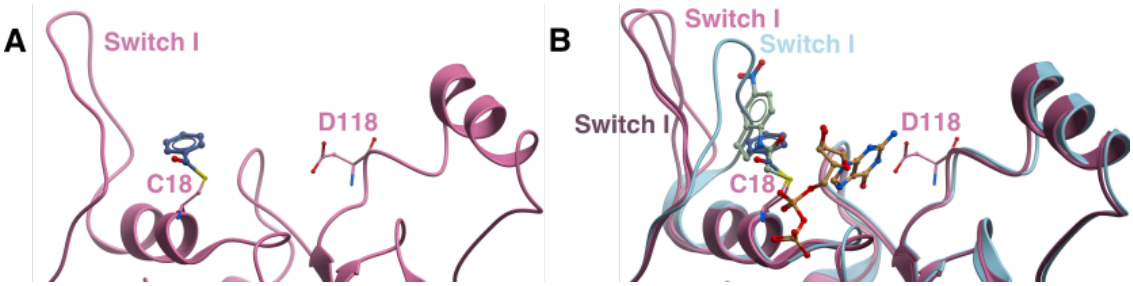
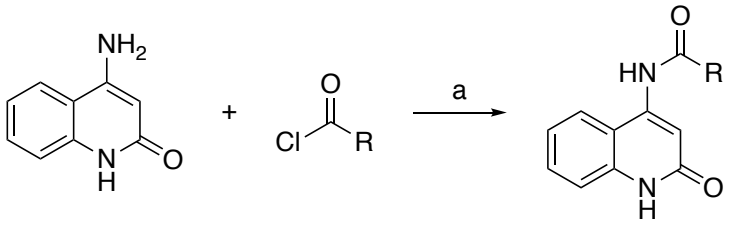


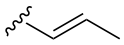
Figure 75. **A**. The solved structure of "**115**-product" (blue sticks) bound to Rac1 (pink) in the Rac1/Kalirin complex indicated that **115**-product bound to C18 rather than K116 as predicted. It was orientated away from D118, unlike its analogue **106**. **B**. Overlay of **115**-product with **99** (green sticks,

Rac1 in purple) and GDP (brown sticks, Rac1 in light blue, PDB code: 5O33) bound structures. **115-**product produces a similar rearrangement of the switch I loop to **99**.

This again supports the theory that while the structure seen in the crystal may not necessarily be representative of the actual loop arrangement, it appears that these compounds are catalysing some movement in the flexible switch I region. The synthesis of further analogues containing these handles is being undertaken by the London group. The structure of **115** bound to Kalirin/Rac1 has been added to the PDB REDO subset in Frank von Delft's group to aid analysis of the switch I loop movement.

Other handles were also synthesised in house, with the acrylamides **116** and **117** and chloropropionamide **118** generated in an acylation reaction using the purchased **106** scaffold and relevant acyl chloride (Scheme 23).



Compound	R	Yield (%)
116		2
117		5
118		48

Scheme 23. Acylation of the Rak-DK-2 core. Reagents and conditions: (a) DMF, r.t., 16h.

5.3.5 Potency Determination of the Covalent Pharmacophores

With co-crystallisation and structure-based design demonstrating that a subset of the *in silico* covalent compounds bound in the desired modality, they were then tested in the nucleotide exchange assay to see if this correlated with activity *in vitro* (Figure 76).

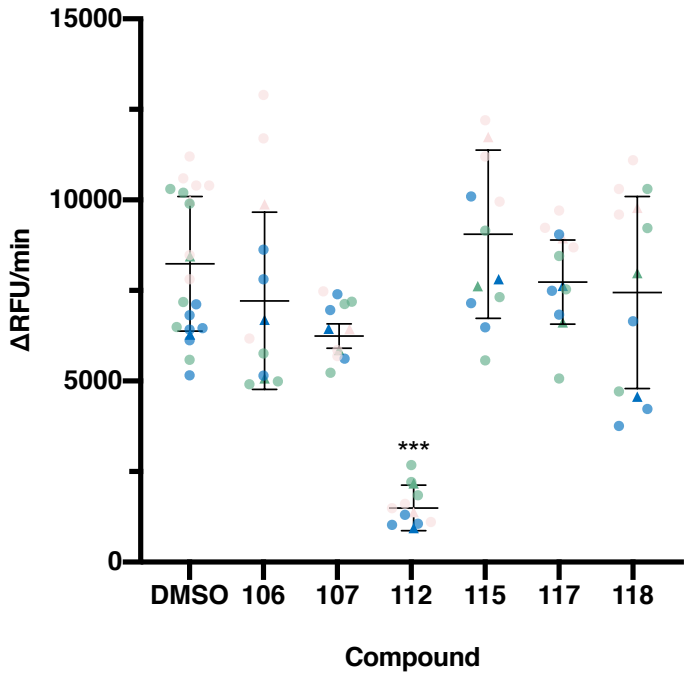
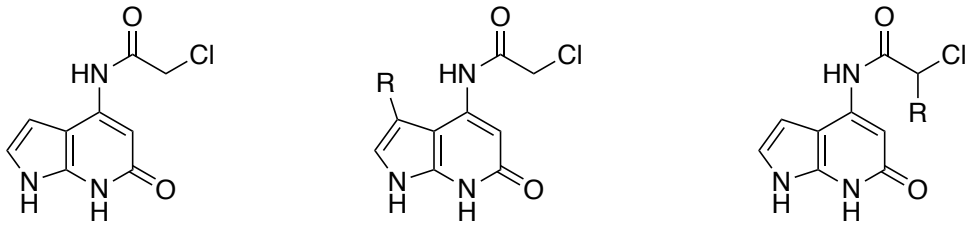


Figure 76. Nucleotide exchange assay for pharmacophore covalent compounds tested at 500 μM (n = 3, *** $p < 0.001$, one-way ANOVA followed by Dunnett's multiple comparisons test). Compounds were incubated for two hours. Blue, green and pink circles represent three technical replicates within each repeat. Triangles represent the mean for each repeat. Black bars represent the overall mean ± SD. All results were normalised using the positive control (GppNHp). **112** shows highly significant inhibition of nucleotide exchange in comparison to the negative control.

The nucleotide exchange assay indicates that **112** was the only compound that significantly inhibited the activation of Rac1 by Kalirin. This signifies that the presence of two hydrogen bond donors interacting with D118 within the heterocyclic ring system is crucial for covalent compound activity. This was an exciting development, with **112** being the first

synthesised compound to possess measurable potency thus far. Analogues of **112** were hence prioritised for further development.

However, during co-crystallisation trials, it was observed that **106** took significantly longer than **107** or **112** to label Kalirin/Rac1 by > 80% by mass spectrometry. The potency of covalent compounds is often time-dependent due to the dual-stage nature of binding; first, the non-covalent scaffold must associate with the protein before the electrophilic handle reacts with the requisite residue. It was hypothesised that the lack of inhibition observed in the assay could be due to a longer association time. However, when the ligands were incubated for 2, 6 12 and 24 hours, no significant inhibition was observed for either scaffold (Appendix 8.13). **115**, as with the other loop-rearrangement covalent compounds, was not active in the assay. Synthesis of follow-up analogues to **112** are being conducted in Paul Brennan's group (Figure 77).



112

Figure 77. Ongoing synthesis of **112** and its analogues in the Brennan group.

5.4 Merging of Fragment Leads 32 and 112

5.4.1 Covalent Urea Synthesis

It was noted that the interactions of **106**, **112** and urea **32** from the XChem screen exhibited similar pharmacophores, with overlapping binding modes and identical hydrogen bonding patterns (Figure 78).

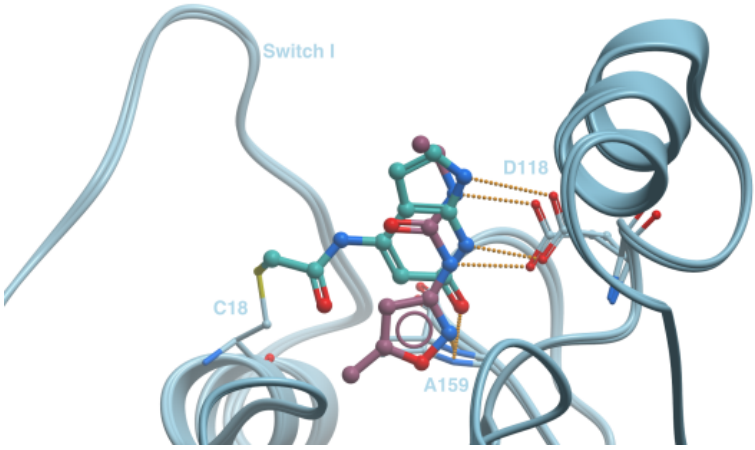
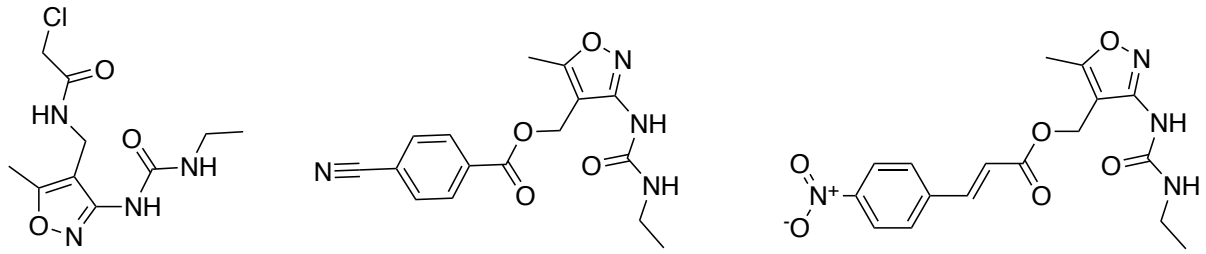


Figure 78. Overlap of **112** (green sticks) and **32** (purple sticks) bound to Rac1 (light blue) in the Kalirin/Rac1 complex. It was hypothesised that the compounds could be merged to form a covalent urea ligand.

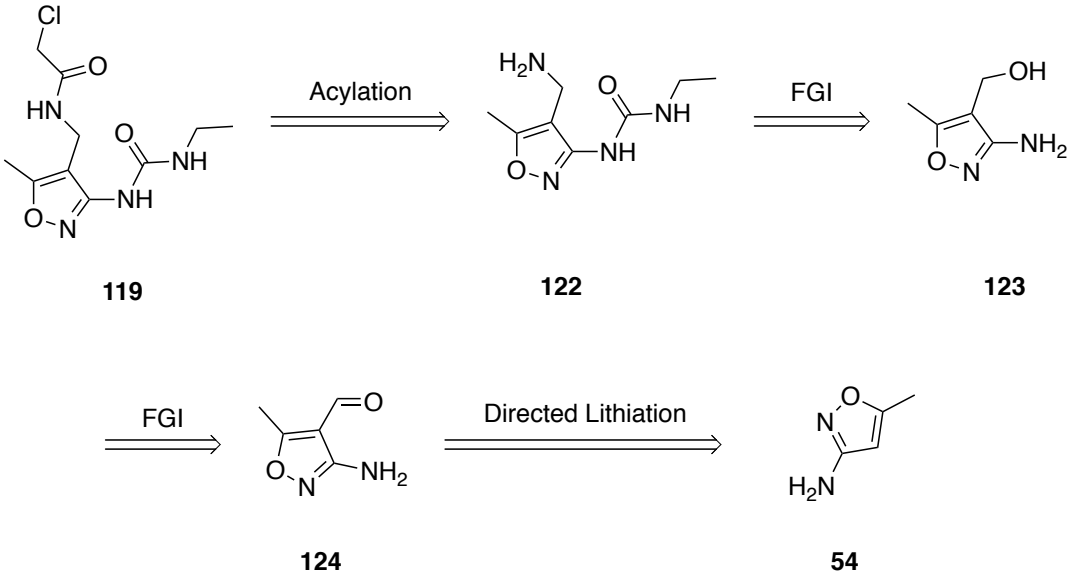
The isoxazole-urea motif was hence proposed as an alternative heterocyclic scaffold to generate a merged covalent fragment. **119** was designed, where **32** was modified to incorporate a chloroacetamide handle on the 4-position of the isoxazole ring (Figure 79). Other electrophilic moieties were desired, although the data obtained from **113** suggested that the warhead needed to be adapted into a better leaving group in order for the urea to bind to C18 instead of the handle. Based on Hammett's substituent constants, it was proposed that nitro or nitrile groups in the para position of a phenyl may stabilise the handle as a leaving group and so promote the formation of a urea adduct with Rac1.[249,250] Hence, **120** and **121** were proposed as urea analogues to **113** and **115**.



119 **120** **121**

Figure 79. Covalent compounds designed by merging **32** with **112, 113**, and **115.**

The retrosynthetic strategy to generate **119** is depicted in Figure 80. It was proposed that **119** could be synthesised from an aminomethyl attached to the 4-position of the isoxazole ring. This could be generated by a sequential functional group interconversion from an intermediate alcohol **123** which could itself be synthesised from the relevant aldehyde **124**. This is turn could be made from the starting material **54** using directed lithiation.



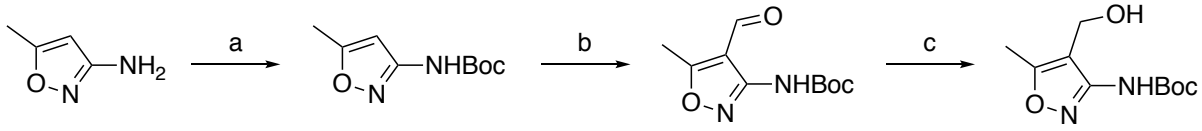
119 **122** **123**

124 **54**

Figure 80. Retrosynthetic strategy to generate covalent analogue **119** from common starting amine **54.**

The steps generating alcohol **127** from **54** were optimised by two rotation students in Paul Brennan's group (Scheme 24). First, **54** was protected using a Boc group to prevent side reactions, to discourage the amine being involved in opening the isoxazole ring and to aid purification of subsequent intermediates. Mild base was used to avoid the doubly protected

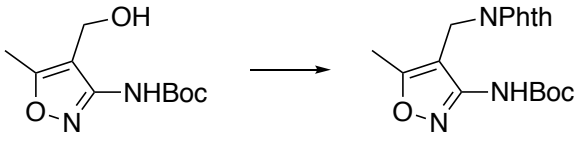
side product. The aldehyde in the 4-position was introduced using a directed lithiation with DMF as an electrophile in high yield. This was subsequently reduced to the alcohol using $NaBH_4$. This reaction was clean and proceeded with reasonable yield, and no opening of the isoxazole ring observed.



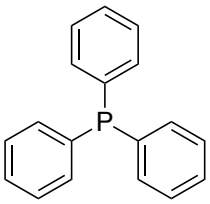
54 **125** **126** **127**

Scheme 24. Optimised of initial steps by Aidan Kerckoffs and Hollie Parsons. Reagents and conditions: (a) Boc anhydride, DMAP, TEA, DCM, r.t., 16 h, 85%; (b) i) *n*-BuLi, THF, -78 °C-r.t., 0.5 h; ii) DMF, -78 °C-r.t., 16 h, 70%; (c) $NaBH_4$, MeOH, r.t., 1 h, 50%.

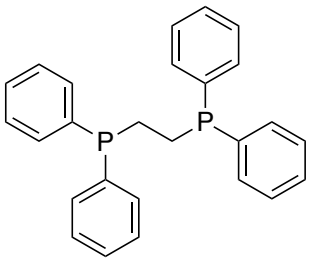
It was conceived that the alcohol might be converted to an amine via the installation of a phthalimide group in a Gabriel-type synthesis. The phthalimide was introduced using Mitsunobu conditions with triphenylphosphine (TPP) and di-*tert*-butyl azodicarboxylate (DtBAD) as reagents. However, these conditions resulted in low yield due to the co-purification of **128** with the triphenylphosphine oxide side product. This necessitated two purifications by flash column chromatography to give pure product. Literature detailing the Mitsunobu reaction documented 1,2-Bis(diphenylphosphino)ethane (DPPE) and diphenyl-2-pyridylphosphine as alternatives to TPP as coupling reagents.[251] The advantages of these compounds are that their side products are easier to purify; the bis-oxidised side product of DPPE is insoluble in THF and the side product of diphenyl-2-pyridylphosphine can be removed by an acid wash. No reaction was observed using diphenyl-2-pyridylphosphine, but the reaction proceeded extremely quickly using DPPE. The acid wash used during literature purification to remove DtBAD side products was unsuitable in this case, as it caused deprotection of the Boc group, with the resulting amine adsorbing to the silica column. Whilst the yields were only mildly improved, the reaction was completed in an hour and only required filtration and a single column purification, simplifying the reaction overall (Scheme 25).



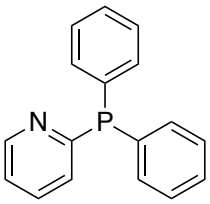
127 **128**



TPP



DPPE



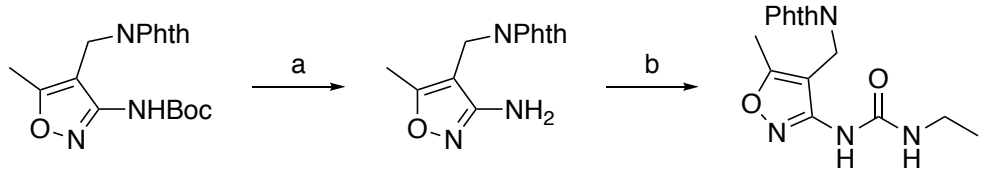
Diphenyl-2-pyridylphosphine

Reagent	**Eq.**	**Yield**	**Comments**
TPP	1.3	10-20%	Requires two columns for purification
DPPE	1.3	N/A	Single P=O is soluble in THF; fewer eq required for bis P=O formation
DPPE	0.6	22	Acid wash hydrolysing Boc group
DPPE	0.6	35	Highest yield and easiest purification
Diphenyl-2-pyridylphosphine	1.3	N/A	No reaction

Scheme 25. Alternative Mitsunobu reagents for the conversion of **127** to **128** resulted in improved purification

Attention then turned to introducing the urea motif, as the phthalimide would act as a protecting group for the aminomethyl in the 4-position and so would not be removed until the final stages of the synthesis. A simple Boc-deprotection using TFA gave the free amine in the 3-position of the isoxazole ring in high yield (**129**, Scheme 26). Literature protocols for similar scaffolds used triphosgene as the reactive reagent and THF and DMA as solvents; these conditions were attempted but the reaction did not reach even 50% completion whilst monitored by LCMS.[223] Neither heating nor the additions of extra equivalents of triphosgene

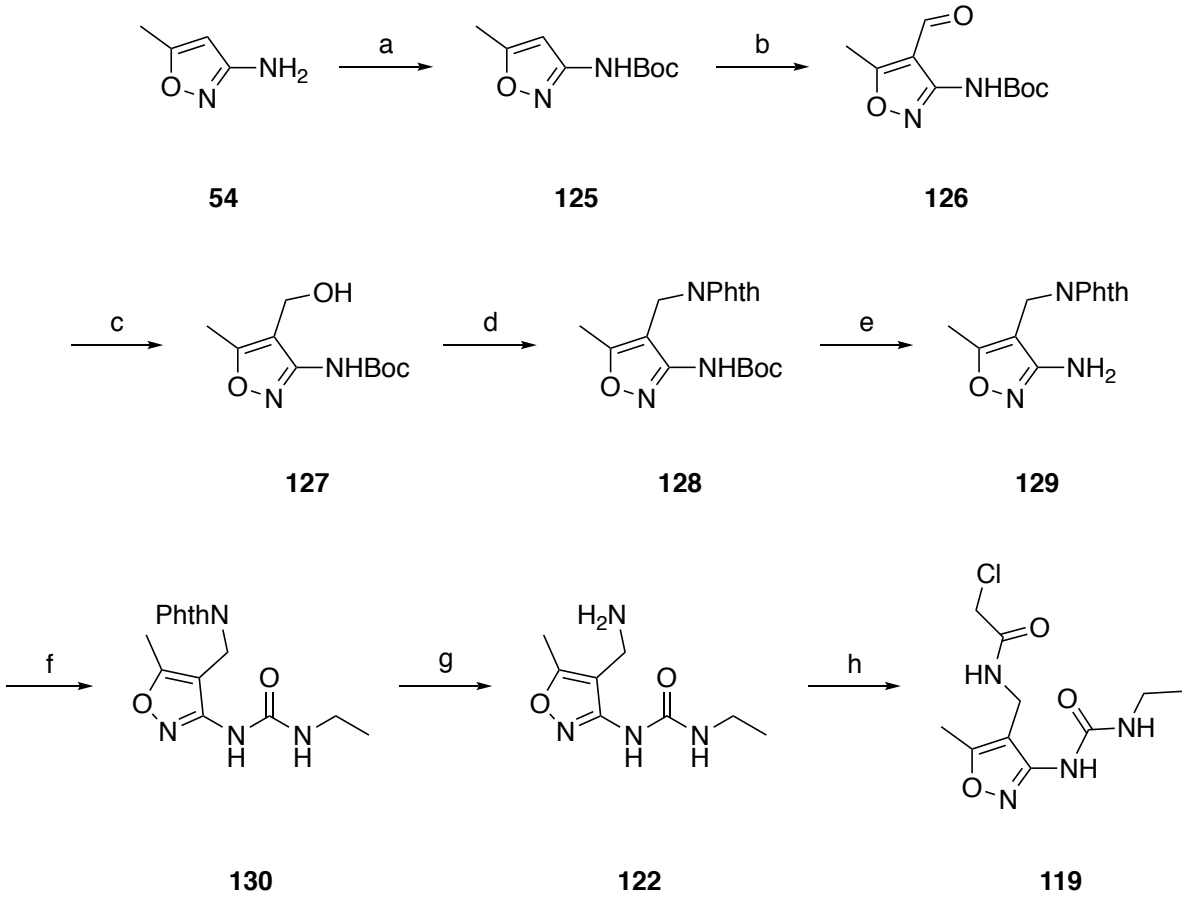
or ethylamine increased the conversion. Changing the solvent instead to DCM significantly improved the reaction, albeit still not to 100% completion. However, attempts to purify by either flash column chromatography or preparative HPLC led to a significant loss of product. Reports in the literature as to the insolubility of ureas in certain solvents prompted trituration of **130** in EtOH. Gratifyingly, the residual starting material was soluble in EtOH, so this method yielded pure product for the final stages of the synthesis.



128 **129** **130**

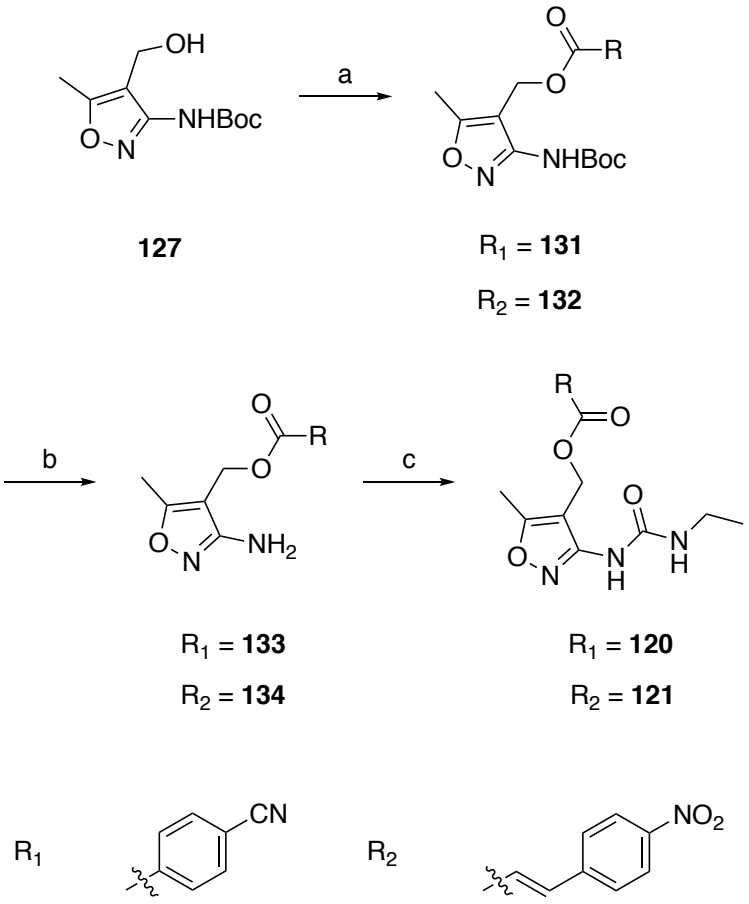
Scheme 26. Generation of the urea motif. Reagents and conditions: (a) TFA, DCM, r.t., 2 h, 99%; (b) i) Triphosgene, DCM, 0 °C-r.t., 0.5 h; (ii) ethylamine, 2 h, 35%.

The subsequent Gabriel-type reaction using hydrazine to convert the phthalimide moiety into the primary amine **122** was low yielding resulting from losses during purification due to the presence of the free amine. The amount of product generated was improved by triturating out the phthalazine side product in MeOH and taking the crude product onto the acylation reaction to avoid significant product loss during purification. Simple acylation conditions using 2-chloroacetyl chloride yielded **119** in reasonable yield. The overall synthetic procedure is shown in Scheme 27.



Scheme 27. Final synthesis of **119**. Reagents and conditions: (a) Boc anhydride, DMAP, TEA, DCM, r.t., 16 h, 85%; (b) i) *n*-BuLi, THF, -78 °C-r.t., 0.5 h; ii) DMF, -78 °C-r.t., 16 h, 70%; (c) $NaBH_4$, MeOH, r.t., 1 h, 50%; (d) HNPhth, DPPE, DtBAD, THF, r.t., 16 h, 35%; (e) TFA, DCM, r.t., 2 h, 99%; (f) i) Triphosgene, DCM 0 °C-r.t., 0.5 h; (ii) ethylamine 2 h, 35%; (g) H_2NNH_2, EtOH, r.t., 16 h; (h) 2-chloroacetyl chloride, TEA, DCM, DMF, r.t. 2 h, 71% (over two steps).

The two analogues of **119**, the ester **120** and enoate ester **121,** were synthesised from the common intermediate **127**. The purchased carboxylic acids were coupled with **127** using the same Mitsunobu conditions with DPPE as previously described. Unfortunately, the products co-eluted during flash column chromatography with the carboxylic acid precursors which reduced yield and purity. However, alternative esterification using EDCI as a coupling partner gave reasonable yields (Scheme 28). Subsequent deprotection of the Boc group and installation of the urea moiety using triphosgene and ethylamine proceeded similarly to **119**.



Scheme 28. Synthesis of **119** analogues **120** and **121**. Reagents and conditions: (a) 4-cyanobenzoic acid or 3-(4-nitrophenyl)prop-2-enoic acid, EDCI, DMAP, DMF, 0 °C-r.t., 16 h, 50-70%. (b) TFA, DCM, r.t., 1-2 h, 91-100%. (c) Triphosgene, ethylamine, TEA, THF, DMA, 0 °C-r.t., 3 h, 8-37%.

5.4.2 Potency Determination of the Merged Covalent Urea Analogues

Attempts to co-crystallise **119** with Kalirin/Rac1 were ultimately unsuccessful, with no labelling observed by mass spectrometry. Subsequent testing in the nucleotide exchange assay showed that none of the analogues inhibited the reaction (Figure 81). No significant inhibition was observed when the assay was repeated over 24 hours (Appendix 8.13). It does appear as though **121** weakly binds, although further mass spectrometry labelling experiments would confirm this hypothesis. It is unknown whether the urea has reduced potency compared to the scaffold of **112**, or if the placement of the handle is sub-optimal for these compounds to covalently bind and interact with D118.

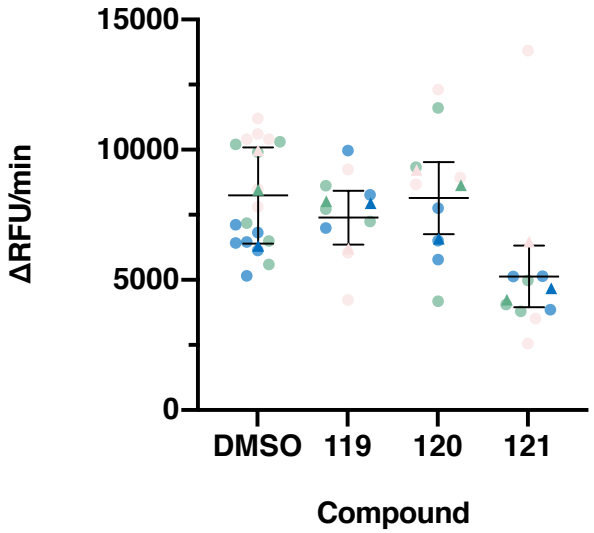


Figure 81. Nucleotide exchange assay for covalent urea analogues tested at 500 μM (n = 3, *** p < 0.001, one-way ANOVA followed by Dunnett's multiple comparisons test). Compounds were incubated for 2 hours. Blue, green and pink circles represent three technical replicates within each repeat. Triangles represent the mean for each repeat. Black bars represent the overall mean ± SD. All results were normalised using the positive control (GppNHp). No compound exhibited significant inhibition of GDP/GTP exchange in comparison to the negative control.

5.5 Conclusions

The covalent fragment screens were overall successful in identifying novel chemical matter targeting to Kalirin/Rac1. The first set of compounds, while less promising for probe or therapeutic development, raised interesting questions regarding the typically static nature of crystal structures. Further experiments to test the impact of these compounds on effector binding and whether they can influence the ability of GTP-bound Rac1 to access 'state 1' or 'state 2' are of interest within the CMD.

The identification of **112** was an exciting development for the Kalirin/Rac1 project. The optimisation of its scaffold synthesis, development of analogues, and adaption of the nucleotide exchange assay to measure k_{inact}/K_I are all intended pursuits within the Brennan group.

Despite the identification of two solvent accessible cysteines within Kalirin/Rac1 (section 3.2.5), covalent compounds were only found to bind to C18 within the guanine nucleotide binding site, perhaps indicative of there being few suitable binding pockets on the surface of the GTPase for interactions with compounds apart from the canonical pocket. It is hoped that despite some conservation of this residue across the Ras superfamily, specific interactions within the Kalirin/Rac1 binding pocket may be exploited for selectivity. A potent irreversible inhibitor may have interesting outcomes on the GEF/GTPase activation cycle *in vivo* if it is able to substantially prolong the rate of nucleotide exchange.

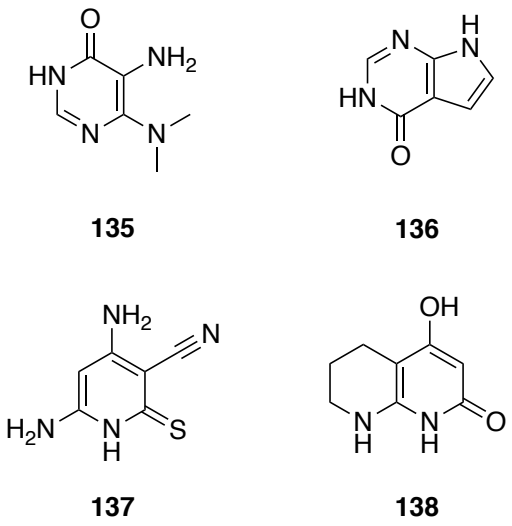
6. Kalirin/Rac1 *In Silico* Pharmacophore Fragment Screening

6.1 Overview

The hit urea **32** identified in Chapter 4 was difficult to optimise. Few commercially available analogues on the Molport website extended into the pocket from the isoxazole or urea motifs. A different strategy for identifying lead compounds was desired. It was decided that further follow-up compounds should be ordered based on a hydrogen bonding pharmacophore model of **32** and GDP, with the inclusion of reactive vectors to elaborate into the pocket. These ligands would be identified using a more stringent *in silico* search of the Enamine Database rather than a manual search of the Molport vendor website. The chapter details the *in silico* experiment conducted by Frank von Delft's group, the subsequent identification of a potent non-covalent compound for Kalirin/Rac1 in the nucleotide exchange assay and subsequent ongoing efforts to produce covalent and non-covalent analogues in Paul Brennan's laboratory.

6.2 Pharmacophore Docking and XChem Screen

The pharmacophore docking was conducted by Rachael Skyner of Frank von Delft's group. The docking approach searched the Enamine Real Database for compounds with high similarity to the pharmacophore model but diverse in structure, based on three sets of starting compounds; the urea hit **32**, GDP or the fragmentation of GDP into its component parts (guanine base, sugar and phosphate groups). The available options for purchase were then docked using the freely available software AutoDock[252] and ranked using two scores, SuCOS[253] and pLIF.[254,255] These scores incorporate the shape similarities and interactions with the protein into the analysis of the docked compound's pose, and are considered better metrics at scoring the poses of bioisosteres than simple RMSD between atom pairs.[253] The top 100 compounds for each subset were then purchased from Enamine and screened in duplicate using the XChem platform at 39 mM concentration. Four hits were identified in the XChem screen (Figure 82, Appendix 8.14).



135 **136**

137 **138**

Figure 82. *In silico* screening and subsequent XChem analysis yielded four hits orientated within the guanine nucleotide binding pocket.

They bound in a similar manner to **32** and the guanine ring of GDP, indicating that the docking approach was successful in maintaining the desired pharmacophore. It also reiterated

how important the interaction of D118 with GDP is for potency within the pocket. Interestingly, **135** bound twice to Rac1, once in the guanine nucleotide binding site, and once in an allosteric pocket formed by P87, F90, T135 and Q141 (Figure 83B). It formed hydrogen bonding interactions with the backbone NH of T135 and amide of Q141.

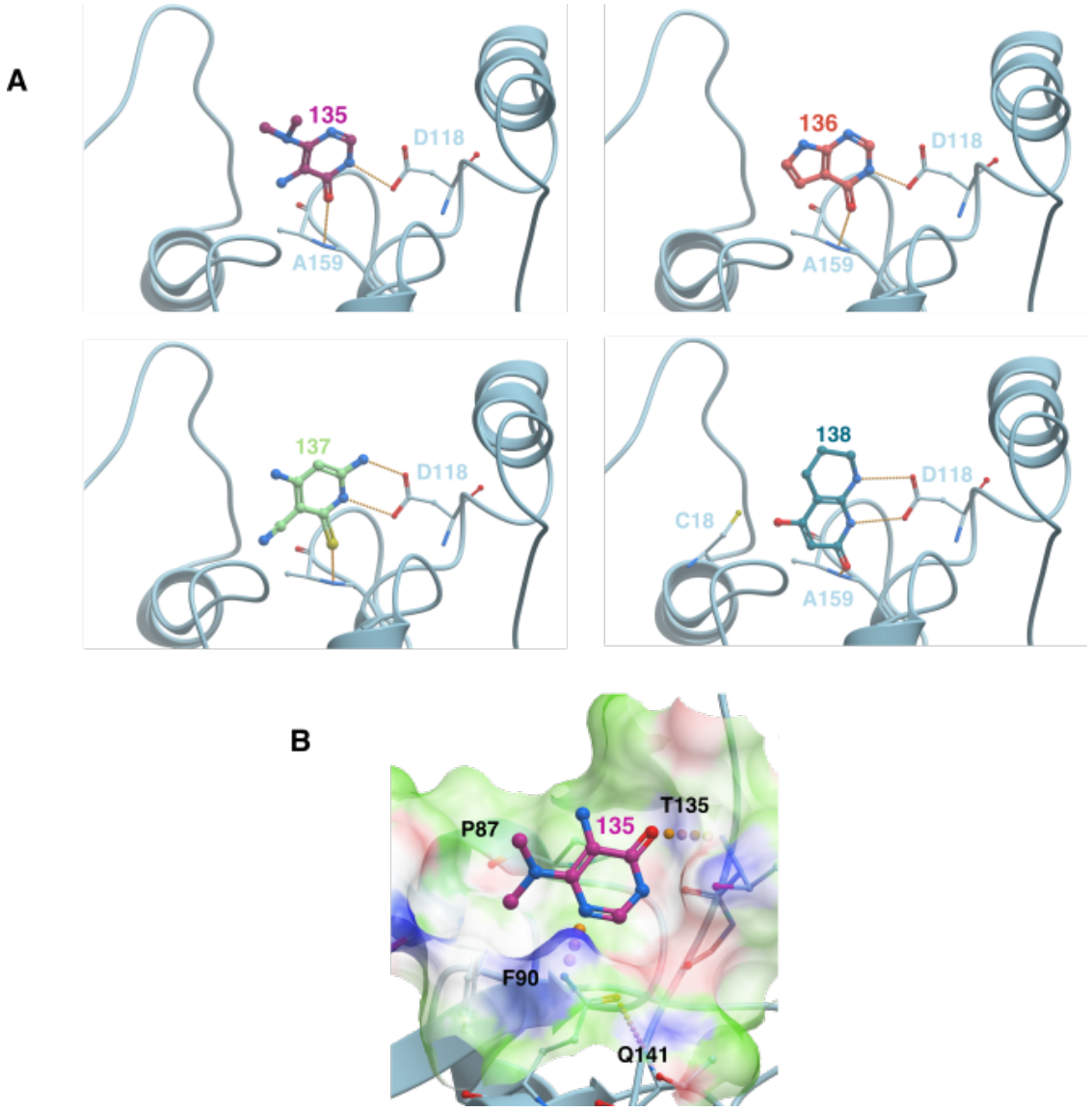


Figure 83. **A**. All four compounds (coloured sticks) from the XChem screen interact with D118 of Rac1 (light blue) with either one or two hydrogen bonds in a similar manner to GDP. **B**. The allosteric binding site for **135**. It forms hydrogen bonds with T135 and Q141.

6.3 Potency Determination of Purchased Pharmacophores

The entire set of compounds was subsequently tested on the nucleotide exchange assay at 500 μM (Appendix 8.15). All compounds were included in the assay due to the potential for false negatives associated with compounds crashing out during XChem. Two ligands, **138** and **139**, significantly inhibited the exchange of GDP for BODIPY-GTP (Figure 84). **139** was not identified in the XChem screen. This could be due to the protonated hydrazine tautomer of **139** being favoured in the low pH crystallisation conditions, which could alter its solubility or ability to form the necessary interactions with D118. **138** was identified in the XChem screen as binding to D118 in a similar manner as **32** and **112.** It was also the first non-covalent compound to show significant inhibition in the exchange assay. As it showed greater potency in the assay, its structure was available for SBDD, and it contained two rather than one hydrogen bond donors, **138** was chosen as the lead compound for further elaboration.

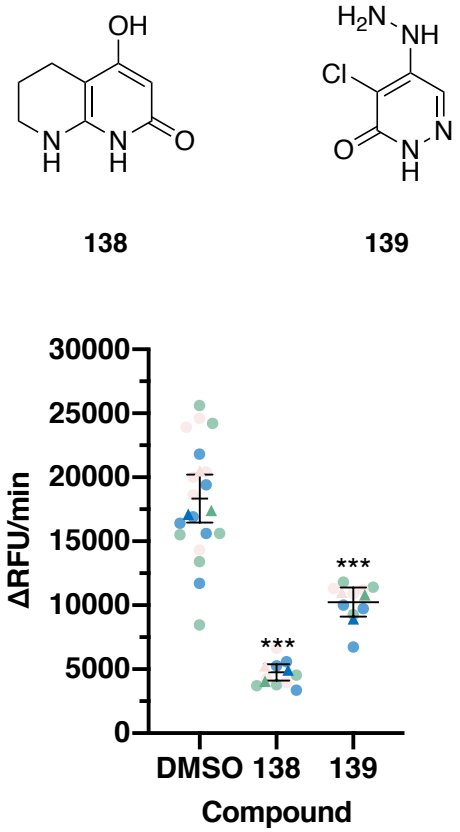


Figure 84. Active pharmacophore analogues **138** and **139** in the nucleotide exchange assay tested at 500 μM (n = 3, *** $p < 0.001$, one-way ANOVA followed by Dunnett's multiple comparisons test). Compounds were incubated for two hours. Blue, green and pink circles represent three technical replicates within each repeat. Triangles represent the mean for each repeat. Black bars represent the overall mean ± SD. All results were normalised using the positive control (GppNHp). **138** and **139** showed highly significant inhibition compared to the negative control.

It was interesting that none of the other compounds observed in the XChem screen were active in the assay. Whilst **135** and **136** were only predicted to form a single hydrogen bond with D118, and thus would likely be less potent, **137** and **138** both formed the hydrogen bond acceptor-donor pattern observed in structures of **112** and urea **32**. It appears that whilst the formation of two hydrogen bonds is a prerequisite for compound activity in the nucleotide exchange reaction, it is not the only factor affecting potency; the electronics in the scaffolds of **32** and **137** must be relatively unfavourable compared to **112** and **138,** leading to weaker

bonding interactions. It could be that a bicyclic ring is required for activity, even if a single ring is capable of forming two hydrogen bonding interactions. However, this is a less promising hypothesis as **139** showed activity in the assay despite not being identified in the XChem screen. It is also unlikely based on its structure to form two hydrogen bonds with D118.

6.4 Synthesis of Analogues of 138

The phenolic oxygen of **138** provided a suitable vector to extend into the guanine nucleotide binding pocket to generate larger, non-covalent analogues. As can be observed in Figure 83, the phenolic OH is also orientated close to C18 within the binding site and could be adapted into a covalent handle. A chloroacetamide warhead was chosen for direct comparison with **112**. The amine starting material was not commercially available and had to be synthesised. The retrosynthetic analysis to generate **140** from an aliphatic amine and malononitrile is shown in Figure 85.

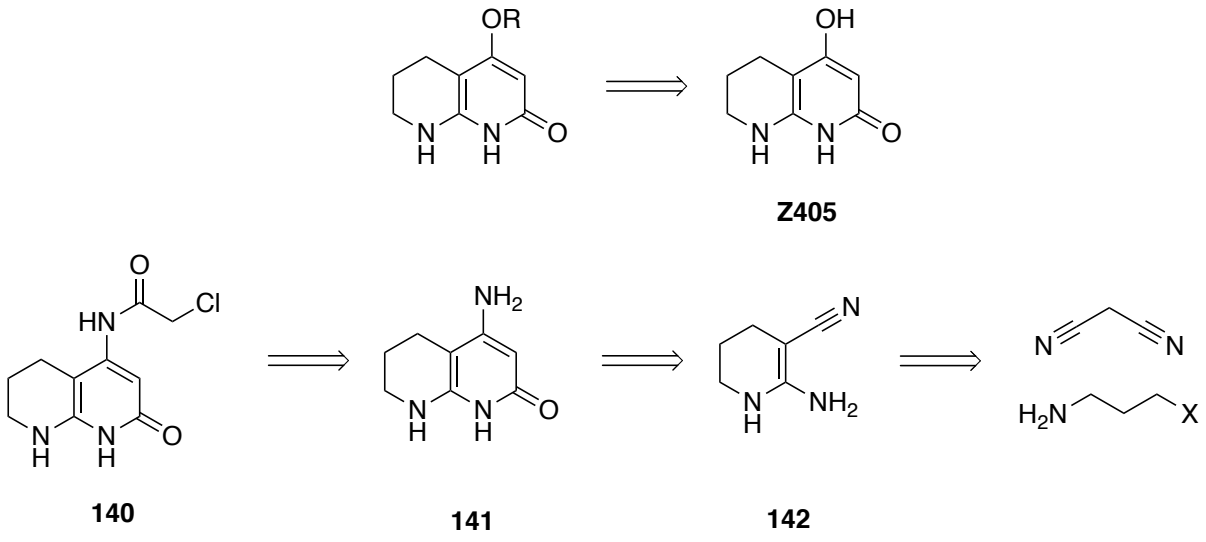
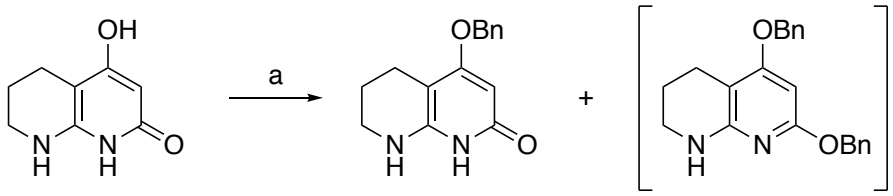


Figure 85. Retrosynthetic analysis for the generation of covalent and non-covalent analogues of **138**.

6.4.1 Synthesis of Non-Covalent Analogues

The first analogue to be generated was a simple benzyl ether to explore the reactivity of the scaffold and whether substitution of the phenolic oxygen would be tolerated within the pocket. Installation of the benzyl group was first attempted using benzyl bromide in an S_N2 reaction. However, despite only 1.1 equivalents of BnBr, the disubstituted product **144** was preferentially formed, resulting in a mixture of starting material and doubly substituted product (Scheme 29). This was consistent with *O*-alkylation of quinolones using benzyl bromides reported in literature.[256] Similar reports indicated that a single alkylation may be possible using the less electrophilic BnCl. The use of minimal BnCl was successful for producing the desired product **143** in low yield, with some disubstituted product still present in the reaction mixture observed by LCMS.

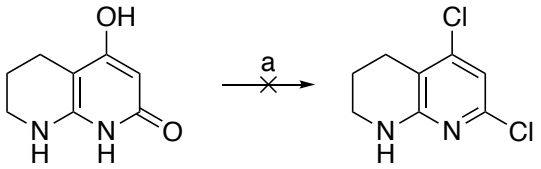


138 **143** **144**

Scheme 29. Alkylation of the lead fragment **138** in an attempt to optimise the synthesis of non-covalent analogues. Regents and conditions: (a) BnCl, K_2CO_3, DMF, 90 °C, 3 h, 4% (**143**).

These initial experiments indicated that optimisation of the alkylation reaction, such as protecting the starting material to avoid overalkylation in the 2-*O* position, was required before expanding the synthesis to an entire library. Subsequent attempts to convert the alcohol into a chlorine as in **145** for coupling reactions using literature protocols for 4-aminoquinoline were also unsuccessful, with a range of side products observed by LCMS (Scheme 30).[257] **143** is

still to be tested in the nucleotide exchange assay. The synthesis of further analogues is ongoing in the Brennan group.



138 **145**

Scheme 30. Unsuccessful chlorination of **138** for subsequent coupling reactions. Reagents and conditions: (a) $POCl_3$, 100 °C, 2 h.

6.4.2 Synthesis of the Covalent Analogue 140

6.4.2.1 Initial Synthetic Strategy

The first synthetic strategy for generating **140** is shown in Figure 86. This route began with the benzyl-protected β-alanine **146**. This route was prioritised as the synthesis of the benzyl-deprotected analogue of **149** directly from β-alanine and malononitrile was documented in the literature.[258] There was also precedence of the synthesis of rings similar to **149** *via* a cyclic anhydride intermediate such as **147** or **148**, which could also be attempted if the single step synthesis of **149** was unsuccessful.[259–261] The formation of the pyridone ring in **150** was predicted to proceed in a similar manner to the pyridone of **112**. An additional step would then be required to reduce the ketone in the piperidone ring, with the use of BH_3·THF in the literature documented for similar scaffolds.[262] The final benzyl deprotection and acylation were predicted to be relatively simple.

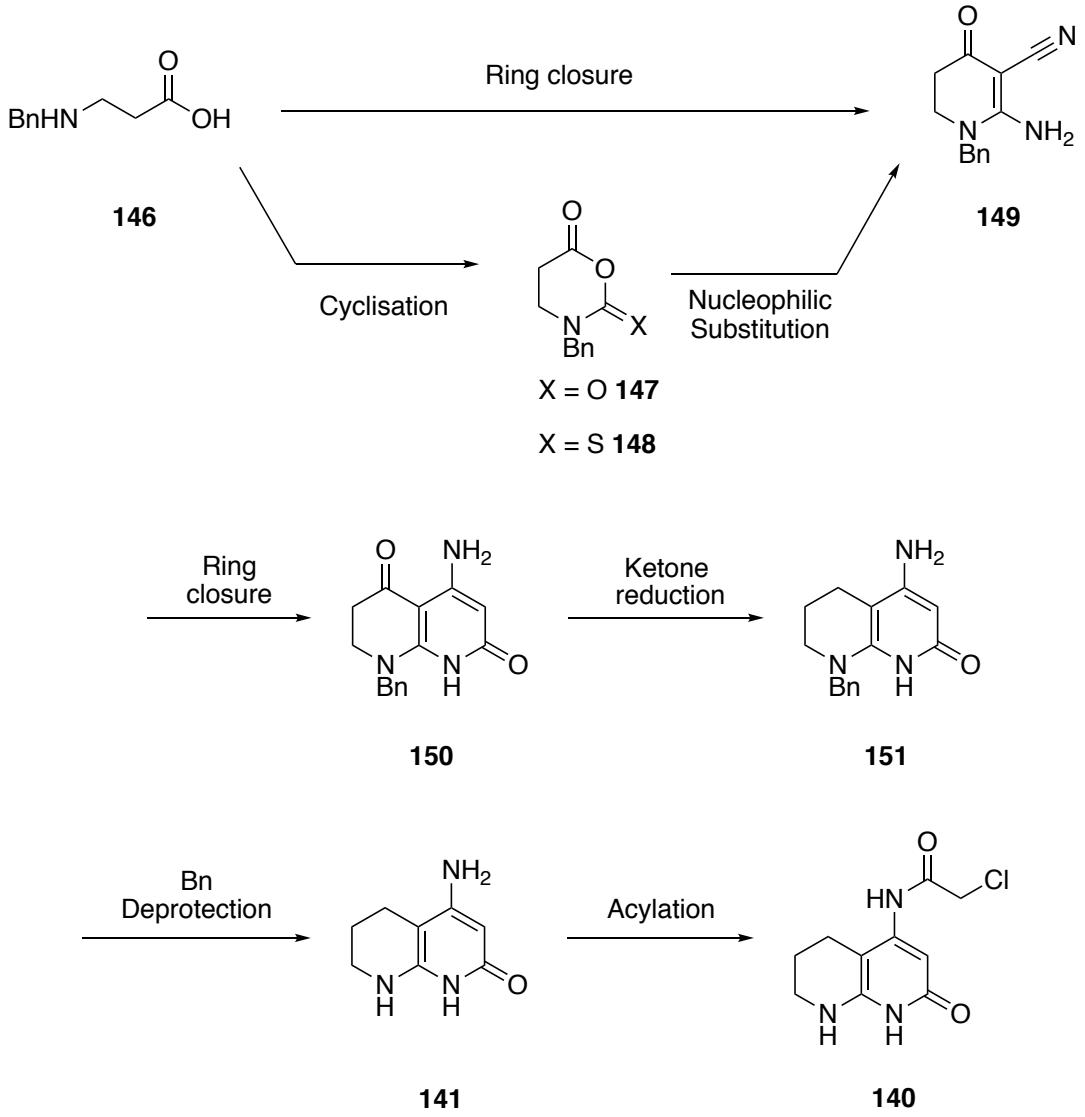
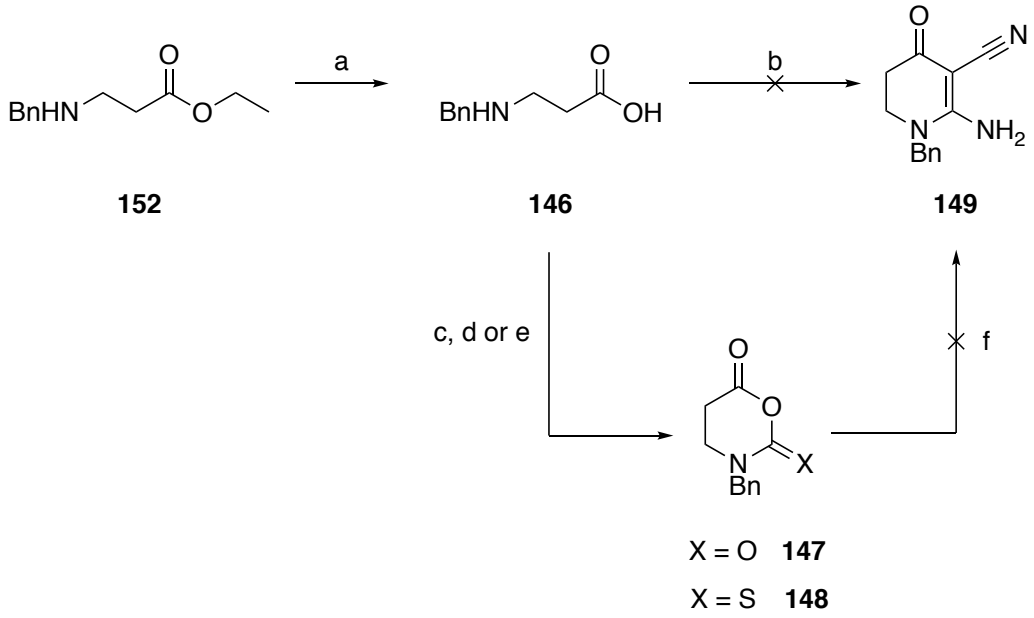


Figure 86. First proposed synthesis of **140** from **146**.

The carboxylic acid **146** was easily obtained via the hydrolysis of the commercially available N-benzyl protected ethyl ester (**152**, Scheme 31). The first attempt at cyclisation used literature conditions which was reported to generate the benzyl-deprotected analogue of **149** in a single step by refluxing the debenzylated version of **146** and malononitrile in acetic anhydride.[258] This reaction was unsuccessful.



Reaction	X	Conditions	Results
c	O	CDI, DIPEA, DMF, r.t., 16 h	No product[a]
d	O	Triphosgene, THF, reflux, 3 h	**147**[b]
e	S	TMSCl, TEA, DCM, DMF, r.t., 1.5 h *then* thiophosgene, r.t., 2 h,	12% yield of **148**

[a]Determined by LCMS [b]Product used directly without purification

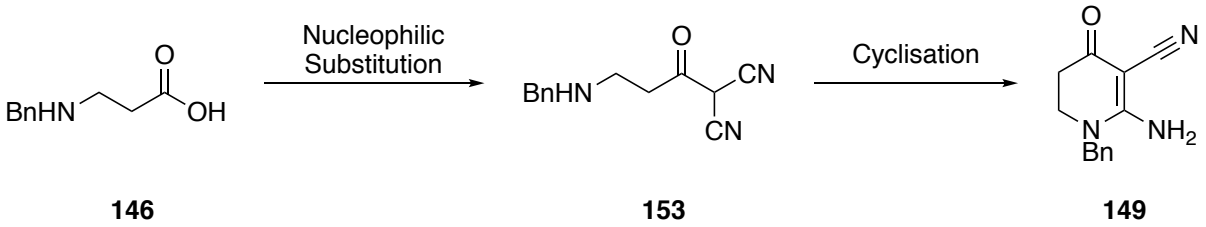
Scheme 31. Unsuccessful attempts to generate the piperidone ring from **146**. Reagents and conditions: (a) NaOH, MeOH, r.t., 16 h, 89%; (b) Malononitrile, Ac_2O, reflux, 16 h; (c) See table; (d) See table; (e) See table; (f) For X = O; Malononitrile, Py, reflux, 16 h.

The alternative proposed route of synthesising **149** in two stages via a cyclic anhydride intermediate was thus attempted. The synthesis of **147** was first pursued by activating the amide for intramolecular attack by the carboxylic acid using CDI in a one-pot process (Scheme 31). This reaction was unsuccessful when using **146** as a starting material due the generation of intermolecular side products.

It was thought that CDI was not a strong enough activating group to catalyse the intramolecular cyclisation. Hence, **146** was reacted with triphosgene to form the activated carbamate, with the carboxylic acid reacting *in situ* to form the cyclised product (Scheme 31).

Whilst this was successful in generating **147**, it was the minor product and purification was challenging as it co-eluted with a major unidentified side product. This resulted in low yields and suggested this reaction may not be suitable on for large-scale production at the beginning of a multi-step synthesis. Nonetheless, the crude **147** was refluxed with malononitrile, but no reaction occurred. It was proposed that the amide moiety of the molecule was not sufficiently reactive to be substituted by malononitrile. The analogous thioamide-anhydride ring **148** was synthesised using thiophosgene; the sulphur was predicted to be a better leaving group than oxygen, thereby increasing the reactivity at this position (Scheme 31). However, **148** was synthesised in poor yields and the reaction was not consistently reproducible, and so this method of forming the piperidone was deemed unsuitable for large-scale synthesis.

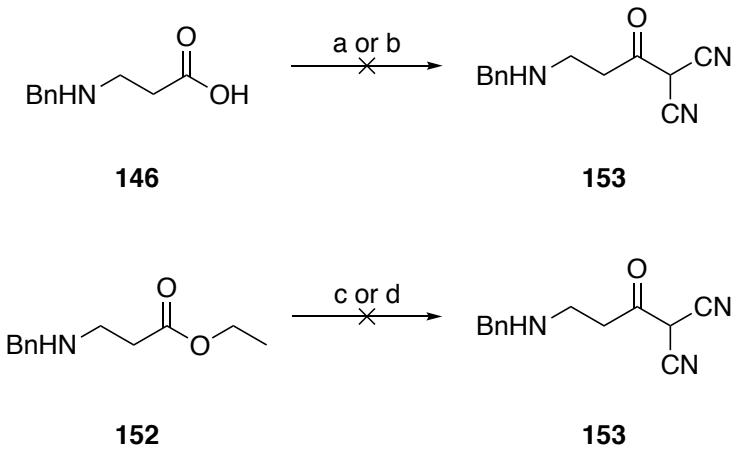
As the installation of malononitrile was the problematic step, it was decided to modify the synthesis and introduce the malononitrile in a substitution reaction, generating intermediate **153**, prior to cyclisation to form **149**, as shown in Figure 87.



146 **153** **149**

Figure 87. Alternative proposed synthesis of **149** *via* aliphatic intermediate **153**.

However, attempts to activate the carboxylic acid for nucleophilic substitution using *N*-hydroxysuccinamide (NHS) or isobutyl chloroformate were unsuccessful, as was direct installation of malononitrile from the ester using piperidine or DABAL-Me3 (Scheme 32). Due to the challenges encountered in introducing malononitrile into this scaffold, as well as the carboxylic acid moiety necessitating the reduction of the resultant piperidone as an additional step, this synthetic route was abandoned.



Scheme 32. Nucleophilic substitution attempts to install the malononitrile at the carbonyl centre. Reagents and conditions: (a) i) Isobutyl chloroformate, TEA, DCM, 0 °C, 1 h; malononitrile, r.t., 16 h; (b) i) NHS-ester, EDCI, DMF r.t., 16 h; ii) Malononitrile; (c) Malononitrile, piperidine, EtOH, reflux, 16 h; (d) Malononitrile, DABAL-M3, THF, reflux, 16 h.

6.4.2.2 An Alternative Route to the Piperidine Ring

Following the lack of success with the initially proposed route, a simpler synthetic procedure was advanced (Figure 88). This route generated the piperidine ring in an S_N2 reaction between malononitrile and an aliphatic protected amine with a suitable leaving group, instead of nucleophilic substitution at a carbonyl centre. This also meant that the route could be shortened by a step, as the change in mechanism meant there was no ketone – and hence no subsequent ketone reduction – present in the intermediates. The second ring closure, benzyl deprotection and final acylation were the same as the previous synthetic route. This route was not originally proposed as there was no literature precedent for the generation of the piperidine ring with malononitrile in an S_N2 reaction.

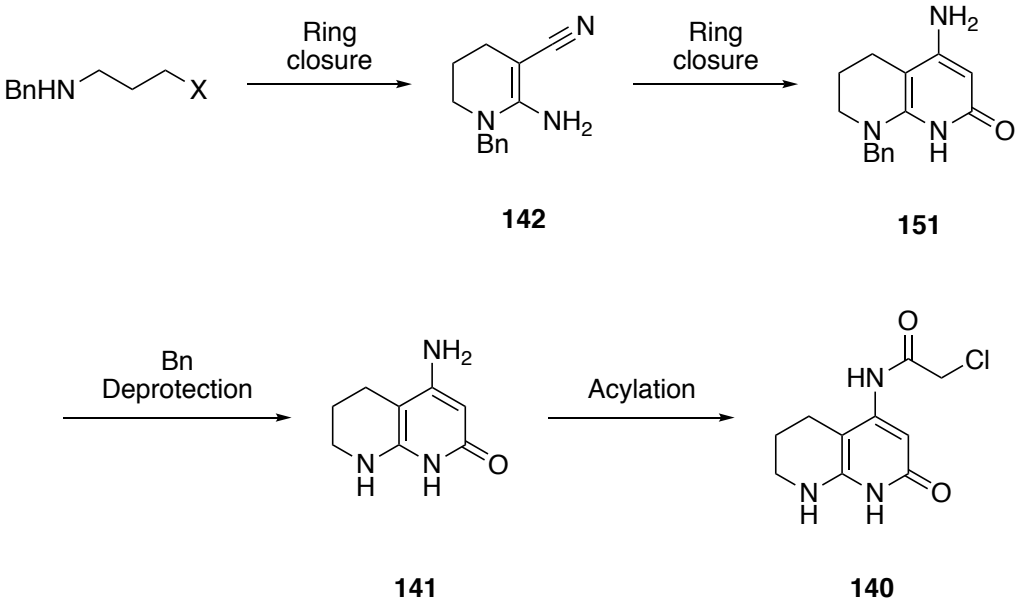
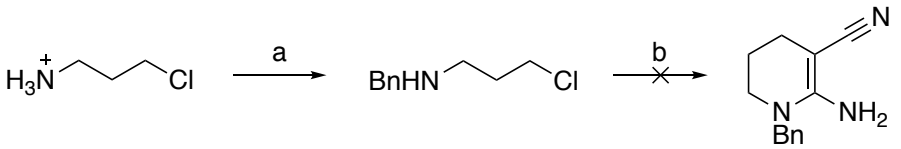
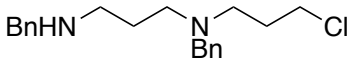


Figure 88. Second synthetic route for the generation of **140**

The commercially available 3-chloropropylamine hydrochloride was chosen as an appropriate starting material, as the installation of the benzyl protecting group via reductive amination from benzaldehyde was well documented in the literature.[263] The reductive amination proceeded with reasonable yields to give **154**, which could be used in S_N2 reaction. Subsequent condensation with malononitrile using a variety of bases either resulted in no reaction or formation of the intermolecular S_N2 product **155** (Scheme 33).



154 **142**



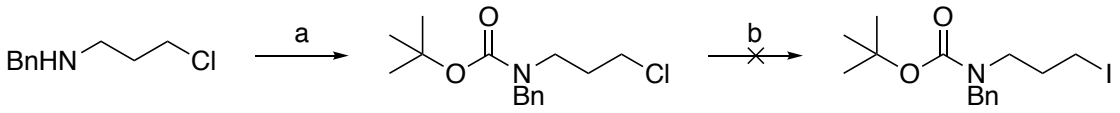
155

Reagents	Solvent	Conditions	Comments
$NCCH_2CN$, K_2CO_3	DMF	90 °C	No product[a]
$NCCH_2CN$, NaHDMS	THF	0 °C - reflux	Intermolecular product **155**[b]
$NCCH_2CN$, KO^{t}Bu	THF	0 °C - reflux	No reaction[a]
$NCCH_2CN$, NaH	THF	0 °C - reflux	Mainly **154**, small amount of **155**[b]
$NCCH_2CN$, Cs_2CO_3	THF	0 °C - reflux	Mainly **154**, small amount of **155**[b]

[a]Determined by LCMS [b] Determined by LCMS and NMR

Scheme 33. Unsuccessful S_N2 reaction using chloride as a leaving group. Reagents and conditions: (a) i) Benzaldehyde, TEA, MeOH, reflux, 16 h; ii) $NaBH_4$, 0 °C-r.t., 2 h, 56%; (b) See Table.

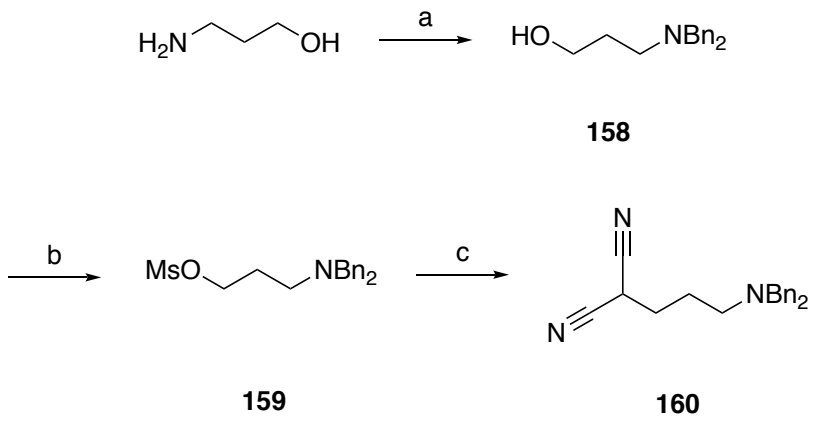
It was hypothesised that the chloride was not a good enough leaving group and conversion into a better one would be necessary, as well as keeping the reaction dilute to avoid intermolecular S_N2 reactions. However, installation of an iodine using Finkelstein's conditions was not successful, with either no reaction or **155** forming despite dilute conditions.[264] From these reactions it was clear that the free NH group was outcompeting the malononitrile in the S_N2 reaction, and that a second protecting group was necessary. A Boc protecting group was installed using Boc anhydride to give **156**, removing the risk of intermolecular S_N2 by-products. However, subsequent substitution was still unsuccessful, even under stringent anhydrous conditions (Scheme 34).



154 **156** **157**

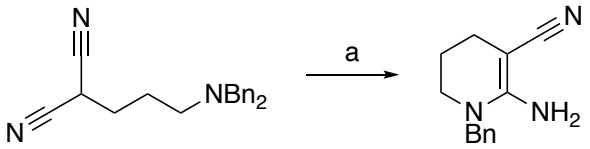
Scheme 34. Unsuccessful attempts to convert the chloride functionality to an iodide. Reagents and conditions: (a) Boc anhydride, DCM, r.t., 16 h, 29%; (b) NaI, anhyd. acetone, 50 °C, 16 h.

It was realised that a better leaving group would have to be installed earlier in the synthesis. It was also decided to use two Bn groups to protect the amine instead of one Bn and one Boc protecting group, as they could be installed in a single reaction. Different benzyl protected compounds were attempted with a variety of leaving groups on the aliphatic chain. Reactions of dibenzylamine with either 3-bromo-1-propanol or 1-bromo-3-chloropropane were not clean, but **158** could be generated in good yield in a double S_N2 reaction using 2 equivalents of BnBr with 3-amino-1-propanol (Scheme 35). Similar reactions in the literature converted **158** to the mesylate analogue to functionalise as a leaving group in S_N2 reactions with amines.[265] These conditions were repeated using malononitrile. Whilst the mesylation occurred quickly to generate **159**, the S_N2 reaction to generate **160** proceeded sluggishly with reduced yields on increased scale. However, this was first successful installation of malononitrile, and so this procedure was used in the synthesis of **140** despite yields of only ~ 30-50% (Scheme 35).

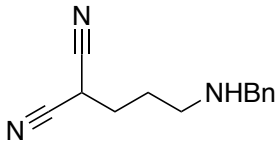


Scheme 35. Successful installation of malononitrile. Reagents and conditions: (a) BnBr, K_2CO_3, anhyd. acetone, r.t., 16 h, 72%; (b) MsCl, DIPEA, DMAP, DCM, 0 °C-r.t., 16h; (c) Malononitrile, DIPEA, MeCN, 60 °C, 16h, 52% (over 2 steps).

In order to form the piperidine ring, the amine would have to be deprotected once so that it could act as a nucleophile in an intramolecular cyclisation reaction with a nitrile. Literature precedent used 1-chloroethyl chloroformate (ACE-Cl) followed by refluxing in MeOH to catalyse a single Bn deprotection when the amine is doubly benzylated.[266] Use of the literature conditions resulted in the formation of a product with the correct mass that could be easily purified. Analysis of the ^{1}H NMR data indicated that the Bn deprotection and cyclisation to create the desired piperidine had happened in a one-pot process, as there was no evidence of a methine proton that would be present if the benzyl deprotected acyclic intermediate **161** was generated (Scheme 36).



160 **142**

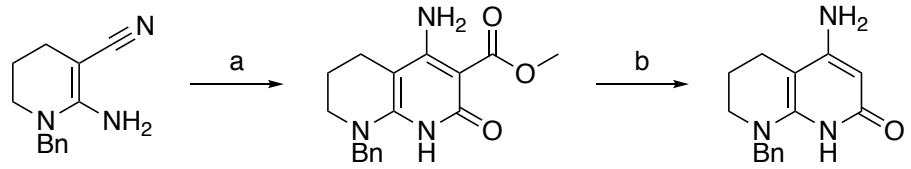


161

Scheme 36. Bn deprotection and cyclisation to form the piperidine ring in a one-pot process. Acyclic intermediate **161** was not observed. Reagents and conditions: (a) i) ACE-Cl, anhyd. DME, reflux, 16 h; ii) MeOH, reflux 16 h, 60%.

6.4.2.3 Formation of the Pyridone Ring

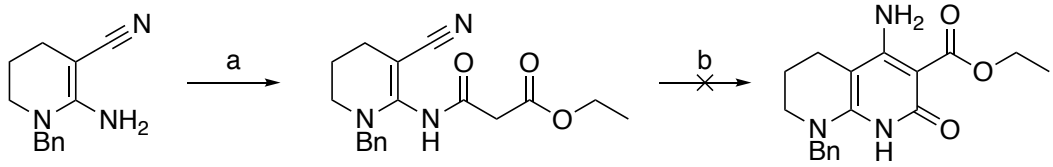
With the piperidine ring successfully synthesised, attention turned to the adjoining ring. The synthesis of **112** by Wuxi achieved the formation of a similar ring system through a malonate condensation and subsequent decarboxylation. An analogous sequence on this scaffold successfully yielded **162** and **151**, although the reaction took significantly longer, did not reach completion and resulted in unwanted side product formation and low purity (Scheme 37). The purification of **151** was also difficult as it did not precipitate on the addition of acid as in the synthesis of other scaffolds, and purification by flash column chromatography for synthesis on a large scale was difficult due to the presence of a free amine. Hence, an alternative robust method was desired.



142 **162** **151**

Scheme 37. Cyclisation of the second ring *via* malonate condensation and subsequent decarboxylation. Reagents and conditions: (a) Dimethyl malonate, NaOMe, MeOH, reflux, 48 h; (b) 15% w.t. NaOH, reflux, 16 h.

It was thought the first stage of the reaction could proceed faster using the more reactive ethyl malonyl chloride. This reaction proceeded quickly forming **163** within two hours. However, the subsequent base-catalysed cyclisation to generate **164** did not occur, despite using a number of strong bases such as NaHMDS, suggesting **163** is relatively unreactive (Scheme 38).

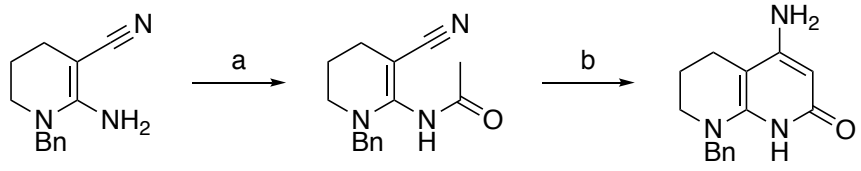


142 **163** **164**

Scheme 38. Alternative condensation reaction to form the pyridone ring. Reagents and conditions: (a) Ethyl malonyl chloride, TEA, DCM, 0 °C-r.t., 1.5 h, 3%; (b) 1 M NaHMDS in THF, 1,4-dioxane, reflux, 16 h.

As the condensation and decarboxylation method was not particularly successful, a different approach was attempted *via* the amide **165** (Scheme 39). A strong base could then be used to form the enolate, catalysing nucleophilic attack into the nitrile to close the ring. This would avoid the need to remove excess groups through hydrolysis and decarboxylation.

Acetylation proceeded smoothly, with **165** generated within one hour and with no side reactions. The subsequent cyclisation was more difficult. No reaction occurred when using NaHMDS, KO^{t}Bu or NaH, but was ultimately successful with LiHMDS acting as the base. With 77% yield across the two steps, this method was deemed adequate to produce the second ring in the synthesis of **140** and further potential analogues.

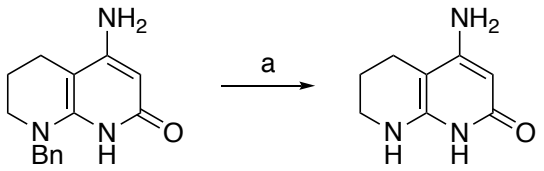


142 **165** **151**

Scheme 39. Successful generation of **151** in high yields. Reagents and conditions: (a) Ac_2O, MeCN, 45 °C, 1 h, 81%; (b) 1 M LiHMDS in THF, THF, reflux, 5 h, 95%.

6.4.2.4 Benzyl Deprotection

With the bicyclic ring synthesised, the final two steps of the proposed synthesis were to remove the remaining Bn protecting group and acylate the resultant free amine. It was originally anticipated based on the synthesis of **112** that the Bn deprotection would be relatively easy using a palladium catalyst. However, the use of identical hydrogenation conditions did not result in any reaction. Subsequent attempts using more active Pd catalysts, such as $Pd(OH)_2$, or the addition of acid were also unsuccessful (Scheme 40). It was hypothesised that the electronics of this ring system were sufficiently different, resulting in the increased stability of the Bn protecting group. Further attempts using transfer hydrogenation instead of gaseous H_2 were ultimately unsuccessful (Scheme 40).[267,268] It was thought that the use of ammonium formate generated **141**, but only trace amounts were visible by LCMS and the yield did not increase despite forcing conditions. Attempts to remove the Bn group earlier in the synthesis on compound **165** by hydrogenation before cyclisation of the second ring were also unsuccessful.



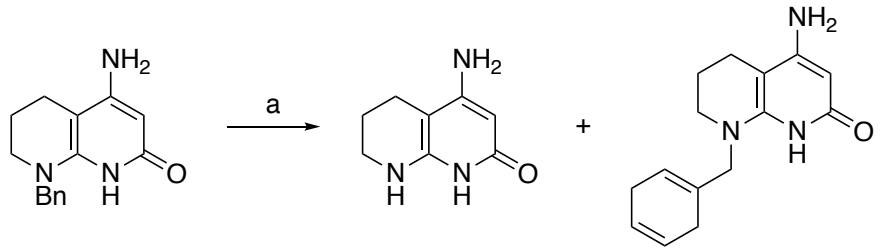
151 **141**

Catalyst	Solvent	Acid	H_2 source	Conditions (a)	Comments
Pd/C	MeOH	None	H_2 balloon	r.t., 16 h	NR
$Pd(OH)_2$	MeOH	None	H_2 balloon	r.t., 16 h	NR
Pd/C	MeOH	HCl	H_2 balloon	r.t., 16 h	NR
Pd/C	MeOH	HCl	H_2 balloon	reflux, 2 h	Unidentified side product
Pd/C	MeOH	None	Ammonium formate	r.t., 16 h	**141** present[a]
Pd/C	EtOH	Acetic	1,4-cyclohexadiene	r.t., 16 h	NR

NR = No reaction [a]Determined by LCMS

Scheme 40. Strategies to deprotect the amine using palladium-catalysed hydrogenation.

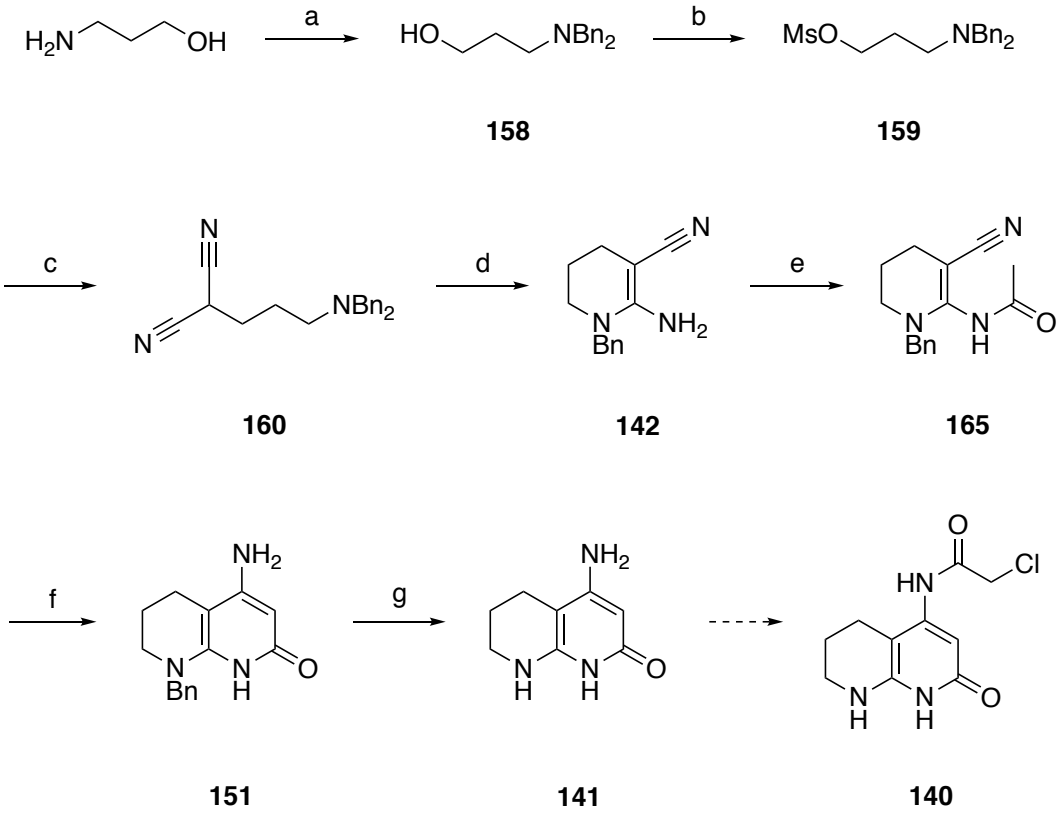
More forcing conditions were required to replace the milder hydrogenation. The Birch reduction has been used in literature for Bn deprotections.[269] This had not been used initially as there were concerns the conditions would also reduce the heterocyclic ring. However, when classic Birch conditions of aqueous ammonia and sodium metal were used, formation of **141** was readily observed, albeit with significant generation of side products (Scheme 41). The major side product **166** contained the benzyl ring reduced to cyclohexadiene whilst still attached to the amine. Purification was also challenging; the high polarity of **141** meant it could not be extracted from the aqueous phase and did not appear to be stable for long periods of time. Attempts at acylating the crude product were unsuccessful.



151 **141** **166**

Scheme 41. The Birch reduction of **151** yielded 141 and proposed major side product **166**. Reagents and conditions: (a) Na, NH_3 (l), -78 °C, 3 h, then r.t., 16 h.

Optimisation of the debenzylation and final synthesis of **140** is still ongoing in Paul Brennan's group. The overall synthetic route achieved thus far is shown in Scheme 42.



Scheme 42. Current synthetic route to generate **140.** Reagents and conditions: (a) BnBr, K_2CO_3, anhyd. acetone, r.t., 16 h, 72%; (b) MsCl, DIPEA, DMAP, DCM, 0 °C-r.t., 16 h; (c) Malononitrile, DIPEA, MeCN, 60 °C, 16 h, 52% (over 2 steps); (d) i) ACE-Cl, DME, reflux, 16 h; ii) MeOH, reflux 16 h, 60% (e) Ac_2O, MeCN, 45 °C, 1 h, 81%; (f) 1 M LiHMDS in THF, THF, reflux, 5 h, 95%; (g) Na, NH_3 (l), -78 °C, 3 h, then r.t., 16 h.

6.5 Conclusions

The discovery of **138** was particularly exciting, as it was the first non-covalent ligand to exhibit appreciable activity in the nucleotide exchange assay. Despite the challenges encountered in the synthesis of its analogues, it remains a promising alternative to **112** in the effort to develop a potent inhibitor for Kalirin/Rac1. Adaptation of the nucleotide exchange assay to identify IC_{50} values is currently underway so quantitative data is available to aid rational synthesis. The other compounds identified in the XChem screen and assay may be investigated in the future but have been deprioritised in favour of the **112** and **138** scaffolds.

Attempts to conclude the synthesis of **140** are ongoing in the Brennan group. Optimisation strategies include using milder Birch conditions, such as the use of crown ethers[270], swapping the synthetic route so that Bn is removed after acylation, or adjusting the synthesis to incorporate a different protecting group. It is hoped that **140** would bind in a similar manner to **112**, forming two hydrogen bonds with D118 and the covalent bond with C18; co-crystallisation experiments following the generation of the compound will provide conclusive data, and aid the rational structure-based and computational design of analogues to extend further into the binding pocket and introduce novel interactions.

7. Conclusions and Future Work

7.1 Summary

This thesis detailed the development of a GEF/GTPase production and crystallisation pipeline, the identification of several novel crystal structures and the development of fragment leads from three different screening methods in an effort to synthesise competitive inhibitors of the small GTPases *via* their regulatory GEF complexes.

The kinetic simulation of reversible and irreversible competitive inhibitors provided support to the hypothesis that the development of compounds that bind to the canonical pocket of the GTPases in the GEF/GTPase complex may provide a novel mechanism of inhibiting these proteins. The GEFs and GTPases were shown to be tractable with numerous constructs cloned, expressed and purified. Extensive optimisation of this process indicated that the GTPases in general were more easily cloned and expressed in *E. coli* than their GEF counterparts. Significant improvement in protein expression and crystallisation was only observed when using a larger range of constructs or if the proteins were expressed with solubility tags. Subsequent crystallisation trials yielded three novel complex structures, with several more being optimised within Frank von Delft's group for future XChem screens.

The apo structure of IQSEC2/Arf1, a member of the Arf subfamily, was determined to 2.35 Å using additive screens to improve the quality of crystals and heavy atom soaking to circumvent phase bias during molecular replacement. However, despite multiple optimisation attempts, the crystals did not reproducibly diffract better than 3 Å, with DSF assays indicating that the complex was not stable under PrepX buffer conditions. This complex was hence deprioritised in order to advance more promising targets.

Two constructs of Kalirin – c001, containing only the GEF DH domain and c002, containing the DH and adjacent PH domain – were crystallised in complex with the GDP-

bound Rho GTPase Rac1. Kalirin is predominantly expressed in the brain and so the Kalirin/Rac1 complex is of interest in neurodegenerative diseases. Despite the low resolution of the Kalirin-c002/Rac1 complex due to the disordered regions of the PH domain, it depicted the same orientation within the guanine nucleotide binding site as the Kalirin-c001/Rac1 complex, suggesting that the conformations of the complex observed in these crystal structures was representative. Kalirin-c001/Rac1 complex crystals could be easily grown with high resolution and so this complex was advanced for ligand screening.

An XChem campaign first identified two novel fragments to bind to Kalirin/Rac1, with one urea hit exhibiting a similar pharmacophore to GDP within the canonical pocket. The development of a GDP-removal assay and subsequent nucleotide-free XChem screen identified the same urea scaffold. The docking and purchase of follow-up analogues yielded ten further ligands that associated with D118 within the guanine nucleotide binding site. Several urea bioisosteres were also synthesised within Paul Brennan's group. A fluorescence intensity assay was developed to measure the ability of compounds to inhibit nucleotide exchange, although none of the urea fragments exhibited significant potency.

Two sets of electrophilic fragments were found to bind to C18 within the guanine nucleotide binding pocket. The first, identified by a mass spectrometry screen, caused significant movement of the switch I loop of Rac1 when co-crystallised within the complex. Whilst these compounds did not show activity in the exchange assay, further analysis of their structures is ongoing, as well as exploration of their potential to inhibit the role of activated Rac1 within the cell by preventing effector recognition. The second subset exhibited a similar pharmacophore to GDP, with optimisation yielding **112** as a potent lead compound for further elaboration.

The final lead compounds were identified using *in silico* docking and subsequent assays, with the compounds exhibiting a similar pharmacophore to previous hits and GDP, but with novel scaffolds. Optimisation of the synthesis of non-covalent and covalent analogues

of potent lead fragment **138** is currently in progress. Extensive synthesis so far has reached the penultimate compound in the synthetic procedure, with optimisation of the final step ongoing in Paul Brennan's group.

As such, the work in this thesis provides novel chemical matter that can be optimised for the synthesis of potent chemical probes or therapeutics. These ligands could be used to elucidate the role of Kalirin/Rac1 in disease, as well as investigating the proposed hypothesis that targeting GEF/GTPase complexes with competitive inhibitors could be a successful strategy for targeting the Ras superfamily. It also provides the starting points to expand this project within the CMD (formerly SGC) to generate data on several members of the entire superfamily.

7.2 Future Work

The future work arising out of this thesis can be defined as follows: establishment of Kalirin-9 as a target; optimisation of promising GEF/GTPase complexes for further XChem screens; deeper analysis of the loop rearrangement observed in Kalirin/Rac1 for identification of compounds that may have a different mode of action; elaboration of lead compounds to expand into the guanine nucleotide pocket; optimisation of the fluorescence intensity assay to provide quantitative potency values to lead compounds; and development of *in vivo* disease models involving Kalirin-7 isoforms for testing of compounds.

7.2.1 Addition of Kalirin-9 as an Additional Target

Another isoform of Kalirin, Kalirin-9, contains a second GEF domain which is selective for RhoA. This domain is also of interest in neurodegenerative disease and is to be added to the GEF/GTPase pipeline.

7.2.2 Optimisation of Crystals for Further XChem Screens

Several complexes, highlighted in Chapter 3, were promising in crystal trials, with either needles, microcrystals or spherulites observed in certain conditions, but were ultimately not optimised to diffract better than 3 Å due to time constraints. Ongoing efforts to improve these for subsequent drug discovery or probe development projects, especially for DOCK proteins which are of particular interest in neurodegenerative diseases, is being conducted by Andrew Thompson of the von Delft and Brennan groups.

7.2.3 Optimisation of Electron Density for the Switch I Loop Rearrangement

As detailed in Chapter 5, several covalent compounds from the electrophilic mass spectrometry screen and non-covalent compounds from the XChem screen caused movement in the switch I loop of Rac1 which made the determination of its orientation, and subsequent identification of the fragment binding sites, extremely difficult. Members of Frank von Delft's group are re-running the modelling of these crystal structures using PDB REDO to improve the loop conformation.

7.2.4 Elaboration of Lead Compounds

As detailed in Chapters 4, 5 and 6, there are several leads which can be elaborated to extend into the guanine nucleotide binding pocket for novel interactions, increasing the potency and selectivity of these initial fragment hits. The covalent lead **112** and non-covalent fragment **138** are being prioritised by Mia Callens in the Brennan group due to their activity in the assay and available crystallisation data.

7.2.5 Optimisation of Nucleotide Exchange Assay to Generate Inhibition Values

The FI assay optimised by Carmen Jimenez Antunez allows the relative prioritisation of compounds based on the amount of inhibition of the initial rate of exchange compared to the negative control DMSO. The assay should be adapted to identify either the IC_{50} for non-covalent compounds or k_{inact}/K_I for covalent ligands, as well as other GEF/GTPase complexes within the pipeline.

7.2.6 *In Vivo* Kalirin Models

Whilst the nucleotide exchange assay assists in the development of potent compounds *in vitro*, it is of paramount importance to test inhibitors in suitable cellular or *in vivo* assays to determine their potency, selectivity and toxicity in a physiological setting. A collaboration has been established with the Department of Psychiatry, University of Oxford, for the development of a suitable disease model in zebrafish where any potent *in vitro* inhibitors of the Kalirin/Rac1 complex can be screened.

8. Appendices

8.1 Construct Data within the SGC GEF/GTPase Pipeline

8.1.1 Construct Data for Pipeline Round 1

Target	Vector	Start	End	Cloned?	M_w (kDa)	Ex. Lvl	Xtal trials	Xtals
AKAP13	pNIC28-10His	S1968	D2338	N - JG	43.4	N/A	N/A	N/A
AKAP13	pNIC-MBP2	S1968	D2338	Y - JG	43.4	-	N/A	N/A
Arf1	pNIC-NStIIT	M18	K181	Y - JG	18.9	H	Y	Y
Arf6	pNIC28-10His	M14	Y173	Y - JG	18.5	H	Y	N
ARFGEF1	pNIC28-10His	Q691	E889	Y - JG	23.2	-	N/A	N/A
ARFGEF1	pNIC-MBP2	Q691	E889	Y - JG	23.2	H	Y	N
ARFGEF2	pNIC28-10His	E635	A828	Y - JG	22.7	Y	Y	N/A
ARHGEF11	pNIC28-10His	Q714	A1081	Y - JG	42.9	-	N/A	N/A
ARHGEF12	pNIC28-10His	P766	S1138	Y - JG	43.5	-	N/A	N/A
Cdc42	pNIC-NStIIT	M1	C188	Y - JG	21.0	H	N	N/A
Cyth1	pNIC28-10His	M60	E248	Y - JG	22.0	M	Y	N
Cyth2	pNIC28-10His	S56	K244	Y - JG	22.3	-	N/A	N/A
Cyth2	pNIC-MBP2	S56	K244	Y - JG	22.3	H	Y	N
Cyth3	pNIC28-10His	P247	K399	Y - JG	18.3	-	N/A	N/A
DENND1B	pNIC28-10His	M1	G410	Y - JG	44.2	-	N/A	N/A
DOCK2	pNIC28-10His	M1192	M1622	Y - JG	51.2	-	N/A	N/A
DOCK3	pNIC28-10His	R1204	C1640	N - JG	52.3	N/A	N/A	N/A
DOCK3	pNIC28-10His	R1204	K1635	N - JG	51.9	N/A	N/A	N/A
DOCK3	pNIC28-10His	E1213	C1640	Y - JG	51.3	-	N/A	N/A
DOCK3	pNIC28-10His	E1213	K1635	Y - JG	50.8	-	N/A	N/A
DOCK9	pNIC28-10His	K1605	Q2067	Y - JG	50.0	-	N/A	N/A
EPAC2	pNIC28-10His	P324	P1011	Y - JG	78.8	-	N/A	N/A
HRas	pNIC-NStIIT	M1	H166	Y - JG	18.9	-	N/A	N/A
IQSEC1	pNIC28-10His	S498	P871	Y - JG	43.8	L	N	N/A
IQSEC2	pNIC28-10His	C730	R1103	N - JG	44.1	N/A	N/A	N/A
IQSEC2	pNIC28-10His	R739	R1193	N - JG	43.0	N/A	N/A	N/A
IQSEC2	pNIC28-10His	C730	K1099	Y - JG	43.6	H	Y	Y
IQSEC2	pNIC28-10His	S741	K1099	Y - JG	42.2	H	Y	N
IQSEC2	pNIC28-10His	F747	K1097	Y - JG	41.3	-	N/A	N/A
IQSEC3	pNIC28-10His	Q320	Q707	Y - JG	44.8	-	N/A	N/A
IQSEC3	pNIC28-10His	E332	Q707	Y - JG	43.4	H	Y	Y
IQSEC3	pNIC28-10His	P334	K697	Y - JG	42.1	-	N/A	N/A
IQSEC3	pNIC28-10His	Q320	K697	Y - JG	43.8	-	N/A	N/A
IQSEC3	pNIC-MBP2	Q320	K697	Y - JG	43.8	L	N	N
ITSN1	pNIC28-10His	D1229	R1581	N - JG	41.1	N/A	N/A	N/A
Kalirin	pNIC28-10His	R1253	V1432	Y - JG	21.2	H	Y	Y
Kalirin	pNIC28-Bsa4	D1259	L1590	Y - JG	38.8	H	Y	Y

Target	Vector	Start	End	Cloned?	M_w (kDa)	Ex. Lvl	Xtal trials	Xtals
Kalirin	pNIC28-Bsa4	D1259	K1587	Y - JG	38.6	H	N	N/A
Kalirin	pNIC28-Bsa4	D1259	E1581	Y - JG	37.8	H	N	N/A
Kalirin	pNIC28-Bsa4	D1259	A1562	Y - JG	35.5	-	N/A	N/A
Kalirin	pNIC28-Bsa4	E1273	L1590	Y - JG	37.1	H	N	N/A
Kalirin	pNIC28-Bsa4	E1273	K1587	Y - JG	36.9	-	N/A	N/A
Kalirin	pNIC28-Bsa4	E1273	E1581	Y - JG	36.1	H	N	N/A
Kalirin	pNIC28-Bsa4	E1273	A1562	Y - JG	33.8	-	N/A	N/A
Kalirin	pNIC28-Bsa4	R1280	L1590	Y - JG	36.3	H	N	N/A
Kalirin	pNIC28-Bsa4	R1280	K1587	Y - JG	36.1	H	N	N/A
Kalirin	pNIC28-Bsa4	R1280	E1581	Y - JG	35.3	-	N/A	N/A
Kalirin	pNIC28-Bsa4	R1280	A1562	Y - JG	33.0	-	N/A	N/A
MSS4	pNIC28-10His	V11	E123	Y - JG	12.9	-	N/A	N/A
NET1	pNIC28-10His	K149	A501	N - JG	41.4	N/A	N/A	N/A
P-Rex1	pNIC28-10His	A38	M408	Y - JG	42.9	-	N/A	N/A
Rab8	pNIC28-10His	M1	G183	Y - JG	21.0	L	N	N/A
Rab21	pNIC-NStIIT	R16	T183	Y - JG	19.4	H	N	N/A
Rab35	pNIC-NStIIT	M1	Q180	Y - JG	20.6	M	N	N/A
RABGEF1	pNIC28-10His	S132	E397	Y - JG	31.3	-	N/A	N/A
Rac1	pNIC-NStIIT	M1	L192	Y - JG	21.5	H	Y	N
Rac1	pNIC-NStIIT	M1	K184	Y - JG	20.5	-	N/A	N/A
Rac1	pNIC-NStIIT	M1	L177	Y - JG	19.8	H	Y	Y
RalA	pNIC-NStIIT	G8	D179	Y - JG	19.9	H	Y	N
Ran	pNIC-NStIIT	M1	L216	Y - JG	24.5	-	N/A	N/A
Ran	pNIC-MBP2	M1	L216	Y - JG	24.5	H	Y	N
RAP1B	pNIC-NStIIT	M1	R167	Y - JG	19.1	L	Y	N
RCC1	pNIC28-10His	V56	S452	Y - JG	42.9	-	N/A	N/A
RCC1	pNIC-MBP2	V56	S452	Y - JG	42.9	H	Y	N
RGL2	pNIC28-10His	V55	G514	Y - JG	51.6	-	N/A	N/A
RGL2	pNIC-MBP2	V55	G514	Y - JG	51.6	H	Y	N
RhoA	pNIC-NStIIT	M1	A181	Y - JG	20.5	-	N/A	N/A
RhoA	pNIC-MBP2	M1	A181	Y - JG	20.5	L	N	N/A
Sos1	pNIC28-10His	M567	N1044	Y - JG	56.2	-	N/A	N/A
Tiam1	pNIC28-10His	R1033	E1406	Y - JG	43.4	-	N/A	N/A
Tiam1	pNIC-MBP2	R1033	E1406	Y - JG	43.4	-	N/A	N/A
Trio	pNIC28-10His	E1285	T1594	Y - JG	36.3	-	N/A	N/A
Trio	pNIC-MBP2	E1285	T1594	Y - JG	36.3	-	N/A	N/A
Vav1	pNIC28-Bsa4	G170	M575	Y - JG	47.6	-	N/A	N/A
IQSEC2/Arf6	Bicistronic			Y - JG		H	Y	N

Y - JG = Cloned by Janine Gray
N - JG = Unsuccessful cloning attempt by Janine Gray
Ex. Lvl = Expression Level.
- = No expression
L = Low expression
M = Medium expression
H = High expression

Xtal trials: ‘Y’ means complex was subjected to crystallisation trials. ‘N’ means protein was not attempted to be crystallised
Xtals: ‘Y’ means potential crystalline conditions observed. ‘N’ means no crystal-like conditions observed

8.1.2 Construct Data for Pipeline Round 2

Target	Vector	Start	End	Cloned?	M_w (kDa)	Ex. Lvl	Xtal trials	Xtals
ARHGEF15	pNIC-NStIIT	S412	E601	Y - BT	22.2	-	N/A	N/A
ARHGEF15	pNIC-NStIIT	R417	E601	Y - BT	21.6	-	N/A	N/A
ARHGEF15	pNIC-NStIIT	S421	E601	Y - BT	21.0	-	N/A	N/A
ARHGEF15	pNIC-NStIIT	S412	K606	Y - BT	22.8	-	N/A	N/A
ARHGEF15	pNIC-NStIIT	R417	K606	Y - BT	22.2	L	N	N/A
ARHGEF15	pNIC-NStIIT	S421	K606	N - BT	21.6	N/A	N/A	N/A
ARHGEF15	pNIC-NStIIT	S412	Q616	Y - BT	24.0	H	N	N/A
ARHGEF15	pNIC-NStIIT	R417	Q616	Y - BT	23.4	-	N/A	N/A
ARHGEF15	pNIC-NStIIT	S421	Q616	Y - BT	22.8	-	N/A	N/A
ARHGEF15	pNIC-NStIIT	S412	E582	Y - BT	20.2	-	N/A	N/A
ARHGEF15	pNIC-NStIIT	R417	E582	Y - BT	19.6	-	N/A	N/A
DEF6	pNIC-NStIIT	M1	N631	Y - BT	74.0	H	Y	N
DEF6	pNIC-NStIIT	M1	Q604	Y - BT	71.3	H	N	N/A
DEF6	pNIC-NStIIT	M1	G592	Y - BT	69.9	H	N	N/A
DEF6	pNIC-NStIIT	M1	P556	Y - BT	66.0	H	N	N/A
DEF6	pNIC-NStIIT	L72	N631	Y - BT	66.0	-	N/A	N/A
DEF6	pNIC-NStIIT	L72	Q604	Y - BT	63.3	-	N/A	N/A
DEF6	pNIC-NStIIT	L72	G592	Y - BT	61.9	-	N/A	N/A
DEF6	pNIC-NStIIT	L72	P556	Y - BT	58.0	-	N/A	N/A
DEF6	pNIC-NStIIT	S203	N631	Y - BT	50.9	H	N	N/A
DEF6	pNIC-NStIIT	S203	Q604	Y - BT	48.1	-	N/A	N/A
DEF6	pNIC-NStIIT	S203	G592	Y - BT	46.8	-	N/A	N/A
DEF6	pNIC-NStIIT	S203	P556	Y - BT	42.8	H	N	N/A
DEF6	pNIC-NStIIT	D216	N631	Y - BT	49.5	H	N	N/A
DEF6	pNIC-NStIIT	D216	Q604	Y - BT	46.8	H	N	N/A
DEF6	pNIC-NStIIT	D216	G592	Y - BT	45.5	-	N/A	N/A
DEF6	pNIC-NStIIT	D216	P556	Y - BT	41.5	H	N	N/A
DEF6	pNIC-NStIIT	K219	N631	Y - BT	49.2	H	N	N/A
DEF6	pNIC-NStIIT	K219	Q604	Y - BT	46.5	M	N	N/A
DEF6	pNIC-NStIIT	K219	G592	Y - BT	45.1	-	N/A	N/A
DEF6	pNIC-NStIIT	K219	P556	Y - BT	41.2	H	N	N/A
DEF6	pNIC-NStIIT	D296	N631	Y - BT	40.1	-	N/A	N/A
DEF6	pNIC-NStIIT	D296	Q604	Y - BT	37.4	-	N/A	N/A
DEF6	pNIC-NStIIT	D296	G592	Y - BT	36.0	-	N/A	N/A
DEF6	pNIC-NStIIT	D296	P556	Y - BT	32.1	-	N/A	N/A
DEF6	pNIC-NStIIT	R312	N631	Y - BT	38.2	H	N	N/A
DEF6	pNIC-NStIIT	R312	Q604	Y - BT	35.5	H	N	N/A
DEF6	pNIC-NStIIT	R312	G592	Y - BT	34.1	H	N	N/A
DEF6	pNIC-NStIIT	R312	P556	Y - BT	30.2	H	N	N/A
DEF6	pNIC-NStIIT	E316	N631	Y - BT	37.7	M	N	N/A

Target	Vector	Start	End	Cloned?	M_w (kDa)	Ex. Lvl	Xtal trials	Xtals
DEF6	pNIC-NStIIT	E316	Q604	Y - BT	35.0	M	N	N/A
DEF6	pNIC-NStIIT	E316	G592	Y - BT	33.6	M	N	N/A
DEF6	pNIC-NStIIT	E316	P556	Y - BT	29.7	H	N	N/A
DEF6	pNIC-NStIIT	D324	N631	Y - BT	36.9	H	N	N/A
DEF6	pNIC-NStIIT	D324	Q604	Y - BT	34.1	H	N	N/A
DEF6	pNIC-NStIIT	D324	G592	Y - BT	32.8	H	N	N/A
DEF6	pNIC-NStIIT	D324	P556	Y - BT	28.8	H	Y	Y
DOCK11	pNIC-NStIIT	A1588	L2050	Y - BT	53.9	H	N	N/A
DOCK11	pNIC-NStIIT	M1590	L2050	N - BT	53.6	N/A	N/A	N/A
DOCK11	pNIC-NStIIT	M1599	L2050	Y - BT	52.5	H	N	N/A
DOCK11	pNIC-NStIIT	A1588	T2059	Y - BT	54.9	H	N	N/A
DOCK11	pNIC-NStIIT	M1590	T2059	Y - BT	54.5	H	N	N/A
DOCK11	pNIC-NStIIT	M1599	T2059	Y - BT	53.4	M	N	N/A
DOCK11	pNIC-NStIIT	E1594	L2050	Y - BT	53.2	H	N	N/A
DOCK11	pNIC-NStIIT	E1594	T2059	Y - BT	54.1	H	N	N/A
DOCK2	pNIC-NStIIT	L1185	F1625	Y - BT	52.5	-	N/A	N/A
DOCK2	pNIC-NStIIT	S1196	F1625	Y - BT	51.2	M	Y	N
DOCK2	pNIC-NStIIT	E1217	F1625	Y - BT	48.7	M	N	N/A
DOCK2	pNIC-NStIIT	L1185	M1622	Y - BT	52.2	L	N	N/A
DOCK2	pNIC-NStIIT	S1196	M1622	Y - BT	50.9	L	N	N/A
DOCK2	pNIC-NStIIT	E1217	M1622	Y - BT	48.4	M	N	N/A
DOCK2	pNIC-NStIIT	L1185	E1616	N - BT	51.4	N/A	N/A	N/A
DOCK2	pNIC-NStIIT	S1196	E1616	Y - BT	50.1	-	N/A	N/A
DOCK2	pNIC-NStIIT	E1217	E1616	Y - BT	47.6	-	N/A	N/A
DOCK3	pNIC-NStIIT	I1217	T1645	Y - BT	51.2	M	N	N/A
DOCK3	pNIC-NStIIT	M1224	T1645	Y - BT	50.4	L	N	N/A
DOCK3	pNIC-NStIIT	K1228	T1645	Y - BT	50.0	L	N	N/A
DOCK3	pNIC-NStIIT	I1217	C1640	Y - BT	50.8	H	N	N/A
DOCK3	pNIC-NStIIT	M1224	C1640	Y - BT	50.0	H	N	N/A
DOCK3	pNIC-NStIIT	K1228	C1640	Y - BT	49.5	L	N	N/A
DOCK3	pNIC-NStIIT	I1217	K1635	Y - BT	50.3	M	N	N/A
DOCK3	pNIC-NStIIT	M1224	K1635	N - BT	49.5	N/A	N/A	N/A
DOCK3	pNIC-NStIIT	K1228	K1635	Y - BT	49.1	L	N	N/A
DOCK6	pNIC-NStIIT	E1570	S2041	N - BT	54.3	N/A	N/A	N/A
DOCK6	pNIC-NStIIT	E1576	S2041	Y - BT	53.6	-	N/A	N/A
DOCK6	pNIC-NStIIT	R1587	S2041	Y - BT	52.2	H	N	N/A
DOCK6	pNIC-NStIIT	D1594	S2041	Y - BT	51.5	H	N	N/A
DOCK6	pNIC-NStIIT	E1570	Q2024	Y - BT	52.5	H	N	N/A
DOCK6	pNIC-NStIIT	E1576	Q2024	Y - BT	51.8	H	N	N/A
DOCK6	pNIC-NStIIT	R1587	Q2024	Y - BT	50.5	H	Y	Y
DOCK6	pNIC-NStIIT	D1594	Q2024	Y - BT	49.7	H	N	N/A
DOCK7	pNIC-NStIIT	D1634	A2088	Y - BT	52.6	H	N	N/A
DOCK7	pNIC-NStIIT	K1647	A2088	Y - BT	51.0	H	Y	N

Target	Vector	Start	End	Cloned?	M_w (kDa)	Ex. Lvl	Xtal trials	Xtals
DOCK7	pNIC-NStIIT	R1656	A2088	N - BT	50.0	N/A	N/A	N/A
DOCK7	pNIC-NStIIT	D1634	Q2084	N - BT	52.1	N/A	N/A	N/A
DOCK7	pNIC-NStIIT	K1647	Q2084	Y - BT	50.5	H	Y	N
DOCK7	pNIC-NStIIT	R1656	Q2084	N - BT	49.5	N/A	N/A	N/A
DOCK7	pNIC-NStIIT	D1634	E2062	Y - BT	49.4	L	N	N/A
DOCK7	pNIC-NStIIT	K1647	E2062	Y - BT	47.8	M	N	N/A
DOCK7	pNIC-NStIIT	R1656	E2062	Y - BT	46.8	M	N	N/A
DOCK8	pNIC-NStIIT	K1564	P2003	Y - BT	51.1	H	N	N/A
DOCK8	pNIC-NStIIT	S1569	P2003	Y - BT	50.5	H	N	N/A
DOCK8	pNIC-NStIIT	R1573	P2003	Y - BT	50.1	H	N	N/A
DOCK8	pNIC-NStIIT	K1564	E1999	Y - BT	50.6	H	N	N/A
DOCK8	pNIC-NStIIT	S1569	E1999	Y - BT	50.0	-	N/A	N/A
DOCK8	pNIC-NStIIT	R1573	E1999	Y - BT	49.6	H	N	N/A
DOCK8	pNIC-NStIIT	K1564	E1994	Y - BT	50.0	H	N	N/A
DOCK8	pNIC-NStIIT	S1569	E1994	Y - BT	49.4	H	N	N/A
DOCK8	pNIC-NStIIT	R1573	E1994	Y - BT	49.0	H	N	N/A
DOCK9	pNIC-NStIIT	K1605	L2068	Y - BT	54.3	-	N/A	N/A
DOCK9	pNIC-NStIIT	S1609	L2068	N - BT	53.9	N/A	N/A	N/A
DOCK9	pNIC-NStIIT	K1605	M2064	Y - BT	53.8	-	N/A	N/A
DOCK9	pNIC-NStIIT	S1609	M2064	Y - BT	53.3	-	N/A	N/A
DOCK9	pNIC-NStIIT	K1605	S2061	Y - BT	53.4	-	N/A	N/A
DOCK9	pNIC-NStIIT	S1609	S2061	Y - BT	53.0	-	N/A	N/A
RASGRF1	pNIC-NStIIT	D630	T1273	Y - BT	73.1	-	N/A	N/A
RASGRF1	pNIC-NStIIT	V632	P1269	Y - BT	72.5	-	N/A	N/A
RASGRF1	pNIC-NStIIT	V632	R1266	Y - BT	72.1	-	N/A	N/A
RASGRF1	pNIC-NStIIT	V632	D1250	Y - BT	70.2	-	N/A	N/A
RASGRF1	pNIC-NStIIT	C643	P1269	Y - BT	71.2	-	N/A	N/A
RASGRF1	pNIC-NStIIT	C643	R1266	Y - BT	70.8	-	N/A	N/A
RASGRF1	pNIC-NStIIT	C643	D1250	Y - BT	68.9	-	N/A	N/A
RASGRF1	pNIC-NStIIT	V653	P1269	Y - BT	70.0	-	N/A	N/A
RASGRF1	pNIC-NStIIT	V653	R1266	Y - BT	69.7	-	N/A	N/A
RASGRF1	pNIC-NStIIT	V653	D1250	Y - BT	67.8	H	N	N/A
RASGRF1	pNIC-NStIIT	N1002	P1269	N - BT	31.3	-	N/A	N/A
RASGRF1	pNIC-NStIIT	N1002	R1266	Y - BT	31.0	-	N/A	N/A
RASGRF1	pNIC-NStIIT	N1002	D1250	Y - BT	29.1	-	N/A	N/A
RASGRF1	pNIC-NStIIT	T1018	P1269	Y - BT	29.5	H	N	N/A
RASGRF1	pNIC-NStIIT	T1018	R1266	Y - BT	29.2	-	N/A	N/A
RASGRF1	pNIC-NStIIT	T1018	D1250	Y - BT	27.3	-	N/A	N/A
RASGRF1	pNIC-NStIIT	S1038	P1269	Y - BT	27.3	H	N	N/A
RASGRF1	pNIC-NStIIT	S1038	R1266	Y - BT	26.9	-	N/A	N/A
RASGRF1	pNIC-NStIIT	S1038	D1250	Y - BT	25.0	-	N/A	N/A
RASGRF1	pNIC-NStIIT	V632	P762	Y - BT	15.3	-	N/A	N/A
RASGRF1	pNIC-NStIIT	V632	S758	Y - BT	14.8	-	N/A	N/A

Target	Vector	Start	End	Cloned?	M_w (kDa)	Ex. Lvl	Xtal trials	Xtals
RASGRF1	pNIC-NStIIT	C643	P762	Y - BT	14.0	-	N/A	N/A
RASGRF1	pNIC-NStIIT	C643	S758	Y - BT	13.6	-	N/A	N/A
RASGRF1	pNIC-NStIIT	V653	P762	Y - BT	12.8	-	N/A	N/A
RASGRF1	pNIC-NStIIT	V653	S758	Y - BT	12.4	-	N/A	N/A
RhoA	pNIC-NStIIT	M1	L179	Y - BT	20.3	H	N	N/A
RhoA	pNIC-NStIIT	A3	L179	Y - BT	20.3	H	N	N/A
RhoA	pNIC-NStIIT	M1	A181	Y - BT	20.5	H	N	N/A
RhoA	pNIC-NStIIT	A3	A181	Y - BT	20.5	H	N	N/A
RhoA	pNIC-NStIIT	M1	C190	Y - BT	21.5	H	N	N/A
RhoA	pNIC-NStIIT	A3	C190	Y - BT	21.5	H	N	N/A
Rit2	pNIC-NStIIT	M1	T217	Y - BT	24.8	-	N/A	N/A
Rit2	pNIC-NStIIT	G12	T217	Y - BT	23.7	L	N	N/A
Rit2	pNIC-NStIIT	G17	T217	Y - BT	23.3	L	N	N/A
Rit2	pNIC-NStIIT	E20	T217	Y - BT	23.0	-	N/A	N/A
Rit2	pNIC-NStIIT	M1	R213	Y - BT	24.3	-	N/A	N/A
Rit2	pNIC-NStIIT	G12	R213	Y - BT	23.2	L	N	N/A
Rit2	pNIC-NStIIT	G17	R213	Y - BT	22.9	-	N/A	N/A
Rit2	pNIC-NStIIT	E20	R213	Y - BT	22.6	-	N/A	N/A
Rit2	pNIC-NStIIT	M1	K198	Y - BT	22.5	H	N	N/A
Rit2	pNIC-NStIIT	G12	K198	Y - BT	21.4	H	N	N/A
Rit2	pNIC-NStIIT	G17	K198	Y - BT	21.1	H	N	N/A
Rit2	pNIC-NStIIT	E20	K198	Y - BT	20.8	M	N	N/A
Rit2	pNIC28-10His	M1	T217	N - JG	24.8	N/A	N/A	N/A
Rit2	pNIC28-10His	M1	K204	Y - JG	23.2	-	N/A	N/A
Rit2	pNIC28-10His	M1	W202	Y - JG	23.0	-	N/A	N/A
Rit2	pNIC28-10His	M1	K196	Y - JG	22.2	-	N/A	N/A
Rit2	pNIC28-10His	M1	K194	Y - JG	22.0	-	N/A	N/A
Rit2	pNIC28-10His	M1	R179	Y - JG	20.1	-	N/A	N/A
Rit2	pNIC28-10His	Y21	T217	Y - JG	22.9	-	N/A	N/A
Rit2	pNIC28-10His	Y21	K204	Y - JG	21.4	-	N/A	N/A
Rit2	pNIC28-10His	Y21	W202	Y - JG	21.1	-	N/A	N/A
Rit2	pNIC28-10His	Y21	K196	Y - JG	20.4	-	N/A	N/A
Rit2	pNIC28-10His	Y21	K194	Y - JG	20.1	-	N/A	N/A
Rit2	pNIC28-10His	Y21	R179	Y - JG	18.3	-	N/A	N/A
Rit2	pNIC-NStIIT	G12	R213	N - JG	23.2	N/A	N/A	N/A
Sos1	pNIC-NStIIT	Q566	T1049	Y - BT	56.9	M	N	N/A
Sos1	pNIC-NStIIT	V574	T1049	Y - BT	56.0	M	N	N/A
Sos1	pNIC-NStIIT	P594	T1049	Y - BT	53.6	-	N/A	N/A
Sos1	pNIC-NStIIT	Q566	R1019	Y - BT	53.6	H	N	N/A
Sos1	pNIC-NStIIT	V574	R1019	N - BT	52.7	N/A	N/A	N/A
Sos1	pNIC-NStIIT	P594	R1019	Y - BT	50.2	-	N/A	N/A
Sos1	pNIC-NStIIT	A572	R1019	Y - BT	52.9	H	N	N/A
Sos1	pNIC-NStIIT	A572	I1016	Y - BT	52.5	-	N/A	N/A

Target	Vector	Start	End	Cloned?	M_w (kDa)	Ex. Lvl	Xtal trials	Xtals
Sos2	pNIC-NStIIT	L565	F1025	Y - BT	54.4	-	N/A	N/A
Sos2	pNIC-NStIIT	E571	R1017	Y - BT	52.8	-	N/A	N/A
Sos2	pNIC-NStIIT	E571	E1015	Y - BT	52.5	-	N/A	N/A
Sos2	pNIC-NStIIT	E571	N1009	Y - BT	51.8	-	N/A	N/A
Sos2	pNIC-NStIIT	N583	R1017	Y - BT	51.3	-	N/A	N/A
Sos2	pNIC-NStIIT	N583	E1015	Y - BT	51.0	-	N/A	N/A
Sos2	pNIC-NStIIT	N583	N1009	Y - BT	50.3	-	N/A	N/A
Sos2	pNIC-NStIIT	G595	R1017	Y - BT	49.9	-	N/A	N/A
Sos2	pNIC-NStIIT	G595	E1015	Y - BT	49.6	-	N/A	N/A
Sos2	pNIC-NStIIT	G595	N1009	Y - BT	48.9	-	N/A	N/A
Sos2	pNIC-NStIIT	T603	R1017	Y - BT	49.2	-	N/A	N/A
Sos2	pNIC-NStIIT	T603	E1015	Y - BT	48.9	-	N/A	N/A
Sos2	pNIC-NStIIT	T603	N1009	Y - BT	48.2	-	N/A	N/A
Sos2	pNIC-NStIIT	F769	R1017	Y - BT	29.5	-	N/A	N/A
Sos2	pNIC-NStIIT	F769	E1015	Y - BT	29.2	-	N/A	N/A
Sos2	pNIC-NStIIT	F769	N1009	Y - BT	28.5	-	N/A	N/A
Sos2	pNIC-NStIIT	P779	R1017	Y - BT	28.3	-	N/A	N/A
Sos2	pNIC-NStIIT	P779	E1015	Y - BT	28.0	-	N/A	N/A
Sos2	pNIC-NStIIT	P779	N1009	Y - BT	27.3	-	N/A	N/A
Sos2	pNIC-NStIIT	E571	E753	Y - BT	21.9	-	N/A	N/A
Sos2	pNIC-NStIIT	E571	A743	Y - BT	20.8	-	N/A	N/A
Sos2	pNIC-NStIIT	E571	K739	Y - BT	20.4	-	N/A	N/A
Sos2	pNIC-NStIIT	N583	E753	Y - BT	20.4	-	N/A	N/A
Sos2	pNIC-NStIIT	N583	A743	Y - BT	19.3	-	N/A	N/A
Sos2	pNIC-NStIIT	N583	K739	Y - BT	18.9	-	N/A	N/A
Sos2	pNIC-NStIIT	G595	E753	Y - BT	19.0	-	N/A	N/A
Sos2	pNIC-NStIIT	G595	A743	Y - BT	17.9	L	N	N/A
Sos2	pNIC-NStIIT	G595	K739	Y - BT	17.5	-	N/A	N/A
Sos2	pNIC-NStIIT	T603	E753	Y - BT	18.3	L	N	N/A
Sos2	pNIC-NStIIT	T603	A743	Y - BT	17.2	L	N	N/A
Sos2	pNIC-NStIIT	T603	K739	Y - BT	16.8	-	N/A	N/A

Y -JG = Cloned by Janine Gray
N - JG = Unsuccessful cloning attempt by Janine Gray
Y - BT = Cloned by the Biotechnology group at the CMD
N - BT = Unsuccessful cloning attempt by the Biotechnology group at the CMD
Ex. Lvl = Expression Level.
- = No expression
L = Low expression
M = Medium expression
H = High expression
Test expression in this pipeline was conducted by the Biotechnology group at the CMD. Experiments were repeated on low and high scale by Janine Gray and Andrew Thompson of the von Delft and Brennan groups.
Xtal trials: 'Y' means complex was subjected to crystallisation trials. 'N' means protein was not attempted to be crystallised
Xtals: 'Y' means potential crystalline conditions observed. 'N' means no crystal-like conditions observed

8.2 Overview of Ligation Independent Cloning (LIC)

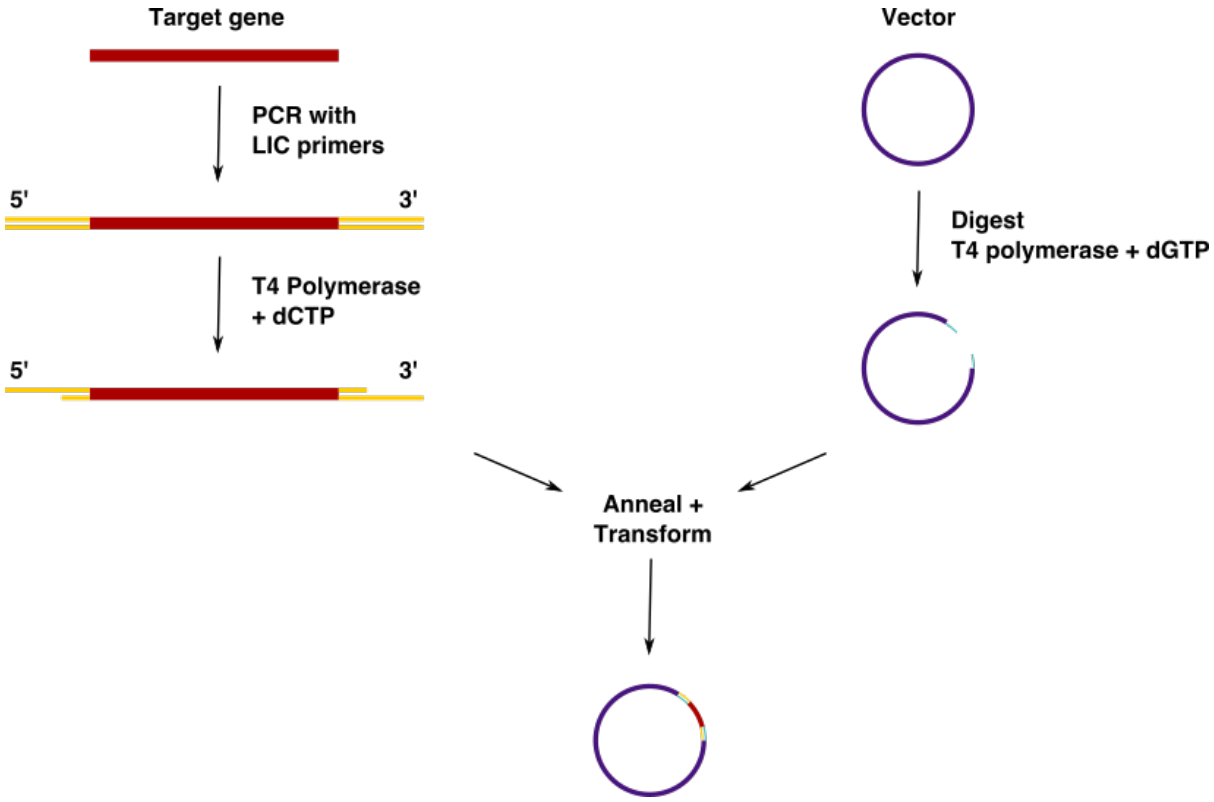


Figure 89. LIC cloning procedure. The target gene is amplified using PCR and LIC-designed primers. T4 polymerase and a single dNTP is added to create the sticky overhangs. The complementary overlaps are created in the vector using dGTP. Mixing of the vector and insert results in enzyme-free annealing of the complementary overhangs. Transformation in *E. coli* seals the nicks in the DNA.

8.3 Summary of Methods to Optimise Protein Expression

Table 10. The various methods used to try and improve the yields of proteins which did not work using the standard PrepX protocol. The colour schemes refer to the following; red: the method did not work at all; orange: cloning was successful but no expression; yellow: moderate improvement in expression; green: successful expression of protein; white: the procedure was not attempted for this protein. 'Bn Alc' denotes Benzyl Alcohol, 'T' denotes Temperature 'T7 express' denotes T7 express lysY competent cells.

	Method			
Target	**MBP2**	**Bn Alc/T**	**T7 express**	**Co-lysis**
AKAP13				
ARFGEF1				
CY2				
Ran				
RCC1				
RhoA				
Tiam1				
Trio				
IQSEC3				
RGL2				
ARHGEF11				
ARHGEF12				
DOCK2				
DOCK9				
IQSEC1				
PRex-1				
DOCK3				

8.4 Data Collection and Refinement Statistics for Novel GEF/GTPase Complexes

PDB code	5O33	6FAE
Complex	Kalirin-c001/Rac1 (holo)	IQSEC2/Arf1 (apo)
Beamline	I04-1	I04-1
Wavelength (Å)	0.92819	0.9159
Space group	P 6_5 2 2	P 1 21 1
a, b, c (Å)	63.255,63.255, 346.704	50.995, 92.744. 66.895
α, β, γ (°)	90, 90, 120	90, 106, 90
Data collection		
Resolution (Å)	57.78-1.64	49.14-2.35
	(1.67-1.64)	(2.41-2.35)
R_{merge}	0.0971 (2.643)	-
I/σ(I)	11.2 (1.2)	8.9 (1.8)
$CC_{1/2}$	0.999 (0.493)	0.995 (0.550)
Completeness (%)	100 (99.7)	-
Redundancy	18.5 (18.8)	3.3
No. reflections	52137 (2571)	25405 (1161)
Refinement		
Resolution (Å)	1.64	2.35
No.reflections	49371	24553
Rwork / Rfree	0.2067/0.2398	0.2405/0.2851
Average B-factor (Å^2)	43.0	44.0
RMSD: bonds (Å)	0.009	0.001
RMSD: bond angles (°)	1.072	0.392
Ramachandran favoured (%)	97	98
Ramachandran outliers (%)	0.3	0.6

Xtal ID	XX02KALRNA-x1376	XX03KALRNA-x002
Complex	Kalirin-c001/Rac1 (apo)	Kalirin-c002/Rac1
Beamline	I04-1	I04
Wavelength (Å)	0.9159	0.9795
Space group	P 6_5 2 2	P 2_1 2_1 2
a, b, c (Å)	62.836,62.836,342.580	115.478,116.383,92.344
α, β, γ (°)	90, 90, 120	90, 90, 90
Data collection		
Resolution (Å)	57.16 – 1.91	72.12 – 2.86
	(1.96 – 1.91)	(2.91 – 2.86)
R_{merge}	0.511 (2.620)	0.243 (4.181)
I/σ(I)	10.44 (1.21)	7.6 (1.1)
$CC_{1/2}$	0.9979 (0.8013)	0.996 (0.576)
Completeness (%)	100 (98.5)	99.1 (95.5)
Redundancy	33.9 (18.5)	13.2 (13.4)
No. reflections	32694 (1554)	29476 (1374)
Refinement		
Resolution (Å)	1.91	2.86
No.reflections	30853	29361
Rwork / Rfree	0.2046/0.2532	0.2493/0.3133
Average B-factor (Å^2)	37.0	82
RMSD: bonds (Å)	0.017	0.008
RMSD: bond angles (°)	1.969	1.105
Ramachandran favoured (%)	96	91
Ramachandran outliers (%)	1.42	1.1

8.5 Conservation of Residue 18 within the Ras Superfamily

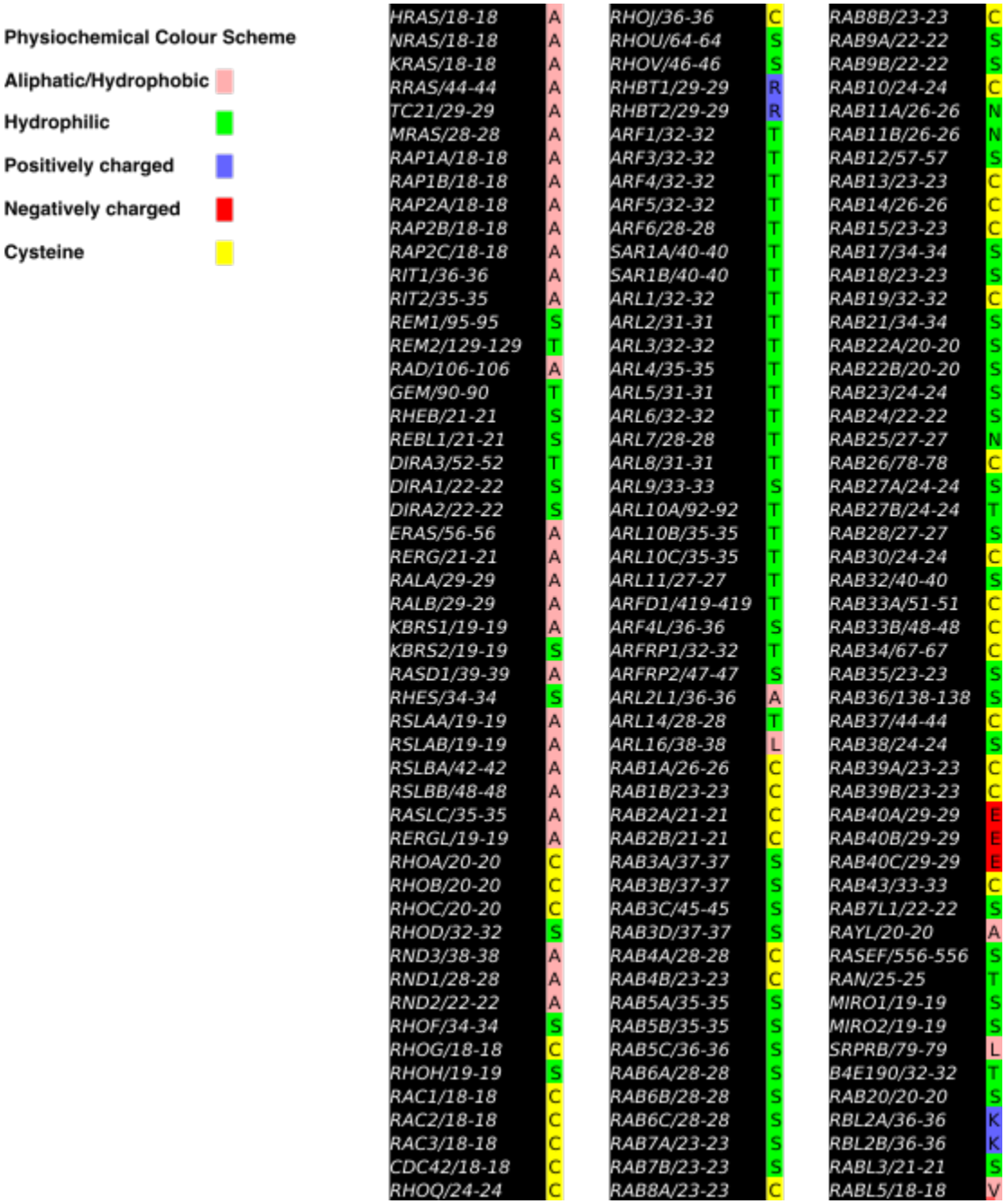


Figure 90. Sequence alignment of 153 members of the Ras superfamily of residue 18 (Rac1 numbering) using ClusterX in Jalview. The conserved residue in this position is serine (S), present in 32% of family members, followed by Cysteine (21%), Alanine (20%) and Threonine (18%).

8.6 HPLC Assay for GDP Removal

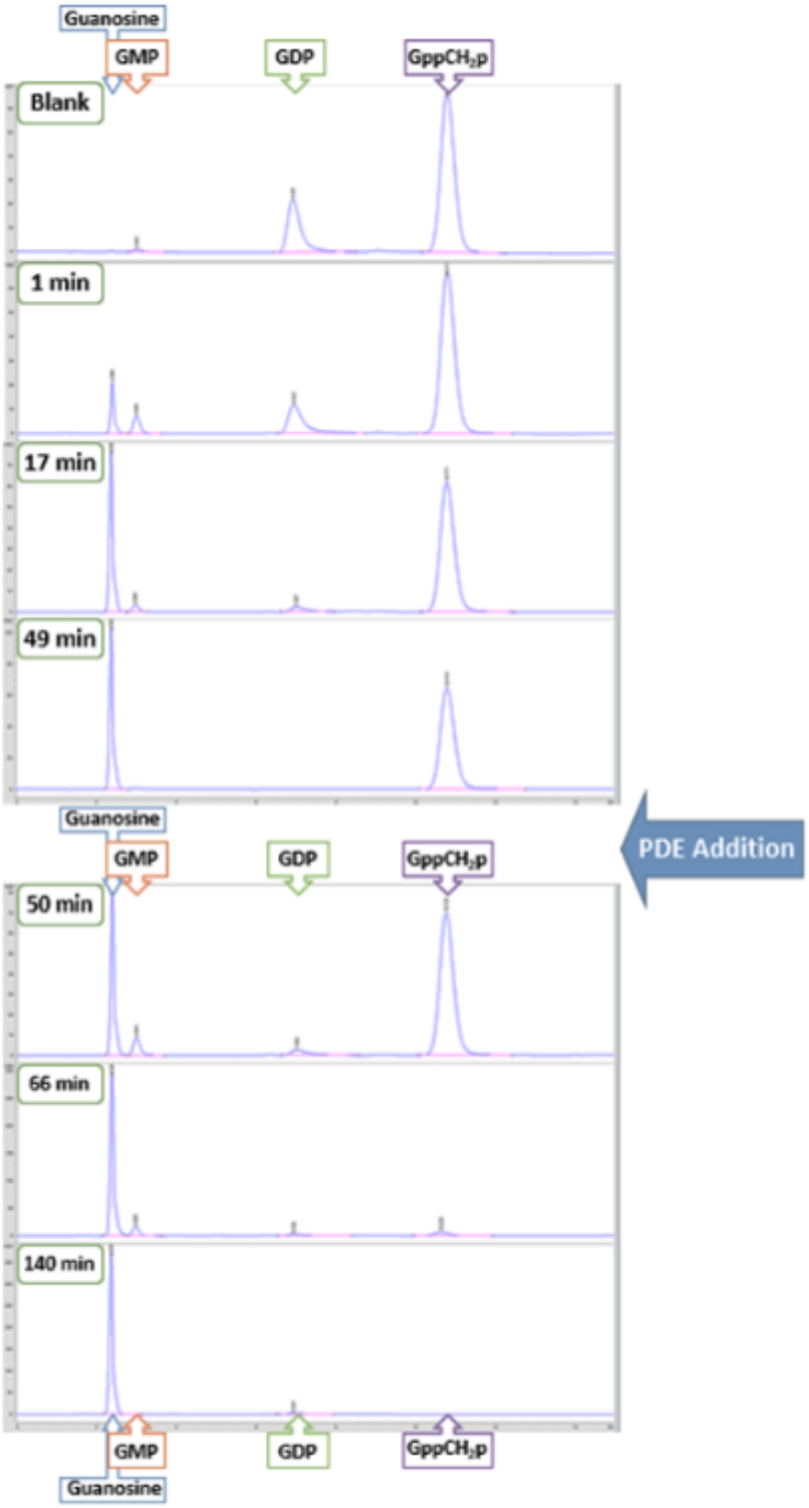


Figure 91. Figure generated by Carmen Jimenez Antunez. HPLC method to monitor the degradation of guanine nucleotides during the reaction.

8.7 Data Collection and Refinement Statistics for XChem Fragments

PDB code	5QQD	TBD	5QQE
Ligand	**32**	**33**	**44**
Beamline	I04-1	I04-1	I04-1
Wavelength	0.91587	0.91587	0.91587
Space group	P 6_5 2 2	P 6_5 2 2	P 6_5 2 2
a, b, c (Å)	62.836, 62.836,342.580	62.730,62.730,342.754	62.807,62.807,342.65
a, b, g (°)	90, 90, 120	90, 90, 120	90,90,120
Data collection			
Resolution	57.10 - 1.91	57.19 – 2.12	57.11 – 1.95
	(1.96 - 1.91)	(2.18 – 2.12)	(2.05 – 1.95)
R_{merge}	0.510 (2.691)	0.227 (2.669)	0.275 (7.137)
I/σ(I)	10.4 (1.6)	9.1 (1.1)	8.8 (1.1)
$CC_{1/2}$	0.998 (0.773)	0.999 (0.474)	0.998 (0.433)
Completeness (%)	99.7 (98.0)	99.3 (98.6)	97.2 (99.9)
Redundancy	33.9 (18.4)	15.3 (17.5)	18.4 (18.5)
No. reflections	32694 (2338)	24312 (1613)	30148 (4307)
Refinement			
Resolution (Å)	1.91	2.12	1.95
No.reflections	30853	22682	28451
Rwork / Rfree	0.2059/0.2586	0.2175/0.2787	0.218/0.284
Average B-factor (Å^2)	35.5	39.8	45.5
RMSD: bonds (Å)	0.009	0.016	0.009
RMSD: bond angles (°)	1.526	1.799	1.524
Ramachandran favoured (%)	96	96	95
Ramachandran outliers (%)	0.3	0.6	0.6
Clashscore	6	-	7
RSRZ outliers (%)	2.2	-	1.7
Ligand RSCC	0.86	-	0.88

PDB code	5QQF	5QQG	5QQH
Ligand	**45**	**46**	**47**
Beamline	I04-1	I04-1	I04-1
Wavelength	0.91587	0.91587	0.91587
Space group	P 6_5 2 2	P 6_5 2 2	P 6_5 2 2
a, b, c (Å)	62.49,62.49,341.271	62.785,62.785, 344.573	62.659,62.659,342.820
α, β, γ (°)	90, 90, 120	90, 90, 120	90,90,120
Data collection			
Resolution	56.88 – 2.26 (2.39 – 2.26)	57.43 – 2.23 (2.29 – 2.23)	57.14 – 2.09 (2.15 – 2.09)
R_{merge}	0.137 (1.575)	0.210 (3.115)	0.373 (4.132)
Mean I/σ(I)	10.4 (1.4)	11.2 (1.2)	9.5 (1.7)
$CC_{1/2}$	0.998 (0.417)	0.999 (0.333)	0.998 (0.619)
Completeness (%)	99.2 (98.1)	100 (100)	99.8 (100)
Redundancy	10.7 (8.6)	15.6 (13.7)	36 (27.7)
No. reflections	19514 (2696)	20947 (2929)	25043 (1740)
Refinement			
Resolution (Å)	2.26	2.23	2.09
No. reflections	18445	19788	22660
Rwork / Rfree	0.1977/0.2699	0.1970/0.2742	0.1995/0.2631
Average B-factor (Å^2)	53.0	52.7	41.3
RMSD: bonds (Å)	0.012	0.012	0.017
RMSD: bond angles (°)	1.570	1.572	1.763
Ramachandran favoured (%)	96	96	98
Ramachandran outliers (%)	0.6	0.6	0.3
Clashscore	4	4	6
RSRZ outliers (%)	1.1	0.3	0.3
Ligand RSCC	0.88	0.93	0.86

PDB code	5QQI	5QQJ	5QQK
Ligand	**48**	**49**	**50**
Beamline	I04-1	I04-1	I04-1
Wavelength	0.91587	0.91587	0.91587
Space group	P 6_5 2 2	P 6_5 2 2	P 6_5 2 2
a, b, c (Å)	62.505,62.505,341.300	62.568,62.568,342.039	62.584,62.584,341.958
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120
Data collection			
Resolution	56.88 – 2.08 (2.13 – 2.08)	54.24 – 1.90 (1.95 – 1.90)	56.99 – 2.24 (2.36 – 2.24)
R_{merge}	0.802 (2.454)	0.197 (3.032)	0.456 (4.818)
Mean I/σ(I)	10.2 (1.7)	9.4 (1.6)	9.2 (1.6)
$CC_{1/2}$	0.998 (0.745)	0.998 (0.697)	0.995 (0.517)
Completeness (%)	99.7 (100)	99.6 (100)	100 (100)
Redundancy	32.7 (18.7)	18.2 (18.6)	17.8 (18.4)
No. reflections	25188 (1772)	32844 (2340)	20517 (2862)
Refinement			
Resolution (Å)	2.08	1.90	2.24
No. reflections	23797	30971	19384
Rwork / Rfree	0.2188/0.2767	0.2118/0.2687	0.1999/0.2656
Average B-factor (Å^2)	41.4	37.6	44.7
RMSD: bonds (Å)	0.016	0.008	0.014
RMSD: bond angles (°)	1.738	1.524	1.618
Ramachandran favoured (%)	97	94	95
Ramachandran outliers (%)	0.3	0.9	0.3
Clashscore	5	7	4
RSRZ outliers (%)	5.3	1.4	0.6
Ligand RSCC	0.80	0.82	0.87

PDB code	5QQL	5QQM	5QQN
Ligand	**51**	**52**	**53**
Beamline	I04-1	I04-1	I04-1
Wavelength	0.91587	0.91587	0.91587
Space group	P 6_5 2 2	P 6_5 2 2	P 6_5 2 2
a, b, c (Å)	62.308,62.308,340.950	62.498,62.498,342.286	62.492,62.492,341.798
α, β, γ (°)	90, 90, 120	90, 90, 120	90,90,120
Data collection			
Resolution	56.83 – 2.25 (2.39 – 2.25)	57.05 – 2.02 (2.08 – 2.02)	54.18 – 2.26 (2.38 – 2.26)
R_{merge}	0.297 (4.116)	0.342 (3.640)	0.439 (3.63)
Mean I/σ(I)	8.6 (1.1)	10.8 (1.8)	11.8 (2.8)
$CC_{1/2}$	0.998 (0.404)	0.946 ()	0.997 (0.830)
Completeness (%)	100 (99.9)	99.8 (98.3)	99.9 (99.9)
Redundancy	18.4 (19.6)	34.6 (17.7)	17.5 (19.4)
No. reflections	19877 (3096)	27501 (2053)	19972 (2787)
Refinement			
Resolution (Å)	2.25	2.02	2.26
No. reflections	18781	25975	18877
Rwork / Rfree	0.1911/0.2702	0.2113/0.2731	0.2050/0.2786
Average B-factor (Å^2)	54.9	38.0	43.4
RMSD: bonds (Å)	0.012	0.017	0.008
RMSD: bond angles (°)	1.516	1.804	1.537
Ramachandran favoured (%)	95	96	97
Ramachandran outliers (%)	0.6	0.3	0.6
Clashscore	3	4	8
RSRZ outliers (%)	0.6	0.6	1.4
Ligand RSCC	0.922	0.88	0.78

8.8 Potency Determination for Urea Analogues and Bioisosteres

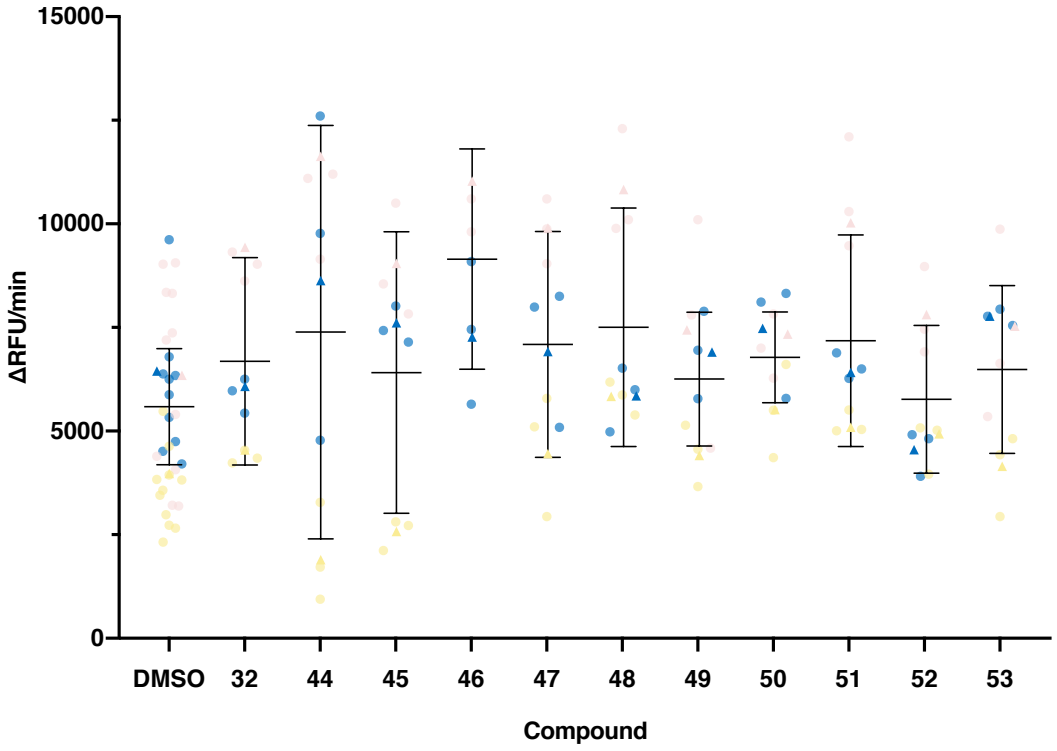


Figure 92. Nucleotide exchange assay for urea hit **32** and analogues ordered from Molport that bound to Kalirin/Rac1 in the XChem screen (compounds **44-53**) tested at 500 μM (n = 3, one-way ANOVA followed by Dunnett's multiple comparisons test). Compounds were incubated for two hours. Blue, yellow and pink circles represent three replicates within each repeat. Triangles represent the mean for each repeat. Black bars represent the overall mean ± SD. All results were normalised using the positive control (GppNHp). No results were significantly different to the negative control.

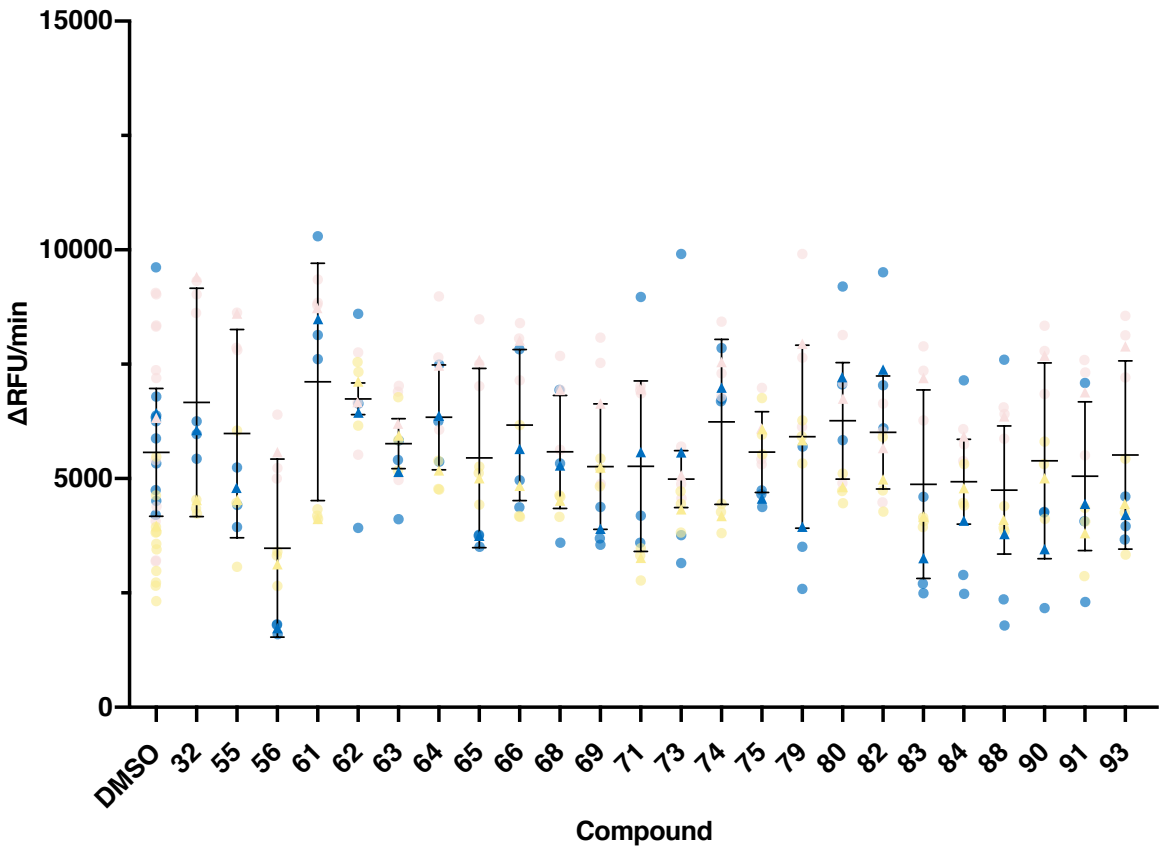
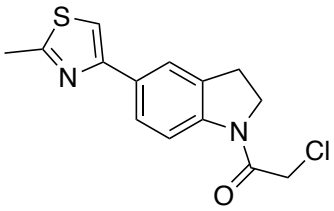
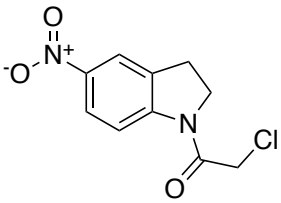
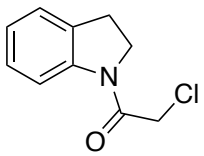
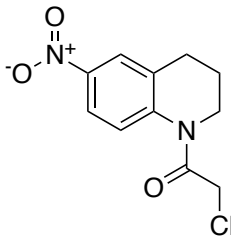
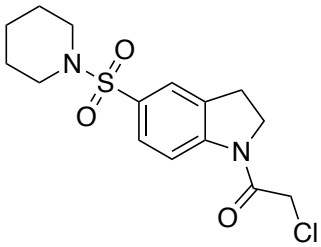
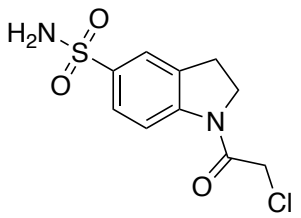
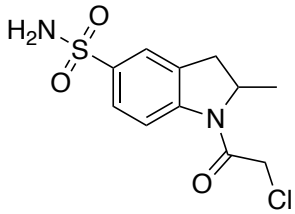
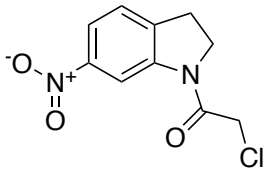
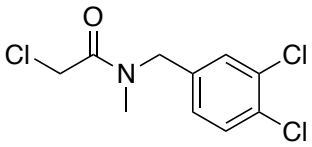
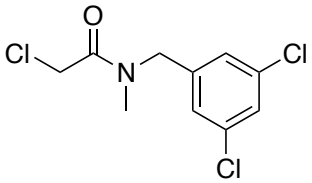
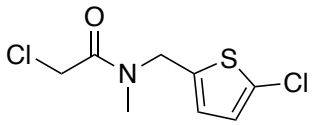
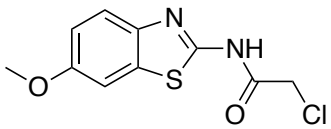
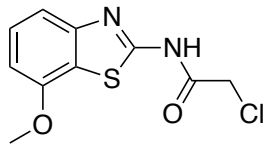
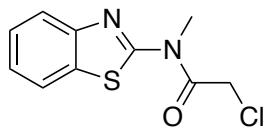
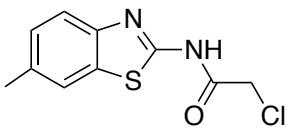


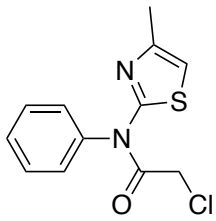
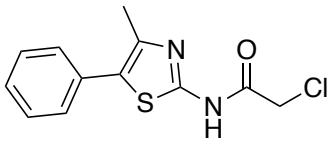
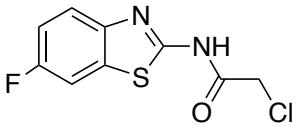
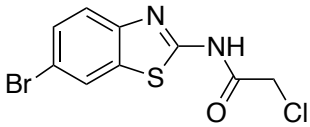
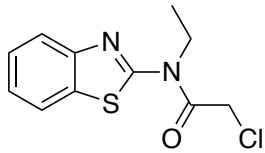
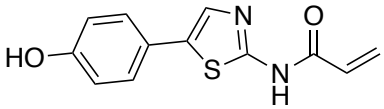
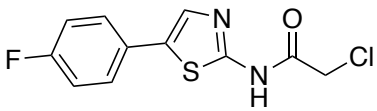
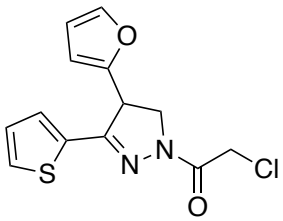
Figure 93. Nucleotide exchange assay for urea hit **32** and synthesised bioisosteres (**55-93**) tested at 500 μM (n = 3, one-way ANOVA followed by Dunnett's multiple comparisons test). Compounds were incubated for two hours. Blue, yellow and pink circles represent three replicates within each repeat. Triangles represent the mean for each repeat. Black bars represent the overall mean ± SD. All results were normalised using the positive control (GppNHp). No results were significantly different to the negative control.

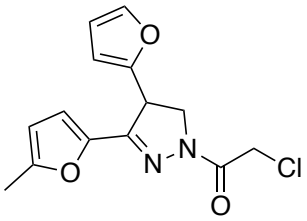
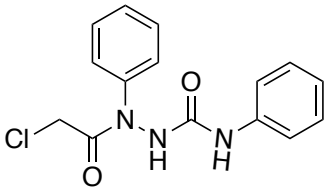
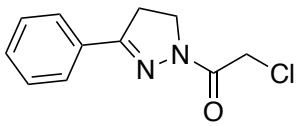
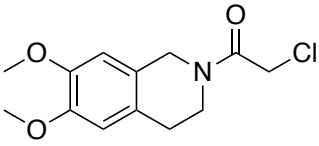
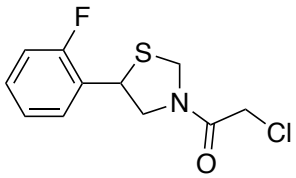
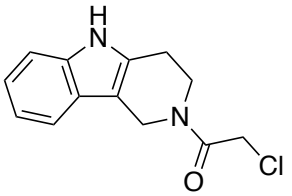
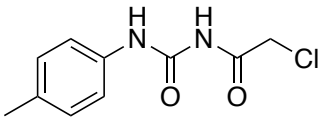
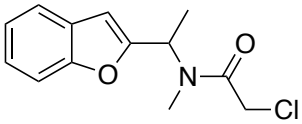
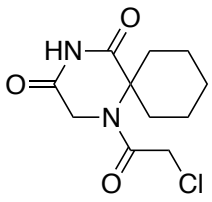
8.9 Electrophilic Mass Spectrometry Screens Conducted by the London Group

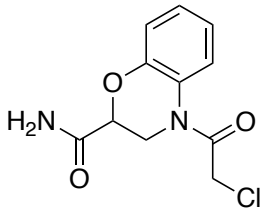
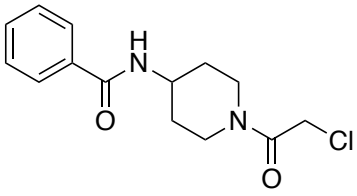
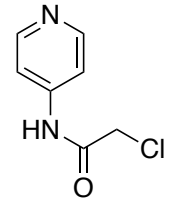
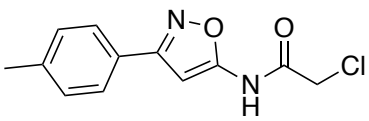
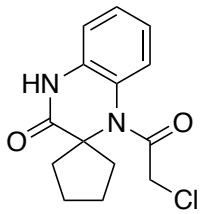
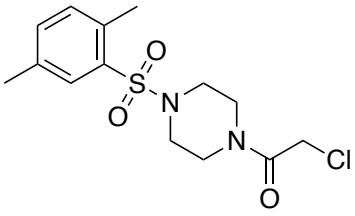
8.9.1 First Electrophilic Mass Spectrometry Screen

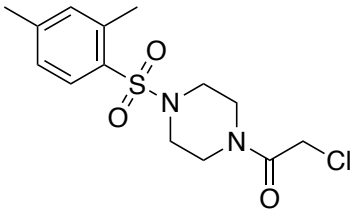
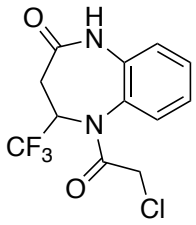
Compound	SMILES	% Labelling
Class I - Indolines		
	O=C(CCl)N1C2=CC=C(C3=CSC(C)=N3)C=C2CC1	100
	O=C(CCl)N1C2=CC=C([N+]([O-])=O)C=C2CC1	91
	O=C(CCl)N1C2=CC=CC=C2CC1	63
	[O-][N+](C1=CC=C(N(C(CCl)=O)CCC2)C2=C1)=O	57
	O=C(CCl)N1C2=CC=C(S(=O)(N3CCCCC3)=O)C=C2CC1	46
	O=C(CCl)N1C2=CC=C(S(=O)(N)=O)C=C2CC1	44

Compound	SMILES	% Labelling
	O=C(CCl)N1C2=CC=C(S(=O)(N)=O)C=C2CC1C	33
	O=C(CCl)N1C2=CC([N+]([O-])=O)=CC=C2CC1	17
Class II - Chlorobenzylamines		
	ClC1=CC=C(CN(C(CCl)=O)C)C=C1Cl	100
	CN(C(CCl)=O)CC1=CC(Cl)=CC(Cl)=C1	74
	CN(C(CCl)=O)CC1=CC=C(Cl)S1	43
Class III - Aminothiazoles (promiscuous binders)		
	O=C(CCl)NC1=NC2=CC=C(OC)C=C2S1	83
	O=C(CCl)NC1=NC2=CC=CC(OC)=C2S1	63
	O=C(CCl)N(C)C1=NC2=CC=CC=C2S1	100
	O=C(CCl)NC1=NC2=CC=C(C)C=C2S1	67

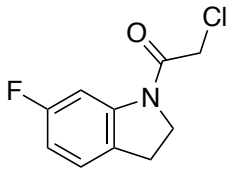
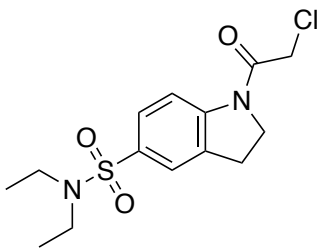
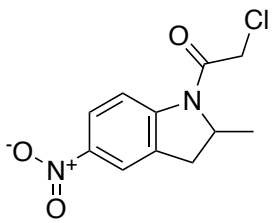
Compound	SMILES	% Labelling
	O=C(CCl)N(C1=NC=CS1)C2=CC=CC=C2	100
	O=C(CCl)NC1=NC(C)=C(C2=CC=CC=C2)S1	50
	O=C(CCl)NC1=NC2=CC=C(F)C=C2S1	71
	O=C(CCl)NC1=NC2=CC=C(Br)C=C2S1	67
	O=C(CCl)N(CC)C1=NC2=CC=CC=C2S1	100
	O=C(C=C)NC1=NC=C(C2=CC=C(O)C=C2)S1	59
	O=C(CCl)NC1=NC=C(C2=CC=C(F)C=C2)S1	47
Class IV - Other promiscuous binders		
	O=C(CCl)N(C1)N=C(C2=CC=CS2)C1C3=CC=CO3	65

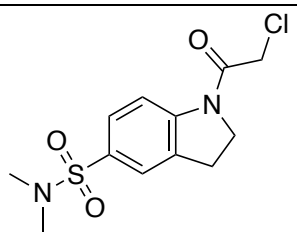
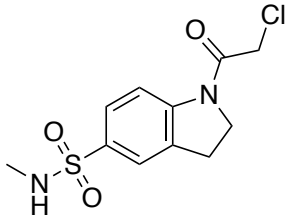
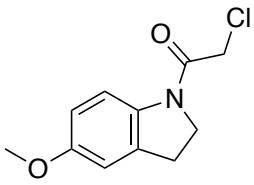
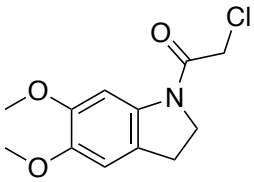
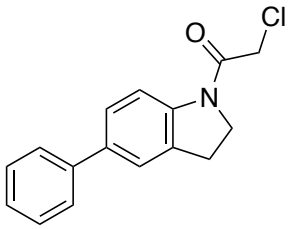
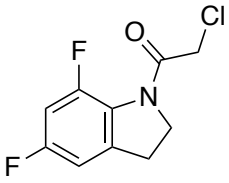
Compound	SMILES	% Labelling
	O=C(CCl)N(C1)N=C(C2=CC=C(C)O2)C1C3=CC=CO3	63
	O=C(CCl)N(NC(NC1=CC=CC=C1)=O)C2=CC=CC=C2	100
	O=C(CCl)N1N=C(C2=CC=CC=C2)CC1	42 (+40 double labelled)
	Class V - Miscellaneous	
	O=C(CCl)N1CCC2=CC(OC)=C(OC)C=C2C1	71
	O=C(CCl)N1CSC(C2=C(F)C=CC=C2)C1	71
	O=C(CCl)N1CCC2=C(C(C=CC=C3)=C3N2)C1	71
	CC1=CC=C(NC(NC(CCl)=O)=O)C=C1	67
	CC(N(C)C(CCl)=O)C1=CC2=CC=CC=C2O1	67
	O=C(CCl)N(C12CCCCC2)CC(NC1=O)=O	63

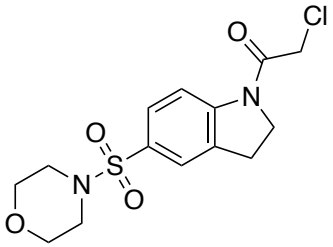
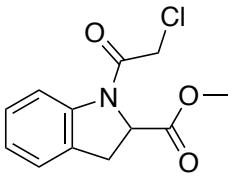
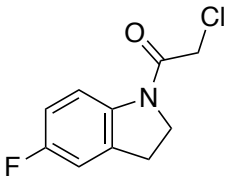
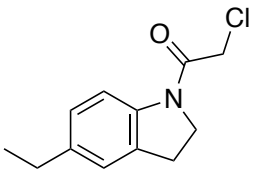
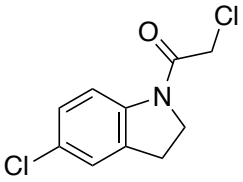
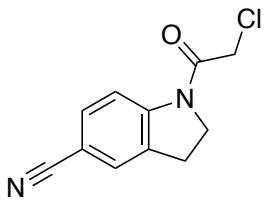
Compound	SMILES	% Labelling
	NC(C1OC2=C(C=CC=C2)N(C(CCl)=O)C1)=O	57
	O=C(C1=CC=CC=C1)NC2CCN(C(CCl)=O)CC2	57
	O=C(CCl)NC1=CC=NC=C1	57
	O=C(CCl)NC1=CC(C2=CC=C(C)C=C2)=NO1	53
	O=C(CCl)N1C2(CCCC2)C(NC3=C1C=CC=C3)=O	50
	O=C(CCl)N(CC1)CCN1S(=O)(C2=C(C)C=CC(C)=C2)=O	43

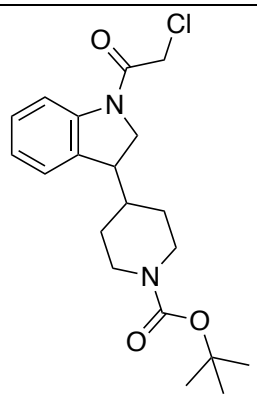
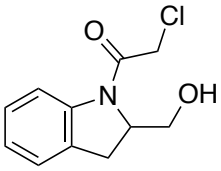
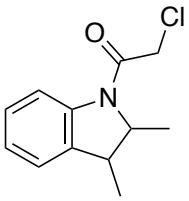
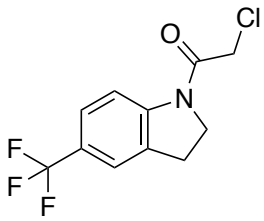
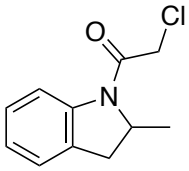
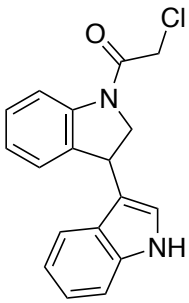
Compound	SMILES	% Labelling
	O=C(CCl)N(CC1)CCN1S(=O)(C2=C(C)C=C(C)C=C2)=O	44
	O=C(CCl)N1C(C(F)(F)F)CC(NC2=C1C=CC=C2)=O	47

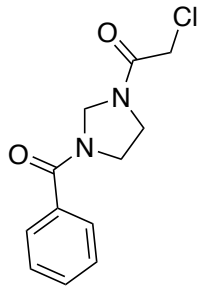
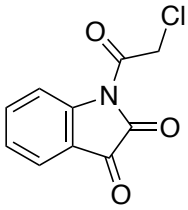
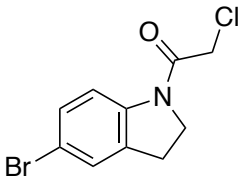
8.9.2 Electrophilic Mass Spectrometry Screen of Indolines on Rac1 and Kalirin/Rac1

Compound	SMILES	% Labelling Rac1	% Labelling Kalirin/Rac1
	O=C(N1CCC2=C1C=C(F)C=C2)CCl	0	25
	O=S(C1=CC2=C(N(C(CCl)=O)CC2)C=C1)(N(CC)CC)=O	25	30
	O=C(N1C(C)CC2=C1C=CC([N+]([O-])=O)=C2)CCl	10	40

Compound	SMILES	% Labelling Rac1	% Labelling Kalirin/Rac1
	O=S(C1=CC2=C(N(C(CCl)=O)CC2)C=C1)(N(C)C)=O	5	15
	O=S(C1=CC2=C(N(C(CCl)=O)CC2)C=C1)(NC)=O	5	20
	O=C(N1CCC2=C1C=CC(OC)=C2)CCl	5	55
	O=C(N1CCC2=C1C=C(OC)C(OC)=C2)CCl	0	0
	O=C(N1CCC2=C1C=CC(C3=CC=CC=C3)=C2)CCl	5	35
	O=C(N1CCC2=C1C(F)=CC(F)=C2)CCl	0	0

Compound	SMILES	% Labelling Rac1	% Labelling Kalirin/Rac1
	O=C(N1CCC2=C1C=CC(S(=O)(N3CCOCC3)=O)=C2)CCl	5	30
	O=C(C1N(C(CCl)=O)C2=C(C=CC=C2)C1)OC	0	10
	O=C(N1CCC2=C1C=CC(F)=C2)CCl	5	35
	O=C(N1CCC2=C1C=CC(CC)=C2)CCl	5	35
	O=C(N1CCC2=C1C=CC(Cl)=C2)CCl	5	55
	N#CC1=CC2=C(N(C(CCl)=O)CC2)C=C1	5	45

Compound	SMILES	% Labelling Rac1	% Labelling Kalirin/Rac1
	O=C(N1CCC(C2CN(C(CCl)=O)C3=C2C=CC=C3)CC1)OC(C)(C)C	5	5
	O=C(N1C(CO)CC2=C1C=CC=C2)CCl	0	0
	O=C(N1C(C)C(C)C2=C1C=CC=C2)CCl	0	0
	O=C(N1CCC2=C1C=CC(C(F)(F)F)=C2)CCl	15	25
	O=C(N1C(C)CC2=C1C=CC=C2)CCl	0	0
	O=C(N1CC(C2=CNC3=C2C=CC=C3)C4=C1C=CC=C4)CCl	0	0

Compound	SMILES	% Labelling Rac1	% Labelling Kalirin/Rac1
	O=C(N1CCN(C(C2=CC=CC=C2)=O)C1)CCl	0	0
	O=C(N1C(C(C2=C1C=CC=C2)=O)=O)CCl	0	0
	O=C(N1CCC2=C1C=CC(Br)=C2)CCl	0	45

8.10 Data Collection and Refinement Statistics for Covalent Compounds

Xtal ID	XX02KALRNA-x1412	XX02KALRNA-x1806	XX02KALRNA-x1802
Ligand	**99**	**103**	**104**
Beamline	I04-1	I04-1	I04-1
Wavelength	0.91587	0.91587	0.91587
Space group	P 6_5 2 2	P 6_5 2 2	P 6_5 2 2
a, b, c (Å)	62.690,62.690,341.801	62.795,62.795,340.101	62.737,62.737,340.5879
α, β, γ (°)	90, 90, 120	90, 90, 120	90,90,120
Data collection			
Resolution (Å)	54.29 – 1.96 (2.06 – 1.96)	54.44 – 1.71 (1.74 – 1.71)	56.88 – 1.97 (2.00 – 1.97)
R_{merge}	0.298 (11.7)	0.113 (3.353)	0.196 (2.42)
I/σ(I)	12.1	16.1	11.1
$CC_{1/2}$	0.997 (0.459)	1.00 (0.357)	0.999 (0.562)
Completeness (%)	99.6 (100)	100 (98.7)	100 (99.9)
Redundancy	18.7 (18.7)	18.4 (18.0)	18.6 (19.5)
No. reflections	32426 (1516)	44513 (1450)	29655 (1414)
Refinement			
Resolution (Å)	1.96	1.71	1.97
No. reflections	30058	43063	28046
Rwork / Rfree	0.2566/0.2798	0.2784/0.3024	0.2636/0.2907
Average B-factor (Å^2)	35.7	32.8	35.6
Ligand B-factor (Å^2)	49.4	45.4	66.7
RMSD: bonds (Å)	0.004	0.010	0.004
RMSD: bond angles (°)	0.664	1.493	1.138
Ramachandran favoured (%)	98	97	97
Ramachandran outliers (%)	0	0.3	0.3

Xtal ID	XX02KALRNA-x1810	XX02KALRNA-x123	XX02KALRNA-5241
Ligand	**105**	**106**	**107**
Beamline	I04-1	I03	I04-1
Wavelength	0.91587	0.97625	0.91587
Space group	P 6_5 2 2	P 6_5 2 2	P 6_5 2 2
a, b, c (Å)	62.917,62.917,340.442	62.475,62.475,339.601	62.280,62.280,339.670
α, β, γ (°)	90, 90, 120	90, 90, 120	90,90,120
Data collection			
Resolution	56.8 – 2.64 (2.69 – 2.64)	56.6 – 1.34 (1.36 – 1.34)	53.9 – 1.65 (1.69 – 1.65)
R_{merge}	0.483 (3.37)	0.196 (13.0)	0.113 (1.789)
I/σ(I)	5.6	19.4 (0.9)	13.1
$CC_{1/2}$	0.989 (0.445)	0.99 (0.352)	0.998 (0.653)
Completeness (%)	100 (100)	100 (100)	100 (100)
Redundancy	17.9 (19.5)	72.0 (63.7)	17.7 (13.7)
No. reflections	12819 (1592)	90013 (4419)	48732 (2362)
Refinement			
Resolution (Å)	2.64	1.34	1.65
No. reflections	12078	89977	46196
Rwork / Rfree	0.2566/0.2897	0.1970/0.2196	0.2607/0.2841
Average B-factor (Å^2)	45.3	19.1	30.3
Ligand B-factor (Å^2)	82.2	19.2	25.1
RMSD: bonds (Å)	0.004	0.009	0.005
RMSD: bond angles (°)	1.266	1.067	1.301
Ramachandran favoured (%)	94	98	96
Ramachandran outliers (%)	1.1	0	0.3

PDB code	XX02KALRNA-x5178	XX02KALRNA-x5429
Ligand	**112**	**115**
Beamline	I04-1	I04-1
Wavelength (Å)	0.91119	0.91587
Space group	P 6_5 2 2	P 6_5 2 2
a, b, c (Å)	62.371,63.371,339.355	62.672,62.672,339.827
α, β, γ (°)	90, 90, 120	90, 90, 120
Data collection		
Resolution	56.52 – 1.68 (1.72 – 1.68)	56.65 – 2.15 (2.19 – 2.15)
R_{merge}	0.121 (1.287)	0.431 (6.942)
$I/\sigma(I)$	12.0	6.1
$CC_{1/2}$	0.991 (0.789)	0.996 (0.175)
Completeness (%)	100 (100)	100 (100)
Redundancy	17.6 (13.2)	36.6 (35.8)
No. reflections	46349 (2229)	22669 (1059)
Refinement		
Resolution (Å)	1.68	2.15
No. reflections	43933	21424
Rwork / Rfree	0.2534/0.2852	0.2560/0.2821
Average B-factor (Å^2)	29.0	46.7
Ligand B-factor (Å^2)	23.2	58.1
RMSD: bonds (Å)	0.005	0.003
RMSD: bond angles (°)	1.276	1.005
Ramachandran favoured (%)	96	97
Ramachandran outliers (%)	0	0.9

8.11 RSCC Analysis of Switch I Loop Rearrangement

These graphs are representative of the analysis conducted by Elliot Nelson of Frank von Delft's group to investigate the rearrangement of the switch I loop. All crystals datasets were solved using Dimple[271] with three different homologous models that varied in the switch I region. The first was the 'normal' homologous model that had the switch I loop organised in its orientation observed in the solved crystal structure of Kalirin/Rac1 (PDB code: 5O33). The second contain the loop only represented in the reorganised conformation which was established in the crystal structure of Kalirin/Rac1 bound to **99**. The third model contained a superposition of these two loop orientations. The RSCC values of the loop for each dataset were then plotted for the switch residues 24-36, with values closest to one predicted to provide the best model for the loop in that dataset. The RSCC values for the 'original' loop are indicated by a dark blue line, the 'rearranged' loop by a green line and the superimposed by light blue ('original') and red ('rearranged') lines.

8.11.1 Dataset with the Switch I Loop in the ‘Original’ Conformation

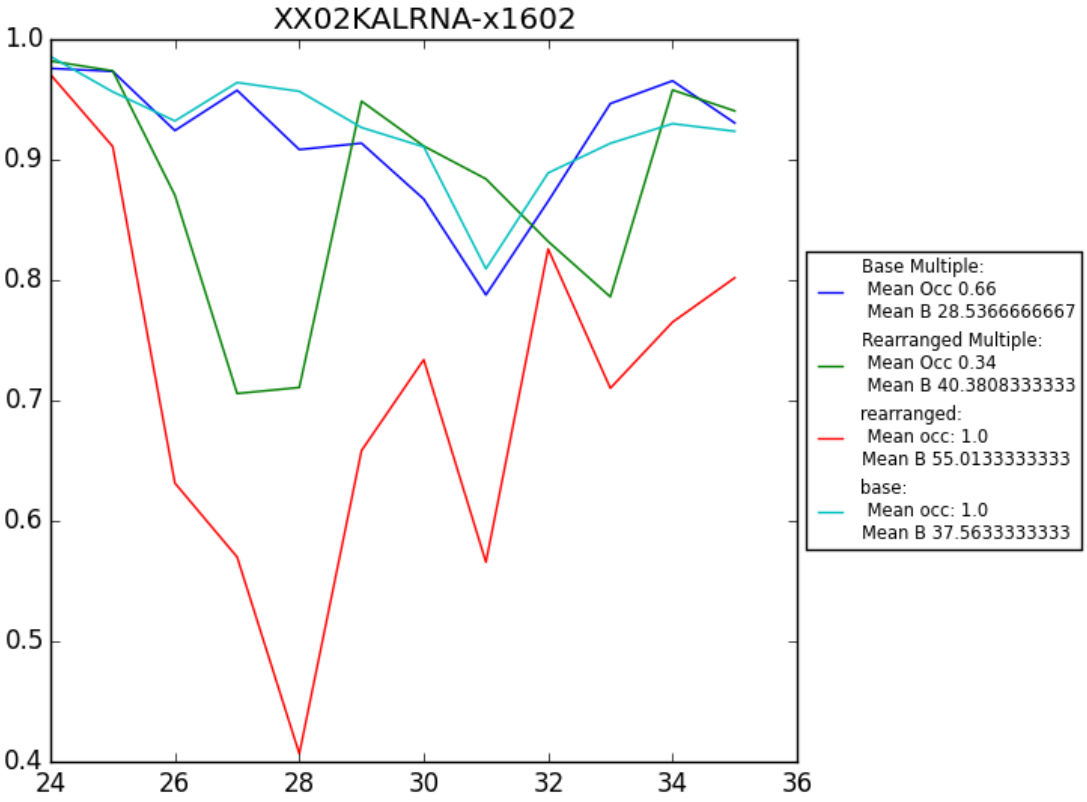


Figure 94. This crystal dataset has values of the ‘original’ loops (dark blue and light blue) close to 1. This indicates the ‘original’ orientation of the switch I region is the best fit for the electron density.

8.11.2 Dataset with the Switch I Loop in the ‘Rearranged’ Conformation

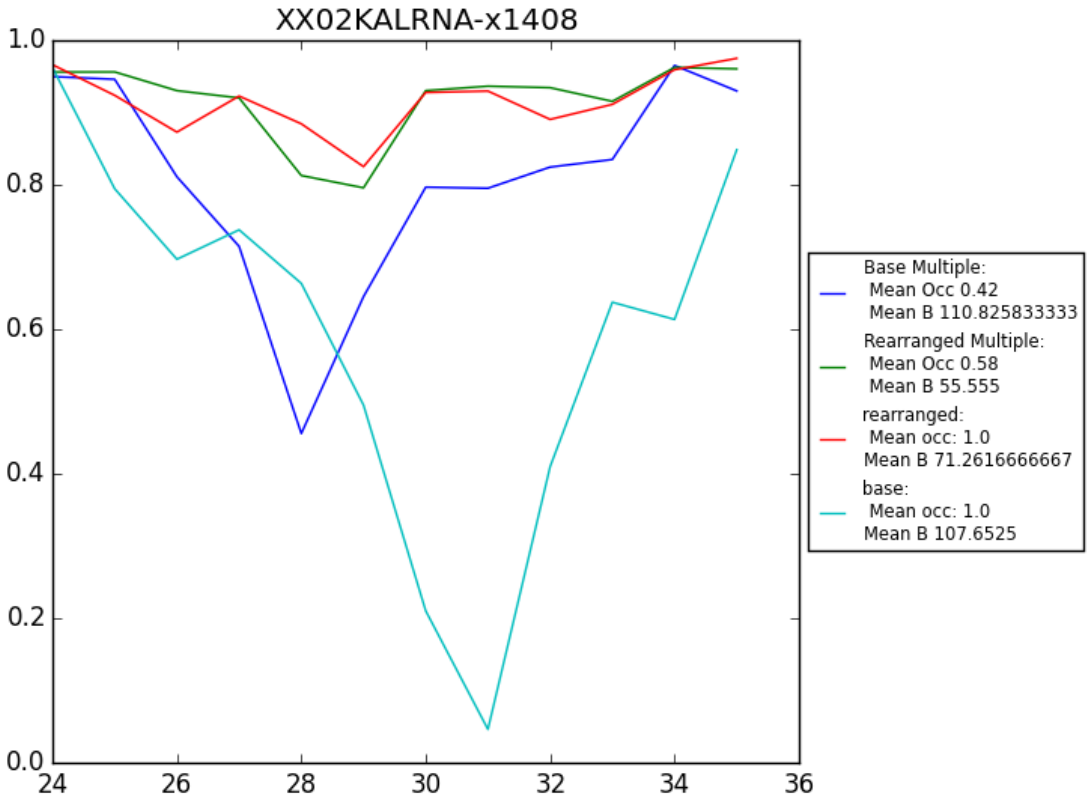


Figure 95. This crystal dataset has values of the ‘rearranged’ loops (red and green) close to 1. This indicates the ‘rearranged’ orientation of the switch I region is the best fit for the electron density.

8.11.3 Dataset with the Switch I Loop in a Mixture of Conformations

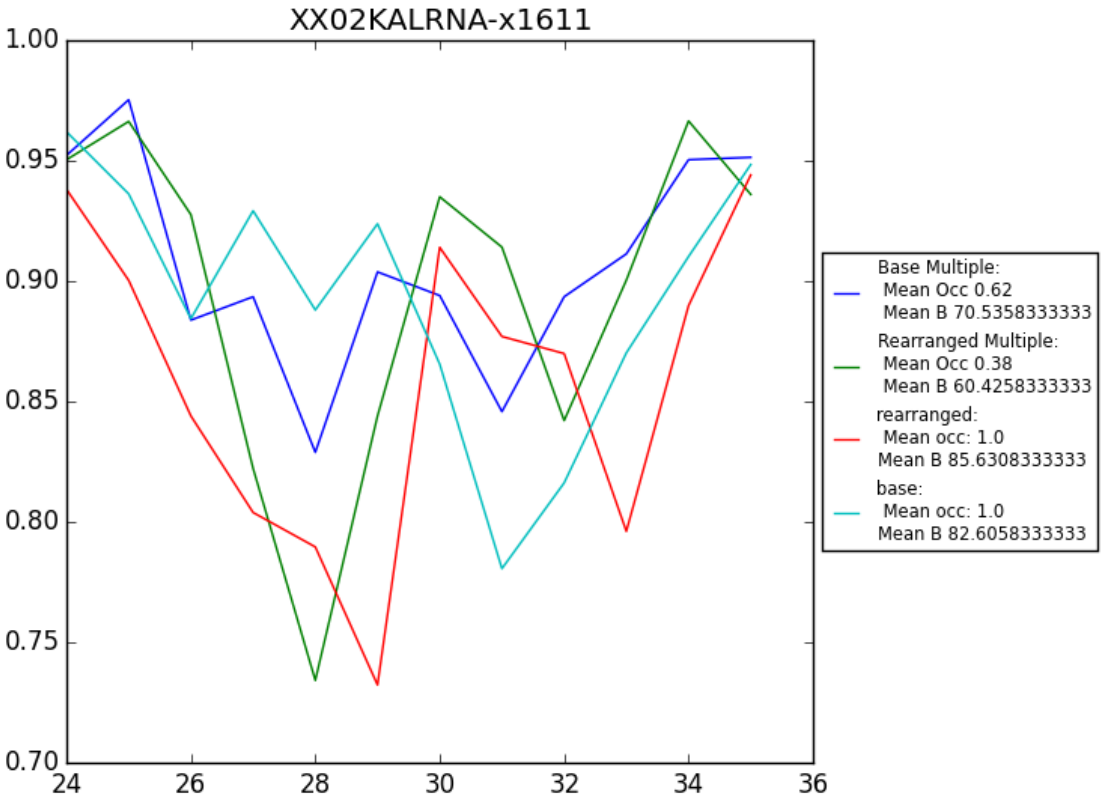


Figure 96. In this crystal, neither the 'original' nor 'rearranged' conformations are a particularly good fit for the dataset. This probably indicates that both conformations exist within the crystal, leading to overlapping electron density which is difficult to fit to a model.

8.12 Dose response assay for 106 and 107

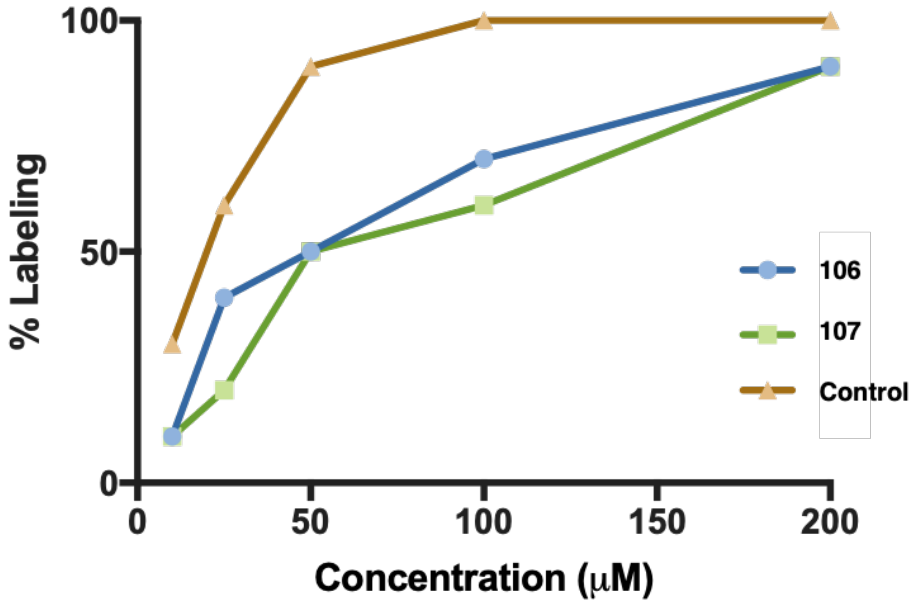


Figure 97. Dose-response assay conducted by the London group. Both **107** and **107** label Kalirin/Rac1, but not as rapidly as the positive control (known promiscuous indoline binder).

Table 11. Reactivity data conducted by the London group. **107** and **106** show similar low reactivity against an in-house protein panel to a selective negative control compared to the reactive indoline positive control.

Compound	**K ($M^{-1}m^{-1}$)**
Positive Control	7.74×10^{-5}
107	1.62×10^{-5}
106	4.10×10^{-6}
Negative Control	3.38×10^{-6}

8.13 Potency Determination over 24 hours for Covalent Compounds

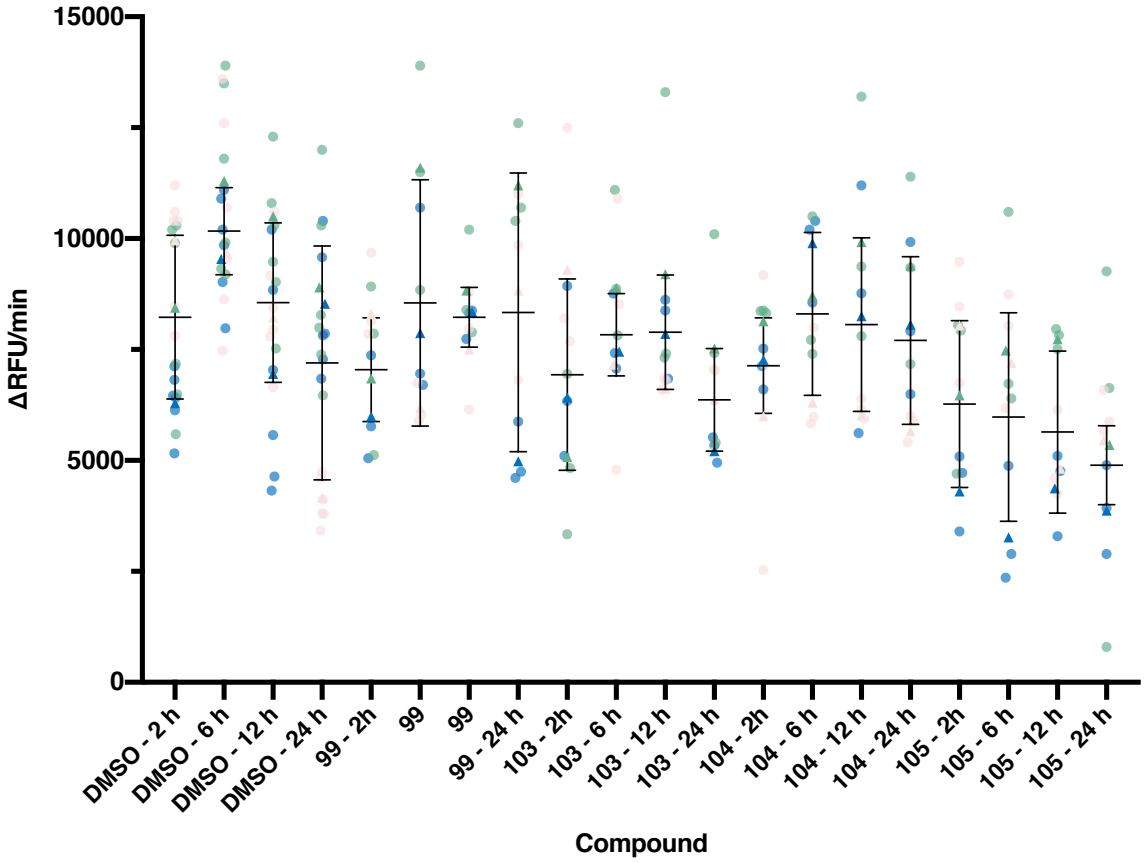


Figure 98. Nucleotide exchange assay over 24 hours for **99**, **103-105** (n = 3, *** $p < 0.001$, one-way ANOVA followed by Dunnett's multiple comparisons test). Blue, green and pink circles represent three technical replicates within each repeat. Triangles represent the mean for each repeat. Black bars represent the overall mean ± SD. All results were normalised using the positive control (GppNHp). No compounds exhibited significant inhibition of nucleotide exchange compared to the negative control.

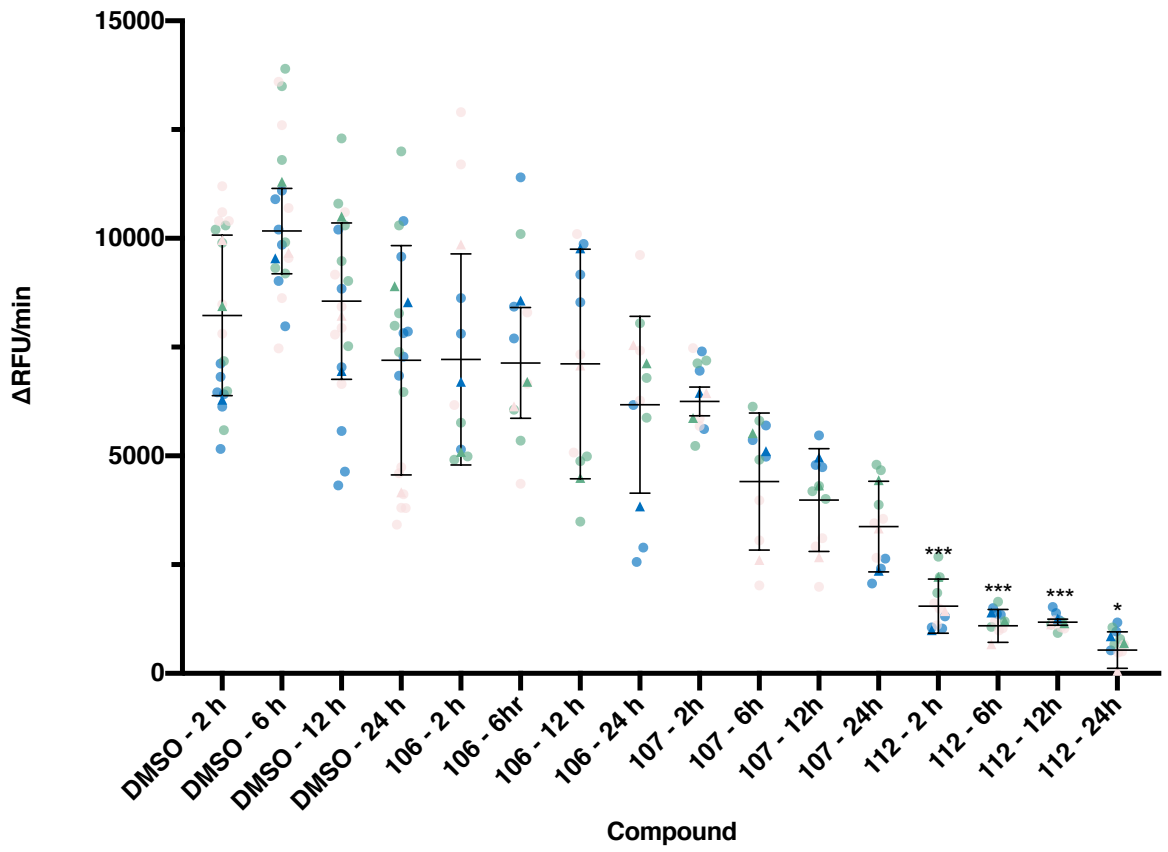


Figure 99. Nucleotide exchange assay over 24 hours for **106**, **107** and **112**. (n = 3, *** $p < 0.001$, one-way ANOVA followed by Dunnett's multiple comparisons test). Blue, green and pink circles represent three technical replicates within each repeat. Triangles represent the mean for each repeat. Black bars represent the overall mean ± SD. All results were normalised using the positive control (GppNHp). **112** exhibited significant inhibition of nucleotide exchange compared to the negative control.

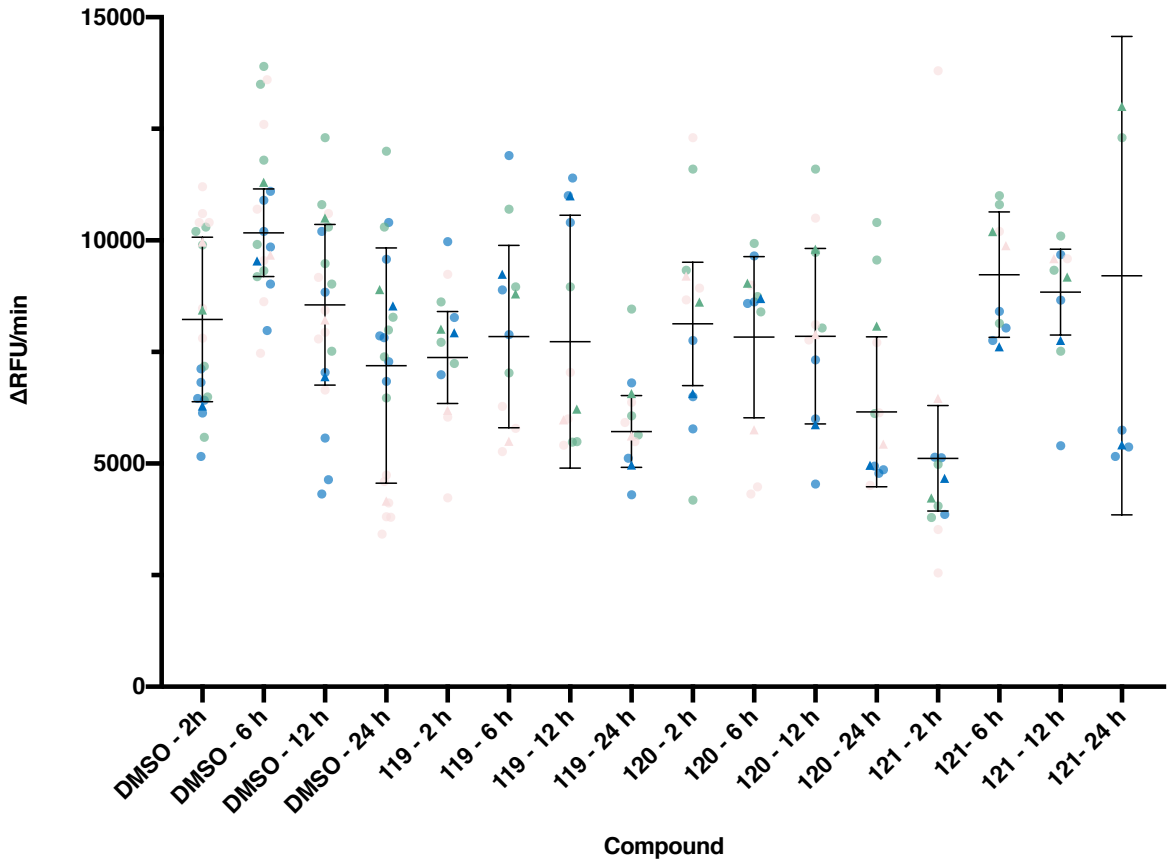


Figure 100. Nucleotide exchange assay over 24 hours for **119-121** (n = 3, *** $p < 0.001$, one-way ANOVA followed by Dunnett's multiple comparisons test). Blue, green and pink circles represent three technical replicates within each repeat. Triangles represent the mean for each repeat. Black bars represent the overall mean ± SD. All results were normalised using the positive control (GppNHp). No compounds exhibited significant inhibition of nucleotide exchange compared to the negative control.

8.14 Data Collection and Refinement Statistics for the Pharmacophore XChem Screen

Xtal ID	XX02KALRNA-x5009	XX02KALRNA-x5035
Ligand	**135**	**136**
Beamline	I04-1	I04-1
Wavelength	0.91188	0.91188
Space group	P 6_5 2 2	P 6_5 2 2
a, b, c (Å)	62.06,62.06,338.960	62.11,62.11,340.360
α, β, γ (°)	90, 90, 120	90, 90, 120
Data collection		
Resolution	56.49 – 2.01	53.82 – 1.95
	(2.04 – 2.01)	(1.98 – 1.95)
R_{merge}	0.197 (2.264)	0.201 (2.497)
I/σ(I)	10.1 (1.1)	10.4 (1.0)
$CC_{1/2}$	0.997 (0.702)	0.998 (0.421)
Completeness (%)	100 (100)	100 (100)
Redundancy	18.8 (19.9)	18.4 (18.4)
No. reflections	27267 (1297)	29926 (1425)
Refinement		
Resolution (Å)	2.01	1.95
No. reflections	25789	28290
Rwork / Rfree	0.2064/0.2653	0.2159/0.2717
Average B-factor (Å^2)	41.7	37.4
RMSD: bonds (Å)	0.008	0.008
RMSD: bond angles (°)	1.496	1.508
Ramachandran favoured (%)	94	96
Ramachandran outliers (%)	0.85	0.57

Xtal ID	XX02KALRNA-5017	XX02KALRNA-x5019
Ligand	**137**	**138**
Beamline	I04-1	I04-1
Wavelength	0.91188	0.91188
Space group	P 6_5 2 2	P 6_5 2 2
a, b, c (Å)	61.65,61.65,339.97	62.06,62.06,339.92
α, β, γ (°)	90, 90, 120	90, 90, 120
Data collection		
Resolution	56.66 – 2.26	56.65 – 1.96
	(2.30 – 2.26)	(1.99 – 1.96)
R_{merge}	0.356 (3.449)	0.168 (1.950)
$I/\sigma(I)$	7.6 (0.9)	12.4 (1.4)
$CC_{1/2}$	0.996 (0.550)	0.998 (0.508)
Completeness (%)	94.3 (99.7)	100 (100)
Redundancy	18.4 (19.1)	17.9 (18.6)
No. reflections	17629 (829)	29418 (1440)
Refinement		
Resolution (Å)	2.26	1.96
No. reflections	17142	27804
Rwork / Rfree	0.2111/0.3046	0.2157/0.2768
Average B-factor (Å^2)	47.3	38.3
RMSD: bonds (Å)	0.007	0.008
RMSD: bond angles (°)	1.495	1.508
Ramachandran favoured (%)	95	93.5
Ramachandran outliers (%)	0.57	0.3

8.15 Potency Determination for Compounds Purchased from Enamine

Table 12. Nucleotide exchange assay for compounds ordered from Enamine from the *in silico* pharmacophore screen. The compound numbers for compounds mentioned in this thesis are shown in brackets following the ID provided by Enamine. Compounds were tested at 500 μM (n = 3, one-way ANOVA followed by Dunnett's multiple comparisons test) and normalised using the positive control (GppNHp). The means and SD for technical repeats are given for each repeat. The p value in comparison to DMSO control is shown, with *** representing statistically highly significant results.

Enamine ID (ref. in thesis)	**SMILES**	**Mean ΔRFU/min n=1**	**Mean ΔRFU/min n=2**	**Mean ΔRFU/min n=3**	**p Value**
DMSO		16967 ± 3441	17128 ± 6567	20300 ± 3753	
PB4058316626	CSC1=NC(N)=C(N=O)C(=O)N1	17533 ± 2108	15667 ± 2558	22033 ± 13413	>0.99
Z104341146 (**139**)	NNC=1C=NNC(=O)C1Cl	8823 ± 1818	10820 ± 1366	11133 ± 208	<0.001 (***)
Z1186073732	CC=1C=C(CNCC(O)C(C)(C)C)NN1	12867 ± 2150	16967 ± 1457	14967 ± 1474	0.92
Z1192665747	CC=1C=C(ON1)C(=O)NCCC#C	23400 ± 2951	19100 ± 2516	19000 ± 625	>0.99
Z1197993156	CN1C=CNC(=O)C1=O	14267 ± 2558	13700 ± 2402	16133 ± 2329	0.82
Z1198160903	CC1=NC(C#N)=C(N)O1	16333 ± 2610	16567 ± 1767	19433 ± 3066	>0.99
Z135511924	CCNC(=O)C1=CC(C)=NO1	14533 ± 2868	15433 ± 777	16367 ± 2178	0.98
Z1366325239	CC=1C=C(ON1)C(=O)NCC(F)F	23367 ± 3591	23467 ± 1002	17900 ± 5268	0.86
Z1366326155	FC(F)CNC(=O)C1=CC=C(Cl)O1	17333 ± 2369	18767 ± 1762	19100 ± 625	>0.99

Enamine ID (ref. in thesis)	SMILES	Mean ΔRFU/min n=1	Mean ΔRFU/min n=2	Mean ΔRFU/min n=3	p Value
Z1367690842	FC(F)CNC(=O)C1=CN=CS1	14667 ± 3450	16767 ± 950	18333 ± 987	>0.99
Z138104752	Cl.CNCC(=O)NC1=NC=CS1	14667 ± 3528	14000 ± 458	17433 ± 2754	0.88
Z138508810	CC(C)N(C)CC(=O)NC1=NC=CS1	17800 ± 3559	16800 ± 2052	20400 ± 265	>0.99
Z1492698428	Cl.OCCN1CCC(=O)CC1	21800 ± 1997	19267 ± 5254	20167 ± 2715	>0.99
Z1587358631	CNC(=O)NCC1CCC(O1)C(=O)N(C)C	17067 ± 5292	21800 ± 300	18533 ± 3754	>0.99
Z1591782821	CN1CC(=O)NCC1=O	18800 ± 954	17967 ± 833	20633 ± 2663	>0.99
Z1630952063	CC(C)NCCC1=NNC(C)=N1	12933 ± 3496	15333 ± 1914	17467 ± 2458	0.95
Z1646602013	C[C@H](NCC(C)(C)O)C=1C=CC=NC1	14200 ± 1652	13200 ± 2972	13500 ± 889	0.24
Z1674923678	COC(C)C(=NOC1=NN=C(C)N1C)N	14433 ± 1026	16233 ± 777	13933 ± 1553	0.83
Z1683068139	O=C1NN=C2CNCCN21	18733 ± 1234	16633 ± 3126	17933 ± 1701	>0.99
Z1728989098	OCC(=O)NCC1(O)CCOCC1	16933 ± 4461	16267 ± 603	15000 ± 608	>0.99
Z1741975957	O[C@H]1CCOC1	15400 ± 854	13033 ± 2386	16700 ± 3780	0.9
Z1741979723	NC1=NC=2N(C=NC2C(=O)N1)[C@@H]3O[C@H](CO)[C@@H](O)[C@H]3O	13833 ± 1858	15567 ± 1537	17333 ± 3544	0.98
Z1741982161	Cl.CC1(N)COC1	16967 ± 1845	13800 ± 2227	22700 ± 1952	>0.99
Z1813248089	CCC1=NNC(C(=O)N)=C1Br	16067 ± 3620	18500 ± 4952	15700 ± 500	>0.99

Enamine ID (ref. in thesis)	SMILES	Mean ΔRFU/min n=1	Mean ΔRFU/min n=2	Mean ΔRFU/min n=3	p Value
Z1845245742	CN(CC#C)CC(=O)NC1=CC(C)=NO1	15367 ± 2818	15800 ± 1682	17467 ± 1069	>0.99
Z1845253275	CN(CC#C)CC(=O)NC1=NC=CS1	15367 ± 4285	14233 ± 929	15967 ± 2113	0.94
Z1873163200	O=C(CCC#C)NC1=NC=CS1	17467 ± 3403	22367 ± 1365	22167 ± 404	>0.99
Z1891776005 (**135**)	CN(C)C=1N=CNC(=O)C1N	14000 ± 3477	20367 ± 5273	17167 ± 4255	>0.99
Z1937730174	CC(O)(CC(=O)NC=1C=CC=NC1)C=C	21600 ± 1836	19967 ± 5727	20367 ± 2466	>0.99
Z195968534	CN(CC#C)CC(=O)NC=1C=C(C)ON1	14033 ± 3412	17567 ± 3287	15167 ± 462	0.98
Z1967989541	OCC1=CC=C(CNCC2(CC2)C#N)O1	13867 ± 1457	13767 ± 2669	12333 ± 929	0.29
Z203038534	NC1=NCC(=O)N1	16133 ± 58	11670 ± 2631	12467 ± 2122	0.21
Z2071714802	CC(NCCO)C=1C=CC(=O)NN1	17200 ± 625	15733 ± 1041	19200 ± 1952	>0.99
Z2143386446	CCNC(=O)C1=CN(C)N=N1	17467 ± 3402	16333 ± 2458	18967 ± 2082	>0.99
Z2146602927	CN1C=C(N=N1)C(=O)NCC(F)F	20800 ± 458	21033 ± 1422	17433 ± 1704	>0.99
Z2163251115	CC(O)C1=NC=C(CNC2CSC2(C)C)S1	14800 ± 2946	18467 ± 5947	19067 ± 2970	>0.99
Z217302304	CC1=CC(NC(=O)CCCl)=NO1	14633 ± 3927	16000 ± 3820	16300 ± 1908	>0.99
Z2303661324	ClC(Cl)CNC(=O)C1=NC=CS1	13067 ± 4629	14467 ± 603	15900 ± 872	0.65
Z2303662513	CC1=NC(=CO1)C(=O)NCC(Cl)Cl	13933 ± 3365	15233 ± 2196	18033 ± 833	>0.99

Enamine ID (ref. in thesis)	SMILES	Mean ΔRFU/min n=1	Mean ΔRFU/min n=2	Mean ΔRFU/min n=3	p Value
Z2409839661	COC(=O)COC1CN(C1)C(=O)C	14267 ± 2928	14967 ± 1537	16767 ± 1320	0.97
Z2442944940	CC1CC(CO1)NC=2C=CC(C)=NC2	15600 ± 2343	16800 ± 1778	15033 ± 3879	>0.99
Z247483960	O=C(NC1=NN=CS1)C=2C=CC=NC2	16967 ± 3439	20333 ± 4043	20133 ± 1818	>0.99
Z2605109997	OC[C@H](NCC(O)C=1C=CC=CN1)C=C	17233 ± 1305	13900 ± 1587	15733 ± 611	0.98
Z2617736988	CCCCNC(=O)C1=NC=CN1	13133 ± 379	22100 ± 6691	21800 ± 1670	>0.99
Z2762999965	CNC(=O)NCC1COCCN1C(=O)NC	13200 ± 1277	16267 ± 2021	17433 ± 1450	0.98
Z28173473	CC(NC(=O)C)C(=O)NC1=NC=CS1	16133 ± 5265	17933 ± 1629	18700 ± 3378	>0.99
Z28173619	CC(C)(C)C(=O)NC1=NC=CS1	18467 ± 569	21667 ± 4067	21600 ± 3830	>0.99
Z2877726607	CN1CCCC1C=2C=C(NN2)C(=O)N	17767 ± 3694	22233 ± 3225	20633 ± 2285	>0.99
Z2897929738	CC(C)(C)C(=O)NOCC=1C=NSN1	14667 ± 3197	19133 ± 2250	19867 ± 3134	>0.99
Z2904479047	CC[C@@H](N(C)C)C(=O)NC1=NC=CS1	10220 ± 624	18367 ± 4148	17933 ± 815	>0.99
Z2915539531	CC(=O)N1CC(O)(C1)C(=O)N	15700 ± 2955	14767 ± 3427	15400 ± 1127	0.96
Z2941304988	FC[C@@H]1CO1	18733 ± 1858	17667 ± 5783	16067 ± 208	>0.99
Z2945855977	NC(=O)C1=CN=C2SCCN12	18100 ± 1952	18167 ± 1097	18267 ± 3710	>0.99
Z3045518085	CCC(F)C(=O)NC1=NC=CS1	16533 ± 6242	20267 ± 2610	19100 ± 625	>0.99
Z32023663	CCNC(=O)C=1C=C(C)ON1	16500 ± 794	16200 ± 1153	18333 ± 1168	>0.99

Enamine ID (ref. in thesis)	SMILES	Mean ΔRFU/min n=1	Mean ΔRFU/min n=2	Mean ΔRFU/min n=3	p Value
Z3205899795	OCCOC1(CNC(=O)C2=NC=CN2)CCSC1	20000 ± 3035	17267 ± 2444	24400 ± 2706	>0.99
Z3209685851	OCC1(CNC(=O)C2=NC=CN2)CC31CCOCC3	17700 ± 6332	24933 ± 7433	18733 ± 2259	>0.99
Z3219876634	NC(=S)CN1C(=O)CNC1=O	13833 ± 586	15400 ± 2651	14067 ± 2743	0.5
Z3227123974	CC(C)(CN)OCCO	12567 ± 1888	11160 ± 3086	17000 ± 1114	0.33
Z3228476506	CCN1C(N)=C(N)C(=O)NC1=O	16067 ± 2113	13733 ± 3523	20067 ± 4222	>0.99
Z3228760866	OCCN1C=C(CNC2CC2)C=N1	14133 ± 2303	17267 ± 1601	18667 ± 851	>0.99
Z3229530769	CC=1NN=C(C(=O)NC=2C=CC=CC2)C1N	19700 ± 1562	14703 ± 4459	19100 ± 2193	>0.99
Z3230564087	NC(=O)CNC=1N=CN=C(Cl)C1N	17733 ± 1739	15433 ± 1060	17733 ± 322	>0.99
Z3234892687	COC[C@@H]1CO1	16100 ± 1253	14567 ± 1877	15353 ± 5679	>0.99
Z3235020306	Cl.Cl.CC=1N=C2CNCC2=C(O)N1	15467 ± 4304	17133 ± 2050	18633 ± 3427	>0.99
Z3235023364	NC(=O)C=1N=CNN1	18300 ± 2152	17033 ± 839	28800 ± 4403	0.93
Z3235031500	O=C1NC=NC=2CCNCC21	21033 ± 4450	17967 ± 4565	19100 ± 866	>0.99
Z3235901391	OC[C@@H]1CO1	18267 ± 3225	19467 ± 2616	15800 ± 3000	>0.99
Z3235903663	OC[C@@H]1CNC(=O)CO1	13433 ± 2810	16200 ± 1732	18067 ± 2031	>0.99
Z3242019008	Cl.Cl.NC1=CNNC1=O	12307 ± 3040	14033 ± 3785	14010 ± 8772	0.23
Z3244890611	Cl.Cl.CN1C=NC(=C1)C(=O)NN	10867 ± 603	16500 ± 2152	14233 ± 2136	0.33

Enamine ID (ref. in thesis)	SMILES	Mean ΔRFU/min n=1	Mean ΔRFU/min n=2	Mean ΔRFU/min n=3	p Value
Z3339765112	CC(CN(C)C)C(=O)NC1=NC=CS1	16267 ± 3868	18633 ± 322	17533 ± 379	>0.99
Z3359241542	CCC(C)(O)CNC(C)C1=CC(C)=NS1	14800 ± 2443	17200 ± 794	15600 ± 2784	>0.99
Z3359278613	CC(C)C(O)CNC(C)C1=CC(C)=NS1	12233 ± 3326	15367 ± 252	16833 ± 1436	0.92
Z3363181425	CC(F)(F)CNCC1=NC(CO)=CS1	14367 ± 1650	15467 ± 3044	15633 ± 1429	0.94
Z3363183005	CC(F)CCNCC=1C=CC=NN1	16033 ± 5523	18167 ± 5607	17300 ± 1136	>0.99
Z3363286159	O=C(CCC#C)NC=1C=CSN1	15933 ± 5132	18167 ± 2754	17367 ± 2026	>0.99
Z3363302081	CC(C)C(=O)NC=1C=CSN1	16267 ± 1930	16600 ± 1836	18967 ± 2873	>0.99
Z3363304847	CCN(CC)CC(=O)NC=1C=CSN1	17867 ± 1767	21767 ± 6918	21867 ± 1815	>0.99
Z3363306927	CCCN(C)CC(=O)NC=1C=CSN1	20000 ± 2022	17567 ± 2919	17400 ± 2691	>0.99
Z3503409087	CC=1SN=CC1CNCCCC#C	12667 ± 1387	22267 ± 2060	17067 ± 1159	>0.99
Z3573589712	CC(O)COC(=O)[C@@H]1C[C@@H](O)CN1C(=O)C	14933 ± 2403	15733 ± 874	17300 ± 872	>0.99
Z3624416697	COC(=O)[C@@H]1C[C@@H](O)CN(C1)C(=O)COl&1:4,6,rl	13433 ± 3250	15833 ± 1137	15633 ± 1012	0.93
Z3630402116	CNC(=S)NCC1(CC2(C1)C(=O)N)OC2	20300 ± 5323	20700 ± 2666	17500 ± 2972	>0.99

Enamine ID (ref. in thesis)	SMILES	Mean ΔRFU/min n=1	Mean ΔRFU/min n=2	Mean ΔRFU/min n=3	p Value
Z3644471598	CS(=O)(=O)N1CCCCNS(=O)(=O)CC1	16500 ± 2088	15100 ± 2352	13000 ± 1082	0.87
Z3649151144	COCC(O)(CO)CNC(=O)CO	19533 ± 5493	18800 ± 1609	16360 ± 7977	>0.99
Z3910572601	Cl.COC[C@H](N)CC(=O)OC	15667 ± 1350	14047 ± 4279	13567 ± 945	0.75
Z3915518427	Cl.COC(=O)C=1C=NN2CCNCC12	16133 ± 1550	20133 ± 3787	16517 ± 6341	>0.99
Z4054722814	CC=1C=CC(NC=2C=NN(CC#C)C2)=NN1	12800 ± 2427	13833 ± 907	14767 ± 379	0.36
Z4058212273	OCCC1=CN(N=N1)C=2C=CC=C(F)N2	14900 ± 3732	16333 ± 3153	21100 ± 4158	>0.99
Z4058224877	Br.CN1C=C(C=N1)C(O)CNC(=N)N	16167 ± 2775	16767 ± 1464	15167 ± 2930	>0.99
Z4058238910	ClC(=C)CNC(=O)C=1C=C(Br)NN1	15833 ± 4244	18167 ± 603	17133 ± 2212	>0.99
Z4058239009	CC=1N=C(O)C=C(CNC2(CC2)C3CC3)N1	15467 ± 3356	16567 ± 1557	20867 ± 751	>0.99
Z4058239015	CC=1N=C(O)C=C(N1)C(F)(F)F	18100 ± 872	13967 ± 1290	15003 ± 7634	0.98
Z4058239021	CC=1N=C2CN(CCOCC(F)F)CC2=C(O)N1	15367 ± 3530	12300 ± 1758	18000 ± 608	0.95
Z4058239084	CN1CCC=2NC(=O)C(CN)=CC2C1	18167 ± 3386	16333 ± 2499	21833 ± 2558	>0.99
Z4058239110	COCCOCCCN1CC=2N=C(C)N=C(O)C2C1	12967 ± 1595	14100 ± 2821	13833 ± 643	0.31

Enamine ID (ref. in thesis)	SMILES	Mean ΔRFU/min n=1	Mean ΔRFU/min n=2	Mean ΔRFU/min n=3	p Value
Z4058239125	COC(=O)CC1(CN2CC=3N=C(C)N=C(O)C3C2)CC1	16700 ± 819	19167 ± 833	19200 ± 2784	>0.99
Z4058239143	COC1(CCOCC1)C(F)(F)CNC(=O)C2=NC=CN2	16833 ± 4535	14933 ± 1069	16600 ± 1652	>0.99
Z4058239167	NC(=NOCCCO)CCO	16200 ± 1389	17033 ± 2237	13467 ± 493	0.98
Z4058239179 (**137**)	NC=1C=C(N)C(C#N)=C(S)N1	14267 ± 551	12467 ± 1890	16267 ± 3007	0.65
Z4058239185	Cl.NCC=1N=CC=NC1O	13533 ± 603	12733 ± 1644	12800 ± 2762	0.15
Z4058239311	OCC1(CNC(=O)C=2NN=CC2Cl)CCOCC1	15033 ± 2108	15100 ± 1300	15733 ± 503	0.98
Z4058239340 (**138**)	OC=1C=C(O)C=2CCCNC2N1	4733 ± 1209	4007 ± 455	4943 ± 1468	<0.001 (***)
Z4058239350	ClC1=NC(=O)C=2C=NNC2N1	17700 ± 1646	13400 ± 2400	17233 ± 2003	>0.99
Z4058239357 (**136**)	O=C1NC=NC=2NC=CC12	16500 ± 1058	13200 ± 656	16867 ± 1595	0.96
Z4058239373	O=C1NN=NC=2NC=NC12	18500 ± 2536	18867 ± 4536	18533 ± 116	>0.99
Z4061381396	ClC=CCNC(=O)C1=CN=CN1	15033 ± 2658	16033 ± 1890	16667 ± 987	>0.99
Z4061389226	CC(CNC(=O)[C@@H]1CC[C@H](CN2C=NC=3C(N)=NC=NC23)O1)S(=O)(=O)N I&1:6,9,rl	21133 ± 1620	21067 ± 1069	19900 ± 3158	>0.99
Z449369066	CC=1C=CC(NCCC(=O)O)=NC1	16333 ± 379	16167 ± 1050	23133 ± 1201	>0.99

Enamine ID (ref. in thesis)	SMILES	Mean ΔRFU/min n=1	Mean ΔRFU/min n=2	Mean ΔRFU/min n=3	p Value
Z46190592	Cl.CNCC(=O)NC =1C=C(C)ON1	11967 ± 1387	14133 ± 2307	18100 ± 1386	0.76
Z927753474	Cl.Cl.NCC(=O)NC =1C=CC=CN1	15333 ± 2483	14367 ± 322	8087 ± 1313	0.13

8.16 General Biology Experimental

8.16.1 Construct Design

8.16.1.1 Construct Design for Round 1 of Pipeline

The boundaries for the expression constructs were chosen based on structures previously crystallised in the PDB, either as GEF/GTPase complexes (Table 13) or the proteins individually (Table 14).

Table 13. GEFs and GTPase which were documented as complexes in the PDB. Their exact construct length was purchased.

GEF	GTPase	PDB code
Sos1	HRAS	1BKD
EPAC2	Rap1B	3CF6
Trio	Rac1	2NZ8
Vav1	Rac1	2VRW
Tiam1	Rac1	1FOE
PRex1	Rac1	5FI0
DOCK2	Rac1	2YIN
DOCK9	Cdc42	2WMN
ITSN1	Cdc42	1KI1
Dbs	RhoA	1LB1
NET1	RhoA	4XH9
ARHGEF11	RhoA	1XCG
ARHGEF12	RhoA	1X86
AKAP13	RhoA	4D0N
MSS4	Rab8	2FU5
DENND1B	Rab35	3TW8
Rabex-5	Rab21	2OT3
RCC1	Ran	1I2M
IQSEC1	Arf1	4C0A
Cyth3	Arf6	4KAX

Table 14. GEFs crystallised alone in the PDB. Their partner GTPase and their construct PDB code are also shown. *Indicates construct present in the PDB as an NMR structure, not present as an X-ray crystallography structure

GEF	PDB code	GTPase	PDB code
ARFGEF1	3LTL	Arf1	4C0A
ARFGEF2	3SWV	Arf1	4C0A
Cyth2	4JMI	Arf1	4C0A
Cyth1	4A4P	Arf1	4C0A
Kalirin*	2KR9	Rac1	2NZ8

For GEFs not present in the PDB, novel constructs were designed based on homologous proteins with high sequence similarity which had been previously crystallised as a complex in the PDB (Table 15)

Table 15. GEFs designed using structure of homologous protein.

GEF	Model GEF	PDB code	GTPase
IQSEC2	IQSEC1	4C0A	Arf1
IQSEC3	IQSEC1	4C0A	Arf1
DOCK3	DOCK2	2YIN	Rac1
Kalirin	Trio	2NZ8	Rac1

8.16.1.2 Construct Design for Round 2 of Pipeline

The GEF domain boundaries were determined using UniprotKB.[272] Bioinformatic servers Crysalis[187], DisMeta[188] or ProteinCCD[189] were used to predict the secondary structure of the relevant domains. Several constructs for each target were chosen by systematically

varying the ends of the N- and C- termini. The end termini were controlled to exclude low complexity regions, terminal hydrophobic residues and to avoid cutting into predicted secondary structures. The likelihood of construct crystallisation was predicted using ProteinCCD and was used to eliminate unfavourable constructs.

Details of all constructs are given in Appendix 8.1.

8.16.2 Molecular Cloning

All synthetic DNA and primers were obtained from Eurofin Genomics.[273] Vectors were obtained from Frank von Delft's group within the SGC. DNA constructs were amplified using PCR. Vectors and constructs were prepared for LIC cloning by creating suitable overhangs using T4 polymerase.

All GEFs were cloned into pNic28-His6 or pNic28-His10 vectors. All GTPases were cloned into pNic28-StIIT vectors. GEFs or GTPases which did not express were also cloned into pNic28-MBP2 vectors. All plasmids were transformed into *E. coli* BL21 DE3 Rosetta strain and stored in 50% glycerol in -80 °C.

8.16.3 Protein Expression and Purification

8.16.3.1 General PrepX Method for the Expression of Proteins

All proteins were overexpressed in *E. coli BL21* DE3 Rosetta cells in growth media (TB or autoinduction TB) supplemented to a final concentration with 20 g/L glucose, 0.01% Antifoam 204 (Sigma), 50 μg/mL kanamycin and 25 μg/mL chloramphenicol. TB was supplemented to a final concentration with additional 1 mM $MgSO_4$, 10 mM $(NH_4)SO_4$ and 0.5

% glucose. The cells were harvested by centrifugation then resuspended in the base buffer (10 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 0.5 mM TCEP) supplemented with 0.5 mg/ml Lysozyme, 0.1 mg/ml benzonase, and 1 in 1000 Protease Inhibitor Cocktail (Merck). 1 % Triton X-100 was added after 30 minutes. Lysis was completed using a freeze-thaw cycle at -80 °C. Imidazole was added to a final concentration of 20 mM post-lysis, the cells centrifuged and the supernatant collected. The supernatant was purified using affinity chromatography with gravity-flow columns. Tag cleavage was achieved with TEV protease overnight at 4 °C. The protein, the affinity tag and TEV protease were separated using reverse IMAC and the proteins further purified using size-exclusion chromatography with base buffer as an elutant (Yarra SEC-2000, Phenomenex). All successfully purified proteins were identified using time-of-flight mass spectroscopy (TOF-MS) and purity confirmed by SDS PAGE analysis.

8.16.3.2 Specifics for the 1st Subset

Proteins were expressed as per the General Procedure. Proteins were grown using Autoinduction TB media (Formedium) incubating at 37 °C for 4 h and then at 25 °C for 16 h. The GEFs were purified using nickel-affinity chromatography (His GraviTrap columns, GE Healthcare) and protein eluted using the base buffer containing imidazole to final concentration of 500 mM before application and elution from a PD-10 desalting column (GE Healthcare). The GTPases were purified using Strep-Tactin-XT columns (IBA Lifesciences) where protein was eluted with buffer containing 50 mM D-Biotin.

8.16.3.3 Specifics for the 2nd Subset

Proteins were expressed as per the General Procedure. All proteins were expression in TB media (Melford) incubating at 37 °C for 4 h and then at 18 °C for 12-16 h. All proteins

were purified using a streptavidin column where protein was eluted with buffer containing 50 mM D-Biotin.

8.16.3.4 Complex Formation

To fluorescently tag the proteins, 0.1 % Alexa-555 and Alexa-488 were added to the GEF and GTPase respectively and incubated for 1 h. Relevant GEF/GTPase pairs were mixed in a 1:2 ratio and left to bind for a minimum of 1 h or overnight at 4 °C. The complex was then purified from the excess GTPase using SEC in base buffer. Complexes were identified using TOF-MS.

8.16.4 GDP-Removal Methods

8.16.4.1 EDTA Method for IQSEC2/Arf1

EDTA was added to the complex to a final concentration of 10 mM and incubated at 4 °C overnight. The complex was purified using SEC in buffer containing 10 mM HEPES pH 7.5, 500 mM NaCl, 10 mM EDTA, 5% glycerol, 0.5 mM TCEP, concentrated to 15 mg/mL, flash frozen in liquid nitrogen and stored at -80 °C.

8.16.4.2 Alkaline Phosphatase Method for Kalirin/Rac1

Kalirin and Rac1 were mixed in a 1:1 ratio and incubated for 1 h before concentrating to 60 mg/mL. 40 µL of 10 mM β,γ-Methyleneguanosine 5'-triphosphate sodium salt (GppCp, Sigma) and 4 Units of Alkaline phosphatase (Roche) were added. 20 µL of the 10x exchange buffer (2M $(NH_4)_3PO_4$, 10 µM ZnCl) was added and the mixture incubated at 4 °C for 3 h. The final concentration of Kalirin/Rac1 was 1 mM. Degradation of GDP to GMP and guanosine was monitored by HPLC (Agilent 1200, reverse-phase C-18) over 15 minutes in buffer

containing 100 mM potassium phosphate pH 6.5, 10 mM tetrabutyl phosphonium bromide and 5% MeOH. Following complete degradation of GDP, 0.008 Units of snake venom phosphodiesterase I (Sigma) was added. The reaction was left overnight at 4 °C and the degradation of GppCp to GMP and guanosine monitored by HPLC. The protein was diluted to 30 mg/mL and purified by size-exclusion chromatography in base buffer. The complex was concentrated to 10.5 mg/mL and stored at -80 °C.

8.16.5 Crystallisation and Structure Determination

8.16.5.1 General Method

Proteins were crystallized using the sitting-drop vapour diffusion method in 96-well, 3 lens SWISSCI plates (SWISSCI AG). Proteins were dispensed in 2:1, 1:1 and 1:2 protein-to-precipitant ratios in 150 nL drops using the TTP Labtech's Mosquito Crystal liquid handling robot. Different coarse screens (JCSG, HCS, HIN, LFS, BCS, Morpheus) obtained from Molecular Dimensions[176] and Hampton Research[274] containing 96 different precipitants were used in each plate. Crystallisation plates were incubated at 20 and 4 °C and regularly imaged over a period of two weeks using an automatic inspection system (Rigaku). If insufficient protein was available, 20 °C and the screens JCSG, HCS and HIN were prioritised. Initial crystals were optimised using follow-up screens designed based on conditions in which crystals or crystalline conditions were observed. The use of microcrystal seeds and additive screens were also used to improve crystal quality. Crystals were cryo-protected in 25% ethylene glycol, and flash-cooled in liquid nitrogen. X-ray diffraction data was collected at Diamond Light Source and was processed using the Diamond autoprocessing pipeline, which utilises xia2[275], DIALS[276], XDS[277], POINTLESS[278] and CCP4[279]. The structures were phased and iteratively refined using PHENIX[280] or CCP4 suites with manual model correction performed using Coot.[281]

Table 16. Commercial custom-designed crystallisation screens purchased by the SGC/CMD.

Crystallisation screen	Supplier
Additive Screen	Hampton Research
Basic ChemSpace (BCS)	Molecular Dimensions
Crystal Screen HT (HCS)	Hampton Research
Index (HIN)	Hampton Research
JCSG+ (JSCG)	Molecular Dimensions
Ligand Friend Screen (LFS)	Molecular Dimensions
Morpheus	Molecular Dimensions

8.16.5.2 Kalirin-c001/Rac1 GDP-bound Structure (PDB code: 5O33)

Crystals of GDP-bound Kalirin/Rac1 were grown by sitting-drop vapour diffusion at 20 °C with 100 nL of protein (15.6 mg/mL) and 50 nL of reservoir solution consisting of 0.1 M citrate pH 4.2, 0.2 M NaCl and 18 % PEG 8K. Molecular replacement was conducted using PHASER from PHENIX software suite with Trio/Rac1 (PDB code: 2NZ8) as a homologous model. Refinement was conducted using Phenix.refine from the PHENIX software suite.[280]

8.16.5.3 IQSEC2/Arf1 Nucleotide-free Structure (PDB code: 6FAE)

1μM thiomersal was incubated with IQSEC2/Arf1 (15 mg/ml) at 4 °C for 1 hour. Crystals of heavy-atom soaked IQSEC2/Arf1 were grown by sitting-drop vapour diffusion at 20 °C with 100 nL of protein (15 mg/mL) and 50 nL of reservoir solution consisting of 0.1 M tris pH 8 and 14 % PEG 8K. SAD was conducted using the CCP4[279] and PHENIX[280] software suites to produce a basic model of IQSEC2/Arf1.

Crystals of nucleotide-free IQSEC2/Arf1 were grown by sitting-drop vapour diffusion at 20 °C with 50 nL of protein (15 mg/mL) and 100 nL of reservoir solution consisting of 0.1 M Tris pH 8, 0.1 M KCl and 16 % PEG 8K. Molecular replacement was conducted using PHASER

from PHENIX software suite using the SAD-determined structure of IQSEC2/Arf1 as a model. Refinement was conducted using Phenix.refine from the PHENIX software suite.

8.16.5.4 Kalirin-c001/Rac1 Nucleotide-free Structure

Crystals of nucleotide-free Kalirin/Rac1 were grown by sitting-drop vapour diffusion at 20 °C with 100 nL of protein (10.4 mg/mL) and 50 nL of reservoir solution consisting of 0.1 M bis-tris pH 5.5, 0.1 M NaCl and 24 % PEG 3350. The XChemExplorer[93] pipeline was used for structure solution and parallel molecular replacement using DIMPLE with GDP-bound Kalirin/Rac1 (PDB code:5O33) as a homologous model. Refinement was conducted using REFMAC software integrated into the XChemExplorer pipeline.

8.16.5 5 Kalirin-c002/Rac1 GDP-bound Structure

Crystals of GDP-bound Kalirin-c002 /Rac1 were grown by sitting-drop vapour diffusion at 20 °C with 100 nL of protein (12.5 mg/mL) and 50 nL of reservoir solution consisting of 0.1 M cacodylate pH 6.5, 0.2 M NaCl and 2 M NH_4SO_4. Molecular replacement was conducted using PHASER from PHENIX software suite with Trio/Rac1 (PDB:2NZ8) as a homologous model. Refinement was conducted using Phenix.refine from the PHENIX software suite.[280]

8.16.6 DSF Buffer Assay

IQSEC2/Arf1 was diluted to 4 mg/ml and 2 μL transferred to each well of a 96-well Multiplate™ PCR plate (Bio Rad). 16 μL of each well from the 96-well buffer screen investigating a range of buffers and pH was dispensed. 1000x concentrate of SYPRO® Orange protein gel stain (ThermoFisher Scientific) was diluted 1 in 100 in MilliQ water and 2 μL was dispensed to each well. The plate was sealed, centrifuged for 30 seconds and fitted

into a RT-PCR (Bio Rad). The temperature was increased from 25 to 95 °C in 1 °C increments every 30 seconds. The fluorescence was measured at each temperature increase using an LED/photodiode matched to the excitation/emission wavelengths of SYPRO orange. Data analysis was performed in Microsoft Excel and Graphpad Prism 9.0 using least squares regression analysis.[183,185]

8.16.7 Co-crystallisation of Kalirin/Rac1 with Covalent Ligands

8.16.7.1 General Method

The compound was incubated with Kalirin/Rac1 at 4 times the concentration of the complex with a maximum 2 % DMSO concentration at 4 °C for 16 hours. Labelling was tested using QTOF intact mass spectrometry. Labelled Kalirin/Rac1 crystals were grown using the general method detailed in section 8.16.5.1 using Kalirin/Rac1 fine screens with 96 precipitants optimised for GDP-bound and nucleotide-free Kalirin/Rac1 crystal formation. Seeding using Kalirin/Rac1 microcrystals was used in all experiments. Crystals were cryo-protected and flash-cooled in liquid nitrogen. Molecular replacement was conducted using PHASER from PHENIX software suite[280] or PHASER MR from CCP4 suite[282]. The covalent bond was modelled using JLigand[283] or REFMAC[284] in CCP4. ACEDRG[285] was used to generate compound coordinates. The structure was iteratively refined using Phenix.refine from the PHENIX software suite or using REFMAC from CCP4.

*8.16.7.2 Co-crystallisation of Kalirin/Rac1 with **99***

99 was incubated with nucleotide-free Kalirin/Rac1 (10.5 mg/mL) to a final concentration of 1 mM. Crystals were grown by mixing 75 nL of labelled Kalirin/Rac1 (10.5 mg/mL) with 75 nL of reservoir solution consisting of 0.1 M citrate pH 5.5 and 20% PEG 3K.

The crystals were cryo-protected using 20% DMSO. Diffraction data was collected at the I04-1 beamline at DLS. Molecular replacement was conducted using PHASER from PHENIX software suite. JLigand from the CCP4 software suite was used to install the covalent bond. The structure was iteratively refined using the PHENIX software suite.

*8.16.7.3 Co-crystallisation of Kalirin/Rac1 with **103***

103 was incubated with GDP-bound Kalirin/Rac1 (14.5 mg/mL) to a final concentration of 1.4 mM. Crystals were grown by mixing 100 nL of labelled Kalirin/Rac1 (14.5 mg/mL) with 50 nL of reservoir solution consisting of 0.1 M citrate pH 4.4, 0.8 M LiCl and 25% PEG 6K. The crystals were cryo-protected using 25% ethylene glycol. Diffraction data was collected at the I04-1 beamline at DLS. Molecular replacement was conducted using PHASER MR from CCP4 suite. ACEDRG and REFMAC from the CCP4 suite was used to install the covalent linkage. The structure was iteratively refined using REFMAC.

*8.16.7.4 Co-crystallisation of Kalirin/Rac1 with **104***

104 was incubated with GDP-bound Kalirin/Rac1 (14.5 mg/mL) to a final concentration of 1.4 mM. Crystals were grown by mixing 100 nL of labelled Kalirin/Rac1 (14.5 mg/mL) with 50 nL of reservoir solution consisting of 0.1 M citrate pH 4.4, 0.8 M LiCl and 25% PEG 6K. The crystals were cryo-protected using 25% ethylene glycol. Diffraction data was collected at the I04-1 beamline at DLS. Molecular replacement was conducted using PHASER MR from CCP4 suite. ACEDRG and REFMAC from the CCP4 suite was used to install the covalent linkage. The structure was iteratively refined using REFMAC.

*8.16.7.5 Co-crystallisation of Kalirin/Rac1 with **105***

105 was incubated with GDP-bound Kalirin/Rac1 (14.5 mg/mL) to a final concentration of 1.4 mM. Crystals were grown by mixing 100 nL of labelled Kalirin/Rac1 (14.5 mg/mL) with 50 nL of reservoir solution consisting of 0.1 M citrate pH 4.4, 0.8 M LiCl and 25% PEG 6K. The crystals were cryo-protected using 25% ethylene glycol. Diffraction data was collected at the I04-1 beamline at DLS. Molecular replacement was conducted using PHASER MR from CCP4 suite. ACEDRG and REFMAC from the CCP4 suite was used to install the covalent linkage. The structure was iteratively refined using REFMAC.

*8.16.7.6 Co-crystallisation of Kalirin/Rac1 with **106***

106 was incubated with nucleotide-free Kalirin/Rac1 (10.4 mg/mL) to a final concentration of 1 mM. Crystals were grown by mixing 50 nL of labelled Kalirin/Rac1 (10.4 mg/mL) with 100 nL of reservoir solution consisting of 0.1 M citrate pH 4.7, 0.8 M LiCl and 25% PEG 6K. The crystals were cryo-protected using 15% ethylene glycol. Diffraction data was collected at the I03 beamline at DLS. Molecular replacement was conducted using PHASER from PHENIX software suite. Jligand from the CCP4 software suite was used to install the covalent linkage. The structure was iteratively refined using the PHENIX[280] software suite.

*8.16.7.7 Co-crystallisation of Kalirin/Rac1 with **107***

107 was incubated with nucleotide-free Kalirin/Rac1 (10.5 mg/mL) to a final concentration of 1 mM. Crystals were grown by mixing 75 nL of labelled Kalirin/Rac1 (10.5 mg/mL) with 75 nL of reservoir solution consisting of 0.1 M bis-tris pH 5.5, 0.2 M NaCl and 25% PEG 3350. The crystals were cryo-protected using 20% DMSO. Diffraction data was collected at the I04-1 beamline at DLS. Molecular replacement was conducted using PHASER

MR from CCP4 suite. ACEDRG and REFMAC from the CCP4 suite was used to install the covalent linkage. The structure was iteratively refined using REFMAC.

*8.16.7.8 Co-crystallisation of Kalirin/Rac1 with **112***

112 was incubated with nucleotide-free Kalirin/Rac1 (10.5 mg/mL) to a final concentration of 1 mM. Crystals were grown by mixing 75 nL of labelled Kalirin/Rac1 (10.5 mg/mL) with 75 nL of reservoir solution consisting of 0.1 M bis-tris pH 5.5, 0.1 M NaCl and 27% PEG 3350. The crystals were cryo-protected using 20% DMSO. Diffraction data was collected at the I04-1 beamline at DLS. Molecular replacement was conducted using PHASER MR from CCP4 suite. ACEDRG and REFMAC from the CCP4 suite was used to install the covalent linkage. The structure was iteratively refined using REFMAC.

*8.16.7.9 Co-crystallisation of Kalirin/Rac1 with **115***

115 was incubated with nucleotide-free Kalirin/Rac1 (10 mg/mL) to a final concentration of 1 mM. Crystals were grown by mixing 100 nL of labelled Kalirin/Rac1 (10 mg/mL) with 50 nL of reservoir solution consisting of 0.1 M bis-tris pH 5.5, 0.1 M NaCl and 27% PEG 3350. The crystals were cryo-protected using 15% DMSO. Diffraction data was collected at the I04-1 beamline at DLS. Molecular replacement was conducted using PHASER MR from CCP4 suite. ACEDRG and REFMAC from the CCP4 suite was used to install the covalent linkage. The structure was iteratively refined using REFMAC.

8.16.8 XChem Screening for Kalirin/Rac1

8.16.8.1 General Method

Hundreds of Kalirin/Rac1 crystals were obtained using the sitting-drop vapour diffusion detailed above at 20 °C. Fragments were added to each drop using an ECHO acoustic liquid dispenser.[91] Compounds were soaked for 1 hour before being harvested using SHIFTER[92] and flash-cooled in liquid nitrogen. Diffraction data was collected automatedly and automatically processed on the I04-1 beamline at DLS. XChemExplorer[93] platform was used for initial structure solvation and parallel molecular replacement using DIMPLE. Subsequent electron density analysis and hit identification was performed using PanDDA[94]. Coot and REFMAC in the XChemExplorer pipeline were used for structure modelling and refinement.

8.16.8.2 XChem – Screen 1 with GDP

GDP-bound Kalirin/Rac1 crystals were made with reservoir solutions of 0.1 M citrate pH 4.2, 0.1-0.4 M NaCl, and 16-26% PEG8K. 37.5 nL of each fragment was dispensed, resulting in final concentration of 20% DMSO per well. Fragments were dispensed from the DSI-Poised[16], Covalent[40], OxXChem[286], Leeds3D[287] and Maybridge[210] libraries, with final concentration of 100 mM. GDP-bound Kalirin/Rac1 (PDB code: 5O33) was used as a homologous model in DIMPLE. The crystals typically diffracted between 1.5 and 2.0 Å.

8.16.8.3 XChem – Screen 2 without GDP

Nucleotide-free Kalirin/Rac1 crystals were made with reservoir solutions of 0.1 M bis-tris pH 5.5, 0-0.2 M NaCl, and 22-27% PEG 3350. 37.5 nL of each fragment from the DSI-Poised[16] library was dispensed, resulting in final concentration of 20% DMSO and 80 mM fragment per well. 27.5 nL of each fragment were dispensed from the Fraglites[214] and

MiniFrags[215] libraries with final concentrations of 15% DMSO and 373 mM fragment per well. Nucleotide-free Kalirin/Rac1 was used as a homologous model in DIMPLE. The crystals typically diffracted between 1.7 and 2.5 Å.

8.16.8.4 XChem – Follow-up Molport Compounds

Hundreds of GDP-bound Kalirin/Rac1 crystals were made with reservoir solution of 0.1M citrate pH 4.4, 0.8 M LiCl and 24% PEG 6K. Hundreds of nucleotide-free Kalirin/Rac1 crystals were made with reservoir solutions of 0.1 M bis-tris pH 5.5, 0-0.2 M NaCl, and 22-27% PEG 3350. 37.5 nL of each fragment from the purchased Molport library was dispensed, resulting in final concentration of 20% DMSO and 20 mM fragment per well. GDP-bound or nucleotide-free Kalirin/Rac1 was used as a homologous model in DIMPLE. The crystals typically diffracted between 1.7 and 2.5 Å.

*8.16.8.5 XChem – **106** Co-crystallisation Screening*

Hundreds of **106**-labelled Kalirin/Rac1 crystals were made with reservoir solutions of 0.1 M bis-tris pH 5.5, 0-0.2 M NaCl, and 22-27% PEG 3350. 27.5 nL of each fragment from the DSI-Poised library was dispensed, resulting in final concentration of 15% DMSO and 62 mM fragment per well. **106** labelled Kalirin/Rac1 was used as a homologous model in DIMPLE. The crystals typically diffracted between 1.8 and 2.5 Å.

8.16.8.6 XChem – Pharmacophore Screening with Enamine Library

Hundreds of nucleotide-free Kalirin/Rac1 crystals were made with reservoir solutions of 0.1 M bis-tris pH 5.5, 0-0.2 M NaCl, and 22-27% PEG 3350. 27.5-37.5 nL of each fragment from the purchased Enamine library was dispensed, resulting in final concentration of 15-20%

DMSO and 38.7 mM fragment per well. Nucleotide-free Kalirin/Rac1 was used as a homologous model in DIMPLE. The crystals typically diffracted between 1.8 and 2.5 Å.

8.16.8.7 XChem Resoaks

Fragments from all screens which were observed to crash out were resoaked at a final 10% DMSO concentration.

8.16.9 Nucleotide Exchange Assay

Rac1 was purified as per section 8.16.3.2 and concentrated to 9.5 mg/mL (480 μM), flash-frozen in liquid nitrogen and stored in 10 μL aliquots at -80 °C. Kalirin-c002 was purified as per section 8.16.3.2 and concentrated to 4.2 mg/mL (108 μM), flash-frozen in liquid nitrogen and stored in 5 μL aliquots at -80 °C in preparation for the nucleotide exchange assay.

Compounds were dispensed using an ECHO acoustic dispenser in triplicate into a 384-polypropylene microplate (Grenier Bio-One), with a maximum final concentration of 2 % DMSO. DMSO was also dispensed into the positive and negative control wells.

10 μL of an assay buffer consisting of 20 mM Tris pH 7.5, 50 mM NaCl, 1 mM $MgCl_2$, 100 μg/mL BSA and 1mM DTT was dispensed into each well except for the control GppNHp wells. 10 μL of 400 μM GppNHp (Jena BioScience) in 20 mM tris pH 7.5, 50 mM NaCl was added to each positive control well. The enzymes were mixed to a final concentration of 4 μM Rac1 and 0.1 μM Kalirin in the assay buffer and 5 μL of this mix was dispensed to every well. The assay was incubated at 20 °C for 2, 6, 12 or 24 h. 5 μL of 2 μM BODIPY™ FL GTP (ThermoFisher Scientific) in assay buffer was dispensed to each well. The plate was then shaken for 10 seconds and FI signal was measured every 30 s for 0.5-1 h using the Pherastar FSX (BMG Labtech, FI 485/520, Gain 300). The initial enzyme rate was calculated using

ICEKAT[245] and normalised with respect to the GppNHp control. Figures of the technical repeats and means from 3 replicates were plotted using GraphPad Prism 9 version 9.0.[185] One-way ANOVA followed by Dunnett's comparison test were conducted using mean values from 3 replicates in comparison to the DMSO control were performed in Prism.

8.16.10 Computational Docking in Molsoft ICM of Follow-up Urea Analogues

In silico docking was performed using Molsoft ICM-Pro software.[288] The receptor was prepared from the structure of Kalirin/Rac1 in complex with fragment **A** (PDB code: 5QQD) or fragment **B**. The protein was converted into an ICM object, the process of which deleted waters and optimised the orientation of Asn, Cys, Gln, His and Pro residues. The receptor pocket was defined surrounding the selected ligand. The original ligand was docked using the standard parameters in ICM to ensure the docking algorithm would produce valid poses. The Molsoft database was manually searched, and suitable compounds downloaded within an SDF file and loaded into ICM as a chemical table. The compounds clustered based on structural similarity and smaller molecular weight compounds were chosen to represent each subset. 3D stereoisomers were generated of the chosen compounds. A rigid docking protocol was performed using the standard parameters in ICM, with three conformations per compound. The compounds were ranked based on their RMSD from the original ligand and the compounds with RMSD < 9 were manually inspected and chosen for purchase.

8.17 General Chemistry Experimental

8.17.1 Materials, Solvents and Reagents

Water was deionised by an Elga DV 25 system. All other solvents and reagents were purchased from commercial sources and used as supplied (analytical or HPLC grade) without prior purification. Anhydrous solvents were purchased from Acros Organics and were stored under a nitrogen atmosphere with activated molecular sieves.

All reactions involving moisture sensitive reagents were carried out under a nitrogen atmosphere. Standard practices were employed when handling moisture- and air-sensitive compounds. Microwave reactions were conducted using a Biotage Initiator+ Microwave system with Robot 8.

Room temperature (r.t.) refers to ambient temperature. Temperatures of 0 °C were maintained using an ice-water bath. Temperatures of -78 °C were maintained using an acetone-cardice bath.

8.17.2 Purification and Chromatography

Yields refer to chromatographically and spectroscopically pure compounds unless otherwise stated. Reactions were monitored by thin layer chromatography (TLC) and/or liquid chromatography-mass spectrometry (LCMS). TLC analysis was performed on commercially prepared aluminium plates coated with 60 F_{254} silica gel. Plates were visualised using UV light (254 nm) or staining with Ninhydrin (1 M, EtOH) or 1% aq. $KMnO_4$. Analytical LCMS was performed on the following system: Kinetex 5μ EVO C18 100A 100 x 3.0 mm column using a linear gradient of 5-95% of solvent A (93 % H_2O, 5 % MeCN, 2 % of 0.5 M NH_4OAc pH 6.0) in

solvent B (18% H_2O, 80% MeCN, 2% of 0.5 M NH_4OAc pH 6.0), eluting at a flow rate of 2 mL/min. Compound formation and/or purity was assessed by either UV absorbance at 254 nm (Waters UV/Visible Detector 2489), ELSD signal (Waters ELS Detector 2424) or ESI+ TIC (SQ Detector 2).

Flash column chromatography was performed on a Biotage Isolera One flash column chromatography platform. Pre-packed SNAP-KP-Sil or SNAP-Ultra columns were used unless otherwise stated (Biotage). Strong cation exchange (SCX) chromatography was conducted using pre-packed sulfonic acid functionalised silica gel columns (Biotage). Purification by HPLC was performed using a Waters SFO with 515 HPLC pump and Waters Binary Gradient 2545 device (5-95% solvent A [18% H_2O, 80% MeCN, 2% NH_4OAc pH 6.0] in solvent B [93% H_2O, 5% MeCN, 2% NH_4OAc pH 6.0]).

8.17.3 Characterisation

Melting points (Mpt) were recorded using a Stuart SMP40 apparatus and are reported uncorrected in °C.

Infrared (IR) spectra were recorded neat on a Thermo Scientific Nicolet iS5 with an iD7 ATR module. Selected absorption maxima (ν_{max}) are reported in wavenumbers (cm^{-1}).

NMR spectra were recorded using a Bruker Avance 400 MHz spectrometer using the deuterated solvent stated. All NMRs were conducted at 298 K unless otherwise stated. The field was locked by external referencing to the relevant deuteron resonance. Chemical shifts (δ) are quoted in parts per million (ppm) and referenced to the residual solvent peak. Proton shifts are quoted to the nearest 0.01 ppm and carbon shifts to the nearest 0.1 ppm. When peak multiplicities are reported, the following abbreviations are used: br = broadened, s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, dd = doublet of doublets, dt = doublet

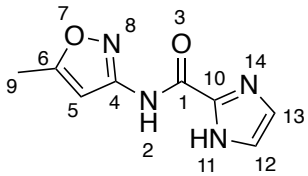
of triplets, dq = doublet of quartets, ddd = doublet of doublet of doublets, tt = triplet of triplets and qd = quartet of doublets. Coupling constants (J) are quoted in Hz and rounded to the nearest 0.1 Hz. Assignments of protons and carbons were supported by two-dimensional NMR experiments (COSY, HSQC, HMBC) or by analogy.

Low resolution mass spectra were recorded by LCMS on a Water SQ Detector 2; data acquisition and processing were performed using Waters FractionLynx software. High Resolution Mass Spectrometry (HRMS) was performed on an Agilent 6530 Accurate Mass Q-TOF; data acquisition and processing were performed using Agilent MassHunter Workstation software.

Compound names were generated using ChemDraw v19 systematic naming. Atom numbering in structures is purely for the purposes of assignment and does not reflect IUPAC numbering conventions.

8.17.4 Synthetic Procedures

N-(5-methyl-1,2-oxazol-3-yl)-1*H*-imidazole-2-carboxamide (55)

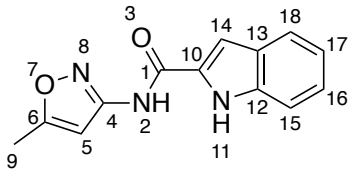


1*H*-imidazole-2-carboxylic acid (100 mg, 0.892 mmol), 3-amino-5-methylisoxazole (105 mg, 1.07 mmol) and anhydrous DIPEA (466 μL, 2.68 mmol) were dissolved in anhydrous DMF (4 mL) and T3P® 50 wt% in EtOAc (789 μL, 1.34 mmol) added dropwise. The reaction was stirred at r.t. for 16 h. The reaction was diluted with DCM and the organic phase washed with water, brine dried using Na_2SO_4 and the solvent removed *in vacuo*. The crude product was purified using flash column chromatography eluting with a gradient of 20-80% EtOAc in Cy. Further

purification using preparative HPLC and subsequent lyophilisation of fractions yielded **55** (12.7 mg, 66.1 μmol, 7%) as a white solid.

Mpt: 241.7-247.0 °C; **ν_{max} (cm^{-1}):** 3270, 3011, 2730, 2627, 1681, 1607, 1544, 1423, 1324, 1266, 1200, 1151, 1109, 991, 873, 807; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 13.36 (s, 1H, H-11), 10.87 (s, 1H, H-2), 7.41 (s, 1H, H-12), 7.16 (s, 1H, H-13), 6.64 (d, J = 0.9 Hz, 1H, H-5), 2.40 (d, J = 0.9 Hz, 3H, H-9); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 169.4 (C-6), 157.5 (C-4), 156.4 (C-1), 139.8 (C-10), 129.6 (C-13), 121.3 (C-12), 96.8 (C-5), 12.1 (C-9); **LR-ESI-MS:** $C_8H_9N_4O_2$ $[M+H]^+$ *m/z* found 193.28 cald 193.07; **HR-ESI-MS:** $C_8H_9N_4O_2$ $[M+H]^+$ *m/z* found 193.0739 cald 193.0720.

***N*-(5-methylisoxazol-3-yl)-1*H*-indole-2-carboxamide (56)**

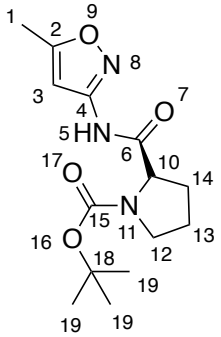


1*H*-indole-2-carboxylic acid (250 mg, 1.55 mmol), 3-amino-5-methylisoxazole (183 mg, 1.86 mmol) and DIPEA (811 μL, 4.65 mmol) were dissolved in anhydrous DCM (10 mL) and T3P® 50 wt% in EtOAc (1.40 mL, 2.33 mmol) added dropwise before stirring at r.t. for 16 h. The reaction was washed with water, dried using Na_2SO_4 and solvent removed *in vacuo*. The crude product was purified using flash column chromatography eluting with a gradient of 0-100% EtOAc in Cy. Further purification using preparative HPLC and subsequent lyophilisation of fractions yielded **56** (6.0 mg, 24.9 μmol, 2%) as a brown solid.

Mpt: 230.7-234.6 °C; **ν_{max} (cm^{-1}):** 3299, 3061, 2923, 1688, 1621, 1557, 1476, 1416, 1343, 1310, 1275, 1244, 1209, 1116, 1028, 871, 731; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 11.80 (s, 1H, H-2), 11.33 (s, 1H, H-11), 7.66 (d, J = 8.2 Hz, 1H, H-18), 7.55 (s, 1H, H-19), 7.46 (d, J = 8.2 Hz, 1H, H-15), 7.29-7.19 (m, 1H, H-17), 7.11-7.02 (m, 1H, H-16), 6.79 (s, 1H, H-5), 2.43 (s, 3H, H-9); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 169.4 (C-6), 159.4 (C-1), 158.4 (C-4), 137.1 (C-

12), 130.1 (C-10), 126.9 (C-13), 124.3 (C-17), 122.1 (C-18), 120.1 (C-16), 112.5 (C-15), 105.4 (C-14), 96.9 (C-5), 12.1 (C-9); **LR-ESI-MS:** $C_{13}H_{12}N_3O_2$ $[M+H]^+$ *m/z* found 242.28 cald 242.09; **HR-ESI-MS:** $C_{13}H_{11}N_3O_2Na$ $[M+Na]^+$ *m/z* found 264.0747 cald 264.0743.

***tert*-butyl (*R*)-2-((5-methylisoxazol-3-yl)carbamoyl)pyrrolidine-1-carboxylate (57)**

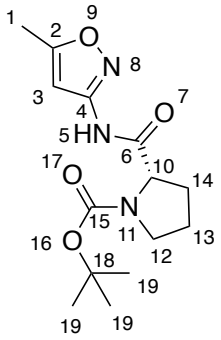


N-Boc-D-proline (250 mg, 1.16 mmol), 3-amino-5-methylisoxazole (137 mg, 1.39 mmol) and DIPEA (607 μL, 3.48 mmol) were dissolved in anhydrous DCM (10 mL) and T3P® 50 wt% in EtOAc (1.00 mL, 1.74 mmol) added dropwise. The reaction was stirred at r.t. for 16 h. The reaction was washed with water, brine, dried using Na_2SO_4 and the solvent removed *in vacuo*. The crude product was purified using flash column chromatography eluting with a gradient of 0-60% EtOAc in Cy. **57** (205 mg, 0.697 mmol, 60%) was obtained as a white solid.

Mpt: 159.2-162.8 °C; **ν_{max} (cm^{-1}):** 3279, 3155, 2978, 2872, 1695, 1682, 1622, 1557, 1478, 1387, 1270, 1154, 1117, 914, 862 ; **^{1}H NMR (400 MHz, 348 K, DMSO-d_6) δ** 10.67 (s, 1H, H-5), 6.59-6.57 (m, 1H, H-3), 4.30 (dd, J = 8.2, 3.5 Hz, 1H, H-10), 3.46-3.31 (m, 2H, H-12), 2.37 (d, J = 0.8 Hz, 3H, H-1), 2.23-2.13 (m, 1H, H-14), 1.93-1.77 (m, 3H, H-13, H-14), 1.34 (s, 9H, H-19); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 171.8 (C-6), 171.3 (C-6), 169.5 (C-2), 169.4 (C-2), 158.1 (C-4), 158.1 (C-4), 153.5 (C-15), 153.0 (C-15), 96.3 (C-3), 96.2 (C-3), 78.8 (C-18), 78.6 (C-18), 59.8 (C-10), 59.5 (C-10), 46.7 (C-12), 46.5 (C-12), 30.8 (C-14), 30.0 (C-14), 28.1 (C-19), 27.9 (C-19), 24.0 (C-13), 23.4 (C-13), 12.1 (C-1); **LR-ESI-MS:** $C_{14}H_{22}N_3O_4$ $[M+H]^+$ *m/z* found 296.51 cald 296.16; **HR-ESI-MS:** $C_{14}H_{21}N_3O_4Na$ $[M+Na]^+$ *m/z* found 318.1259 cald 318.1424.

Reduced rotation of Boc group at 298 K resulted in two rotamer values per carbon in the ^{13}C NMR spectra

***tert*-butyl (*S*)-2-((5-methylisoxazol-3-yl)carbamoyl)pyrrolidine-1-carboxylate (58)**

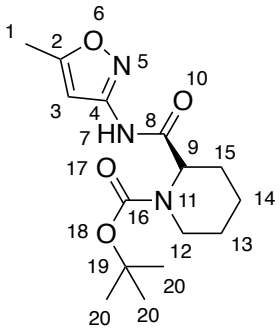


N-Boc-L-proline (250 mg, 1.16 mmol), 3-amino-5-methylisoxazole (137 mg, 1.39 mmol) and anhydrous DIPEA (607 mL, 3.48 mmol) were added to anhydrous DCM (10 mL) and T3P® 50 wt% in EtOAc (1.00 mL, 1.74 mmol) was added dropwise. The reaction was stirred at r.t. for 16 h before the organic layer was washed with water, brine, dried using Na_2SO_4 and the solvent removed *in vacuo*. The crude product was purified using flash chromatography eluting with a gradient of 0-60% EtOAc in Cy. **58** (168 mg, 0.568 mmol, 49%) was obtained as a white solid.

Mpt: 171-174 °C; **ν_{max} (cm^{-1}):** 3212, 3144, 3088, 2983, 2881, 1712, 1667, 1620, 1563, 1475, 1424, 1366, 1270, 1157, 1125, 910, 854; **^{1}H NMR (400 MHz, 348 K, DMSO-d_6) δ** 10.67 (s, 1H, H-5), 6.58 (d, J = 0.9 Hz, 1H, H-3), 4.30 (dd, J = 8.3, 3.5 Hz, 1H, H-10), 3.47-3.30 (m, 2H, H-12), 2.37 (d, J = 0.9 Hz, 3H, H-1), 2.23-2.14 (m, 1H, H-14), 1.93-1.76 (m, 3H, H-13, H-14), 1.34 (s, 9H, H-19); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 171.8 (C-6), 171.3 (C-6), 169.5 (C-2), 169.5 (C-2), 158.1 (C-4), 158.1 (C-4), 153.6 (C-15), 153.0 (C-15), 96.3 (C-3), 96.2 (C-3), 78.8 (C-18), 78.6 (C-18), 59.8 (C-10), 59.5 (C-10), 46.7 (C-12), 46.5 (C-12), 30.8 (C-14), 30.0 (C-14), 28.1 (C-19), 27.9 (C-19), 24.0 (C-13), 23.4 (C-13), 12.1 (C-1); **LR-ESI-MS:** $C_{14}H_{22}N_3O_4$ $[M+H]^+$ *m/z* found 296.45 cald 296.16; **HR-ESI-MS:** $C_{14}H_{21}N_3O_4Na$ $[M+Na]^+$ *m/z* found 318.1259 cald 318.1424.

Reduced rotation of Boc group at 298 K resulted in two rotamer values per carbon in the ^{13}C NMR spectra

***tert*-butyl (*R*)-2-((5-methylisoxazol-3-yl)carbamoyl)piperidine-1-carboxylate (59)**

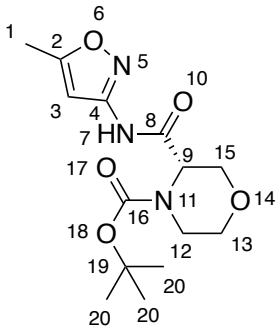


(2*R*)-1-[(*tert*-butoxy)carbonyl]piperidine-2-carboxylic acid (257 mg, 1.12 mmol), 3-amino-5-methylisoxazole (132 mg, 1.35 mmol) and anhydrous DIPEA (586 μL, 3.36 mmol) were dissolved in anhydrous DCM (10 mL) and T3P® 50 wt% in EtOAc (991 μL, 1.68 mmol) added dropwise. The reaction was stirred at r.t. for 72 h. Additional T3P® (991 μL, 1.68 mmol was added and the reaction stirred at r.t. for 16 h. The crude reaction was diluted with water, and the organic phase was washed with water, brine, dried using Na_2SO_4 and the solvent removed *in vacuo*. The crude product was purified using flash column chromatography eluting with a gradient of 0-60% EtOAc in Cy to give **59** (34.3 mg, 0.111 mmol, 10%) as a white solid.

Mpt: 69.5-75.3 °C; **ν_{max} (cm^{-1}):** 3430, 3205, 3145, 2951, 1703, 1671, 1622, 1563, 1478, 1415, 1377, 1269, 1180, 1020, 933, 899; **^{1}H NMR (400 MHz, 348 K, DMSO-d_6) δ** 10.61 (s, 1H, H-7), 6.57 (d, J = 0.9 Hz, 1H, H-3), 4.69 (dd, J = 5.9, 2.3 Hz, 1H, H-9), 3.87-3.68 (m, 1H, H-12), 3.22 (td, J = 12.5, 3.4 Hz, 1H, H-12), 2.37 (d, J = 0.9 Hz, 3H, H-1), 2.14-1.95 (m, 1H, H-15), 1.81-1.49 (m, 3H, H-13, H-14, H-15), 1.38 (s, 9H, H-20), 1.35-1.21 (m, 2H, H-13, H-14); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 171.7 (C-8), 170.0 (C-2), 158.5 (C-4), 155.6 (C-16), 96.8 (C-3), 79.5 (C-19) , 55.3 (C-9), 54.1 (C-9), 42.4 (C-12), 41.6 (C-12), 28.5 (C-20), 27.7 (C-15), 24.7 (C-13), 24.5 (C-13), 19.9 (C-14), 12.6 (C-1); **LR-ESI-MS:** $C_{15}H_{24}N_3O_4$ $[M+H]^+$ *m/z* found 310.56 cald 310.17; **HR-ESI-MS:** $C_{15}H_{23}N_3O_4Na$ $[M+Na]^+$ *m/z* found 332.1409 cald 332.1581.

Reduced rotation of Boc group at 298 K resulted in two rotamer values per carbon in the ^{13}C NMR spectra

***tert*-butyl (*S*)-3-((5-methylisoxazol-3-yl)carbamoyl)morpholine-4-carboxylate (60)**

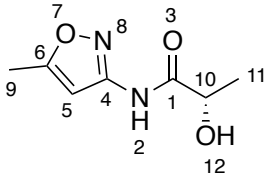


(3*S*)-4-[(*tert*-butoxy)carbonyl]morpholine-3-carboxylic acid (250 mg, 1.08 mmol), 3-amino-5-methylisoxazole (127 mg, 1.30 mmol) and DIPEA (565 μL, 3.24 mmol) were dissolved in anhydrous DCM (10 mL) and T3P® 50 wt% in EtOAc (956 μL, 1.62 mmol) added dropwise. The reaction was stirred for 72 h at r.t. and the solvent removed *in vacuo*. The crude product was purified using flash chromatography eluting with a gradient of 20-80% EtOAc in Cy. **60** (124 mg, 0.399 mmol, 37%) was obtained as a white solid.

Mpt: 142.6-147.9 °C; **ν_{max} (cm^{-1}):** 3404, 3219, 3091, 2984, 2871, 1706, 1651, 1557, 1421, 1378, 1161, 1110, 982, 867 ; **^{1}H NMR (400 MHz, Methanol-*d*$_4$) δ** 6.60 (br s, 1H, H-3), 4.49 (br s, 1H, H-9), 4.25 (br s, 1H, H-15), 3.90 (br s, 1H, H-13), 3.76 (dd, J = 12.3, 4.2 Hz, 1H, H-15), 3.72-3.63 (m, 1H, H-12), 3.57-3.43 (m, 2H, H-12, H-13), 2.40 (s, 3H, H-1), 1.44 (br s, 9H, H-20); **^{13}C NMR (101 MHz, Methanol-*d*$_4$) δ** 170.0 (C-2), 169.5 (C-8), 158.0 (C-4), 156.1 (C-16), 95.8 (C-3), 80.7 (C-19), 67.7 (C-15), 67.2 (C-15), 66.2 (C-13), 55.8 (C-9), 54.7 (C-9), 42.0 (C-12), 41.0 (C-12), 27.1 (C-20), 10.9 (C-1); **LR-ESI-MS:** $C_{14}H_{22}N_3O_5$ $[M+H]^+$ *m/z* found 312.48 cald 312.15; **HR-ESI-MS:** $C_{14}H_{21}N_3O_5Na$ $[M+Na]^+$ *m/z* found 334.1200 cald 334.1373.

Reduced rotation of Boc group at 298 K resulted in two rotamer values per carbon in the ^{13}C NMR spectra

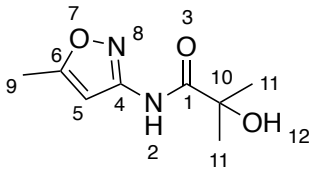
(*S*)-2-hydroxy-*N*-(5-methylisoxazol-3-yl)propenamide (61)



3-amino-5-methylisoxazole (142 mg, 1.45 mmol) was added to a stirred solution of DABAL-Me3 (370 mg, 1.45 mmol) in anhydrous THF (9 mL). The solution was stirred at 40 °C for 1.5 h before methyl (2*S*)-2-hydroxypropanoate (92.0 μL, 0.963 mmol) was added and the solution refluxed for 16 h. The reaction was cooled to r.t., quenched by dropwise addition of aq. HCl (1 M, 10 mL) and extracted with EtOAc. The organic phase was washed with brine, dried using Na_2SO_4 and the solvent removed *in vacuo*. The crude product was purified using flash column chromatography eluting with a gradient of 20-70% EtOAc in Cy. Further purification using preparative HPLC and subsequent lyophilisation of fractions yielded **61** (48.4 mg, 0.284 mmol, 30%) as a white solid.

Mpt: 63.9-69.3 °C; **ν_{max} (cm^{-1}):** 3273, 2981, 1693, 1615, 1540, 1431, 1270, 1122, 799; **^{1}H NMR (400 MHz, Methanol-*d*$_4$) δ** 6.61 (d, J = 0.9 Hz, 1H, H-7), 4.27 (q, J = 6.8 Hz, 1H, H-10), 2.40 (d, J = 0.9 Hz, 3H, H-9), 1.41 (d, J = 6.8 Hz, 3H, H-11); **^{13}C NMR (101 MHz, Methanol-*d*$_4$) δ** 176.0 (C-1), 171.5 (C-6), 159.1 (C-4), 97.1 (C-5), 69.2 (C-10), 21.0 (C-11), 12.3 (C-9); **LR-ESI-MS:** $C_7H_{11}N_2O_3$ $[M+H]^+$ *m/z* found 171.16 cald 171.07; **HR-ESI-MS:** $C_7H_{10}N_2O_3Na$ $[M+Na]^+$ *m/z* found 193.1090 cald 193.5840.

2-hydroxy-2-methyl-*N*-(5-methylisoxazol-3-yl)propenamide (62)

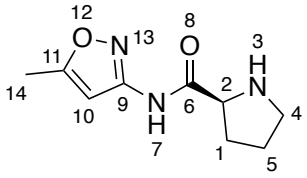


3-amino-5-methylisoxazole (62.3 mg, 0.635 mmol) was added to a stirred solution of DABAL-Me$_3$ (163 mg, 0.635 mmol) in anhydrous THF (4 mL). The solution was stirred at 40 °C for 1

h before methyl 2-hydroxy-2-methylpropanoate (49.0 μL, 0.423 mmol) was added. The solution was refluxed for 16 h. The reaction was cooled to r.t., quenched by dropwise addition of aq. HCl (1 M, 10 mL) and extracted with EtOAc. The organic phase was washed with brine, dried using Na_2SO_4 and the solvent removed *in vacuo*. The crude product was purified using flash column chromatography eluting with a gradient of 10-60% EtOAc in Cy. **62** (16.2 mg, 88.0 μmol, 21%) was obtained as a white solid.

Mpt: 114.7-118.6 °C ; **ν_{max} (cm^{-1}):** 3396, 3215, 2981, 2937, 1677, 1613, 1535, 1421, 1272, 1199, 908, 634; **^{1}H NMR (400 MHz, $CDCl_3$) δ** 9.41 (s, 1H, H-2), 6.72 (d, J = 0.9 Hz, 1H, H-5), 2.40 (d, J = 0.9 Hz, 3H, H-9), 1.53 (s, 6H, H-11); **^{13}C NMR (101 MHz, $CDCl_3$) δ** 174.9 (C-1), 170.3 (C-6), 157.8 (C-4), 96.2 (C-5), 74.2 (C-10), 27.9 (C-11), 12.8 (C-9); **LR-ESI-MS:** $C_8H_{13}N_2O_3$ $[M+H]^+$ *m/z* found cald 185.09; **HR-ESI-MS:** $C_8H_{12}N_2O_3Na$ $[M+Na]^+$ *m/z* found 207.0619 cald 207.0740.

(*S*)-*N*-(5-methylisoxazol-3-yl)pyrrolidine-2-carboxamide (63)

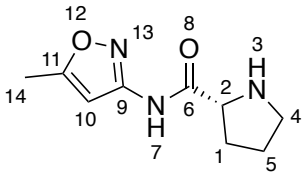


58 (86.7 mg, 0.294 mmol) was dissolved in anhydrous DCM (1.7 mL). TFA (1.5 mL, 19.7 mmol) was added, and the reaction stirred at r.t. for 1.5 h. The volatile reagents were removed *in vacuo*. The crude oil was redissolved in minimal DCM and washed with saturated aq. Na_2CO_3.The organic phases were combined, dried using Na_2SO_4, and the solvent removed *in vacuo*. **63** (53.2 mg, 0.273 mmol, 93%) was obtained as a yellow oil.

ν_{max} (cm^{-1}): 3219, 2966, 2870, 1694, 1612, 1531, 1421, 1267, 1200, 1101, 1024, 914, 881; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 10.53 (s, 1H, H-7), 6.64 (d, J = 1.0 Hz, 1H, H-10), 3.75-3.70 (m, 1H, H-2), 2.86 (t, J = 6.5 Hz, 2H, H-4), 2.37 (d, J = 1.0 Hz, 3H, H-14), 2.08-1.97 (m, 1H, H-1), 1.78-1.69 (m, 1H, H-1), 1.67-1.60 (m, 2H, H-5); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 173.6

(C-6), 169.7 (C-11), 157.5 (C-9), 95.9 (C-10), 60.4 (C-2), 46.7 (C-4), 30.2 (C-1), 25.9 (C-5), 12.1 (C-14); **LR-ESI-MS:** $C_9H_{14}N_3O_2$ $[M+H]^+$ *m/z* found 196.26 cald 196.10; **HR-ESI-MS:** $C_9H_{14}N_3O$ $[M+H]^+$ *m/z* found 196.1098 cald 196.1081.

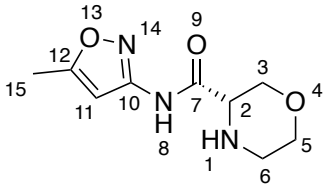
(*R*)-*N*-(5-methylisoxazol-3-yl)pyrrolidine-2-carboxamide (64)



57 (123 mg, 0.416 mmol) was dissolved in DCM (2.4 mL). TFA (2.1 mL, 27.8 mmol) was added, and the reaction stirred at r.t. for 3 h. The volatile reagents were removed *in vacuo*. The crude oil was redissolved in minimal DCM and washed with saturated aq. Na_2CO_3. The organic phases were combined, dried using Na_2SO_4, and the solvent removed *in vacuo*. **64** (61.2 mg, 0.313 mmol, 75%) was obtained as a yellow oil.

ν_{max} (cm^{-1}): 3363, 3120, 2969, 2820, 1681, 1615, 1521, 1420, 1274, 1117, 1032, 915, 881, 836; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 6.64 (d, J = 1.0 Hz, 1H, H-10), 3.76 (dd, J = 8.8, 5.7 Hz, 1H, H-2), 2.88 (t, J = 6.6 Hz, 2H, H-4), 2.37 (d, J = 1.0 Hz, 3H, H-14), 2.09-2.00 (m, 1H,H-1), 1.79-1.71 (m, 1H, H-1), 1.68-1.61 (m, 2H, H-5); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 173.3 (C-6), 169.8 (C-11), 157.6 (C-9), 95.9 (C-10), 60.3 (C-2), 46.7 (C-4), 30.2 (C-1), 25.8 (C-5), 12.1 (C-14); **LR-ESI-MS:** $C_9H_{14}N_3O_2$ $[M+H]^+$ *m/z* found 196.26 cald 196.10; **HR-ESI-MS:** $C_9H_{14}N_3O_2$ $[M+H]^+$ *m/z* found 196.1099 cald 196.1081.

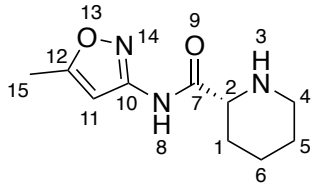
(*S*)-*N*-(5-methylisoxazol-3-yl)morpholine-3-carboxamide (65)



60 (124 mg, 0.399 mmol) was dissolved in DCM (2.3 mL). TFA (2.1 mL, 26.7 mmol) was added, and the reaction was stirred at r.t. for 1 h. The volatile reagents were removed *in vacuo*. The crude product was redissolved in minimal DCM and washed with saturated aq. Na_2CO_3. The organic phases were combined, dried using Na_2SO_4, and the solvent removed *in vacuo*. The crude product was purified using preparative HPLC and subsequent lyophilisation of fractions. **65** (14.7 mg, 69.6 μmol, 17%) was obtained as a white solid.

Mpt: 115.7-118.4 °C; **ν_{max} (cm^{-1}):** 3306, 3148, 2983, 2853, 1678, 1612, 1532, 1468, 1419, 1348, 1307, 1150, 1098, 1013, 910, 837; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 10.64 (s, 1H, H-8), 6.61 (s, 1H, H-11), 3.83-3.75 (m, 1H, H-3), 3.62-3.56 (m, 1H, H-5), 3.55-3.47 (m, 2H, 2, H-3), 3.44-3.37 (m, 1H, H-5), 2.86-2.79 (m, 1H, H-6), 2.75-2.65 (m, 1H, H-6), 2.37 (s, 3H, H-15); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 170.2 (C-7), 170.1 (C-12), 158.2 (C-10), 96.7 (C-11), 68.7 (C-3), 67.3 (C-5), 58.0 (C-2), 44.0 (C-6), 12.5 (C-15); **LR-ESI-MS:** $C_9H_{14}N_3O_3$ $[M+H]^+$ *m/z* found 212.29 cald 212.10; **HR-ESI-MS:** $C_9H_{13}N_3O_3Na$ $[M+Na]^+$ *m/z* found 234.0889 cald 234.0849.

(*R*)-*N*-(5-methylisoxazol-3-yl)piperidine-2-carboxamide (66)

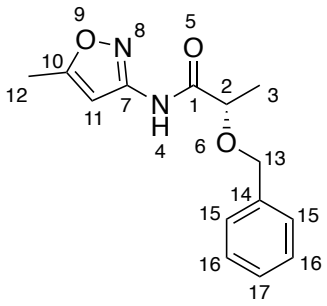


59 (50.1 mg, 0.162 mmol) was dissolved in DCM (1 mL). TFA (836 μL, 10.9 mmol) was added, and the reaction was stirred at r.t. for 3 h. The volatile reagents were removed *in vacuo*. The crude oil was redissolved in minimal DCM and washed with saturated aq. Na_2CO_3. The organic

phases were combined, dried using Na_2SO_4, and the solvent removed *in vacuo*. **66** (30.9 mg, 0.148 mmol, 91%) was obtained as a yellow amorphous solid.

ν_{max} (cm^{-1}): 3280, 3129, 2933, 2834, 2793, 1682, 1622, 1517, 1443, 1317, 1208, 1132, 1113, 1025, 929, 836; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 6.61 (d, J = 1.0 Hz, 1H, H-11), 3.33-3.25 (m, 1H, H-2), 2.97-2.87 (m, 1H, H-4), 2.57-2.49 (m, 1H, H-4), 2.36 (d, J = 1.0 Hz, 3H, H-15), 1.79-1.69 (m, 2H, H-1, H-6), 1.50-1.21 (m, 4H, H-1, H-5, H-6); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 172.8 (C-7), 169.9 (C-12), 158.3 (C-10), 96.7 (C-11), 59.8 (C-2), 45.4 (C-4), 29.7 (C-1), 26.2 (C-5), 24.2 (C-6), 12.6 (C-15); **LR-ESI-MS:** $C_{10}H_{16}N_3O_2$ $[M+H]^+$ *m/z* found 210.21 cald 210.12; **HR-ESI-MS:** $C_{10}H_{16}N_3O_2$ $[M+H]^+$ *m/z* found 210.1259 cald 210.1237.

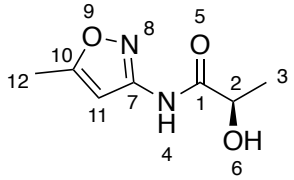
(*S*)-2-(benzyloxy)-*N*-(5-methylisoxazol-3-yl)propanamide (67)



3-amino-5-methylisoxazole (65.4 mg, 0.667 mmol), (2*S*)-2-(benzyloxy)propanoic acid (87.0 μL, 0.555 mmol) and anhydrous DIPEA (290 μL, 1.67 mmol) were dissolved in anhydrous DCM (2 mL) and T3P® 50 wt% in EtOAc (491 μL, 0.833 mmol) was added dropwise. The reaction was stirred at r.t. for 72 h. The reaction was diluted with water and the organic phase extracted, washed with brine, dried using Na_2SO_4, and the solvent removed *in vacuo*. The crude product was purified using flash column chromatography eluting with a gradient of 0-50% EtOAc in Cy. The crude product **67** was used directly without further purification.

LR-ESI-MS: $C_{14}H_{17}N_2O_3$ $[M+H]^+$ *m/z* found 261.33 cald 261.12.

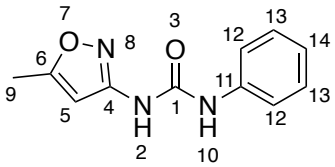
(*R*)-2-hydroxy-*N*-(5-methylisoxazol-3-yl)propanamide (68)



3-amino-5-methylisoxazole (142 mg, 1.45 mmol) was added to a solution of DABAL-Me3 (370 mg, 1.45 mmol) in anhydrous THF (9 mL). The solution was stirred at 40 °C for 1.5 h before methyl (2*R*)-2-hydroxypropanoate (92.0 μL, 0.963 mmol) was added and the reaction refluxed for 16 h. The reaction was cooled to r.t., quenched by dropwise addition of aq. HCl (1 M, 10 mL) and extracted with EtOAc. The organic phase was washed with brine, dried using Na_2SO_4 and the solvent removed *in vacuo*. The crude product was purified using flash column chromatography eluting with a gradient of 20-70% EtOAc in Cy. Further purification using preparative HPLC and subsequent lyophilisation of fractions yielded **68** (57.1 mg, 0.336 mmol, 35%) as a white solid.

Mpt: 61.8-66.4 °C; **ν_{max} (cm^{-1}):** 3264, 2980, 2918, 1686, 1615, 1541, 1472, 1431, 1270, 1122, 1048, 1015, 917, 881; **^{1}H NMR (400 MHz, MeOD) δ** 6.61 (d, J = 0.9 Hz, 1H, H-11), 4.27 (q, J = 6.8 Hz, 1H, H-2), 2.40 (d, J = 0.9 Hz, 3H, H-12), 1.41 (d, J = 6.8 Hz, 3H, H-3); **^{13}C NMR (101 MHz, MeOD) δ** 176.0 (C-1), 171.5 (C-10), 159.1 (C-7), 97.1 (C-11), 69.2 (C-2), 21.0 (C-3), 12.3 (C-12); **LR-ESI-MS:** $C_7H_{11}N_2O_3$ $[M+H]^+$ *m/z* found 171.28 cald 171.07; **HR-ESI-MS:** $C_7H_{10}N_2O_3Na$ $[M+Na]^+$ *m/z* found 192.0848 cald 192.0584.

1-(5-methylisoxazol-3-yl)-3-phenylurea (69)

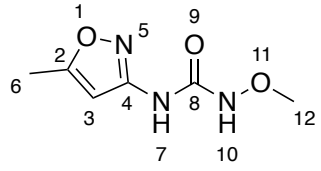


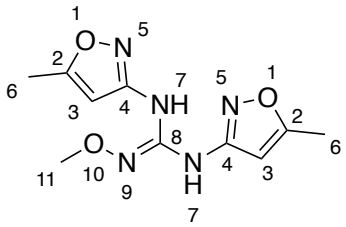
Phenyl isocyanate (118 mg, 1.20 mmol), 3-amino-5-methylisoxazole (109 μL, 1.00 mmol) and anhydrous DIPEA (63.0 μL, 0.362 mmol) were dissolved in anhydrous MeCN (3 mL) and the

reaction was stirred at 75 °C for 4 h. The reaction was quenched with water, extracted with DCM, the organic phase dried using Na_2SO_4 and the solvent removed *in vacuo*. The crude product was purified using flash column chromatography using a gradient of 0-4% [7N NH_3 in MeOH] in DCM. The crude product was further purified using preparative HPLC and subsequent lyophilisation of fractions. **69** (39.8 mg, 0.183 mmol, 18%) was obtained as a white solid.

Mpt: 185.4-190.6 °C; **ν_{max} (cm^{-1}):** 3360, 3310, 3097, 2986, 1683, 1633, 1598, 1526, 1484, 1445, 1314, 1274, 1212, 1084, 1007, 925, 880, 808; **^{1}H NMR (400 MHz, MeOD) δ** 7.46-7.41 (m, 2H, H-13), 7.33-7.27 (m, 2H, H-12), 7.06 (m, 1H, H-14), 6.40-6.38 (m, 1H, H-5), 2.38 (d, J = 0.8 Hz, 3H, H-9); **^{13}C NMR (101 MHz, MeOD) δ** 171.1 (C-6), 160.3 (C-4), 153.8 (C-1), 139.8 (C-11), 130.0 (C-12), 124.6 (C-14), 120.7 (C-13), 96.4 (C-5), 12.2 (C-9); **LR-ESI-MS:** $C_{11}H_{12}N_3O_2$ $[M+H]^+$ *m/z* found 218.36 cald 218.09; **HR-ESI-MS:** $C_{11}H_{12}N_3O_2$ $[M+H]^+$ *m/z* found 218.0947 cald 218.0924.

1-methoxy-3-(5-methylisoxazol-3-yl)urea (71) & 2-methoxy-1,3-bis(5-methylisoxazol-3-yl)guanidine (73)





3-amino-5-methylisoxazole (100 mg, 1.02 mmol) was dissolved in anhydrous THF (4 mL) and cooled to 0 °C. Triphosgene (151 mg, 0.510 mmol) was dissolved in anhydrous THF (1 mL) and added dropwise. After 5 mins, anhydrous TEA (426 μL, 3.06 mmol) was added dropwise, and the reaction was stirred at 0 °C for 30 mins. *O*-methylhydroxylamine hydrochloride (128 mg, 1.53 mmol) was dissolved in anhydrous 1:1 THF:DMSO (1 mL) and added dropwise. The reaction was stirred at r.t. for 16 h. The organic phase was washed with water and brine, dried using Na_2SO_4 and the solvent removed *in vacuo*. The crude products were purified by

preparative HPLC. **71** (14.9 mg, 87.1 μmol, 9%) was obtained as a yellow solid. **73** (21.6 mg, 86.0 μmol, 8%) was obtained as a brown solid.

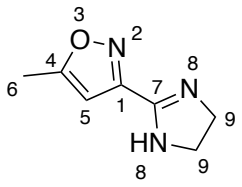
*Data for 1-methoxy-3-(5-methylisoxazol-3-yl)urea (**71**)*

Mpt: 136.4-142.6 °C; **ν_{max} (cm^{-1}):** 3153, 3081, 2856, 1687, 1672, 1615, 1542, 1433, 1268, 1098. 961, 846; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 9.78 (s, 1H, H-10), 9.72 (s, 1H, H-7), 6.50 (d, J = 1.0 Hz, 1H, H-3), 3.58 (s, 3H, H-12), 2.34 (d, J = 1.0 Hz, 3H, H-6); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 168.9 (C-2), 158.5 (C-4), 155.7 (C-8), 96.2 (C-3), 64.1 (C-12), 12.1 (C-6); **LR-ESI-MS:** $C_6H_{10}N_3O_3$ $[M+H]^+$ *m/z* found 172.19 cald 172.07; **HR-ESI-MS:** $C_6H_{10}N_3O_3$ $[M+H]^+$ *m/z* found 172.0840 cald 172.0717.

*Data for 2-methoxy-1,3-bis(5-methylisoxazol-3-yl)guanidine (**73**)*

Mpt: 204.7-208.4 °C; **ν_{max} (cm^{-1}):** 3198, 2967, 1731, 1676, 1614, 1535, 1426, 1290, 1198, 963, 842; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 10.85 (s, 2H, H-7), 6.59 (d, J = 1.0 Hz, 2H, H-3), 3.84 (s, 3H, H-11), 2.40 (d, J = 0.9 Hz, 6H, H-6); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 170.1 (C-2), 157.4 (C-4), 150.34 (C-8), 96.6 (C-3), 64.6 (C-11), 12.1 (C-6); **LR-ESI-MS:** $C_{10}H_{14}N_5O_3$ $[M+H]^+$ *m/z* found 252.30 cald 252.11; **HR-ESI-MS:** $C_{10}H_{13}N_5O_3Na$ $[M+Na]^+$ *m/z* found 274.0763 cald 274.0911.

3-(4,5-dihydro-1*H*-imidazol-2-yl)-5-methylisoxazole (74)

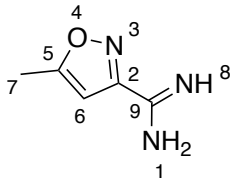


78 (30.0 mg, 0.211 mmol) and ethane-1,2-diamine (20.0 μL, 0.299 mmol) were dissolved in anhydrous THF and refluxed for 16 h. The solvent was removed *in vacuo*. Water was added and the pH adjusted to 11 by the dropwise addition of aq. NaOH (1 M). The aqueous phase

was extracted with DCM and EtOAc. The solvent was removed *in vacuo*, the crude product redissolved in minimal MeOH, and applied to an SCX column. The column was washed twice with 10 CV of MeOH and the product eluted using 2 M NH_3 in MeOH. The solvent was removed *in vacuo*. **74** (16.3 mg, 0.108 mmol, 51%) was obtained as a white solid.

Mpt: 146.5-151.4 °C; **ν_{max} (cm^{-1}):** 3140, 2932, 2871, 1624, 1592, 1526, 1421, 1292, 1206, 981, 912, 817; **^{1}H NMR (400 MHz, $CDCl_3$) δ** 6.42 (s, 1H, H-5), 3.72 (s, 4H, H-9), 2.40 (s, 3H, H-6); **^{13}C NMR (101 MHz, $CDCl_3$) δ** 170.8 (C-4), 157.1 (C-7), 155.3 (C-1), 101.0 (C-5), 55.2 (C-9) 12.1 (C-6); **LR-ESI-MS:** $C_7H_{10}N_3O$ $[M+H]^+$ *m/z* found 152.13 cald 152.08; **HR-ESI-MS:** $C_7H_{10}N_3O$ $[M+H]^+$ *m/z* found 152.0831 cald 152.0818.

5-methyl-1,2-oxazole-3-carboximidamide (75)

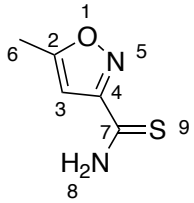


5-methyl-1,2-oxazole-3-carbonitrile (110 mg, 1.02 mmol) was dissolved in MeOH (5 mL) and sodium methoxide (1.4 mL, 10.2 mmol) added. The reaction was stirred for 3 h before NH_4Cl (544 mg, 10.2 mmol) and additional MeOH (5 mL) were added. The reaction was stirred for 48 h, concentrated and filtered. The filtrate was extracted with EtOAc and the solvent removed *in vacuo*. The crude product was redissolved in minimal MeOH and applied to an SCX column. The column was washed twice with 10 CV of MeOH and the product eluted using 2 M NH_3 in MeOH. The solvent was removed *in vacuo*. **75** (41.6 mg, 0.332 mmol, 33%) was obtained as a yellow solid.

Mpt: 127.3-132.8 °C; **ν_{max} (cm^{-1}):** 3456, 3390, 3310, 3200, 3130, 2920, 2851, 1688, 1656, 1598, 1488, 1453, 1340, 1251, 1195, 1136, 1010, 919, 815; **^{1}H NMR (400 MHz, MeOD) δ** 6.52-6.50 (m, 1H, H-6), 2.46 (d, J = 0.8 Hz, 3H, H-7); **^{13}C NMR (101 MHz, MeOD) δ** 172.8 (C-

5), 160.2 (C-9), 158.0 (C-2), 101.4 (C-6), 12.0 (C-7); **LR-ESI-MS:** $C_5H_8N_3O$ $[M+H]^+$ *m/z* found 126.24 cald 126.06; **HR-ESI-MS:** $C_5H_8N_3O$ $[M+H]^+$ *m/z* found 126.0676 cald 126.0662.

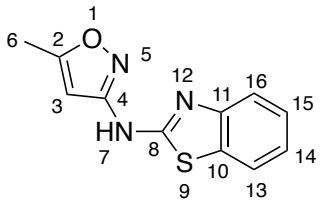
5-methylisoxazole-3-carbothioamide (78)



5-methylisoxazole-3-carboxamide (100 mg, 0.793 mmol) and Lawesson's reagent (353 mg, 0.872 mmol) were dissolved in anhydrous toluene (3 mL) and refluxed for 16 h. The reaction was quenched with water and the crude product extracted with EtOAc. The organic phase was washed with water, brine, dried using Na_2SO_4 and the solvent removed *in vacuo*. The crude solid was purified using flash column chromatography eluting with a gradient of 0-60% EtOAc in Cy. **78** was used directly without further purification.

^{1}H NMR (400 MHz, $CDCl_3$, crude) δ 6.63 (d, J = 1.1 Hz, 1H, H-3), 3.88 (m, 2H, H-8), 2.59-2.45 (m, 3H, H-6); **LR-ESI-MS:** $C_5H_7N_2OS$ $[M+H]^+$ *m/z* found 143.19 cald 143.02

***N*-(benzo[*d*]thiazol-2-yl)-5-methylisoxazol-3-amine (79)**

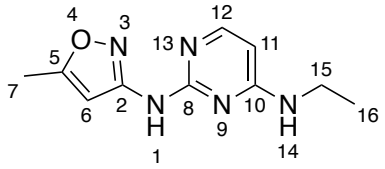


NaH 60 wt% dispersion in mineral oil (48.9 mg, 1.22 mmol) was dissolved in anhydrous DMF (2 mL) and cooled to 0 °C. 3-amino-5-methylisoxazole (100 mg, 1.02 mmol) in anhydrous DMF (1.5 mL) was added and the solution stirred for 1 h at 0 °C. 2-bromo-1,3-benzothiazole (218 mg, 1.02 mmol) in anhydrous DMF (1.5 mL) was added and the reaction stirred at r.t. for 16 h. The reaction was quenched with water and extracted 3 times with DCM. The organic phase was washed with brine, dried with Na_2SO_4 and solvent removed *in vacuo*. The crude

product was purified using flash column chromatography eluting with a gradient of 0-50% EtOAc in Cy. The product was further purified using preparative HPLC and subsequent lyophilisation of fractions. **79** (35.5 mg, 0.154 mmol, 15%) was obtained as a white solid.

Mpt: 195.4-200.3 °C; **ν_{max} (cm^{-1}):** 3404, 3211, 3053, 2914, 1692, 1597, 1557, 1436, 1362, 1245, 1163, 1017, 894; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 11.35 (s, 1H, H-7), 7.89 (d, J = 7.8 Hz, 1H, H-16), 7.62 (d, J = 8.0 Hz, 1H, H-13), 7.49-7.30 (m, 1H, H-14), 7.30-7.06 (m, 1H, H-15), 6.34 (s, 1H, H-3), 2.40 (d, J = 0.7 Hz, 3H, H-6); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 169.6 (C-2), 160.6 (C-8), 158.9 (C-4), 149.3 (C-11), 131.5 (C-10), 126.0 (C-14), 122.4 (C-15), 121.5 (C-16), 119.3 (C-13), 95.3 (C-3), 12.1 (C-6); **LR-ESI-MS:** $C_{11}H_{10}N_3OS$ $[M+H]^+$ *m/z* found 232.22 cald 232.05; **HR-ESI-MS:** $C_{11}H_{10}N_3OS$ $[M+H]^+$ *m/z* found 232.0408 cald 232.0539.

N^{4}-ethyl-*N*^{2}-(5-methylisoxazol-3-yl)pyrimidine-2,4-diamine (80)

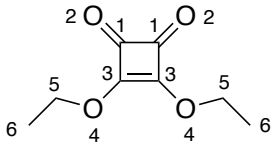


3-amino-5-methylisoxazole (41.2 mg, 0.420 mmol) and 2-chloro-*N*-ethylpyrimidin-4-amine (55.2 mg, 0.350 mmol) were dissolved in anhydrous 2-propanol (3 mL). A drop of 37% aq. HCl was added. The reaction was irradiated in a microwave reactor at 100 °C for 2 h. The solvent was removed *in vacuo* and the crude product purified using preparative HPLC and subsequent lyophilisation of fractions. **80** (7.5 mg, 34.2 μmol, 10%) was obtained as a white solid.

Mpt: 180.8-186.3 °C; **ν_{max} (cm^{-1}):** 3418, 3285, 2973, 2871, 1637, 1555, 1429, 1344, 1227, 1130, 983, 901, 812; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 9.58 (s, 1H, H-1), 7.78 (br s, 1H, H-12), 7.22 (br s, 1H, H-14), 6.79 (br s, 1H, H-6), 5.96 (d, J = 5.9 Hz, 1H, H-11), 3.29 (br s, 2H, H-15), 2.34 (d, J = 0.9 Hz, 3H, H-7), 1.14 (t, J = 7.2 Hz, 3H, H-16); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 167.9 (C-5), 162.6 (C-10), 159.9 (C-2), 158.7 (C-8), 154.2 (C-12), 98.6 (C-11), 96.9 (C-

6), 34.9 (C-15), 14.5 (C-16), 12.2 (C-7); **LR-ESI-MS:** $C_{10}H_{14}N_5O$ $[M+H]^+$ *m/z* found 220.33 cald 220.12; **HR-ESI-MS:** $C_{10}H_{14}N_5O$ $[M+H]^+$ *m/z* found 220.1075 cald 220.1193.

3,4-diethoxycyclobut-3-ene-1,2-dione (81)

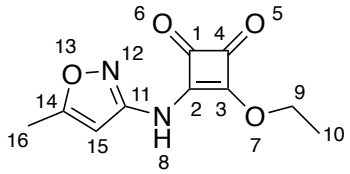


1,2-Dihydroxycyclobutenedione (500 mg, 4.38 mmol) was dissolved in EtOH (5 mL) and the mixture refluxed for 3 hours. The solvent was removed *in vacuo*, further EtOH (5 mL) added, and the mixture refluxed for another hour. This process was repeated 4 times. The solvent was removed *in vacuo* to give **81** (686 mg, 4.03 mmol, 92%) as a pink oil.

^{1}H NMR (400 MHz, $CDCl_3$) δ 4.67 (q, J = 7.1 Hz, 4H, H-5), 1.41 (t, J = 7.1 Hz, 6H, H-6); **^{13}C NMR (101 MHz, $CDCl_3$) δ** 189.3 (C-1), 184.2 (C-3), 70.6 (C-5), 15.5 (C-6); **LR-ESI-MS:** $C_8H_{11}O_4$ $[M+H]^+$ *m/z* found 171.28 cald 171.06.

Analytical data consistent with that previously reported.[289]

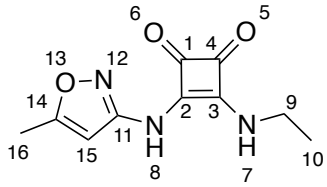
3-ethoxy-4-((5-methylisoxazol-3-yl)amino)cyclobut-3-ene-1,2-dione (82)



81 (295 μL, 1.99 mmol) and 3-amino-5-methylisoxazole (160 mg, 1.63 mmol) were dissolved in EtOH (15 mL). DIPEA (694 μL, 3.99 mmol) was added and the reaction was stirred for 120 h at 45 °C. The temperature was then increased to 55 °C for 4 h before being allowed to cool to r.t. The precipitate was collected by filtration, washed with ice cold EtOH and dried *in vacuo*. **82** (153 mg, 0.689 mmol, 42%) was obtained as a white solid.

Mpt: 227.3-232.5 °C; **ν_{max} (cm^{-1}):** 3196, 3112, 2998, 2907, 2866, 1804, 1717, 1644, 1599, 1549, 1487, 1469, 1427, 1349, 1241, 1192, 1107, 1066, 990, 934, 844; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 11.49 (s, 1H, H-8), 6.47 (s, 1H, H-15), 4.75 (q, J = 7.1 Hz, 2H, H-9), 2.38 (s, 3H, H-16), 1.40 (t, J = 7.1 Hz, 3H, H-10); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 187.0 (C-1), 185.2 (C-4), 179.6 (C-3), 170.5 (C-14), 169.6 (C-2), 158.4 (C-11), 96.5 (C-15), 69.8 (C-9), 15.6 (C-10), 12.2 (C-16); **LR-ESI-MS:** $C_{10}H_{11}N_2O_4$ $[M+H]^+$ *m/z* found 223.28 cald 223.07; **HR-ESI-MS:** $C_{10}H_{10}N_2O_4Na$ $[M+Na]^+$ *m/z* found 245.0461 cald 245.0533.

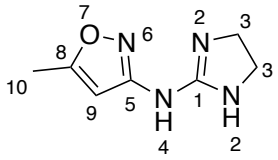
3-(ethylamino)-4-((5-methylisoxazol-3-yl)amino)cyclobut-3-ene-1,2-dione (83)



82 (36.7 mg, 0.165 mmol) and ethylamine (17.0 μL, 0.330 mmol) were dissolved in EtOH (1.2 mL) and stirred at r.t. for 16 h. The precipitated solid was collected by filtration and washed with ice cold EtOH. **83** (30.3 mg, 0.137 mmol, 83%) was obtained as a white solid.

Mpt: 230.1-235.4 °C; **ν_{max} (cm^{-1}):** 3211, 3162, 3071, 2938, 1803, 1694, 1635, 1557, 1441, 1342, 1143; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 10.68 (s, 1H, H-8), 7.80 (t, J = 5.5 Hz, 1H, H-7), 6.37 (s, 1H, H-15), 3.69-3.57 (m, 2H, H-9), 2.36 (s, 3H, H-16), 1.18 (t, J = 7.2 Hz, 3H, H-10); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 185.6 (C-4), 180.8 (C-1), 170.6 (C-14), 169.2 (C-3), 162.1 (C-2), 158.9 (C-11), 95.4 (C-15), 38.7 (C-9), 16.6 (C-10), 12.1 (C-16); **LR-ESI-MS:** $C_{10}H_{12}N_3O_3$ $[M+H]^+$ *m/z* found 222.25 cald 222.08; **HR-ESI-MS:** $C_{10}H_{12}N_3O_3$ $[M+H]^+$ *m/z* found 222.0901 cald 222.0873.

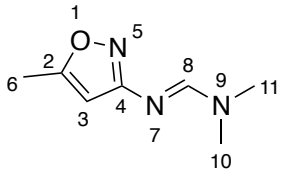
***N*-(4,5-dihydro-1*H*-imidazol-2-yl)-5-methylisoxazol-3-amine (84)**



92 (99.2 mg, 0.271 mmol) was dissolved in anhydrous DCM (2.5 mL) and TFA (1.4 mL, 18.1 mmol) added dropwise. The reaction was stirred at r.t. for 16 h. The volatile reagents were removed *in vacuo*. The crude oil was redissolved in minimal DCM and washed with saturated aq. Na_2CO_3. The organic phases were combined, dried using Na_2SO_4, and the solvent removed *in vacuo*. **84** (30.5 mg, 0.184 mmol, 68%) was obtained as a light brown solid.

Mpt: 169.1-175.9 °C; **ν_{max} (cm^{-1}):** 3342, 3123, 2868, 1633, 1609, 1511, 1469, 1407, 1372, 1285, 1257, 1201, 1085, 1008, 896; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 6.76 (s, 1H, H-4), 5.63 (d, J = 1.0 Hz, 1H, H-9), 3.42 (s, 4H, H-3), 2.24 (d, J = 1.0 Hz, 3H, H-10); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 167.5 (C-8), 166.9 (C-5), 160.6 (C-1), 99.6 (C-9), 42.0 (C-3), 12.0 (C-10); **LR-ESI-MS:** $C_7H_{11}N_4O$ $[M+H]^+$ *m/z* found 167.19 cald 167.09; **HR-ESI-MS:** $C_7H_{11}N_4O$ $[M+H]^+$ *m/z* found 167.0830 cald 167.0927.

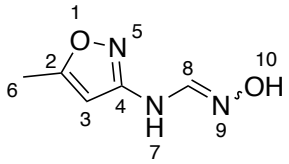
(*E*)-*N*,*N*-dimethyl-*N*'-(5-methylisoxazol-3-yl)formimidamide (87)



3-amino-5-methylisoxazole (2.00 g, 20.4 mmol) and DMF-DMA (3.0 mL, 22.4 mmol) were dissolved in anhydrous toluene (20 mL) and refluxed for 2 h. The solvent was removed *in vacuo* and the crude product **87** was used directly without further purification.

LR-ESI-MS: $C_7H_{12}N_3O$ $[M+H]^+$ *m/z* found 154.14 cald 154.09.

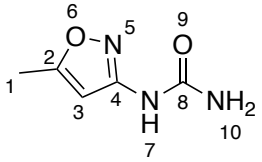
***N*-methyl-*N*-(5-methylisoxazol-3-yl)formimidamide (88)**



87 (3.12 g, 20.4 mmol) was dissolved in MeOH (100 mL) and NH_2OH 50% in water (1.0 mL, 24.5 mmol) added. The reaction was stirred at r.t. for 2 h. The precipitated solid was collected by filtration and washed with ice cold MeOH. **88** (2.14 g, 15.1 mmol, 74%) was obtained as a white solid.

Mpt: 189.6-195.5 °C; **ν_{max} (cm^{-1}):** 3156, 2778, 1688, 1615, 1547, 1484, 1357, 1310, 1276, 1230, 1020, 902, 888; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 10.14 (s, 1H, H-10), 9.18 (d, J = 10.4 Hz, 1H, H-7), 7.21 (d, J = 10.4 Hz, 1H, H-8), 5.96 (d, J = 1.0 Hz, 1H, H-3), 2.29 (d, J = 1.0 Hz, 3H, H-6); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 169.0 (C-2), 159.4 (C-4), 136.4 (C-8), 94.1 (C-3), 12.0 (C-6); **LR-ESI-MS:** $C_5H_8N_3O_2$ $[M+H]^+$ *m/z* found 142.10 cald 142.06; **HR-ESI-MS:** $C_7H_{11}N_2O_3$ $[M+H]^+$ *m/z* found 142.0112 cald 142.0611.

1-(5-methylisoxazol-3-yl)urea (89)

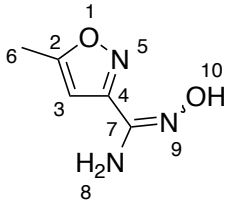


88 (296 mg, 2.09 mmol) was suspended in anhydrous DCE (5.9 mL) and MsCl (178 μL, 2.30 mmol) was added. The reaction was refluxed for 1 h, anhydrous TEA (23.0 μL, 0.165 mmol) was added and the reaction refluxed for a further 2 h. The solution was allowed to cool to r.t., the precipitated solid collected by filtration and dried *in vacuo*. **89** (52.6 mg, 0.373 mmol, 18%) was obtained as a white solid.

Mpt: 211.2-213.8 °C; **ν_{max} (cm^{-1}):** 3410, 3195, 3107, 2983, 2923, 1686, 1646, 1614, 1570, 1494, 1441, 1338, 1266, 1138, 1009, 930; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 9.23 (s, 1H, H-7), 6.38 (d, J = 1.0 Hz, 1H, H-3), 6.25 (s, 2H, H-10), 2.31 (d, J = 1.0 Hz, 3H, H-1); **^{13}C NMR (101**

MHz, DMSO-*d*$_6$) δ 168.7 (C-2), 159.2 (C-4), 154.5 (C-8), 95.4 (C-3), 12.0 (C-1); **LR-ESI-MS:** $C_5H_8N_3O_2$ $[M+H]^+$ *m/z* found 142.10 cald 142.06; **HR-ESI-MS:** $C_7H_{11}N_2O_3$ $[M+H]^+$ *m/z* found 142.0081 cald 142.0611.

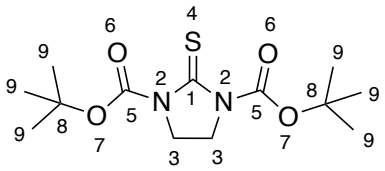
***N*-hydroxy-5-methylisoxazole-3-carboximidamide (90)**



5-methyl-1,2-oxazole-3-carbonitrile (233 mg, 2.16 mmol) was dissolved in EtOH (2.3 mL). NH_2OH 50% in water (198 μL, 3.23 mmol) was added and the reaction refluxed for 3 h. The reaction was allowed to cool to r.t., the precipitated solid collected by filtration and washed with ice cold EtOH. **90** (60.0 mg, 0.425 mmol, 20%) was obtained as a white solid.

Mpt: 162.3-165.9 °C; **ν_{max} (cm^{-1}):** 3495, 3337, 3144, 3056, 2783, 1652, 1574, 1435, 1315, 1232, 1101, 1002, 960, 916; **^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ** 10.08 (s, 1H, H-10), 6.32 (d, J = 0.9 Hz, 1H, H-3), 5.75 (s, 2H, H-8), 2.41 (d, J = 0.9 Hz, 3H, H-6); **^{13}C NMR (101 MHz, DMSO-*d*$_6$) δ** 169.6 (C-2), 157.5 (C-4), 143.9 (C-7), 99.3 (C-3), 11.7 (C-7); **LR-ESI-MS:** $C_5H_8N_3O_2$ $[M+H]^+$ *m/z* found 142.10 cald 142.06; **HR-ESI-MS:** $C_7H_{11}N_2O_3$ $[M+H]^+$ *m/z* found 142.1030 cald 142.0611.

Di-*tert*-butyl 2-thioxoimidazolidine-1,3-dicarboxylate (91)

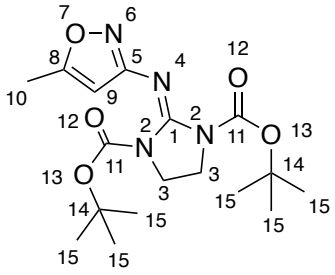


Imidazolidine-2-thione (605 mg, 5.92 mmol) was dissolved in anhydrous THF (90 mL) and cooled to 0 °C before NaH 60 wt% dispersion in mineral oil (1.07 g, 26.6 mmol) was added. After 5 mins, the ice bath was removed, and the reaction stirred at r.t. for 10 mins. The reaction

was then cooled to 0 °C and Boc anhydride (1.07 g, 13.0 mmol) added. The reaction was stirred at 0 °C for 30 minutes and then stirred at r.t. for 16 h. The reaction was quenched by the dropwise addition of saturated aq. $NaHCO_3$ (100 mL). The mixture was poured into water and extracted 3 times with EtOAc. The combined organic phases were washed with brine, dried using Na_2SO_4 and the solvent removed *in vacuo*. The crude product was purified using flash column chromatography eluting with a gradient of 0-60% EtOAc in Cy. **91** (1.57 g, 5.19 mmol, 88%) was obtained as a yellow solid.

Mpt: 113.6-117.7 °C; **ν_{max} (cm^{-1}):** 2975, 2930, 1741, 1711, 1474, 1388, 1377, 1266, 1241, 1139, 949, 847; **^{1}H NMR (400 MHz, $CDCl_3$) δ** 3.90 (s, 4H, H-3), 1.54 (s, 18H, H-9); **^{13}C NMR (101 MHz, $CDCl_3$) δ** 175.6 (C-1), 150.5 (C-5), 84.1 (C-8), 44.6 (C-3), 28.2 (C-9); **LR-ESI-MS:** $C_{13}H_{23}N_2O_4S$ $[M+H]^+$ *m/z* found 303.26 cald 303.13; **HR-ESI-MS:** $C_{13}H_{22}N_2O_4SNa$ $[M+Na]^+$ *m/z* found 325.1582 cald 325.1192.

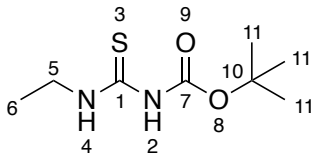
Di-*tert*-butyl 2-((5-methylisoxazol-3-yl)imino)imidazolidine-1,3-dicarboxylate (92)



91 (1.37 g, 4.53 mmol) and 3-amino-5-methylisoxazole (444 mg, 4.53 mmol) were dissolved in anhydrous DCM (23.4 mL) and anhydrous TEA (2.1 mL, 14.9 mmol) was added. The mixture was cooled to 0 °C and $CuCl_2$ (640 mg, 4.76 mmol) was added. The reaction was stirred at 0 °C for 30 mins and then stirred at r.t. for 16 h. The reaction was diluted with EtOAc and filtered twice through a pad of celite. Water was added, and the organic phase extracted, washed with brine, dried using Na_2SO_4 and the solvent removed *in vacuo*. The crude product was purified using flash column chromatography eluting with a gradient of 0-60% EtOAc in Cy. **92** (678 mg, 1.85 mmol, 41%) was obtained as a white solid.

Mpt: 148.6-152.1 °C; **ν_{max} (cm^{-1}):** 2977, 2936, 1747, 1708, 1666, 1614, 1482, 1369, 1287, 1246, 1136, 975, 908, 842; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 5.87 (d, J = 1.0 Hz, 1H, H-9), 3.76 (s, 4H, H-3), 2.27 (d, J = 1.0 Hz, 3H, H-10), 1.35 (s, 18H, H-15); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 167.9 (C-8), 165.7 (C-5), 149.4 (C-11), 144.0 (C-1), 97.4 (C-9), 82.1 (C-14), 43.1 (C-3), 27.5 (C-15), 12.1 (C-10); **LR-ESI-MS:** $C_{17}H_{27}N_4O_5$ $[M+H]^+$ *m/z* found 367.49 cald 367.19; **HR-ESI-MS:** $C_{17}H_{26}N_4O_5Na$ $[M+Na]^+$ *m/z* found 389.1582 cald 389.1191.

***tert*-butyl *N*-(ethylcarbamothioyl)carbamate (93)**

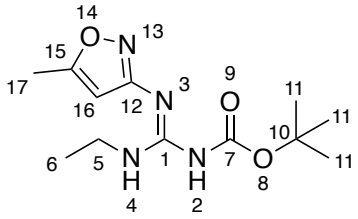


Thiourea (500 mg, 6.57 mmol) was dissolved in anhydrous THF (120 mL) and cooled to 0 °C. NaH 60 wt% dispersion in mineral oil (1.18 g, 29.6 mmol) was added and the reaction stirred at r.t. for 45 mins. The reaction was cooled again to 0 °C and Boc anhydride (3.3 mL, 14.5 mmol) added. The reaction was stirred at r.t. for 16 h. The reaction was then cooled to 0 °C and a second portion of NaH 60 wt% dispersion in mineral oil (441 mg, 11.0 mmol) added. The reaction was stirred at r.t. for 1 h before the addition of TFAA (1.4 mL, 10.1 mmol) at 0 °C. The reaction was stirred at r.t. for 2 h before the addition of ethylamine (5.1 mL, 10.1 mmol) and stirred at r.t. for 16 h. The reaction was quenched with water and extracted 3 times with EtOAc. The organic phase was washed with brine, dried using Na_2SO_4 and the solvent was removed *in vacuo*. The crude product was purified using flash column chromatography eluting with a gradient of 0-20% EtOAc in Cy. **93** (147 mg, 0.720 mmol, 11%) was obtained as a white amorphous solid.

ν_{max} (cm^{-1}): 2923, 2853, 1753, 1710, 1456, 1368, 1249, 1146, 1023; **^{1}H NMR (400 MHz, $CDCl_3$) δ** 9.63 (s, 1H, H-2), 7.81 (br s, 1H, H-4), 3.68 (qd, J = 7.3, 5.3 Hz, 2H, H-5), 1.49 (s, 9H, H-11), 1.28 (t, J = 7.3 Hz, 3H, H-6); **^{13}C NMR (101 MHz, $CDCl_3$) δ** 179.4 (C-7), 152.0 (C-

1), 83.8 (C-10), 40.6 (C-5), 28.2 (C-11), 13.8 (C-6); **LR-ESI-MS:** $C_8H_{17}N_2O_2S$ $[M+H]^+$ *m/z* found 205.07 cald 205.10; **HR-ESI-MS:** $C_8H_{16}N_2O_2SNa$ $[M+Na]^+$ *m/z* found 227.0180 cald 227.0825.

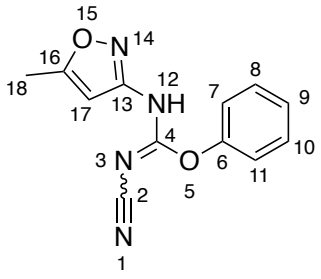
***tert*-butyl *N*-[(*E*)-*N*-ethyl-*N'*-(5-methylisoxazol-3-yl)carbamimidoyl]carbamate (94)**



93 (126 mg, 0.615 mmol) and 3-amino-5-methylisoxazole (60.4 mg, 0.615 mmol) were dissolved in anhydrous DCM (3.2 mL) and anhydrous TEA (283 μL, 2.03 mmol) added. The solution was cooled to 0 °C and $CuCl_2$ (86.9 mg, 0.646 mmol) added. The reaction was stirred at 0 °C for 30 mins and then stirred at r.t. for 16 h. The reaction mixture was diluted with EtOAc and filtered twice through a pad of celite. Water was added, and the organic phase extracted, washed with brine, dried using Na_2SO_4 and the solvent removed *in vacuo*. The crude product was purified using flash column chromatography eluting with a gradient of 0-40% EtOAc in Cy. **94** was used directly without further purification.

LR-ESI-MS: $C_{12}H_{21}N_4O_3$ $[M+H]^+$ *m/z* found 269.30 cald 269.16.

Phenyl *N*-cyano-*N*-(5-methylisoxazol-3-yl)carbamimidate (96)

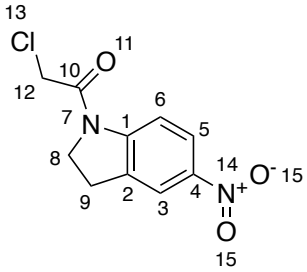


3-amino-5-methylisoxazole (41.2 mg, 0.420 mmol) and *N*-Cyano-diphenyl imidocarbonate (100 mg, 0.420 mmol) were dissolved in 2-propanol (5 mL) and the reaction was stirred at r.t. for 16 h. The reaction was filtered, quenched with water and extracted using DCM. The organic

phase was washed with water, brine, dried using Na_2SO_4, and the solvent removed *in vacuo*. **96** was used directly without further purification.

LR-ESI-MS: $C_{12}H_{11}N_4O_2$ $[M+H]^+$ *m/z* found 243.30 cald 243.08

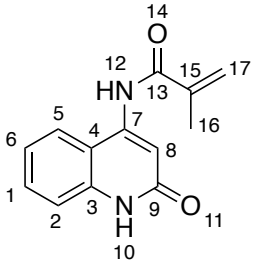
2-chloro-1-(5-nitroindolin-1-yl)ethan-1-one (99)



5-nitro-2,3-dihydro-1*H*-indole (100 mg, 0.609 mmol) was dissolved in anhydrous DCM (5 mL) and anhydrous TEA (170 μL, 1.22 mmol) added. The solution was cooled to 0 °C and 2-chloroacetyl chloride (58 μL, 0.728 mmol) added. The reaction was allowed to warm to r.t. and stirred for 16h. The reaction was quenched with water and the organic phase washed with water, brine, dried using Na_2SO_4 and the solvent removed *in vacuo*. The crude product was purified using flash chromatography using a gradient of 0-50% EtOAc in Cy. **99** (110 mg, 0.458 mmol, 75%) was obtained as a yellow solid.

Mpt: 164.8-168.1 °C; **ν_{max} (cm^{-1}):** 3139, 3086, 2985, 2940, 1681, 1598, 1507, 1476, 1400, 1329, 1267, 1165, 1109, 1068, 894, 855; **^{1}H NMR (400 MHz, $CDCl_3$) δ** 8.32 (d, J = 8.1 Hz, 1H, H-6), 8.17 (dd, J = 8.1, 2.3 Hz, 1H, H-5), 8.10-8.07 (m, 1H, H-3), 4.32 (t, J = 8.6 Hz, 2H, H-8), 4.20 (s, 2H, H-12), 3.35 (t, J = 8.6 Hz, 2H, H-9); **^{13}C NMR (101 MHz, $CDCl_3$) δ** 164.9 (C-10), 147.8 (C-1), 144.3 (C-4), 132.5 (C-2), 124.8 (C-5), 120.4 (C-3), 116.8 (C-6), 48.6 (C-8), 42.8 (C-12), 27.6 (C-9); **LR-ESI-MS:** $C_{10}H_{10}ClN_2O_3$ $[M+H]^+$ *m/z* found 241.38 cald 241.03; **HR-ESI-MS:** $C_{10}H_9ClN_2O_3Na$ $[M+Na]^+$ *m/z* found 263.0194 cald 263.0455.

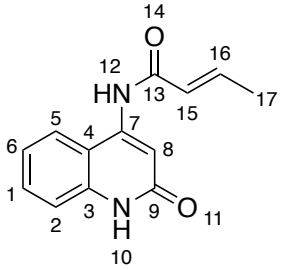
***N*-(2-oxo-1,2-dihydroquinolin-4-yl)methacrylamide (116)**



4-amino-1,2-dihydroquinolin-2-one (45.8 mg, 0.286 mmol) was dissolved in anhydrous DMF (2.2 mL) and Py (69.0 μL, 0.427 mmol) added. 2-methylprop-2-enoyl chloride (37.0 μL, 0.375 mmol) was added to the solution and stirred at r.t. for 16 h. Water was added and the reaction stirred at r.t. for 10 mins. The aqueous phase was extracted 3 times with DCM. The combined organic phases were washed with 5% LiCl and brine, and the solvent removed *in vacuo*. The crude product was purified using preparative HPLC and subsequent lyophilisation of fractions. **116** (1.2 mg, 5 μmol, 2%) was obtained as an off-white solid.

Mpt: 254.3-257.1 °C; **ν_{max} (cm^{-1}):** 3303, 3095, 2959, 2841, 1682, 1634, 1558, 1494, 1423, 1222, 941, 854; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 11.64 (s, 1H, H-10), 9.78 (s, 1H, H-12), 7.80 (dd, J = 8.2, 1.3 Hz, 1H, H-2), 7.57-7.44 (m, 1H, H-6), 7.32 (dd, J = 8.2, 1.3 Hz, 1H, H-5), 7.23-7.12 (m, 1H, H-1), 6.88 (s, 1H, H-8), 5.99-5.94 (m, 1H, H-17), 5.66-5.60 (m, 1H, H-17), 2.00 (s, 3H, H-16); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 167.7 (C-13), 162.2 (C-9), 144.1 (C-7), 139.6 (C-15), 138.9 (C-3), 130.7 (C-6), 123.6 (C-2), 121.6 (C-17), 121.2 (C-1), 115.5 (C-5), 115.2 (C-4), 112.4 (C-8), 18.6 (C-16); **LR-ESI-MS:** $C_{13}H_{13}N_2O_2$ $[M+H]^+$ *m/z* found 229.25 cald 229.09; **HR-ESI-MS:** $C_{13}H_{13}N_2O_2$ $[M+H]^+$ *m/z* found 229.0261 cald 229.0972.

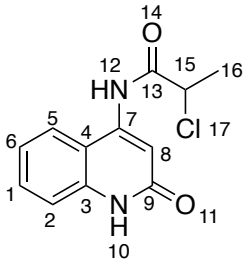
(*E*)-N-(2-oxo-1,2-dihydroquinolin-4-yl)but-2-enamide (117)



4-amino-1,2-dihydroquinolin-2-one (22.2 mg, 0.139 mmol), (*E*)-but-2-enoyl chloride (18.0 μL, 0.183 mmol) and Py (33.0 μL, 0.206 mmol) were dissolved in anhydrous DMF (1.5 mL) and stirred at r.t. for 16 h. The solvent was removed *in vacuo*. Water was added and the precipitated solid was collected by filtration and washed with ice cold water. **117** (1.5 mg, 6.60 μmol, 5%) was obtained as an off-white solid.

Mpt: 245.2-247.8 °C; **ν_{max} (cm^{-1}):** 3248, 3014, 2849, 1639, 1560, 1523, 1437, 1390, 1221; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 11.54 (s, 1H, H-10), 9.74 (s, 1H, H-12), 8.04 (d, J = 8.2 Hz, 1H, H-2), 7.57-7.49 (m, 1H,H- 6), 7.32 (d, J = 8.0 Hz, 1H, H-5), 7.25 (s, 1H, H-8), 7.24-7.19 (m, 1H, H-1), 6.91 (m, 1H, H-16), 6.51 (dd, J = 15.3, 1.9 Hz, 1H, H-15), 1.91 (dd, J = 6.9, 1.6 Hz, 3H, H-17); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 164.8 (C-13), 162.4 (C-9), 143.7 (C-7), 142.1 (C-16), 139.0 (C-3), 130.7 (C-6), 125.5 (C-15), 122.7 (C-2), 121.2 (C-1), 115.7 (C-15), 114.1 (C-4), 109.2 (C-8), 17.7 (C-17); **LR-ESI-MS:** $C_{13}H_{13}N_2O_2$ $[M+H]^+$ *m/z* found 229.25 cald 229.09; **HR-ESI-MS:** $C_{13}H_{13}N_2O_2$ $[M+H]^+$ *m/z* found 229.0104 cald 229.0972.

2-chloro-*N*-(2-oxo-1,2-dihydroquinolin-4-yl)propenamide (118)

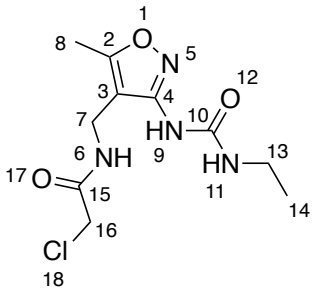


4-amino-1,2-dihydroquinolin-2-one (21.2 mg, 0.132 mmol) was dissolved in anhydrous DMF (0.5 mL) and Py (32.0 μL, 0.197 mmol) added. 2-chloropropanoyl chloride (17 μL, 0.175 mmol)

was added to the solution and stirred at r.t. for 16 h. The solvent was removed *in vacuo*. MeOH was added and a white precipitate formed. The solvent was removed *in vacuo* and the solid resuspended in water, collected by filtration and washed with cold water. **118** (15.8 mg, 6.3 μmol, 48%) was obtained as an off-white solid.

Mpt: 282.3-285.6 °C; **ν_{max} (cm^{-1}):** 3250, 2991, 2825, 1661, 1537, 1444, 1224, 1075, 912, 852; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 11.65 (s, 1H, H-10), 10.16 (br s, 1H, H-12), 8.01-7.95 (m, 1H, H-2), 7.55 (ddd, J = 8.3, 7.2, 1.2 Hz, 1H, H-6), 7.33 (dd, J = 8.3, 1.2 Hz, 1H, H-5), 7.24 (ddd, J = 8.3, 7.2, 1.2 Hz, 1H, H-1), 7.07 (d, J = 1.9 Hz, 1H, H-8), 5.02 (q, J = 6.6 Hz, 1H, H-15), 1.65 (d, J = 6.6 Hz, 3H, H-16); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 169.4 (C-13), 162.6 (C-9), 143.7 (C-7), 139.4 (C-3), 131.4 (C-6), 123.2 (C-2), 121.8 (C-1), 116.3 (C-5), 114.6 (C-4), 111.0 (C-8), 54.8 (C-15), 21.3 (C-16); **LR-ESI-MS:** $C_{12}H_{12}ClN_2O_2$ $[M+H]^+$ *m/z* found 251.10 cald 251.05; **HR-ESI-MS:** $C_{12}H_{12}ClN_2O_2$ $[M+H]^+$ *m/z* found 250.9804 cald 251.0582.

2-chloro-*N*-((3-(3-ethylureido)-5-methylisoxazol-4-yl)methyl)acetamide (119)

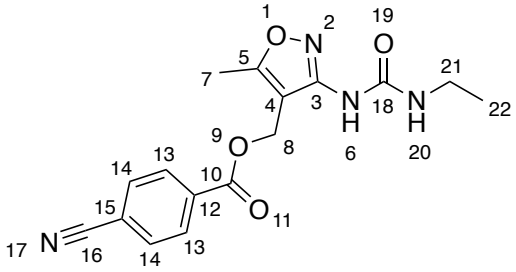


122 (72.3 mg, 0.365 mmol) was dissolved in anhydrous DCM (3 mL) and anhydrous DMF (6 mL) and anhydrous TEA (102 μL, 0.729 mmol) added. The solution was cooled to 0 °C and 2-chloroacetyl chloride (35.0 μL, 0.438 mmol) added. The solution was stirred at r.t. for 2 h. The solvent was removed *in vacuo* and the crude product purified using preparative HPLC and subsequent lyophilisation of fractions. **119** (71.0 mg, 0.258 mmol, 71%) was obtained as a yellow solid.

Mpt: 189.7-195.3 °C; **ν_{max} (cm^{-1}):** 3357, 3300, 3151, 3031, 2974, 1688, 1663, 1570, 1518, 1429, 1405, 1358, 1293, 1257, 1205, 1062, 972, 907, 827; **^{1}H NMR (400 MHz, DMSO-d_6) δ**

8.93 (s, 1H, H-9), 8.58 (t, J = 5.8 Hz, 1H, H-6), 7.25 (t, J = 5.4 Hz, 1H, H-11), 4.13 (s, 2H, H-16), 4.04 (d, J = 5.8 Hz, 2H, H-7), 3.18 (qd, J = 7.2, 5.4 Hz, 2H, H-13), 2.36 (s, 3H, H-8), 1.07 (t, J = 7.2 Hz, 3H, H-14); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 167.3 (C-2), 166.5 (C-15), 157.9 (C-4), 153.6 (C-10), 105.5 (C-3), 42.5 (C-16), 34.4 (C-13), 30.5 (C-7), 15.2 (C-14), 10.9 (C-8); **LR-ESI-MS:** $C_{10}H_{16}ClN_4O_3$ $[M+H]^+$ *m/z* found 275.27 cald 275.08; **HR-ESI-MS:** $C_{10}H_{15}ClN_4O_3Na$ $[M+Na]^+$ *m/z* found 297.0571 cald 297.0725.

(3-(3-ethylureido)-5-methylisoxazol-4-yl)methyl 4-cyanobenzoate (120)

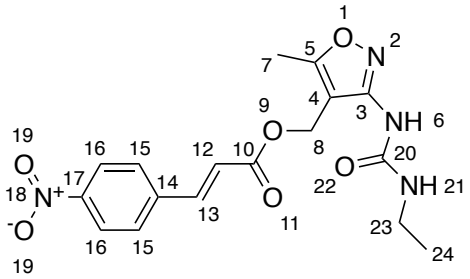


133 (51.3 mg, 0.199 mmol) was dissolved in anhydrous THF (5 mL) and DMA (1 mL) and cooled to 0 °C. Triphosgene (29.6 mg, 0.100 mmol) was dissolved in anhydrous THF (0.5 mL) and added dropwise. The reaction was stirred for 5 mins before anhydrous TEA (83.0 μL, 0.598 mmol) was added dropwise. The reaction was stirred at 0 °C for 30 mins. Ethylamine (150 μL, 0.299 mmol) was added dropwise. The reaction was allowed to warm to r.t. and stirred for 2 h. The reaction was poured into ice water and diluted with DCM. The organic phase was extracted, washed with brine, dried using Na_2SO_4 and the solvent was removed *in vacuo*. MeOH was added, the precipitated solid collected by filtration, washed with ice cold MeOH, and dried *in vacuo*. **120** (24.1 mg, 73.4 μmol, 37%) was obtained as a white solid.

Mpt: 216.2-220.1 °C; **ν_{max} (cm^{-1}):** 3338, 3244, 3130, 3013, 2932, 2226, 1733, 1683, 1650, 1577, 1444, 1379, 1303, 1266, 1175, 1102, 1016, 941, 885, 859; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 9.20 (s, 1H, H-6), 8.10-8.03 (m, 2H, H-13), 8.02-7.97 (m, 2H, H-14), 7.08 (t, J = 5.6 Hz, 1H, H-20), 5.21 (s, 2H, H-8), 3.14 (qd, J = 7.2, 5.6 Hz, 2H, H-21), 2.43 (s, 3H, H-7), 1.03 (t, J = 7.2 Hz, 3H, H-22); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 169.1 (C-5), 164.4 (C-10), 158.1 (C-3),

153.7 (C-18), 133.7 (C-15), 132.8 (C-14), 129.8 (C-13), 118.0 (C-16), 115.5 (C-12), 103.2 (C-4), 56.3 (C-8), 34.4 (C-21), 15.6 (C-22), 11.0 (C-7); **LR-ESI-MS:** $C_{16}H_{17}N_4O_4$ $[M+H]^+$ *m/z* found 329.40 cald 329.12; **HR-ESI-MS:** $C_{16}H_{16}N_4O_4Na$ $[M+Na]^+$ *m/z* found 351.0878 cald 351.1064.

(3-(3-ethylureido)-5-methylisoxazol-4-yl)methyl (*E*)-3-(4-nitrophenyl)acrylate (121)

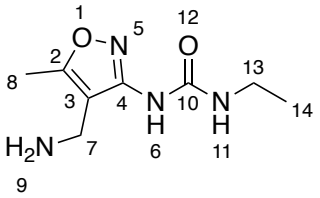


134 (55.4 mg, 0.183 mmol) was dissolved in anhydrous THF (5 mL) and DMA (1 mL) and cooled to 0 °C. Triphosgene (27.1 mg, 0.0913 mmol) was dissolved in anhydrous THF (0.5 mL) and added dropwise. The reaction was stirred for 5 mins at 0 °C before anhydrous TEA (76.0 μL, 0.548 mmol) was added dropwise. The reaction was stirred at 0 °C for 30 mins. Ethylamine (137 μL, 0.274 mmol) was added dropwise. The reaction was stirred at r.t. for 2 h. The reaction was poured onto ice and extracted 3 times with DCM. The organic phase was washed with brine, dried using Na_2SO_4 and solvent removed *in vacuo*. The crude product was purified using preparative HPLC and subsequent lyophilisation of fractions. **121** (5.2 mg, 13.9 μmol, 8%) was obtained as a white solid.

Mpt: 219.6-223.9 °C; **ν_{max} (cm^{-1}):** 3331, 3132, 3025, 1729, 1681, 1568, 1514, 1340, 1300, 1165, 1107, 950, 842; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 9.13 (s, 1H, H-6), 8.27-8.19 (m, 2H, H-16), 8.06-7.98 (m, 2H, H-15), 7.76 (d, J = 16.1 Hz, 1H, H-13), 7.09 (t, J = 5.6 Hz, 1H, H-21), 6.87 (d, J = 16.1 Hz, 1H, H-12), 5.10 (s, 2H, H-8), 3.16 (qd, J = 7.2, 5.6 Hz, 2H, H-23), 2.41 (s, 3H, H-7), 1.06 (t, J = 7.2 Hz, 3H, H-24); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 168.9 (C-5), 165.5 (C-10), 158.0 (C-3), 153.5 (C-20), 148.0 (C-17), 142.0 (C-13), 140.3 (C-14), 129.4 (C-15), 123.8 (C-16), 122.1 (C-12), 103.3 (C-4), 55.0 (C-8), 34.3 (C-23), 15.1 (C-24), 10.8 (C-7);

LR-ESI-MS: $C_{17}H_{19}N_4O_6$ $[M+H]^+$ *m/z* found 375.36 cald 375.13; **HR-ESI-MS:** $C_{17}H_{18}N_4O_6Na$ $[M+Na]^+$ *m/z* found 397.0912 cald 397.1119.

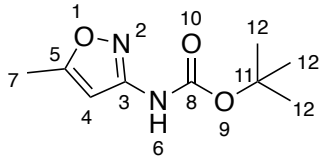
1-(4-(aminomethyl)-5-methylisoxazol-3-yl)-3-ethylurea (122)



130 (118 mg, 0.360 mmol) was suspended in EtOH (5 mL) and hydrazine monohydrate (118 μL, 1.95 mmol) added. The reaction was stirred at r.t. for 16 h. The reaction was filtered, the filtrate collected, and the solvent removed *in vacuo*. The crude product **122** was used directly without further purification.

LR-ESI-MS: $C_8H_{15}N_4O_2$ $[M+H]^+$ *m/z* found 199.26 cald 199.12.

***tert*-butyl (5-methylisoxazol-3-yl)carbamate (125)**

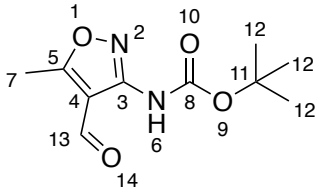


3-amino-5-methylisoxazole (10.0 g, 102 mmol), Boc anhydride (28.9 g, 133 mmol), DMAP (125 mg, 1.02 mmol) and anhydrous TEA (18.5 mL, 133 mmol) were dissolved in anhydrous DCM (160 mL) and stirred at r.t. for 16 h. The solvent was removed *in vacuo*. The crude product was purified using flash column chromatography eluting with a gradient of 0-60% EtOAc in Cy. **125** (17.1 g, 86.5 mmol, 85%) was obtained as a white solid.

Mpt: 82-89 °C; **ν_{max} (cm^{-1}):** 3341, 3214, 2986, 2871, 1735, 1610, 1476, 1413, 1365, 1277, 1244, 1146, 1084, 989, 898; **^{1}H NMR (400 MHz, $CDCl_3$) δ** 7.00 (s, 1H, H-6), 6.48 (br s, 1H, H-4), 2.37 (d, J = 0.9 Hz, 3H, H-7), 1.48 (s, 9H, H-12); **^{13}C NMR (101 MHz, $CDCl_3$) δ** 169.8 (C-5), 158.5 (C-3), 151.4 (C-8), 95.4 (C-4), 84.0 (C-11), 27.8 (C-12), 12.6 (C-7); **LR-ESI-MS:**

$C_9H_{15}N_2O_3$ $[M+H]^+$ *m/z* found 199.08 cald 199.10; **HR-ESI-MS:** $C_9H_{14}N_2O_3Na$ $[M+Na]^+$ *m/z* found 221.0776 cald 221.0897.

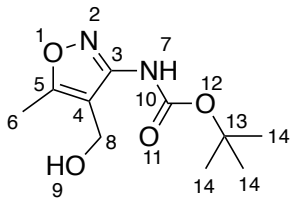
***tert*-butyl (4-formyl-5-methylisoxazol-3-yl)carbamate (126)**



125 (3.00 g, 15.1 mmol) was dissolved in anhydrous THF (75 mL) and cooled to -78 °C. *n*-BuLi 2.5 M in hexane (14.5 mL, 36.3 mmol) was added dropwise and stirred at -78 °C for 5 mins, before being allowed to warm to r.t. and stirred for 30 mins. The reaction was once again cooled to -78 °C and anhydrous DMF (2.3 mL, 30.3 mmol) was added dropwise. The reaction was stirred at r.t. for 16 h. The reaction was quenched by the dropwise addition of saturated aq. NH_4Cl and stirred for 30 mins. The crude product was extracted 3 times with EtOAc. The combined organic phases were washed with brine, dried using Na_2SO_4 and the solvent removed *in vacuo*. The crude product was purified using flash column chromatography eluting with a gradient of 0-50% EtOAc in Cy. **126** (2.04 g, 9.01 mmol, 60%) was obtained as a yellow oil.

ν_{max} (cm^{-1}): 3363, 2980, 2935, 2867, 1744, 1665, 1620, 1541, 1458, 1418, 1367, 1283, 1246, 1150, 1011, 943, 852; **^{1}H NMR (400 MHz, $CDCl_3$) δ** 9.89 (s, 1H, H-13), 8.38 (s, 1H, H-6), 2.67 (s, 3H, H-7), 1.53 (s, 9H, H-12); **^{13}C NMR (101 MHz, $CDCl_3$) δ** 184.6 (C-13), 176.8 (C-5), 156.5 (C-3), 150.3 (C-8), 109.2 (C-4), 82.7 (C-11), 28.2 (C-12), 12.0 (C-7); **LR-ESI-MS:** $C_{10}H_{15}N_2O_4$ $[M+H]^+$ *m/z* found 227.59 cald 227.10; **HR-ESI-MS:** $C_{10}H_{14}N_2O_4Na$ $[M+Na]^+$ *m/z* found 249.0712 cald 249.0846.

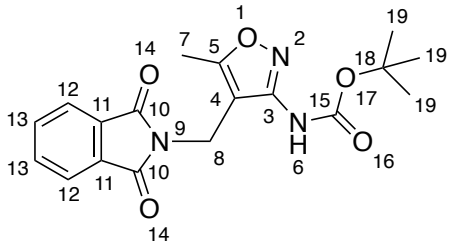
***tert*-butyl (4-(hydroxymethyl)-5-methylisoxazol-3-yl)carbamate (127)**



127 (1.86 g, 8.22 mmol) was dissolved in anhydrous MeOH (40 mL) and cooled to 0 °C. $NaBH_4$ (373 mg, 9.87 mmol) was added, and the reaction stirred at r.t. for 1 h. The reaction was quenched with water and extracted 3 times with DCM. The combined organic phases were washed with brine, dried using Na_2SO_4 and the solvent removed *in vacuo*. The crude product was purified using flash column chromatography eluting with a gradient of 10-60% EtOAc in Cy. **127** (946 mg, 4.14 mmol, 50%) was obtained as a white solid.

Mpt: 120.0-124.9 °C; **ν_{max} (cm^{-1}):** 3451, 3167, 3056, 2975, 1703, 1643, 1561, 1477, 1435, 1367, 1293, 1254, 1213, 1156, 1009, 917, 893; **^{1}H NMR (400 MHz, $CDCl_3$) δ** 7.00 (s, 1H, H-7), 4.37 (d, J = 6.8 Hz, 2H, H-8), 3.79 (t, J = 6.8 Hz, 1H, H-9), 2.41 (s, 3H, H-6), 1.51 (s, 9H, H-14); **^{13}C NMR (101 MHz, $CDCl_3$) δ** 169.3 (C-5), 156.9 (C-3), 153.3 (C-10), 109.6 (C-4), 82.7 (C-13), 53.3 (C-8), 28.1 (C-14), 11.4 (C-6); **LR-ESI-MS:** $C_{10}H_{17}N_2O_4$ $[M+H]^+$ *m/z* found 229.28 cald 229.11; **HR-ESI-MS:** $C_{10}H_{16}N_2O_4Na$ $[M+Na]^+$ *m/z* found 251.0868 cald 251.1002.

***tert*-butyl (4-((1,3-dioxoisoindolin-2-yl)methyl)-5-methylisoxazol-3-yl)carbamate (128)**

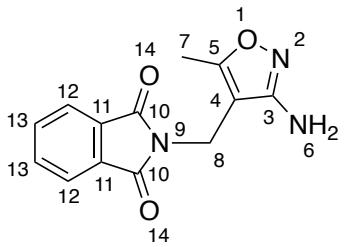


127 (736 mg, 3.23 mmol), phthalimide (617 mg, 4.19 mmol) and DPPE (771 mg, 1.94 mmol) were dissolved in anhydrous THF (29 mL) and cooled to 0 °C. DtBAD (965 mg, 4.19 mmol) was added, and the reaction stirred at r.t. for 16 h. The reaction was filtered, water added to the filtrate and extracted 3 times with DCM. The combined organic phases were washed with

brine, dried using Na_2SO_4 and the solvent removed *in vacuo*. The crude product was purified using flash column chromatography eluting with a gradient of 0-40 % EtOAc in Cy. **128** (401 mg, 1.12 mmol, 35%) was obtained as a white solid.

Mpt: 150.1-154.6 °C; **ν_{max} (cm^{-1}):** 3289, 3221, 2981, 2931, 1768, 1697, 1653, 1534, 1510, 1435, 1396, 1340, 1251, 1012, 911; **^{1}H NMR (400 MHz, $CDCl_3$) δ** 8.04 (s, 1H, H-6), 7.91-7.82 (m, 2H, H-12), 7.80-7.70 (m, 2H, H-13), 4.55 (s, 2H, H-8), 2.51 (s, 3H, H-7), 1.56 (s, 9H, H-19); **^{13}C NMR (101 MHz, $CDCl_3$) δ** 169.0 (C-5), 168.6 (C-10), 156.9 (C-3), 151.4 (C-15), 134.6 (C-13), 131.9 (C-11), 123.9 (C-12), 103.8 (C-4), 81.7 (C-18), 29.1 (C-8), 28.3 (C-19), 11.5 (C-7); **LR-ESI-MS:** $C_{18}H_{20}N_3O_5$ $[M+H]^+$ *m/z* found 358.37 cald 358.14; **HR-ESI-MS:** $C_{18}H_{19}N_3O_5Na$ $[M+Na]^+$ *m/z* found 380.1026 cald 380.1217.

2-((3-amino-5-methylisoxazol-4-yl)methyl)isoindoline-1,3-dione (129)

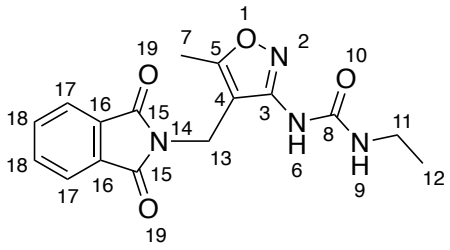


128 (229 mg, 0.639 mmol) was dissolved in anhydrous DCM (5.9 mL) and TFA (3.3 mL, 42.8 mmol) was added dropwise. The reaction was stirred at r.t. for 1.5 h. The volatile reagents were removed *in vacuo*. The crude oil was redissolved in minimal DCM and washed with saturated aq. Na_2CO_3. The organic phases were combined, dried using Na_2SO_4, and the solvent removed *in vacuo*. **129** (164 mg, 0.638 mmol, 100%) was obtained as a light brown solid.

Mpt: 169.7-172.8 °C; **ν_{max} (cm^{-1}):** 3411, 3196, 2928, 1773, 1697, 1661, 1529, 1481, 1465, 1433, 1398, 1352, 1323, 1217, 1043, 921; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 7.93-7.81 (m, 4H, H-12, H-13), 5.33 (s, 2H, H-6), 4.45 (s, 2H, H-8), 2.32 (s, 3H, H-7); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 168.1 (C-10), 167.3 (C-5), 162.7 (C-3), 134.7 (C-13), 131.5 (C-11), 123.3 (C-12), 102.2

(C-4), 29.1 (C-8), 11.1 (C-7); **LR-ESI-MS:** $C_{13}H_{12}N_3O_3$ $[M+H]^+$ *m/z* found 258.31 cald 258.08; **HR-ESI-MS:** $C_{13}H_{11}N_3O_3Na$ $[M+Na]^+$ *m/z* found 280.0546 cald 280.0693.

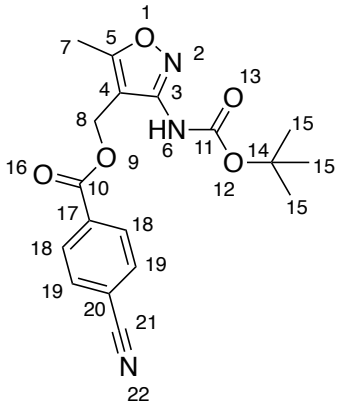
1-(4-((1,3-dioxoisoindolin-2-yl)methyl)-5-methylisoxazol-3-yl)-3-ethylurea (130)



129 (262 mg, 1.02 mmol) was dissolved in anhydrous DCM (17 mL) and cooled to 0 °C. Triphosgene (151 mg, 0.508 mmol) was dissolved in anhydrous DCM (1 mL) and added dropwise. The reaction was stirred for 5 mins before anhydrous TEA (425 μL, 3.05 mmol) was added dropwise. The reaction was stirred at 0 °C for 20 mins and ethylamine (763 μL, 1.53 mmol) was added dropwise. The reaction was allowed to warm to r.t. and stirred for 2 h. Additional ethylamine (112 μL, 0.383 mmol) was added and stirred at r.t for 1 h. The reaction was poured into ice water and diluted with DCM. The organic phase was extracted, washed with brine, dried using Na_2SO_4 and the solvent was removed *in vacuo*. The crude product was purified by trituration in EtOH. **130** was used directly without further purification.

LR-ESI-MS: $C_{16}H_{17}N_4O_4$ $[M+H]^+$ *m/z* found 329.38 cald 329.12.

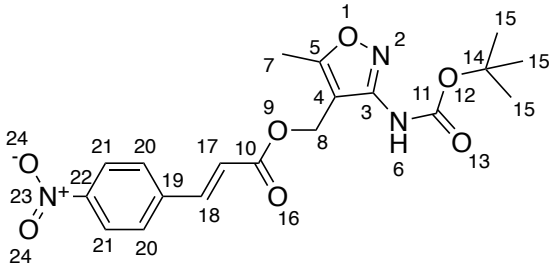
(3-((*tert*-butoxycarbonyl)amino)-5-methylisoxazol-4-yl)methyl 4-cyanobenzoate (131)



127 (216 mg, 0.944 mmol), 4-cyanobenzoic acid (125 mg, 0.850 mmol) and DMAP (20.8 mg, 0.170 mmol) were dissolved in anhydrous DMF (8.9 mL) and cooled to 0 °C. EDCI (195 mg, 1.02 mmol) was added, and the reaction was stirred at r.t. for 16 h. The reaction was quenched with water and extracted 3 times with DCM. The combined organic phases were washed with brine, dried using Na_2SO_4 and solvent removed *in vacuo*. The crude product was purified using flash column chromatography eluting with a gradient of 0-50% EtOAc in Cy. **131** (153 mg, 0.427 mmol, 50%) was obtained as a white solid.

Mpt: 172.5-176.9 °C; **ν_{max} (cm^{-1}):** 3310, 2986, 2232, 1698, 1480, 1432, 1366, 1324, 1245, 1152, 1051, 1020, 931, 899, 864; **^{1}H NMR (400 MHz, $CDCl_3$) δ** 8.14-8.10 (m, 2H, H-18), 7.77-7.73 (m, 2H, H-19), 5.24 (s, 2H, H-8), 2.47 (s, 3H, H-7), 1.50 (s, 9H, H-15); **^{13}C NMR (101 MHz, $CDCl_3$) δ** 170.2 (C-5), 165.5 (C-10), 157.0 (C-3), 151.7 (C-11), 133.5 (C-20), 132.4 (C-19), 130.4 (C-18), 117.9 (C-21), 117.0 (C-17), 104.8 (C-4), 82.2 (C-14), 56.8 (C-8), 28.3 (C-15), 11.8 (C-7); **LR-ESI-MS:** $C_{18}H_{20}N_3O_5$ $[M+H]^+$ *m/z* found 358.39 cald 358.14; **HR-ESI-MS:** $C_{18}H_{19}N_3O_5Na$ $[M+Na]^+$ *m/z* found 380.1024 cald 380.1217.

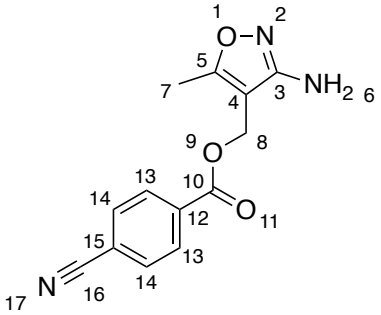
(3-((*tert*-butoxycarbonyl)amino)-5-methylisoxazol-4-yl)methyl (*E*)-3-(4-nitrophenyl)acrylate (132)



127 (216 mg, 0.944 mmol), 3-(4-nitrophenyl)prop-2-enoic acid (164 mg, 0.850 mmol) and DMAP (30.8 mg, 0.252 mmol) were dissolved in anhydrous DMF (8.9 mL) and cooled to 0 °C. EDCI (195 mg, 1.02 mmol) was added and the reaction was stirred at r.t. for 16 h. The reaction was quenched with water and extracted 3 times with EtOAc. The organic phase was washed with 5% LiCl, brine, dried using Na_2SO_4 and the solvent removed *in vacuo*. The crude product was purified using flash column chromatography eluting with a gradient of 0-50% EtOAc in Cy. **132** (241 mg, 0.597 mmol, 70%) was obtained as a white solid.

Mpt: 139.9-143.4 °C; **ν_{max} (cm^{-1}):** 3316, 2989, 2940, 1720, 1701, 1634, 1538, 1513, 1429, 1366, 1270, 1150, 1068, 961, 861; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 9.70 (s, 1H, H-6), 8.27-8.22 (m, 2H, H-20), 8.03-7.99 (m, 2H, H-20), 7.75 (d, J = 16.1 Hz, 1H, H-18), 6.86 (d, J = 16.0 Hz, 1H, H-17), 5.10 (s, 2H, H-8), 2.46 (s, 3H, H-7), 1.42 (s, 9H, H-15); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 169.4 (C-5), 165.5 (C-10), 157.3 (C-3), 152.3 (C-11), 148.1 (C-22), 142.2 (C-18), 140.4 (C-19), 129.5 (C-21), 124.0 (C-20), 122.1 (C-17), 105.6 (C-4), 80.1 (C-14), 55.7 (C-8), 27.9 (C-15), 11.2 (C-7); **LR-ESI-MS:** $C_{19}H_{20}N_3O_7$ $[M-H]^-$ *m/z* found 402.20 cald 402.14; **HR-ESI-MS:** $C_{19}H_{21}N_3O_7Na$ $[M+Na]^+$ *m/z* found 426.1054 cald 426.1272.

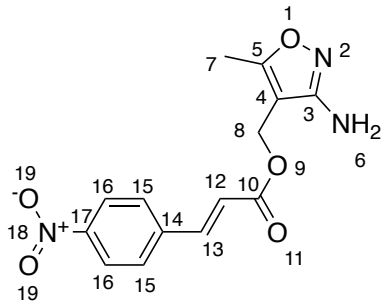
(3-amino-5-methylisoxazol-4-yl)methyl 4-cyanobenzoate (133)



131 (78.2 mg, 0.219 mmol) was dissolved in DCM (1.4 mL) and TFA (1.1 mL, 14.7 mmol) added dropwise. The reaction was stirred at r. t. for 1 h. The volatile reagents were removed *in vacuo*. The crude oil was redissolved in minimal DCM and washed with saturated aq. Na_2CO_3. The organic phases were combined, dried using Na_2SO_4, and the solvent removed *in vacuo*. **133** (51.3 mg, 0.199 mmol, 91%) was obtained as a white gum.

ν_{max} (cm^{-1}): 3399, 3321, 3212, 2929, 2890, 2234, 1652, 1525, 1477, 1422, 1297, 1276, 1204, 1121, 988, 897; **^{1}H NMR (400 MHz, MeOD) δ** 8.21-8.12 (m, 2H, H-13), 7.92-7.82 (m, 2H, H-14), 5.19 (s, 2H, H-8), 2.38 (s, 3H, H-7); **^{13}C NMR (101 MHz, MeOD) δ** 170.2 (C-5), 166.6 (C-10), 165.0 (C-3), 135.2 (C-15), 133.6 (C-14), 131.3 (C-13), 118.9 (C-16), 117.8 (C-12), 107.7 (C-4), 57.1 (C-8), 11.2 (C-7); **LR-ESI-MS:** $C_{13}H_{12}N_3O_3$ $[M+H]^+$ *m/z* found 258.15 cald 258.08; **HR-ESI-MS:** $C_{13}H_{11}N_3O_3Na$ $[M+Na]^+$ *m/z* found 280.0543 cald 280.0693.

(3-amino-5-methylisoxazol-4-yl)methyl (*E*)-3-(4-nitrophenyl)acrylate (134)

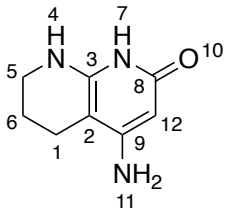


132 (80.5 mg, 0.200 mmol) was dissolved in DCM (1.2 mL) and TFA (1.0 mL, 13.4 mmol) added dropwise. The reaction was stirred at r.t. for 1.5 h. The volatile reagents were removed *in vacuo*. The crude oil was redissolved in minimal DCM and washed with saturated aq.

Na_2CO_3. The organic phases were combined, dried using Na_2SO_4, and the solvent removed *in vacuo*. **134** (60.4 mg, 0.199 mmol, 100%) was obtained as a light brown solid.

Mpt: 204.7-209.6 °C; **ν_{max} (cm^{-1}):** 3405, 3321, 3214, 2930, 1708, 1658, 1640, 1597, 1515, 1455, 1378, 1333, 1189, 1108, 983, 841; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 8.24 (d, J = 8.9 Hz, 2H, H-16), 8.02 (d, J = 8.9 Hz, 2H, H-15), 7.77 (d, J = 16.1 Hz, 1H, H-13), 6.87 (d, J = 16.1 Hz, 1H, H-12), 5.60 (s, 2H, H-6), 4.99 (s, 2H, H-8), 2.29 (s, 3H, H-7); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 167.5 (C-5), 165.9 (C-10), 163.1 (C-3), 148.1 (C-17), 142.1 (C-14), 140.5 (C-13), 129.5 (C-15), 124.0 (C-16), 122.3 (C-12), 101.6 (C-4), 55.2 (C-8), 11.0 (C-7); **LR-ESI-MS:** $C_{14}H_{14}N_3O_5$ $[M+H]^+$ *m/z* found 304.34 cald 304.09; **HR-ESI-MS:** $C_{14}H_{14}N_3O_5$ $[M+H]^+$ *m/z* found 304.0762 cald 304.0928.

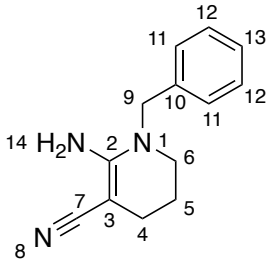
4-amino-5,6,7,8-tetrahydro-1,8-naphthyridin-2(1*H*)-one (141)



151 (340 mg, 1.332mmol) was dissolved in anhydrous THF (5 mL) and cooled to -78 °C. Gaseous NH_3 was condensed into the flask (approx. 10 mL), and sodium metal added in portions until the deep blue colour persisted for more than 15 mins. The reaction was stirred at -78 °C for 1 h. The reaction was quenched with NH_4Cl and stirred at -78 °C for 1 h before being allowed to r.t. and stirred for 16 h. Ice water was added, and the reaction stirred for 30 mins. The aqueous phase was washed 3 times with EtOAc. The aqueous phase was extracted, and the solvent removed *in vacuo*. The crude product was then redissolved in minimal MeOH, collected by filtration and the solvent removed *in vacuo*. This process was repeated twice. **141** was used directly without further purification.

LR-ESI-MS: $C_8H_{12}N_3O$ $[M+H]^+$ *m/z* found 166.11 cald 166.09

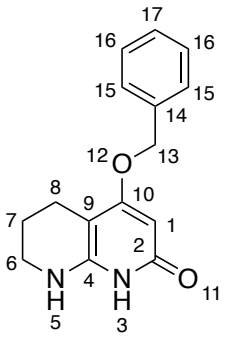
2-amino-1-benzyl-1,4,5,6-tetrahydropyridine-3-carbonitrile (142)



160 (1.01 g, 3.32 mmol) was dissolved in anhydrous DCE (50 mL) and 1-chloroethyl chloroformate (1.4 mL, 13.4 mmol) added. The reaction was refluxed for 16 h. Upon reaction completion as monitored by LCMS, the solvent was removed *in vacuo* and MeOH (50 mL) added. The reaction was refluxed for 16 h. The reaction was quenched using saturated aq. Na_2CO_3 and stirred for 30 mins. The reaction was extracted 3 times using DCM and the organic phase washed with brine, dried using Na_2SO_4 and concentrated *in vacuo*. The crude product was purified using flash column chromatography eluting with a gradient of 0-10% MeOH in DCM. **142** (425 mg, 1.99 mmol, 60%) was obtained as a yellow oil.

ν_{max} (cm^{-1}): 3308, 2937, 2360, 1601, 1495, 1453, 1355, 1287, 1117, 1028; **^{1}H NMR (400 MHz, MeOD) δ** 7.40-7.22 (m, 7H, H-11, H-12, H-13, H-14), 4.47 (s, 2H, 9), 3.20-3.13 (m, 2H, H-6), 2.23 (t, J = 6.3 Hz, 2H, H-4), 1.74 (dq, J = 7.1, 5.8 Hz, 2H, H-5); **^{13}C NMR (101 MHz, Methanol-*d*$_4$) δ** 159.5 (C-2), 138.9 (C-10), 129.8 (C-12), 128.4 (C-11), 128.1 (C-13), 127.4 (C-7), 54.1 (C-9), 51.7 (C-3), 50.2 (C-6), 23.9 (C-4), 23.2 (C-5); **LR-ESI-MS:** $C_{13}H_{16}N_3$ $[M+H]^+$ *m/z* found 214.29 cald 214.12; **HR-ESI-MS:** $C_{13}H_{16}N_3$ $[M+H]^+$ *m/z* found 214.0633 cald 214.1339.

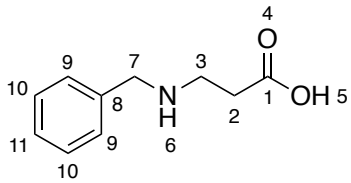
4-(benzyloxy)-5,6,7,8-tetrahydro-1,8-naphthyridin-2(1*H*)-one (143)



138 (40.5 mg, 0.244 mmol), benzyl chloride (28.0 μL, 0.246 mmol) and K_2CO_3 (40.4 mg, 0.292 mmol) were suspended in anhydrous DMF (2 mL) and heated to 90 °C for 3 h. The solvent was removed *in vacuo*, and the crude product purified by preparative HPLC. **143** (2.2 mg, 8.6 μmol, 4%) was obtained as a dark red gum.

ν_{max} (cm^{-1}): 3246, 3062, 2927, 2847, 1594, 1493, 1465, 1336, 1189, 1023; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 9.82 (br s, 1H, H-3), 7.43-7.24 (m, 5H, H-15, H-16, H-17), 5.95 (d, J = 2.8 Hz, 1H, H-5), 5.48 (s, 1H, H-1), 5.15 (s, 2H, H-13), 3.20-3.12 (m, 2H, H-6), 2.41 (t, J = 6.4 Hz, 2H, H-8), 1.76-1.67 (m, 2H, H-7); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 161.4 (C-10), 155.2 (C-2), 144.4 (C-4), 136.8 (C-14), 128.2 (C-16), 127.7 (C-15), 127.4 (C-17), 93.5 (C-9), 84.4 (C-1), 65.7 (C-13), 41.0 (C-6), 21.1 (C-7), 19.6 (C-8); **LR-ESI-MS:** $C_{15}H_{17}N_2O_2$ $[M+H]^+$ *m/z* found 257.35 cald 257.12; **HR-ESI-MS:** $C_{15}H_{17}N_2O_2$ $[M+H]^+$ *m/z* found 257.0489 cald 257.1285.

3-(benzylamino)propanoic acid (146)

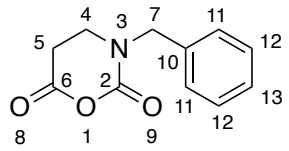


Ethyl 3-(benzylamino)propanoate) (1.00 g, 4.83 mmol) was dissolved in MeOH (20 mL) and 1M aq. NaOH (5.8 mL, 5.79 mmol) was added dropwise. The reaction was stirred at r.t. for 16 h. The reaction mixture was filtered, and the precipitated residue discarded. The filtrate was concentrated *in vacuo* to a volume of 1 mL, acidified using 1 M HCl and the solvent removed

in vacuo. The residue was suspended in MeOH, collected by filtration and washed twice with ice cold MeOH. **146** (765 mg, 4.27 mmol, 89%) was obtained as a white solid.

Mpt: 174.3-177.6 °C; **ν_{max} (cm^{-1}):** 3074, 2933, 2812, 1750, 1466, 1354, 1212, 1024, 998, 893; **^{1}H NMR (400 MHz, MeOD) δ** 7.50-7.44 (m, 5H, H-9, H-10, H-11), 4.21 (s, 2H, H-7), 3.18 (t, J = 6.4 Hz, 2H, H-3), 2.51 (t, J = 6.4 Hz, 2H, H-2); **^{13}C NMR (101 MHz, MeOD) δ** 177.8 (C-1), 132.9 (C-8), 130.7 (C-10), 130.6 (C-9), 130.3 (C-11), 51.8 (C-7), 45.5 (C-3), 32.7 (C-2); **LR-ESI-MS:** $C_{10}H_{14}NO_2$ $[M+H]^+$ *m/z* found 180.22 cald 180.10; **HR-ESI-MS:** $C_{10}H_{14}NO_2$ $[M+H]^+$ *m/z* found 180.0439 cald 180.1025.

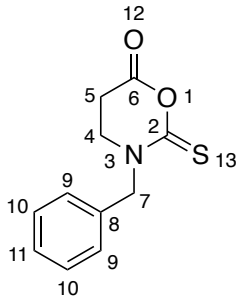
3-benzyl-1,3-oxazinane-2,6-dione (147)



146 (40.8 mg, 0.228 mmol) was dissolved in anhydrous THF (0.7 mL) and stirred for 5 minutes. Triphosgene (45.3 mg, 0.153 mmol) was dissolved in anhydrous THF (0.7 mL) and added dropwise. The reaction was stirred at r.t. for 5 mins and then refluxed for 3 h. The reaction was allowed to cool to r.t. before poured onto ice water and stirred for 30 mins. The solution was then extracted 3 times with DCM. The combined organic phases were then washed with brine, dried using Na_2SO_4 and the solvent removed *in vacuo*. The crude product was purified using preparative HPLC and subsequent lyophilisation of fractions. **147** was used directly without further purification.

LR-ESI-MS: $C_{11}H_{12}NO_3$ $[M+H]^+$ *m/z* found 206.27 cald 206.08.

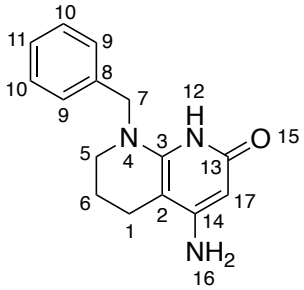
3-benzyl-2-thioxo-1,3-oxazinan-6-one (148)



146 (53.2 mg, 0.297 mmol) was dissolved in anhydrous DCM (2.5 mL) and anhydrous DMF (2.5 mL), then anhydrous TEA (103 μL, 0.740 mmol) was added. TMSCl (74.0 μL, 0.587 mmol) was added and the reaction was stirred at r.t. for 1.5 h. Thiophosgene (21.0 μL, 0.278 mmol) was added and the reaction was stirred for at r.t. for 2 h. The reaction was quenched with water and the organic phase extracted. The organic phase was washed with 5% LiCl and brine, dried using Na_2SO_4 and the solvent removed *in vacuo*. The crude product was purified by flash column chromatography eluting with a gradient of 0-50% EtOAc in Cy. **148** (8.0 mg, 36.2 μmol, 12%) was obtained as a yellow oil.

ν_{max} (cm^{-1}): 2935, 2739, 1703, 1655, 1415, 1232, 1076, 1032; **^{1}H NMR (400 MHz, $CDCl_3$) δ** 7.41-7.28 (m, 5H, H-9, H-10, H-11), 4.74 (s, 2H, H-7), 3.53-3.48 (m, 2H, H-4), 2.77-2.72 (m, 2H, H-5); **^{13}C NMR (101 MHz, $CDCl_3$) δ** 196.9 (C-6), 163.2 (C-2), 135.6 (C-8), 129.2 (C-10), 128.5 (C-11), 128.3 (C-9), 52.3 (C-7), 43.4 (C-4), 41.10 (C-5); **LR-ESI-MS:** $C_{11}H_{12}NO_2S$ $[M+H]^+$ *m/z* found 222.23 cald 222.05; **HR-ESI-MS:** $C_{11}H_{12}NO_2S$ $[M+H]^+$ *m/z* found 222.0440 cald 222.0583.

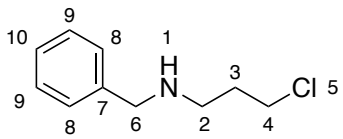
4-amino-8-benzyl-5,6,7,8-tetrahydro-1,8-naphthyridin-2(1*H*)-one (151)



165 (38.9 mg, 0.152 mmol) was dissolved in anhydrous THF (1.4 mL) and 1.0 M LiHMDS in THF (914 μL, 0.914 mmol) was added and the reaction refluxed for 5 h. The reaction was then poured into water and the product extracted using DCM and EtOAc. The combined organic layers were washed with brine, dried using Na_2SO_4 and the solvent removed *in vacuo*. **151** (37.0 mg, 0.145 mmol, 95%) was obtained as a brown oil.

ν_{max} (cm^{-1}): 3339, 3226, 3031, 1621, 1576, 1329, 1207, 1127, 1024; **^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ** 7.32-7.26 (m, 2H, H-10), 7.24-7.17 (m, 3H, H-9, H-11), 5.31 (s, 2H, H-16), 5.07 (s, 1H, H-17), 4.67 (s, 2H, H-7), 3.11 (t, J = 5.5 Hz, 2H, H-5), 2.27 (t, J = 6.4 Hz, 2H, H-1), 1.76 (m, 2H, H-6); **^{13}C NMR (101 MHz, DMSO-*d*$_6$) δ** 161.5 (C-13), 156.1 (C-14), 151.8 (C-3), 139.7 (C-8), 128.7 (C-10), 127.8 (C-9), 127.0 (C-11), 89.0 (C-2), 83.5 (C-17), 51.7 (C-7), 47.5 (C-5), 21.6 (C-6), 21.1 (C-1); **LR-ESI-MS:** $C_{15}H_{18}N_3O$ $[M+H]^+$ *m/z* found 256.33 cald 256.14; **HR-ESI-MS:** $C_{15}H_{18}N_3O$ $[M+H]^+$ *m/z* found 256.0664 cald 256.1444.

***N*-benzyl-3-chloropropan-1-amine (154)**

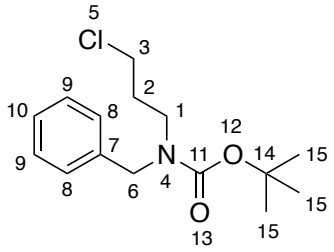


3-chloropropan-1-amine hydrochloride (520 mg, 4.00 mmol), benzaldehyde (407 μL, 4.00 mmol) and anhydrous TEA (1.1 mL, 8.00 mmol) were dissolved in MeOH (20 mL) and refluxed for 16 h. The reaction mixture was then cooled to 0 °C and $NaBH_4$ (303 mg, 8.00 mmol) was added in portions. The reaction was stirred at r.t. for 2 h before the reaction was quenched

with water. The organic solvent was removed *in vacuo*, and the aqueous layer extracted with DCM. The organic phase was dried using Na_2SO_4 and the solvent removed *in vacuo*. The crude product was purified using flash column chromatography (SNAP-KP-NH column, Biotage) eluting with a gradient of 20-80% EtOAc in Cy. **154** (415 mg, 2.26 mmol, 56%) was obtained as a white gum.

ν_{max} (cm^{-1}): 3056, 2941, 2777, 1447, 1304, 1018, 918; **^{1}H NMR (400 MHz, MeOD) δ** 7.36-7.29 (m, 4H, H-8, H-9), 7.28-7.22 (m, 1H, H-6), 3.62 (t, J = 6.5 Hz, 2H, H-4), 2.76-2.68 (m, 2H, H-2), 2.03-1.91 (m, 2H, H-3); **^{13}C NMR (101 MHz, Methanol-d_4) δ** 140.5 (C-7), 129.5 (C-9), 129.5 (C-8), 128.2 (C-10), 54.5 (C-6), 47.2 (C-2), 43.8 (C-4), 33.4 (C-3); **LR-ESI-MS:** $C_{10}H_{15}ClN$ $[M+H]^+$ *m/z* found 184.18 cald 184.08; **HR-ESI-MS:** $C_{10}H_{15}ClN$ $[M+H]^+$ *m/z* found 184.0303 cald 184.0893.

***tert*-butyl benzyl(3-chloropropyl)carbamate (156)**

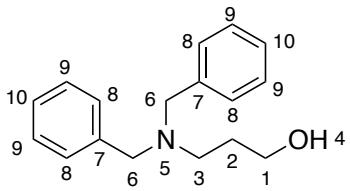


154 (37.4 mg, 0.204 mmol) was dissolved in anhydrous DCM (0.5 mL) and a solution of Boc anhydride (44.4 mg, 0.204 mmol) in anhydrous DCM (0.5 mL) was added dropwise. The reaction was stirred at r.t. for 16 h. The solvent was then removed *in vacuo* and the crude product purified by preparative HPLC. **156** (16.6 mg, 58.5 μmol, 29%) was obtained as a white gum.

ν_{max} (cm^{-1}): 3249, 2975, 1686, 1453, 1413, 1365, 1242, 1157, 1115; **^{1}H NMR (400 MHz, $CDCl_3$) δ** 7.36-7.18 (m, 5H, H-8, H-9, H-10), 4.45 (s, 2H, H-6), 3.51 (t, J = 6.2 Hz, 2H, H-3), 3.32 (s, 2H, H-1), 2.02-1.90 (m, 2H, H-2), 1.52-1.43 (m, 9H, H-15); **^{13}C NMR (101 MHz, $CDCl_3$) δ** 157.0 (C-11), 138.3 (C-7), 128.6 (C-9), 127.8 (C-10), 127.3 (C-8), 80.0 (C-14), 51.2 (C-6),

44.3 (C-1), 42.5 (C-3), 31.2 (C-2), 28.4 (C-15); **LR-ESI-MS:** $C_{15}H_{23}ClNO_2$ $[M+H]^+$ *m/z* found 284.12 cald 284.14; **HR-ESI-MS:** $C_{15}H_{22}ClNO_2Na$ $[M+Na]^+$ *m/z* found 306.0298 cald 306.1231.

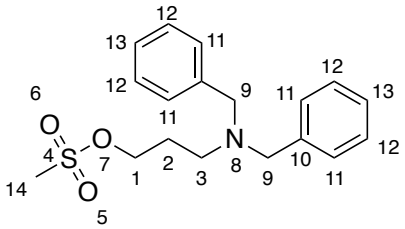
3-(dibenzylamino)propan-1-ol (158)



3-aminopropan-1-ol (102 μL, 1.33 mmol) and K_2CO_3 (920 mg, 6.66 mmol) were added to anhydrous acetone (2 mL) and benzyl bromide (333 μL, 2.80 mmol) added. The reaction was stirred at r.t. for 16 h, filtered and the solvent removed *in vacuo*. The crude product was purified using flash column chromatography eluting with 0-50% EtOAc in Cy. **158** (244 mg, 0.956 mmol, 72%) was obtained as a yellow oil.

ν_{max} (cm^{-1}): 3319, 3026, 2940, 2799, 1600, 1494, 1451, 1365, 1126, 1068, 909; **^{1}H NMR (400 MHz, $CDCl_3$) δ** 7.37-7.23 (m, 10H, H-8, H-9, H-10), 4.64-4.53 (m, 1H, H-4), 3.65 (t, J = 5.3 Hz, 2H, H-1), 3.58 (s, 4H, H-6), 2.65 (t, J = 5.8 Hz, 2H, H-3), 1.81-1.72 (m, 2H, H-2); **^{13}C NMR (101 MHz, $CDCl_3$) δ** 138.4 (C-7), 129.3 (C-8), 128.6 (C-9), 127.5 (C-10), 64.0 (C-1), 58.8 (C-6), 53.3 (C-3), 28.1 (C-2); **LR-ESI-MS:** $C_{17}H_{22}NO$ $[M+H]^+$ *m/z* found 256.39 cald 256.17; **HR-ESI-MS:** $C_{17}H_{22}NO$ $[M+H]^+$ *m/z* found 256.0907 cald 256.1696.

3-(dibenzylamino)propyl methanesulfonate (159)



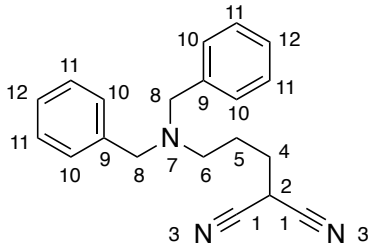
Methanesulfonyl chloride (540 μL, 6.98 mmol) was dissolved in anhydrous DCM (7 mL) and cooled to 0 °C. **158** (885 mg, 3.46 mmol), anhydrous DIPEA (3.0 mL, 17.3 mmol) and DMAP

(42.3 mg, 0.346 mmol) were dissolved in anhydrous DCM (3.5 mL) and added dropwise. The reaction was stirred at 0 °C for 15 mins before being stirred at r.t. for 16 h. The reaction was poured into a mixture of ice water and DCM. The organic layer was extracted, washed with brine, dried using Na_2SO_4 and the solvent removed *in vacuo* to give **159** was used directly without further purification.

N.B. In synthesis of **160**, product was purified using flash column chromatography eluting with a gradient of 0-50% EtOAc in Cy. **159** was obtained as a yellow oil.

ν_{max} (cm^{-1}): 3001, 2929, 1643, 1454, 1182, 1039; **^{1}H NMR (400 MHz, $CDCl_3$) δ** 7.33-7.18 (m, 10H, H-11, H-12, H-13,), 4.16 (t, J = 6.5 Hz, 2H, 1), 3.48 (s, 4H, H-9), 2.74 (s, 3H, H-14), 2.48 (t, J = 6.5 Hz, 2H, H-3), 1.83 (t, J = 6.5 Hz, 2H, H-2); **^{13}C NMR (101 MHz, $CDCl_3$) δ** 139.4 (C-10), 129.0 (C-11), 128.4 (C-12), 127.2 (C-13), 68.3 (C-1), 58.9 (C-9), 49.4 (C-3), 37.3 (C-14), 27.2 (C-2); **LR-ESI-MS:** $C_{18}H_{23}NO_3S$ $[M+H]^+$ *m/z* found 334.42 cald 334.14; **HR-ESI-MS:** $C_{18}H_{23}NO_3S$ $[M+H]^+$ *m/z* found 334.0500 cald 334.1477.

2-(3-(dibenzylamino)propyl)malononitrile (160)

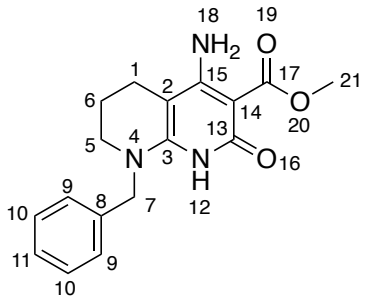


Malononitrile (203 mg, 3.08 mmol) and anhydrous DIPEA (1.20 mL, 4.13 mmol) were dissolved in anhydrous MeCN (10 mL) and stirred at r.t. for 1 h. **159** (685 mg, 2.05 mmol) was dissolved in anhydrous MeCN (2 mL) and added to the reaction. The reaction was heated to 60 °C and stirred for 16 h. The solvent was removed *in vacuo* before water and DCM were added. The organic layer was extracted and washed with brine, dried using Na_2SO_4 and the solvent removed *in vacuo*. The crude product was purified using flash column chromatography

eluting with a gradient of 0-50% EtOAc in Cy. **160** (326 mg, 1.07 mmol, 52%) was obtained as a yellow oil.

ν_{max} (cm^{-1}): 2890, 2799, 2255, 1600, 1494, 1452, 1127, 1052, 1026, 968; **^{1}H NMR (400 MHz, $CDCl_3$) δ** 7.41-7.27 (m, 10H, H-10, H-11, H-12), 3.54 (s, 4H, H-8), 3.26 (t, J = 7.3 Hz, 1H, H-2), 2.50-2.43 (t, J = 5.9 Hz, 2H, H-6), 2.06-1.96 (m, 2H, H-4), 1.78-1.67 (m, 2H, H-5); **^{13}C NMR (101 MHz, $CDCl_3$) δ** 139.2 (C-9), 129.1 (C-11), 128.7 (C-10), 127.5 (C-12), 112.9 (C-1), 58.9 (C-8), 50.5 (C-6), 28.6 (C-4), 23.5 (C-5), 21.3 (C-2); **LR-ESI-MS:** $C_{20}H_{22}N_3$ $[M+H]^+$ *m/z* found 304.40 cald 304.18; **HR-ESI-MS:** $C_{20}H_{22}N_3$ $[M+H]^+$ *m/z* found 304.0924 cald 304.1808.

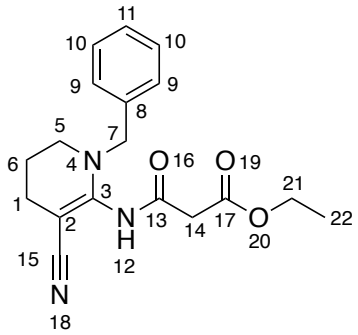
Methyl 4-amino-8-benzyl-2-oxo-1,2,5,6,7,8-hexahydro-1,8-naphthyridine-3-carboxylate (162)



Dimethyl malonate (162 μL, 1.39 mmol) and 0.5 M NaOMe in MeOH (438 μL, 1.64 mmol) were dissolved in MeOH (0.7 mL) and stirred at r.t. for 10 mins. **142** (74.4 mg, 0.349 mmol) was added and the reaction refluxed for 16 h. The reaction mixture was filtered and concentrated. **162** was used directly without purification.

LR-ESI-MS: $C_{17}H_{20}N_3O_3$ $[M+H]^+$ *m/z* found 314.43 cald 314.5.

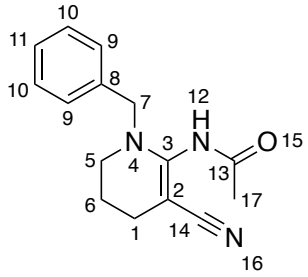
Ethyl 3-((1-benzyl-3-cyano-1,4,5,6-tetrahydropyridin-2-yl)amino)-3-oxopropanoate (163)



142 (53.6 mg, 0.251 mmol) was dissolved in anhydrous DCM (0.5 mL) and cooled to 0 °C. 2-ethylpropanedioyl dichloride (39.0 μL, 0.305 mmol) was added and the reaction stirred at 0 °C for 30 mins. Anhydrous TEA (35.0 μL, 0.254 mmol) was added and the reaction warmed to r.t. and stirred for 1 h. The solvent was removed *in vacuo* and the crude reaction was dissolved in anhydrous 1,4-dioxane (0.2 mL). 1.0 M NaHMDS in THF (377 μL, 0.377 mmol) was added and the reaction was refluxed overnight. The solvent was removed *in* vacuo and the reaction was filtered and purified by preparative HPLC. **163** (2.4 mg, 7.30 *μ*mol, 3%) was obtained as a brown oil.

ν_{max} (cm^{-1}): 3172, 2978, 2362, 1734, 1653, 1559, 1506, 1455, 1363, 1250, 1162, 1029; **^{1}H NMR (400 MHz, $CDCl_3$) δ** 8.85 (s, 1H, H-12), 7.38-7.28 (m, 3H, H-9, H-11), 7.24-7.18 (m, 2H, H-10), 4.32 (s, 2H, H-7), 4.19 (q, J = 7.1 Hz, 2H, H-21), 3.45 (s, 2H, H-14), 3.17 (t, J = 5.6 Hz, 2H, H-5), 2.35 (t, J = 6.3 Hz, 2H, H-1), 1.76-1.70 (m, 2H, H-6), 1.28 (dt, J = 17.3, 7.2 Hz, 3H, H-22); **^{13}C NMR (101 MHz, $CDCl_3$) δ** 169.4 (C-17), 164.5 (C-13), 147.5 (C-3), 137.0 (C-8), 129.0 (C-10), 127.9 (C-11), 127.7 (C-9), 121.7 (C-15), 73.4 (C-2), 62.3 (C-21), 54.1 (C-7), 47.3 (C-5), 41.1 (C-14), 24.0 (C-1), 20.9 (C-6), 14.1 (C-22); **LR-ESI-MS:** $C_{18}H_{22}N_3O_3$ $[M+H]^+$ *m/z* found 328.44 cald 328.16; **HR-ESI-MS:** $C_{18}H_{22}N_3O_3$ $[M+H]^+$ *m/z* found 328.0658 cald 328.1656.

***N*-(1-benzyl-3-cyano-1,4,5,6-tetrahydropyridin-2-yl)acetamide (165)**



142 (585 mg, 2.74 mmol) was dissolved in anhydrous MeCN (2.5 mL) and acetic anhydride (409 μL, 4.34 mmol) was added. The reaction was stirred at 45 ºC for 1 h. The reaction was allowed to cool to r.t., and DCM and water added. The organic phase was extracted and washed with brine, dried using Na_2SO_4 and the solvent removed *in vacuo*. **165** (565 mg, 2.21 mmol, 81%) was obtained as an off-white solid.

Mpt: 114.6-117.8 °C; **ν_{max} (cm^{-1}):** 3230, 2934, 2852, 2183, 1673, 1602, 1494, 1424, 1357, 1247, 1191, 1025; **^{1}H NMR (400 MHz, DMSO-d_6) δ** 9.76 (s, 1H, H-12), 7.34 (m, 2H, H-10), 7.31-7.21 (m, 3H, H-9, H-11), 4.26 (s, 2H, H-7), 3.08 (t, J = 5.4 Hz, 2H, H-5), 2.21 (t, J = 6.3 Hz, 2H, H-1), 1.96 (s, 3H, H-17), 1.62-1.54 (m, 2H, H-6); **^{13}C NMR (101 MHz, DMSO-d_6) δ** 168.9 (C-13), 148.6 (C-3), 137.4 (C-8), 128.4 (C-10), 127.7 (C-9), 127.3 (C-11), 122.5 (C-14), 71.7 (C-2), 52.6 (C-7), 46.5 (C-5), 23.6 (C-1), 22.6 (C-17), 20.5 (C-6); **LR-ESI-MS:** $C_{15}H_{18}N_3O$ $[M+H]^+$ *m/z* found 256.33 cald 256.14; **HR-ESI-MS:** $C_{15}H_{17}N_3ONa$ $[M+Na]^+$ *m/z* found 278.0412 cald 278.1515.

9. References

1. G. M. Keserü and G. M. Makara, *Nat. Rev. Drug Discov.*, 2009, **8**, 203–212.
2. J. Robson-Tull, *Biosci. Horizons Int. J. Student Res.*, 2018, **11**, hzy015.
3. W. P. Jencks, *Proc. Natl. Acad. Sci.*, 1981, **78**, 4046–4050.
4. C. Jacquemard and E. Kellenberger, *Expert Opin. Drug Discov.*, 2019, **14**, 413–416.
5. G. Bollag, J. Tsai, J. Zhang, et al., *Nat. Rev. Drug Discov.*, 2012, **11**, 873–886.
6. T. Oltersdorf, S. W. Elmore, A. R. Shoemaker, et al., *Nature*, 2005, **435**, 677–681.
7. A. J. Souers, J. D. Leverson, E. R. Boghaert, et al., *Nat. Med.*, 2013, **19**, 202–208.
8. C. Tse, A. R. Shoemaker, J. Adickes, et al., *Cancer Res.*, 2008, **68**, 3421–3428.
9. C. W. Murray, D. R. Newell and P. Angibaud, *MedChemCommun*, 2019, **10**, 1509–1511.
10. D. A. Erlanson, S. W. Fesik, R. E. Hubbard, et al., *Nat. Rev. Drug Discov.*, 2016, **15**, 605–619.
11. C. A. Lipinski, F. Lombardo, B. W. Dominy, et al., *Adv. Drug Deliv. Rev.*, 2001, **46**, 3–26.
12. M. Congreve, R. Carr, C. Murray, et al., *Drug Discov. Today*, 2003, **8**, 876–877.
13. P. Kirsch, A. M. Hartman, A. K. H. Hirsch, et al., *Molecules*, 2019, **24**, 4309.
14. G. M. Keserü, D. A. Erlanson, G. G. Ferenczy, et al., *J. Med. Chem.*, 2016, **69**, 8189–8206.
15. J. B. Baell and G. A. Holloway, *J. Med. Chem.*, 2010, **53**, 2719–2740.
16. O. B. Cox, T. Krojer, P. Collins, et al., *Chem. Sci.*, 2016, **7**, 2322–2330.
17. F. Lovering, J. Bikker and C. Humblet, *J. Med. Chem.*, 2009, **52**, 6752–6756.
18. A. D. Morley, A. Pugliese, K. Birchall, et al., *Drug Discov. Today*, 2013, **18**, 1221–1227.
19. M. M. Hann, A. R. Leach and G. Harper, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 856–864.

20. T. E. Nielsen and S. L. Schreiber, *Angew. Chem. Int. Ed. Engl.*, 2007, **47**, 48–56.

21. C. John Harris, R. D. Hill, D. W. Sheppard, et al., *Comb. Chem. High Throughput Screen.*, 2011, **14**, 521–531.

22. R. Santos, O. Ursu, A. Gaulton, et al., *Nat. Rev. Drug Discov.*, 2016, **16**, 19–34.

23. ChemBridge Libraries, https://www.chembridge.com/screening_libraries/targeted_libraries/.

24. Enamine Covalent Fragment Library, https://enamine.net/fragments/plated-libraries/covalent-fragment-library.

25. BIONET Fragment Libraries, https://www.keyorganics.net/services/bionet-products/fragment-libraries/.

26. A. Keeley, L. Petri, P. Ábrányi-Balogh, et al., *Drug Discov. Today*, 2020, **25**, 983–996.

27. D. E. Scott, A. G. Coyne, S. A. Hudson, et al., *Biochemistry*, 2012, **51**, 4990–5003.

28. S. B. Shuker, P. J. Hajduk, R. P. Meadows, et al., *Science*, 1996, **274**, 1531–1534.

29. E. Lionta, G. Spyrou, D. Vassilatis, et al., *Curr. Top. Med. Chem.*, 2014, **14**, 1923–1938.

30. M. Saur, M. J. Hartshorn, J. Dong, et al., *Drug Discov. Today*, 2020, **25**, 485–490.

31. D. E. Scott, A. G. Coyne, S. A. Hudson, et al., *Biochemistry*, 2012, **51**, 4990–5003.

32. H. Chen, X. Zhou, A. Wang, et al., *Drug Discov. Today*, 2015, **20**, 105–113.

33. D. S. H. Chan, A. J. Whitehouse, A. G. Coyne, et al., *Essays Biochem.*, 2017, **61**, 465–473.

34. J. R. Huth, C. Park, A. M. Petros, et al., *Chem. Biol. Drug Des.*, 2007, **70**, 1–12.

35. P. A. Brough, X. Barril, J. Borgognoni, et al., *J. Med. Chem.*, 2009, **52**, 4794–4809.

36. N. Howard, C. Abell, W. Blakemore, et al., *J. Med. Chem.*, 2006, **49**, 1346–1355.

37. C. E. Quevedo, A. Cruz-Migoni, N. Bery, et al., *Nat. Commun.*, 2018, **9**, 1–12.

38. L. J. Martin, M. Koegl, G. Bader, et al., *J. Med. Chem*, 2016, **59**, 4462–4475.

39. A. J. Whitehouse, S. E. Thomas, K. P. Brown, et al., *J. Med. Chem.*, 2019, **62**, 7210–

7232.

40. E. Resnick, A. Bradley, J. Gan, et al., *J. Am. Chem. Soc.*, 2019, **141**, 8951–8968.

41. J. Lyu, S. Wang, T. E. Balius, et al., *Nature*, 2019, **566**, 224–229.

42. C. J. Radoux, T. S. G. Olsson, W. R. Pitt, et al., *J. Med. Chem*, 2016, **59**, 4314–4325.

43. O. Ichihara, J. Barker, R. J. Law, et al., *Mol. Inform.*, 2011, **30**, 298–306.

44. B. Lamoree and R. E. Hubbard, *Essays Biochem.*, 2017, **61**, 453–464.

45. S. M. Bertrand, N. Ancellin, B. Beaufils, et al., *J. Med. Chem.*, 2015, **58**, 7140–7163.

46. Y. Kamada, N. Sakai, S. Sogabe, et al., *J. Med. Chem.*, 2017, **60**, 4358–4368.

47. S. M. Bronner, J. Murray, F. Anthony Romero, et al., *J. Med. Chem*, 2017, **60**, 10151–10171.

48. I. D. Kuntz, K. Chen, K. A. Sharp, et al., *Proc. Natl. Acad. Sci.*, 1999, **96**, 9997–10002.

49. A. L. Hopkins, C. R. Groom and A. Alex, *Drug Discov. Today*, 2004, **9**, 430–431.

50. P. N. Mortenson and C. W. Murray, *J. Comput. Aided. Mol. Des.*, 2011, **25**, 663–667.

51. M. L. Verdonk and D. C. Rees, *ChemMedChem*, 2008, **3**, 1179–1180.

52. C. W. Murray, M. L. Verdonk and D. C. Rees, *Trends Pharmacol. Sci.*, 2012, **33**, 224–232.

53. C. G. Wermuth, D. Aldous, P. Raboisson, et al., *The practice of medicinal chemistry*, Academic Press, London, 2015.

54. L. H. Jones, *Expert Opin. Drug Discov.*, 2016, 11, 623–625.

55. J. P. Hughes, S. S. Rees, S. B. Kalindjian, et al., *Br. J. Pharmacol.*, 2011, **162**, 1239–1249.

56. J. Blagg and P. Workman, *Curr. Opin. Pharmacol.*, 2014, **17**, 87–100.

57. M. E. Bunnage, E. L. P. Chekler and L. H. Jones, *Nat. Chem. Biol.*, 2013, **9**, 195–199.

58. M. Moustakim and P. E. Brennan, in *Chemical and Biological Synthesis: Enabling Approaches for Understanding Biology*, eds. N. J. Westwood and A. Nelson, RSC, 2018, pp. 114–137.

59. P. Workman and I. Collins, *Chem. Biol.*, 2010, **17**, 561–577.

60. C. H. Arrowsmith, J. E. Audia, C. Austin, et al., *Nat. Chem. Biol.*, 2015, **11**, 536–541.

61. G. E. Winter, D. L. Buckley, J. Paulk, et al., *Science*, 2015, **348**, 1376–1381.

62. J. Blagg and P. Workman, *Cancer Cell*, 2017, **32**, 9–25.

63. Eurofins Pharmacology Panels, https://www.eurofinsdiscoveryservices.com/services/safety-and-efficacy/safety-pharmacology/safety-panels/.

64. J. Bowes, A. J. Brown, J. Hamon, et al., *Nat. Rev. Drug Discov.*, 2012, **11**, 909–922.

65. A. S. Reddy and S. Zhang, *Expert Rev. Clin. Pharmacol.*, 2013, **6**, 41–47.

66. E. Gregori-Puigjané, V. Setola, J. Hert, et al., *Proc. Natl. Acad. Sci.*, 2012, **109**, 11178–11183.

67. Chemical Probes Portal, https://www.chemicalprobes.org/.

68. S. De Cesco, J. Kurian, C. Dufresne, et al., *Eur. J. Med. Chem.*, 2017, **138**, 96–114.

69. A. K. Ghosh, I. Samanta, A. Mondal, et al., *ChemMedChem*, 2019, **14**, 889–906.

70. R. Lonsdale and R. A. Ward, *Chem. Soc. Rev.*, 2018, **47**, 3816–3830.

71. R. A. Bauer, *Drug Discov. Today*, 2015, **20**, 1061–1073.

72. J. Singh, R. C. Petter, T. A. Baillie, et al., *Nat. Rev. Drug Discov.*, 2011, **10**, 307–317.

73. M. H. Potashman and M. E. Duggan, *J. Med. Chem.*, 2009, **52**, 1231–1246.

74. M. Gehringer and S. A. Laufer, *J. Med. Chem.*, 2019, **62**, 5673–5724.

75. J. S. Martin, C. J. MacKenzie, D. Fletcher, et al., *Bioorganic Med. Chem.*, 2019, **27**, 2066–2074.

76. B. Rupp, *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*, Garland Science, New York, 2010.

77. PDB Statistics, https://www.rcsb.org/stats/summary.

78. P. Savitsky, J. Bray, C. D. O. Cooper, et al., *J. Struct. Biol.*, 2010, **172**, 3–13.

79. A. McPherson and J. A. Gavira, *Acta Crystallogr. Sect. F Struct. Biol. Commun.*,

2014, **70**, 2–20.

80. M. Spiliopoulou, A. Valmas, D. P. Triandafillidis, et al., *Crystals*, 2020, **10**, 54.
81. G. Taylor, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2003, **59**, 1881–1890.
82. I. Müller, *Acta Crystallogr. Sect. D Struct. Biol.*, 2017, **73**, 79–92.
83. M. Congreve, C. W. Murray and T. L. Blundell, *Drug Discov. Today*, 2005, **10**, 895–907.
84. R. L. M. Van Montfort and P. Workman, *Essays Biochem.*, 2017, **61**, 431–437.
85. D. Patel, J. D. Bauman and E. Arnold, *Prog. Biophys. Mol. Biol.*, 2014, **116**, 92–100.
86. Practical Fragments poll results in 2019, http://practicalfragments.blogspot.com/2019/12/poll-results-affiliation-and-fragment.html.
87. C. W. Murray and T. L. Blundell, *Curr. Opin. Struct. Biol.*, 2010, **20**, 497–507.
88. P. G. Wyatt, A. J. Woodhead, V. Berdini, et al., *J. Med. Chem.*, 2008, **51**, 4986–4999.
89. S. Howard, V. Berdini, J. A. Boulstridge, et al., *J. Med. Chem.*, 2009, **52**, 379–388.
90. XChem - DLS, https://www.diamond.ac.uk/Instruments/Mx/Fragment-Screening.html.
91. P. M. Collins, J. T. Ng, R. Talon, et al., *Acta Crystallogr. Sect. D Struct. Biol.*, 2017, **73**, 246–255.
92. N. D. Wright, P. Collins, L. Koekemoer, et al., *Acta Crystallogr. Sect. D Struct. Biol.*, 2021, **77**, 62–74.
93. T. Krojer, R. Talon, N. Pearce, et al., *Acta Crystallogr. Sect. D Struct. Biol.*, 2017, **73**, 267–278.
94. N. M. Pearce, T. Krojer, A. R. Bradley, et al., *Nat. Commun.*, 2017, **8**, 1–8.
95. B. N. Atkinson, D. Steadman, Y. Zhao, et al., *MedChemComm*, 2019, **10**, 1361–1369.
96. J. L. Gray, F. Delft and P. E. Brennan, *Angew. Chemie Int. Ed.*, 2020, **59**, 6342–6366.
97. K. Wennerberg, K. L. Rossman and C. J. Der, *J. Cell Sci.*, 2005, **118**, 843–846.
98. A. M. Rojas, G. Fuentes, A. Rausell, et al., *J. Cell Biol.*, 2012, **196**, 189–201.

99. D. T. Leicht, V. Balan, A. Kaplun, et al., *Biochim. Biophys. Acta*, 2007, **1773**, 1196–1212.

100. A. L. Bishop and A. Hall, *Biochem. J.*, 2000, **348**, 241–255.

101. C. L. Jackson, *Front. cell Dev. Biol.*, 2018, **6**, 29.

102. I. R. Vetter, in *Ras Superfamily Small G Proteins: Biology and Mechanisms 1*, Springer, Vienna, 2014, pp. 25–50.

103. I. R. Vetter and A. Wittinghofer, *Science*, 2001, **294**, 1299–1304.

104. J. L. Bos, H. Rehmann and A. Wittinghofer, *Cell*, 2007, **129**, 865–877.

105. A. G. Stephen, D. Esposito, R. K. Bagni, et al., *Cancer Cell*, 2014, **25**, 272–281.

106. J. Cherfils and M. Zeghouf, *Physiol. Rev.*, 2013, **93**, 269–309.

107. A. Wittinghofer, *Ras superfamily small G proteins: Biology and mechanisms Volume 1*, Springer, Vienna, 2014.

108. E. Sztul, P. W. Chen, J. E. Casanova, et al., *Mol. Biol. Cell*, 2019, **30**, 1249–1271.

109. S. Toma-Fukai and T. Shimizu, *Molecules*, 2019, **24**, 3308–3338.

110. G. A. Hobbs, C. J. Der and K. L. Rossman, *J. Cell Sci.*, 2016, **129**, 1287–1292.

111. I. A. Prior, P. D. Lewis and C. Mattos, *Cancer Res.*, 2012, **72**, 2457–2467.

112. G. A. Hobbs, C. J. Der and K. L. Rossman, *J. Cell Sci.*, 2016, **129**, 1287–1292.

113. A. Fernández-Medarde and E. Santos, *Genes Cancer*, 2011, **2**, 344–358.

114. B. Boettner and L. Van Aelst, *Gene*, 2002, **286**, 155–174.

115. K. A. Reedquist and P. P. Tak, *Open Rheumatol. J.*, 2012, **6**, 259–272.

116. E. Seixas, M. Barros, M. C. Seabra, et al., *Traffic*, 2013, **14**, 871–885.

117. D. Vigil, J. Cherfils, K. Rossman, et al., *Nat Rev Cancer*, 2010, **10**, 842–857.

118. A. K. Mishra and D. G. Lambright, *Biopolymers*, 2016, **105**, 431–448.

119. M. A. Harding and D. Theodorescu, *Eur. J. Cancer*, 2010, **46**, 1252–1259.

120. K. L. Rossman, C. J. Der and J. Sondek, *Nat Rev Mol Cell Biol*, 2005, **6**, 167–180.

121. J. G. Donaldson and C. L. Jackson, *Nat. Rev. Mol. Cell Biol.*, 2011, **12**, 362–375.

122. A. Blangy, N. Bouquier, C. Gauthier-Rouvière, et al., *Biol. Cell*, 2006, **98**, 511–522.

123. Y. Gao, J. Xing, M. Streuli, et al., *J. Biol. Chem.*, 2001, **276**, 47530–47541.

124. A. Czuchra, X. Wu, H. Meyer, et al., *Mol. Biol. Cell*, 2005, **16**, 4473–4484.

125. R. Wolin, D. Wang, J. Kelly, et al., *Bioorganic Med. Chem. Lett.*, 1996, **6**, 195–200.

126. J. Spiegel, P. M. Cromm, G. Zimmermann, et al., *Nat. Chem. Biol.*, 2014, **10**, 613–622.

127. T. W. Traut, *Mol. Cell. Biochem.*, 1994, **140**, 1–22.

128. P. M. Cromm, J. Spiegel, T. N. Grossmann, et al., *Angew. Chemie Int. Ed.*, 2015, **54**, 13516–13537.

129. A. T. Baines, D. Xu and C. J. Der, *Future Med. Chem.*, 2011, **3**, 1787–1808.

130. A. D. Cox and C. J. Der, *Small GTPases*, 2010, **1**, 2–27.

131. M. Shen, P. Pan, Y. Li, et al., *Drug Discov. Today*, 2015, **20**, 267–276.

132. N. Berndt, A. D. Hamilton and S. M. Sebti, *Nat. Rev. Cancer*, 2011, **11**, 775–791.

133. R. R. Mattingly, *ISRN Oncol.*, 2013, **2013**, 536529.

134. Y. J. Zhu, B. Zheng, H. Y. Wang, et al., *Acta Pharmacol. Sin.*, 2017, **38**, 614–622.

135. L. H. Araujo, B. M. Souza, L. R. Leite, et al., *BMC Cancer*, 2021, **21**, 193.

136. J. Cherfils and M. Zeghouf, *Physiol. Rev.*, 2013, **93**, 269–309.

137. J. M. Ostrem, U. Peters, M. L. Sos, et al., *Nature*, 2013, **503**, 548–551.

138. M. R. Janes, J. Zhang, L.-S. Li, et al., *Cell*, 2018, **172**, 578–589.

139. ARS-3248 Clinical Trials, https://clinicaltrials.gov/ct2/show/NCT04006301.

140. M. Zeng, J. Lu, L. Li, et al., *Cell Chem. Biol.*, 2017, **24**, 1005–1016.

141. J. Canon, K. Rex, A. Y. Saiki, et al., *Nature*, 2019, **575**, 217–223.

142. Y. Shin, J. W. Jeong, R. P. Wurz, et al., *ACS Med. Chem. Lett.*, 2019, **10**, 1302–1308.

143. B. A. Lanman, J. R. Allen, J. G. Allen, et al., *J. Med. Chem.*, 2020, **63**, 52–65.

144. J. Hallin, L. D. Engstrom, L. Hargis, et al., *Cancer Discov*, 2019, **10**, 54–71.

145. J. B. Fell, J. P. Fischer, B. R. Baer, et al., *J. Med. Chem.*, 2020, **63**, 6679–6693.

146. MRXT849 Clinical trials, https://www.newswise.com/articles/targeted-inhibitor-of-mutated-kras-gene-shows-promise-in-early-trial-for-lung-bowel-and-other-solid-tumors?sc=sphr&xy=10021790.

147. M. J. Bond, L. Chu, D. A. Nalawansha, et al., *ACS Cent. Sci.*, 2020, **6**, 1367–1375.

148. E. de Vita, M. Maneiro and E. W. Tate, *ACS Cent. Sci.*, 2020, **6**, 1281–1284.

149. N. Tanaka, J. J. Lin, C. Li, et al., *Cancer Discov.*, 2021, DOI:10.1158/2159-8290.CD-21-0365.

150. S. Min Lim, K. D. Westover, S. B. Ficarro, et al., *Angew Chem Int Ed*, 2014, **3**, 199–204.

151. J. C. Hunter, D. Gurbani, S. B. Ficarro, et al., *Proc. Natl. Acad. Sci.*, 2014, **111**, 8895–8900.

152. R. C. Hillig, B. Sautier, J. Schroeder, et al., *Proc. Natl. Acad. Sci.*, 2019, **116**, 2551–2560.

153. D. Kessler, M. Gmachl, A. Mantoulidis, et al., *Proc. Natl. Acad. Sci.*, 2019, **116**, 15823–15829.

154. T. H. Tran, P. Alexander, S. Dharmaiah, et al., *Proc. Natl. Acad. Sci.*, 2020, **117**, 3363–3364.

155. D. Kessler, A. Gollner, M. Gmachl, et al., *Proc. Natl. Acad. Sci.*, 2020, **117**, 3365–3367.

156. D. P. Bondeson and C. M. Crews, *Annu. Rev. Pharmacol. Toxicol.*, 2017, **57**, 107–123.

157. BI 1701963 - Clinical trials, https://clinicaltrials.gov/ct2/show/NCT04111458.

158. M. H. Hofmann, M. Gmachl, J. Ramharter, et al., *Cancer Discov.*, 2021, **11**, 142–157.

159. S. J. Ross, A. S. Revenko, L. L. Hanson, et al., *Sci. Transl. Med.*, 2017, **9**, eaal5253.

160. Antisense KRas Clinical trials, https://www.evaluate.com/vantage/articles/news/trial-results/astras-first-attempt-fails-theres-no-giving-kras.

161. P. A. Randazzo, X. Jian, P. W. Chen, et al., *Cell. Logist.*, 2013, **3**, e27609.

162. M. P. Müller, S. Jeganathan, A. Heidrich, et al., *Sci. Rep.*, 2017, **7**, 3687.

163. R. S. Goody, *FEBS J.*, 2014, **281**, 593–600.

164. J. John, R. Sohmen, J. Feuerstein, et al., *Biochemistry*, 1990, **29**, 6058–6065.

165. C. Klebe, H. Prinz, A. Wittinghofer, et al., *Biochemistry*, 1995, **34**, 12543–12552.

166. K. A. Johnson, Z. B. Simpson and T. Blom, *Anal. Biochem.*, 2009, **387**, 20–29.

167. M. Beck, A. Schmidt, J. Malmstroem, et al., *Mol. Syst. Biol.*, 2011, **7**, 549.

168. R. S. Goody and W. Hofmann-Goody, *Eur. Biophys. J.*, 2002, **31**, 268–274.

169. J. K. Northup, X. Jian and P. A. Randazzo, *Cell. Logist.*, 2012, **2**, 140–146.

170. A. B. Goryachev and A. V Pokhilko, *PLOS Comput. Biol.*, 2006, **2**, 1511–1521.

171. L. Shi, *Commun. Integr. Biol.*, 2013, **6**, e26839.

172. C. Shoubridge, P. S. Tarpey, F. Abidi, et al., *Nat Genet*, 2010, **42**, 486–488.

173. M. K. Chhatriwala, L. Betts, D. K. Worthylake, et al., *J. Mol. Biol.*, 2007, **368**, 1307–1320.

174. A. De Marco, L. Vigh, S. Diamant, et al., *Cell Stress Chaperones*, 2005, **10**, 329–339.

175. T7 Express lysY competent cells, https://international.neb.com/products/c3010-t7-express-lysy-competent-e-coli-high-efficiency#Product Information.

176. Molecular Dimensions, https://www.moleculardimensions.com/products/all-crystallization-screens.

177. BLAST: Basic Local Alignment Search Tool, https://blast.ncbi.nlm.nih.gov/Blast.cgi.

178. D. K. Worthylake, K. L. Rossman and J. Sondek, *Nature*, 2000, **408**, 682–688.

179. B. Debreceni, Y. Gao, F. Guo, et al., *J. Biol. Chem.*, 2004, **279**, 3777–3786.

180. E. Sztul, P. W. Chen, J. E. Casanova, et al., *Mol. Biol. Cell*, 2019, **30**, 1249–1271.

181. J. E. Casanova, *Traffic*, 2007, **8**, 1476–1485.

182. J. Cherfils, J. Ménétrey, M. Mathieu, et al., *Nature*, 1998, **392**, 101–105.

183. G. Mlynek, J. Kostan, S. Leeb, et al., *Cryst. Growth Des.*, 2020, **20**, 984–994.

184. A. Chari, D. Haselbach, J. M. Kirves, et al., *Nat. Methods*, 2015, **12**, 859–865.

185. GraphPad, https://www.graphpad.com/.

186. N. Pankratz, G. W. Beecham, A. L. Destefano, et al., *Ann. Neurol.*, 2012, **71**, 370–384.

187. H. Wang, L. Feng, Z. Zhang, et al., *Sci. Rep.*, 2016, **6**, 1–14.

188. Y. J. Huang, T. B. Acton and G. T. Montelione, *Methods Mol. Biol.*, 2014, **1091**, 3–16.

189. W. T. M. Mooij, E. Mitsiki and A. Perrakis, *Nucleic Acids Res.*, 2009, **37**, W402-405.

190. A. R. Bradley, A. Echalier, M. Fairhead, et al., *Essays Biochem.*, 2017, **61**, 495–503.

191. J. L. Gray, C. J. Antunez, T. Krojer, et al., Human Kalirin/RAC1 GEF/GTPase Complex, https://zenodo.org/record/3266934.

192. Q. L. Ma, F. Yang, S. A. Frautschy, et al., *Cell. Logist.*, 2012, **2**, 117–125.

193. D. R. Cook, K. L. Rossman and C. J. Der, *Oncogene*, 2014, **33**, 4021–4035.

194. J. DeGeer and N. Lamarche-Vane, *Exp. Cell Res.*, 2013, **319**, 2384–2394.

195. A. P. Thompson, C. Bitsina, J. L. Gray, et al., *J. Biol. Chem.*, 2021, **296**, 100521.

196. M. A. Sarangdhar, D. Chaubey, A. Bhatt, et al., *Front. Mol. Neurosci.*, 2017, **10**, 95.

197. P. Penzes, R. C. Johnson, V. Kambampati, et al., *J. Neurosci.*, 2001, **21**, 8426–8434.

198. S. Schmidt and A. Debant, *Small GTPases*, 2014, **5**, e983880.

199. P. Penzes, R. C. Johnson, M. R. Alam, et al., *J. Biol. Chem.*, 2000, **275**, 6395–6403.

200. C. Remmers, R. A. Sweet and P. Penzes, *Brain Res. Bull.*, 2014, **103**, 29–38.

201. P. Penzes and C. Remmers, *Mol. Neurobiol.*, 2012, **45**, 109–118.

202. J. D. Paskus, C. Tian, E. Fingleton, et al., *Cell Rep.*, 2019, **29**, 2944–2952.

203. F. Lemtiri-Chlieh, L. Zhao, D. D. Kiraly, et al., *BMC Neurosci.*, 2011, **12**, 126–135.

204. Y. Bai, X. Xiang, C. Liang, et al., *Biomed Res. Int.*, 2015, **2015**, 632540.

205. A. Hayashi-Takagi, M. Takaki, N. Graziane, et al., *Nat. Neurosci.*, 2010, **13**, 327–332.

206. T. G. G. Battye, L. Kontogiannis, O. Johnson, et al., *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2011, **67**, 271–281.

207. T. Krojer, J. S. Fraser and F. von Delft, *Curr. Opin. Struct. Biol.*, 2020, **65**, 209–216.

208. B. Zhang, Y. Zhang, Z.-X. Wang, et al., *J. Biol. Chem.*, 2000, **275**, 25299–25307.

209. D. J. Foley, P. G. E. Craven, P. M. Collins, et al., *Chem. - A Eur. J.*, 2017, **23**, 15227–15232.

210. Maybridge Ro3 library, https://www.maybridge.com/images/pdfs/ro3frag.pdf.

211. P. Peveri, P. G. Heyworth and J. T. Curnutte, *Biochemistry*, 1992, **89**, 2494–2498.

212. K. M. Welsh, K. A. Trach, C. Folger, et al., *J. Bacteriol.*, 1994, **203**, 7161–7168.

213. A. Eberth and M. R. Ahmadian, in *Current Protocols in Cell Biology*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2009, vol. 43, pp. 14.9.1-14.9.25.

214. D. J. Wood, J. D. Lopez-Fernandez, L. E. Knight, et al., *J. Med. Chem.*, 2019, **62**, 3741–3752.

215. M. O'Reilly, A. Cleasby, T. G. Davies, et al., *Drug Discov. Today*, 2019, **24**, 1081–1086.

216. S. Y. Huang and X. Zou, *Int. J. Mol. Sci.*, 2010, **11**, 3016–3034.

217. I. A. Guedes, F. S. S. Pereira and L. E. Dardenne, *Front. Pharmacol.*, 2018, **9**, 1089.

218. M. Wirth, V. Zoete, O. Michielin, et al., *Nucleic Acids Res.*, 2013, **41**, D113–D1143.

219. A. K. Ghosh and M. Brindisi, *J. Med. Chem*, 2019, **63**, 2751–2788.

220. N. Dubois, D. Glynn, T. McInally, et al., *Tetrahedron*, 2013, **69**, 9890–9897.

221. R. Eligeti, R. Saini and N. Dharavath, *J. Sulfur Chem.*, 2012, **33**, 427–438.

222. I. J. Kang, L. W. Wang, C. C. Lee, et al., *Bioorganic Med. Chem. Lett.*, 2009, **19**, 1950–1955.

223. WO2005023818A2, 2005.

224. WO2016116061A1, 2016.

225. A. Cornia, F. Felluga, V. Frenna, et al., *Tetrahedron*, 2012, **68**, 5863–5881.

226. J. L. Collins, S. G. Blanchard, G. E. Boswell, et al., *J. Med. Chem.*, 1998, **41**, 5037–5054.

227. D. S. Surry and S. L. Buchwald, *Chem. Sci.*, 2011, **2**, 27–50.

228. M. De Rosa, D. Arnold and D. Hartline, *J. Org. Chem.*, 2013, **78**, 8614–8623.

229. A. Choudhury, H. Chen, C. N. Nilsen, et al., *Tetrahedron Lett.*, 2008, **49**, 102–105.

230. M. P. Dwyer, Y. Yu, J. Chao, et al., *J. Med. Chem.*, 2006, **49**, 7603–7606.

231. B. Stanovnik, A. Štimac, M. Tišler, et al., *J. Heterocycl. Chem.*, 1982, **19**, 577–583.

232. C. C. Lin, T. H. Hsieh, P. Y. Liao, et al., *Org. Lett.*, 2014, **16**, 892–895.

233. K. S. Kim and L. Qian, *Tetrahedron Lett.*, 1993, **34**, 7677–7680.

234. C. Dardonville, P. Goya, I. Rozas, et al., *Bioorganic Med. Chem.*, 2000, **8**, 1567–1577.

235. J. O'Brien-Brown, A. Jackson, T. A. Reekie, et al., *Eur. J. Med. Chem.*, 2017, **130**, 433–439.

236. M. Buchholz, A. Hamann, S. Aust, et al., *J. Med. Chem.*, 2009, **52**, 7069–7080.

237. W. Lin, L. Yang, S. C. Chai, et al., *Eur. J. Med. Chem.*, 2016, **108**, 505–528.

238. H. A. Albrecht, J. F. Blount, F. M. Konzelmann, et al., *J. Org. Chem.*, 1979, **44**, 4191–4194.

239. WO2016141035A1, 2016.

240. T. Maurer, L. S. Garrenton, A. Oh, et al., *Proc. Natl. Acad. Sci.*, 2012, **109**, 5299–5304.

241. M. C. Burns, J. E. Howes, Q. Sun, et al., *Anal. Biochem.*, 2018, **548**, 44–52.

242. C. R. Evelyn, X. Duan, J. Biesiada, et al., *Chem. Biol.*, 2014, **21**, 1618–1628.

243. J. Korlach, D. W. Baird, A. A. Heikal, et al., *Proc. Natl. Acad. Sci.*, 2004, **101**, 2800–2805.

244. GTP Exchange Factor (GEF) Assays, https://www.cytoskeleton.com/gef-compound-screen.

245. M. D. Olp, K. S. Kalous and B. C. Smith, *BMC Bioinformatics*, 2020, **21**, 186.

246. S. J. Lord, K. B. Velle, R. Dyche Mullins, et al., *J. Cell Biol.*, 2020, **219**, e202001064.

247. K. Zhu, K. W. Borrelli, J. R. Greenwood, et al., *J. Chem. Inf. Model.*, 2014, **54**, 1932–1940.

248. S. E. Dalton, L. Dittus, D. A. Thomas, et al., *J. Am. Chem. Soc.*, 2018, **140**, 932–939.

249. L. P. Hammett, *J. Am. Chem. Soc.*, 1937, **59**, 96–103.

250. C. Hansch, A. Leo and R. W. Taft, *Chem. Rev*, 1991, **91**, 165–195.

251. S. Fletcher, *Org. Chem. Front*, 2015, **2**, 739–752.

252. G. M. Morris, H. Ruth, W. Lindstrom, et al., *J. Comput. Chem.*, 2009, **30**, 2785–2791.

253. S. Leung, M. Bodkin, F. von Delft, et al., *ChemRxiv*, 2019, 1–47.

254. G. Marcou and D. Rognan, *J. Chem. Inf. Model.*, 2007, **47**, 195–207.

255. S. Salentin, S. Schreiber, V. J. Haupt, et al., *Nucleic Acids Res.*, 2015, **43**, 443–447.

256. M. M. Abdou, *Arab. J. Chem.*, 2017, **10**, S3324–S3337.

257. W. Engen, T. E. O'Brien, B. Kelly, et al., *Bioorganic Med. Chem.*, 2010, **18**, 5995–6005.

258. L. I. . Ibrahim, S. M. . Eldin and F. A. Attaby, *Pharmazie*, 1992, **47**, 792–794.

259. F. Khoramdelan, A. Davoodnia, M. R. . Bozorgmehr, et al., *Heterocycl. Lett.*, 2017, **7**, 947–952.

260. G. M. Coppola and G. E. Hardtmann, *J. Heterocycl. Chem.*, 1979, **16**, 1605–1610.

261. K. D. Luderman, P. Jain, R. Benjamin Free, et al., *Bioorganic Med. Chem. Lett.*, 2021, **31**, 127696.

262. M. Wijtmans, D. A. Pratt, J. Brinkhorst, et al., *J. Org. Chem.*, 2004, **69**, 9215–9223.

263. Y. Zhang, D. Duan, Y. Zhong, et al., *Org. Lett.*, 2019, **21**, 4960–4965.

264. US 10,844,003 BI, 2020.

265. E. Jagu, S. Pomel, A. Diez-Martinez, et al., *Eur. J. Med. Chem.*, 2018, **150**, 655–666.

266. R. A. Olofson, J. T. Martz, J. P. Senet, et al., *J. Org. Chem.*, 1984, **49**, 2081–2082.

267. T. Kiguchi, N. Kuninobu, Y. Takahashi, et al., *Synthesis*, 1989, **10**, 778–781.

268. J. S. Bajwa, J. Slade and O. Repic, *Tetrahedron Lett.*, 2000, **41**, 6025–6028.

269. E. A. Meade and L. M. Beauchamp, *J. Heterocycl. Chem.*, 1996, **33**, 303–308.

270. P. Lei, Y. Ding, X. Zhang, et al., *Org. Lett*, 2018, **20**, 3439–3442.

271. M. Wojdyr, R. Keegan, G. Winter, et al., *Acta Crystallogr. Sect. A Found. Adv.*, 2014, **70**, C1447.

272. UniProt, https://www.uniprot.org/.

273. Eurofins Genomics, https://eurofinsgenomics.eu/.

274. Hampton Research, https://hamptonresearch.com/categories.php.

275. G. Winter, *J. Appl. Crystallogr.*, 2010, **43**, 186–190.

276. D. G. Waterman, G. Winter, R. J. Gildea, et al., *Acta Crystallogr. Sect. D Struct. Biol.*, 2016, **72**, 558–575.

277. W. Kabsch, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2010, **66**, 125–132.

278. P. Evans, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2006, **62**, 72–82.

279. M. D. Winn, C. C. Ballard, K. D. Cowtan, et al., *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2011, **67**, 235–242.

280. P. D. Adams, P. V Afonine, G. Bunkóczi, et al., *Acta Crystallogr. Sect. D*, 2010, **66**, 213–221.

281. P. Emsley, B. Lohkamp, W. G. Scott, et al., *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2010, **66**, 486–501.

282. L. Potterton, J. Agirre, C. Ballard, et al., *Acta Crystallogr. Sect. D Struct. Biol.*, 2018, **74**, 68–84.

283. A. A. Lebedev, P. Young, M. N. Isupov, et al., *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2012, **68**, 431–440.

284. G. N. Murshudov, P. Skubák, A. A. Lebedev, et al., *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2011, **67**, 355–367.

285. F. Long, R. A. Nicholls, P. Emsley, et al., *Acta Crystallogr. Sect. D, Struct. Biol.*, 2017, **73**, 112–122.

286. OxXChem, https://xchem.github.io/oxxchem/.

287. A. Aimon, G. Karageorgis, J. Masters, et al., *Org. Biomol. Chem.*, 2018, **16**, 3160–3167.

288. R. Abagyan, M. Totrov and D. Kuznetsov, *J. Comput. Chem.*, 1994, **15**, 488–506.

289. T. Mejuch, G. Garivet, W. Hofer, et al., *Angew. Chemie Int. Ed.*, 2017, **56**, 6181–6186.