

Chemical Probes for Rho GEF/GTPases

Mia Callens

Balliol College

Supervisor: Professor Paul Brennan

A thesis submitted for the degree of

Doctor of Philosophy

Synthesis for Biology and Medicine Centre for Doctoral Training

Department of Chemistry, University of Oxford

Declaration

The work presented in this thesis is my own except where explicitly stated otherwise.

No parts of this thesis have been submitted for any other degree.

Mia Callens

Balliol College

Abstract

The Rho subfamily of GTPases are nucleotide dependent molecular switches that govern diverse cytoskeletal cellular processes. Despite their implication across many disease areas, modulating their activity is historically challenging due to tight nucleotide binding and limited opportunities for allosteric control. An alternative approach is targeting the GEF/GTPase complex: the GEFs catalyse nucleotide exchange from GDP to GTP, thus activating the GTPase. It is hypothesised that inhibition of these complexes could regulate GTPase signalling and reveal new druggable targets for GTPase-associated diseases.

This thesis focuses on validating complexes based on splice variants of GEF Kalirin. Studies found variants Kalirin-7 and Kalirin-9 – the former containing a Rac1 activating domain, and the latter containing an additional RhoA activating domain – were dysregulated in post-mortem schizophrenic brains. A promising covalent inhibitor was previously identified for Kalirin-GEF(1)/Rac1, providing a starting point for probing the Rho GEF/GTPases at the nucleotide binding pocket.

Chapter 1 of this thesis describes the importance of early target validation, strategies for identifying small molecule probes, and an in-depth background of the Rho GEF/GTPases. Chapter 2 explores the inhibition of Kalirin-GEF(1)/Rac1 with fragments by providing an SAR analysis, corroborating the binding mode of fragments, and investigating selectivity in focused screens and across the proteome. This resulted in the deeper understanding and improved validation of Kalirin-GEF(1)/Rac1 inhibition, and demonstrated that selectivity between GEF/GTPase complexes is possible. Chapter 3 centres on the second Kalirin target of interest, Kalirin-GEF(2)/RhoA. The

design of active constructs, the generation of novel structural data, and a proof-of-concept that the target is amenable to modulation by small molecules is described. Most notably, the novel crystal structure of the Kalirin DH2 domain was solved, and differences in the GEF activity of the DH2 and DH2-PH2 domains were observed. This key information alluded to autoinhibition of Kalirin via the PH2 domain, an important consideration for future screening campaigns. Finally, Chapter 4 investigates the conserved activation mechanisms of the Rho GEFs in a structural alignment study, and explores using GTPase-specificity markers to aid the prediction of new GEF/GTPase targets. With this information, it was expected new targets could be considered for inhibition by small molecules as described in Chapters 2 and 3. 70 Rho Dbl-GEF domains were compared, and spatially conserved residues were concluded to be significant in Rho GTPase binding. GTPase-specificity markers were identified and used to predict GTPase partners for uncharacterised GEFs, which was explored experimentally in a nucleotide exchange assay. This resulted in the identification of new GEF/GTPase complexes *in vitro*, which could be considered as alternative targets in GEF-associated diseases.

Acknowledgements

I would first like to thank my supervisor Professor Paul Brennan, for not only his support, but also for his unwavering commitment to his students and their ideas. I would not have the multidisciplinary view on target validation I have now without the freedom to explore biology, biochemistry, crystallography, organic synthesis, and medicinal chemistry. He has not only been an absolute rock of knowledge to learn from, but also a place of constant positive encouragement.

I would like to say a special thanks to the amazing scientists I was surrounded by during these projects: the brilliant Dr. Janine Gray, for paving the way for an amazing DPhil project; Dr. Andrew Thompson, for teaching me all things protein production and crystallography, and being very patient in my initial attempts to mount crystals (sorry about all the ones I crushed!); my fabulous Part II student Anya Robinson, for her contributions to this thesis, hard work, and being a joy to supervise; and Amber Truepenny and Mira Rothweiler, for all things proteomics. Finally, to the entire Brennan Group and the extended NDMRB crew, without their humour, kindness, and incredible expertise, I would not have gotten through this DPhil!

Thank you to the other extraordinary people around me for keeping me sane; my incredible Leeds circle: Kathy, Alicia, Car, Dan, Frank, Rose, Clare, for all the fun and reminding me I have an incredible life outside of my work! To Denmark Street: Morgan, Ben, Harry, for all the silliness and having a comfortable place to go home; and to Leah, who has relentlessly backed me for the last 10 years.

It is near impossible I would be writing these acknowledgements without my family. Whenever I was close to packing it in they reminded me that even starting a doctorate, let alone at Oxford, was something that never seemed possible, so finishing it could be too. This one's for you!

I am also very grateful for funding from the Synthesis for Biology and Medicine CDT and the Oxford-RE Jones Scholarship, support from the Centre for Medicines Discovery, and for the invaluable advice offered by my industrial mentors Dr. Gemma Liwicki and Chris Wray.

1 Table of Contents

1 Introduction 1

1.1 Early target validation 1

1.1.1 The importance of target validation 1

1.1.2 Chemical probes 2

1.1.3 Covalent vs. reversible probes 5

1.2 Drug discovery approaches 8

1.2.1 High-throughput screening 8

1.2.2 Fragment based drug discovery 9

1.3 X-ray crystallography 13

1.3.1 Protein production 13

1.3.2 Protein crystallisation 15

1.3.3 Data acquisition 17

1.3.4 Ligand screening 19

1.4 Computational docking and visualisation 22

1.5 Protein structure prediction 24

1.6 Targeting the GEF/GTPases 26

1.6.1 The Ras superfamily 26

1.6.2 The GTPase activation cycle 27

1.6.3 GTPases in drug discovery 28

1.6.4 Targeting Rho GEF/GTPases 31

1.6.5 Rac1 and RhoA 33

1.6.6 Kalirin 34

1.6.7 Previous Kalirin-GEF(1)/Rac1 progress 36

1.6.8 Previous Kalirin-GEF(2)/RhoA progress 44

1.6.9 Alternative Rho GEF/GTPase complexes as targets 44

1.7 Project Objectives 46

1.7.1 Kalirin-GEF(1)/Rac1 46

1.7.2 Kalirin-GEF(2)/RhoA 47
1.7.3 Predicting new Rho GEF/GTPase targets 47

2 Kalirin-GEF(1)/Rac1 49

2.1 Quantitative analysis of previous hits 49

2.1.1 IC_{50} 49
2.1.2 k_{inact}/K_i 51

2.2 Core SAR analysis 53

2.2.1 Design of core analogues 53
2.2.2 Synthesis of core analogues 56
2.2.3 Activity of core analogues 61
2.2.4 Binding mode of MC-278 64

2.3 Warhead SAR analysis 66

2.3.1 Scalable synthesis of MC-278 66
2.3.2 Design of non-covalent analogues 69
2.3.3 Synthesis of non-covalent analogues 71
2.3.4 Activity of non-covalent analogues 72
2.3.5 Design of covalent analogues 73
2.3.6 Synthesis of covalent analogues 74

2.4 Selectivity screening 78

2.4.1 Limited Rho GEF/GTPase panel 78
2.4.2 Activity-based proteomic profiling 80

2.5 Potency at cellular nucleotide concentrations 83

2.5.1 Fluorescence intensity assay 83
2.5.2 Mass spectrometry assay 85

2.6 Fragment elaboration 90

2.6.1 XChem screen 90
2.6.2 Elaboration of MC-278 91
2.6.3 Elaboration of RS-009 94

2.7 Conclusions 98

3 Kalirin-GEF(2)/RhoA 100

3.1 Production of active Kalirin-GEF(2) 101

3.1.1 Construct design 101

3.1.2 Expression and purification 101

3.1.3 Kalirin-GEF(2) activity analysis 103

3.2 Crystallisation of Kalirin-GEF(2)/RhoA 105

3.2.1 Kalirin-GEF(2)/RhoA complex purification 105

3.2.2 Crystallisation screening 108

3.2.3 Novel structure of the Kalirin DH2 domain 111

3.2.4 Encouraging Kalirin-GEF(2)/RhoA crystallisation 112

3.3 Fragment screening 114

3.3.1 Limited fragment screen 114

3.4 Conclusions 118

4 Predicting new GEF/GTPase targets 120

4.1 Rho Dbl GEF structure analysis 121

4.1.1 Selection criteria and alignment protocols 121

4.1.2 Highly conserved residues 122

4.1.3 Substrate specificity markers 127

4.1.4 Rac1 selectivity 129

4.1.5 RhoA selectivity 132

4.1.6 Cdc42 selectivity 136

4.1.7 Activation of multiple GEFs 138

4.2 Validation of GEF/GTPase predictions 140

4.2.1 GEF activity of selected GEFs 140

4.3 Conclusions 143

5 Conclusions and Future Work 144

6 Appendices 148

7 Experimental 163

7.1 General chemistry experimental 163

7.1.1 Solvents and reagents 163

7.1.2 Purification and chromatography 163

7.1.3 Characterisation 164

7.1.4 Synthetic procedures 165

7.2 Computational docking and visualisation 203

7.3 BODIPY-GTP nucleotide exchange assay 204

7.3.1 General protocol 204

7.3.2 K_{inact}/Ki 205

7.3.3 Limited selectivity panel 205

7.4 Activity-based protein profiling 206

7.4.1 Competitive-ABPP with MC-278 206

7.4.2 MS sample preparation 208

7.4.3 Sample analysis by LC-MS/MS 208

7.4.4 Data evaluation by MaxQuant 209

7.5 GppNHp competition assay 210

7.6 Production of Kalirin-GEF(2) constructs 211

7.6.1 Expression 211

7.6.2 Purification 211

7.7 Crystallisation and structure determination 213

7.7.1 General experimental 213

7.7.2 Co-crystallisation of MC-278 with Kalirin-GEF(1)/Rac1 213

7.7.3 Crystallisation of the Kalirin DH2 domain 214

7.8 Rho Dbl-GEF Alignment 215

7.8.1 General protocol 215

8 References 216

List of Abbreviations

Å	Angstroms
ACE-Cl	1-Chloroethyl chloroformate
AD	Alzheimer's disease
Akap13	A-Kinase anchoring protein 13
ALS	Amyotrophic lateral sclerosis
APP	Amyloid precursor protein
Arf	ADP-ribosylation factor
ARHGEF	Ras homology guanine nucleotide exchange factor
ATP	Adenosine triphosphate
Aβ	Amyloid beta
Bcl-2	B-cell lymphoma 2
Bcl-xL	B-cell lymphoma-extra large
BET	Bromo- and extra-terminal
Bn	benzyl
Boc	*tert*-butoxycarbonyl
BODIPY-GTP	Guanosine 5'-*O*-(3-Thiotriphosphate), BODIPY™ FL Thioester
BTK	Bruton's tyrosine kinase
Bu	Butyl
Cdc42	Cell division control protein 42 homolog
CLL	Chronic lymphocytic leukemia
CMD	Centre for Medicines Discovery
CNS	Central nervous system
Cryo-EM	Cryogenic electron microscopy

Da	Daltons
Dbl	Diffuse B-cell lymphoma
Dbs	Dbl's big sister
DCE	1,2-Dichloroethane
DCM	Dichloromethane
DH	Dbl homology
DIPEA	*N,N*-diisopropylethylamine
DISC1	Disrupted-in-schizophrenia 1
DMAP	4-Dimethylaminopyridine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DOCK	Dedicator of cytokinesis
DTT	Dithiothreitol
FBDD	Fragment based drug design
FDA	Food and drug administration
FI	Fluorescence intensity
GAP	GTPase activating protein
GDI	Guanine nucleotide dissociation inhibitor
GDP	Guanosine diphosphate
GEF	Guanine nucleotide exchange factor
GppNHp	Guanosine-5'-[(β,γ)-imido]triphosphate
GTP	Guanosine triphosphate
HD	Huntington's disease
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HMDS	Bis(trimethylsilyl)amide

HNSCC	Head and neck squamous cell carcinoma
HPLC	High performance liquid chromatography
HTS	High-throughput screening
IC_{50}	Half maximal inhibitory concentration
IMAC	Immobilised metal affinity chromatography
ITSN1	Intersectin 1
Kd	Dissociation constant
KRas	Kirsten Rat sarcoma
L	Litre
LCMS	Liquid chromatography mass spectrometry
M	Molar
MR	Molecular replacement
MS	Mass spectrometry
MW	Molecular weight
NCE	New chemical entity
Net1	Neuroepithelial cell-transforming 1
nM	nanomolar
NMR	Nuclear magnetic resonance
PAINs	Pan-Assay interference compounds
PanDDA	Pan-dataset density analysis
PBS	Phosphate buffered saline
PD	Parkinson’s disease
PDB	Protein data bank
PEG	Polyethylene glycol
PH	Pleckstrin homology

PK	Pharmacokinetic
pM	picomolar
PPI	Protein-protein interaction
Prex1	Phosphatidylinositol-3,4,5-trisphosphate dependent rac
QTOF	Quantitative time of flight
r.t.	room temperature
Rab	Ras-related in brain
Rac1	Ras-related C3 botulinum toxin substrate 1
Ran	Ras-like nuclear
Ras	Rat sarcoma
RASGRF1	Ras-specific Guanine Nucleotide-Releasing Factor 1
Rho	Ras homologous
RhoA	Ras homology family member A
RMSD	Root-mean-square-deviation
RSCC	Real-space correlation coefficient
RSRZ	Real-Space R-value Z-score
SAR	Structure-activity relationship
SBDD	Structure-based drug design
SDS-PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SGC	Structural Genetics Consortium
SH3	Src homology 3
S_NAr	Substitution nucleophilic aromatic
Sos	Son of Sevenless

Strep	Streptavidin
t	time
TB	Terrific broth
TBTA	Tris(benzyltriazolylmethyl)amine
TEA	Triethylamine
TEAB	Tetraethylammonium bromide
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
Tiam1	T-cell lymphoma invasion and metastasis-inducing protein 1
Trio	Triple functional domain protein

1 Introduction

1.1 Early target validation

1.1.1 The importance of target validation

In the mid-2000s, a clinical trial was underway for the treatment of Alzheimer's disease (AD).[1] The drug, semagacestat, was developed to target the accumulation of amyloid beta (Aβ) peptides; the prevailing theory at the time was that excessive accumulation of Aβ peptides in the brain played a central role in the disease, known as the 'amyloid hypothesis'.[2] More specifically, semagacestat was an inhibitor of the γ-secretase enzyme, an enzyme involved in the cleavage of the amyloid precursor protein (APP).[3] The drug was expected to reduce the production of Aβ-42, a more toxic Aβ peptide associated with AD.

Unfortunately, the clinical trial was terminated before completion. The drug showed a lack of efficacy, and even more concerning, patients experienced cognitive worsening and skin-related side effects. There were several key factors that contributed to this failure, much of it weighing on the incomplete validation of the target. Whilst the 'amyloid hypothesis' suggested that inhibiting γ-secretase could slow the progression of AD, the biological complexity of the process was not fully understood. There were also unintended consequences based on poor target selection; semagacestat's indiscriminate inhibition affected other crucial processes regulated by the γ-secretase enzyme, including Notch receptor-related nuclear signalling. There was also a lack of Aβ-40 and Aβ-42 reduction in the cerebrospinal fluid despite there being a reduction in the plasma, suggesting the target was engaged peripherally rather than in the brain. It was not known whether the study truly tested the amyloid hypothesis due to a lack

of understanding of the target, and the effects seen in the clinical trial could not be confidently explained.[4]

This story is one of many in which inadequate target validation has been attributed as a major factor in drug attrition. It was predicted over 10 years ago that validation of targets in the early stages of discovery could reduce attrition in Phase II trials by ~24%, and reduce the cost of developing new chemical entities (NCEs) by ~30%.[5] Briefly, target validation involves confirming that a specific biological molecule or pathway is a valid and relevant therapeutic target for a particular indication.[6] There are many approaches that can be used in combination to form this proof-of-concept; from biological rationale, in which the researcher should identify the molecular mechanisms that contribute to the onset and progression of the disease, and genetic evidence, such as the use of animal models and genetic knockouts. A valid target should also be druggable, meaning it is amenable to manipulation by small molecules or alternative therapeutic reagents.[7] Employing biological approaches alone, however, may lead to off-target effects and poor reproducibility without orthogonal validation.[8] A more common approach is integrating biological methods of validation with small molecule tools, known as chemical probes, to more thoroughly investigate a target and its association with the proposed disease.

1.1.2 Chemical probes

A chemical probe is a specialised compound used to investigate the specific function of biological targets. The probes are typically small molecules designed to interact selectively with a target to generate observable phenotypes through *in vitro* or *in vivo*

modulation.[8] A well-characterised probe is generally made openly available and accessible to allow the community to generate a deeper knowledge of the target.[9]

The presence of high-quality probes in research is noticeably beneficial; bromodomain and extra-terminal domain (BET) probes JQ1[10], I-BET[11] and PFI-1[12] triggered a more in-depth understanding of bromodomain biology and thus the implication of the BET family in distinct disease areas (Figure 1). There have also been successes using chemical probes across other disease areas including immunology.[13]

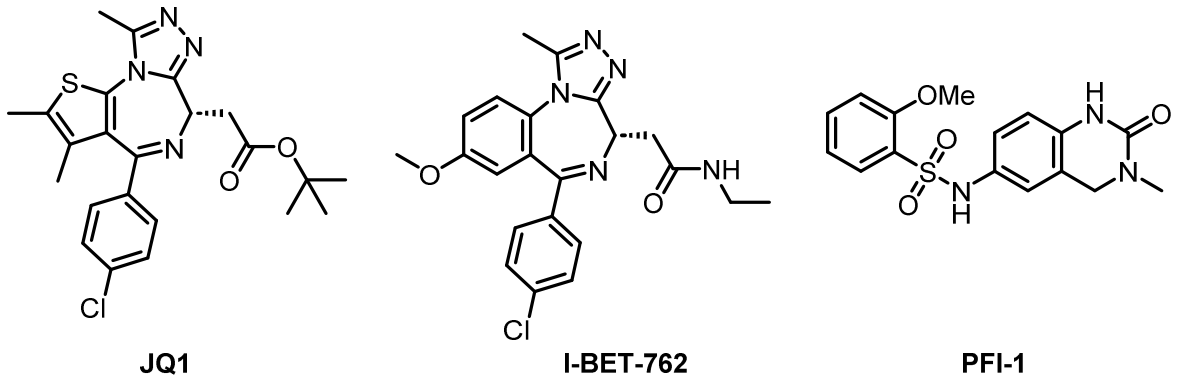
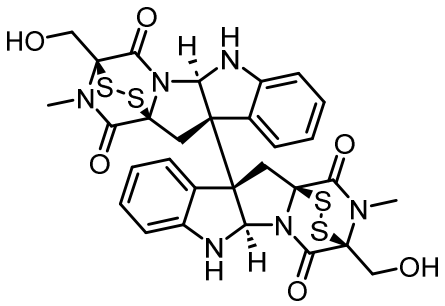


Figure 1 – *Structures of BET bromodomain inhibitors JQ1, I-BET-762, and PFI-1.*

There are also reports of poor-quality and insufficiently selective probes. The use of these compounds may perpetuate more misleading data, resulting a trail of inaccurate results.[8, 13] Chaetocin is a natural product that was reported to be a selective inhibitor of *Drosophila* histone methyltransferase SU(VAR)3-9,[14] but the scientific community found chaetocin to be active against many proteins; it was ultimately reported that the natural product formed covalent adducts indiscriminately through a pair of promiscuous disulfide bonds (Figure 2).[15]



Chaetocin

Figure 2 – *Structure of Chaetocin.*

It is unsurprising, therefore, that there was a desire to discern the most appropriate chemical probes for specific targets, leading to the peer-reviewed Chemical Probes Portal.[16]

More stringent methods of characterisation have also been suggested in the literature. Pfizer found that successful drug discovery programmes achieved the 'three pillars of survival' - 1) sufficient exposure at the site of action, 2) proof of target engagement and 3) expression of functional pharmacological activity. They subsequently applied this to cell-based target validation using chemical probes, adding proof of phenotypic perturbation as the 'fourth pillar'.[17] Specific criteria have also been proposed to maintain the standard of chemical probes to encourage accurate and high-quality data (Table 1).[9]

Characteristic	Criteria
Potency	< 100 nM (biochemical) < 1 μM (cellular)
Selectivity	> 100 fold
Controls	An inactive compound close in structure to that of the probe (e.g. an enantiomer)

Table 1 – *Specific criteria for high-quality chemical probes.*

The development of a chemical probe may appear synonymous with the drug discovery pipeline, but there are fundamental differences. Whilst a drug must be safe, effective, and have appropriate physicochemical and pharmaceutical properties, a chemical probe focuses on being able to accurately answer a biological question; it must have a defined mode of action and be selective to the target, and has less of a focus on drug-like properties. These include oral bioavailability, half-life and the optimisation of other pharmacokinetic (PK) metrics.

1.1.3 Covalent vs. reversible probes

The majority of drugs approved by the FDA are reversible binders that interact with their target under equilibrium binding states. This reversibility affects the duration of interaction between the drug and the target and therefore the therapeutic response. To prolong this interaction, researchers are turning to covalent binders more often, which now make up 30% of the approved drug market.[18-20]

A covalent compound encompasses an electrophilic centre that can form a covalent bond with protein targets, most often through cysteine side chains.[21] Aspirin was the first covalent drug approved in 1899; the mode of action was unknown until 1971, in which it was found to acetylate the substrate binding channel of cyclooxygenase 1.[22] Now, the purposeful addition of covalent moieties to pre-existing reversible ligands is an appreciated method to improve potency and selectively, such as in the discovery of ibrutinib for Bruton's tyrosine kinase (BTK).[23] Moreover, covalent ligands are being discovered from specialised libraries of electrophilic compounds.[24-26] The approach is further encouraged by the rapid target identification and profiling that is enabled by chemoproteomic platforms.[27-28]

A modern example of a successful covalent therapy is the first approved oral treatment for COVID-19. Paxlovid, which contains the drug nirmatrelvir, was designed with an 'electrophile first' approach (Figure 3). Nirmatrelvir started as an electrophile-containing peptide that was developed using a structure-based approach, whilst still considering the properties of its covalent nature.[19, 29-30] The combination therapy was found to reduce the risk of hospitalisation or death by 89% in high-risk, ambulatory adults with COVID-19.[31]

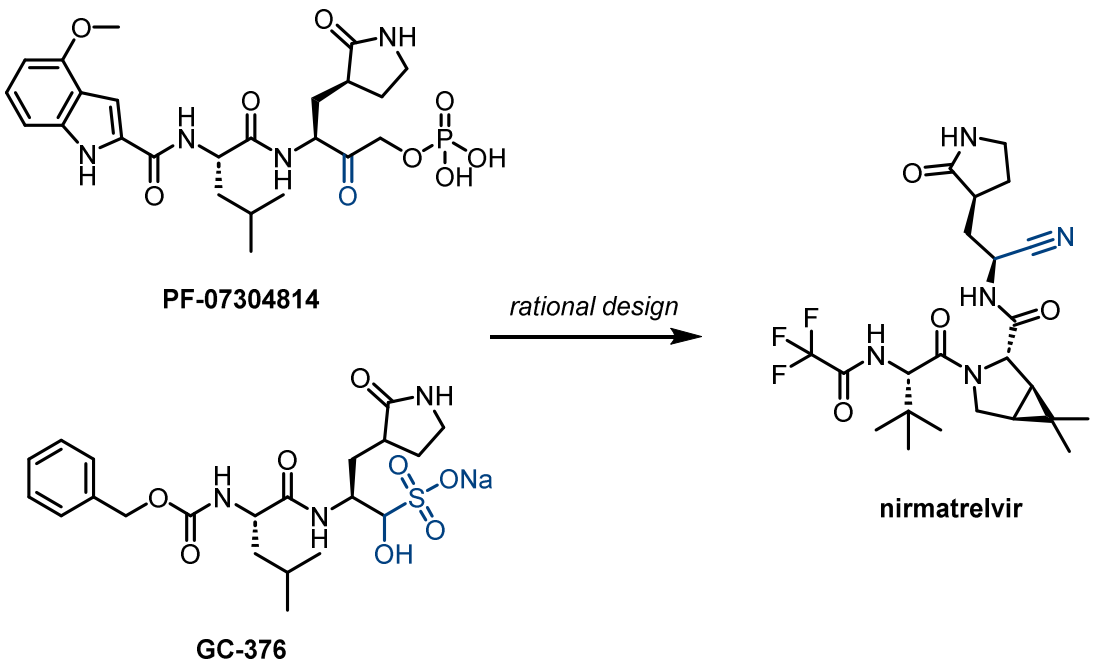


Figure 3 – *The rational design of nirmatrelvir, a covalent drug used as part of an oral combination treatment for COVID-19 (Paxlovid). PF-07304814 and GC-376 were known covalent M^{pro} inhibitors and used as starting points. Covalent moieties in each compound are highlighted (blue).*

In light of these successes, the benefits of using a covalent drug are generally supported by the modern community: they can increase the duration of action, therefore requiring less frequent dosing, reduce PK sensitivity, and offer the ability to target shallow binding sites.[32-33] Many of these benefits can be applied in the generation of chemical probes, in which the increased selectivity and potency is useful for biological utility.[34] The disadvantages of using a covalent compound should still be carefully understood, such as the potential for drug-induced toxicity, or their suitability for targets that are rapidly turned over by enzymes.[20]

1.2 Drug discovery approaches

1.2.1 High-throughput screening

The generation of bioactive compounds has historically been initiated through high-throughput screening (HTS).[35-36] HTS is a target-based approach in which large compound libraries comprising thousands to millions of compounds are screened against a biological target. The approach is popular where there is limited prior knowledge about potential ligands or druggability, as there is less bias toward specific chemical structures. The success of HTS has encouraged many companies to develop their own multimillion pound libraries.[37-38] Indeed, tyrosine kinase inhibitors gefitinib (AstraZeneca) and sorafenib (Bayer/Onyx Pharmaceuticals), both successful treatments for cancer, were developed using HTS libraries.[36]

Despite these progressions, there are concerns about the volume of new chemical space explored during HTS campaigns. The field of organic chemistry is forever expanding, and the chemist's synthetic toolkit continues to allow access to previously inaccessible scaffolds.[39] However, many libraries are based on 'reliable' chemistry and historical targets.[39-41] This in turn creates a snowball effect, in which a framework is more likely to be used if it has been used before, limiting the exploration of new chemical space.[41-42]

Orthogonal screening is also required to confirm true hits from false positives. These include problematic chemotypes, known as pan-assay interference compounds (PAINS), which are defined by their misleading behaviour in biochemical assays.[43-45] The deceptive outputs may be a result of promiscuous binders non-specifically interacting with targets, or compounds that contain redox-sensitive functionality that

generate reactive species.[46] There are also structures that directly interfere with the assay, for example due to their inherent fluorescence, or chelators that interfere with assays involving metal-dependent species.[47] Historical HTS libraries are more likely to contain PAINS, as many problematic chemotypes have been discovered in more recent years. These should be considered when using historical HTS libraries.

1.2.2 Fragment based drug discovery

An alternative approach is fragment-based drug discovery (FBDD).[38] FBDD is a rational and structure-based approach in which small, low-molecular weight compounds known as 'fragments' are screened; these are typically smaller and simpler than drug-like molecules found in HTS libraries.[40] The broader success of FBDD relies on the idea that small fragments can bind to target proteins with weak affinity but low specificity.[48] To detect these weak interactions, which may not be observable under standard assay conditions, techniques such as nuclear magnetic resonance (NMR) and X-ray crystallography are used.[49-50] Fragment binders can then be used as starting points for further elaboration to generate more potent compounds. FBDD is an iterative process, where systematic optimisation to improve affinity and selectivity leads to the development of drug-like molecules.[51]

There are several benefits to using a FBDD approach. The technique is highly sensitive to weak interactions, making it ideal for challenging drug targets.[37-38] It is also highly efficient; smaller compound libraries are required, reducing both screening and synthesis costs.[52] Finally, it offers focused hit optimisation, where a clear starting point can lead to higher success rates in lead development.[53] On the other hand, the same

arguments can be drawbacks; the iterative nature of FBDD can be time consuming, as connecting fragments to form drug-like compounds can be challenging.

Connecting or growing these fragments is a crucial step in the FBDD hit-to-lead process and can take many forms (Figure 4).[48] Fragment growing involves expanding the initial hit by adding substituents to enhance affinity and specificity.[51] Ideally, this will result in a larger complex molecule that maintains favourable binding interactions. A structure-activity relationship (SAR) analysis can also be obtained, in which researchers can understand how specific chemical changes affect the activity.[54] Fragment linking is another common approach, in which multiple fragments with complementary binding modes are connected to yield a larger molecule.[55] The desired outcome is 'superadditivity', in which the potency of the resulting compound is significantly greater than the sum of affinities of the individual fragment components. This, however, is a relatively rare occurance.[56] The final strategy is fragment merging, in which complementary fragments are identified, but overlap within the binding pocket. The overlapping portion may be replaced by a linker or structural moiety that retains the original interactions of both fragments.[57]

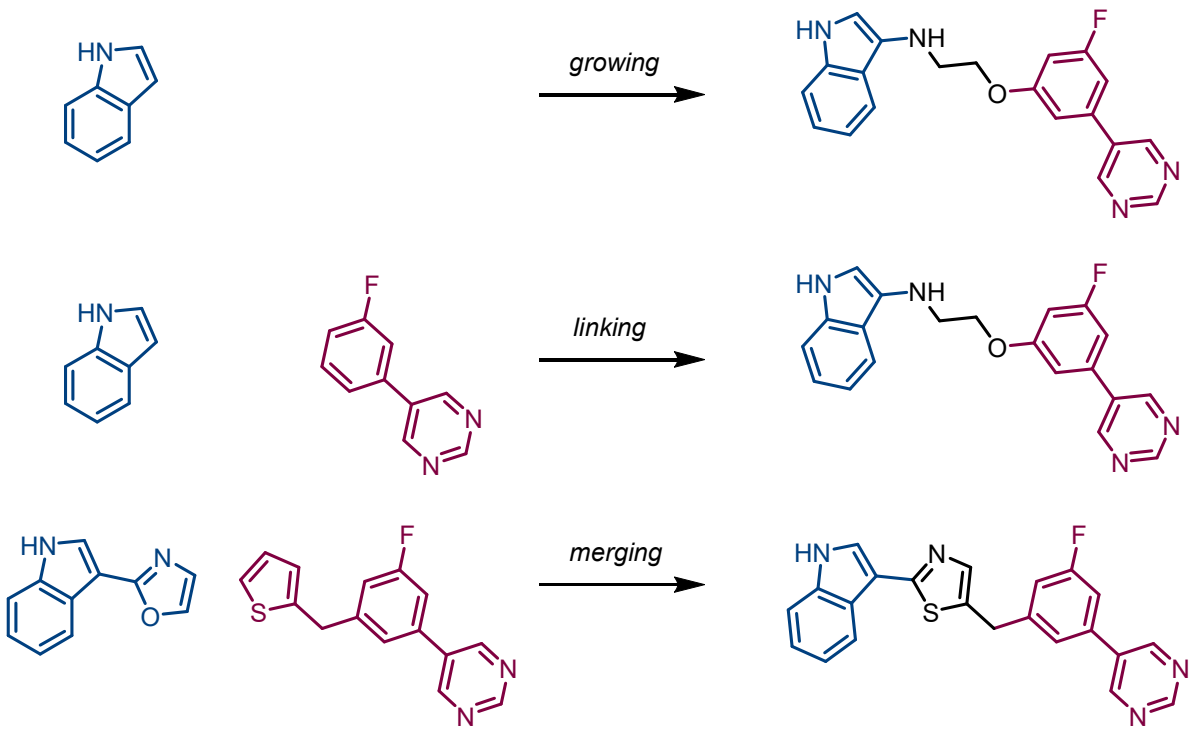
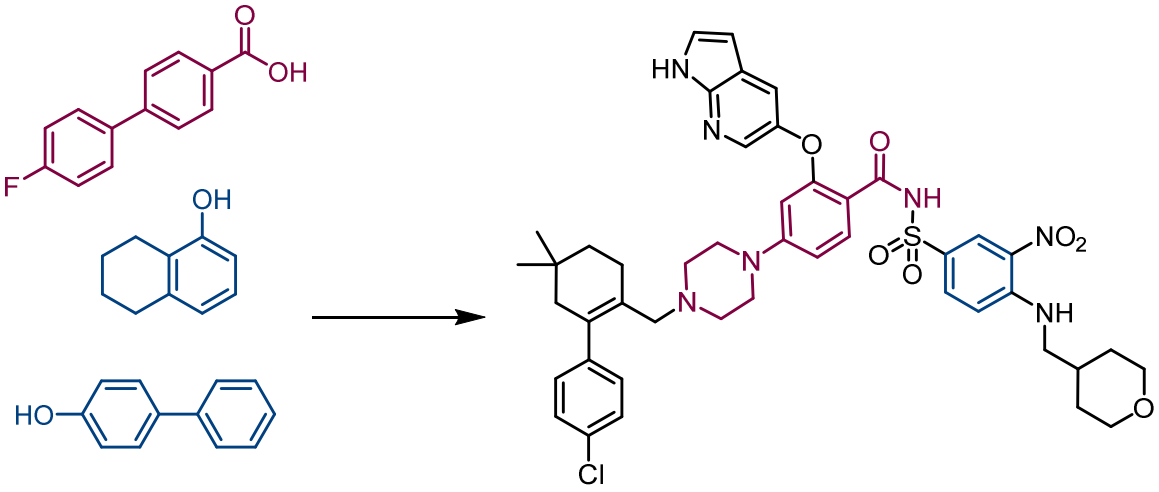


Figure 4 *– Methods of fragment elaboration used in FBDD.*

As of 2022, FBDD has been applied in the identification of over 40 compounds in clinical trials and six FDA approved drugs; pexidartinib, vemurafenib, erdafitinib, venetoclax, sotorasib and asciminib.[58-63] In the case of venetoclax, a B-cell lymphoma 2 (Bcl-2) inhibitor used for the treatment of chronic lymphocytic leukemia (CLL), an 'SAR by NMR' approach was used in early discovery (Figure 5).[64-65] This is a specialised technique in which NMR is used to study the atomic-level interactions between small molecules and target proteins.[66-67] The information is critical for understanding how specific chemical features of the fragments contribute to binding affinity.



fragments identified by NMR

venetoclax
selective Bcl-2 inhibitor

Figure 5 *– The use of FBDD in the discovery of Bcl-2 inhibitor venetoclax. Fragments (left) were identified by NMR for a related family member Bcl-xL and used as a starting point in discovery. The corresponding parts of the molecules with respect to binding are highlighted (blue and red).*

Overall, taking an FBDD approach vs. HTS can be entirely project specific. FBDD is well suited for challenging targets with poorly defined or unknown binding sites and excels at detecting weak interactions that may be missed in HTS campaigns. HTS, however, might be more appropriate for simpler targets or those with known ligands. FBDD is also less resource intensive due to the libraries generally being smaller, making it more accessible for smaller institutions.[68-69] Ultimately, FBDD has cemented itself as a promising alternative for drug and probe development and only continues to progress with new methodologies.[38]

1.3 X-ray crystallography

One of the major techniques used in FBDD is X-ray crystallography.[50, 70] This determines the three-dimensional atomic and molecular structure of a crystalline material such as a biological macromolecule. It relies on the principles of X-ray diffraction, in which X-rays directed at a crystal will result in diffraction patterns that provide information about the spatial arrangement of atoms within the crystal lattice.[71-72] Compared with NMR and cryogenic electron microscopy (cryo-EM), X-ray crystallography is historically used more often, accounting for 84% of all structures in the PDB.[73-75] High resolutions can be obtained, and it is not generally limited by the molecular weight of the sample.[76] Cryo-EM and NMR do not require crystallisation, and can be used to study dynamic molecules closer to their native state.[77] However, proteins <50 kDa are generally too small to be imaged at high resolution by cryo-EM in isolation, and NMR can be more complex in sample preparation and analysis for larger proteins or screens.[78-79] To perform X-ray crystallography, one must obtain a high-quality protein crystal from a gene; the interactions between electrons and X-rays are generally weak, so it is necessary to examine a vast number of molecules simultaneously. The next sections will describe the basis of obtaining a crystal structure.

1.3.1 Protein production

The design of a protein crystallisation system must first begin with the production of the protein. The goal is to create a functional protein that can be expressed, purified, and crystallised successfully. A ‘construct’ is described as an artificially engineered or designed version of a protein or protein segment, and construct design was found to be a major driver for successful protein crystallisation.[80] The first important factor is

function; if a specific function of a protein is being studied, it is crucial that the active domains are included in the design. Secondary structure should also be considered, as excluding part of a secondary structure can lead to protein folding issues. Secondary structure prediction servers are commonly used for guidance, such as crystallisation construct designer 2 (CCD^2) and the more recently published AlphaFold2.[81-82]

Extended disordered regions at each terminus are also undesired. These dynamic regions could contribute to poor crystallisation due to suboptimal crystal packing or could inhibit the function of the protein. Finally, hydrophobic residues at the termini can lead to aggregation of the protein and thus should be avoided.

The construct must then be cloned into an expression vector, where factors such as the expression system and purification tag should be considered.[80] Expression systems include *Escherichia coli* (*E. coli*), yeast, Baculovirus/insect cells or mammalian cells.[83-85] *E. coli* is generally used where possible, given the cost effectiveness and availability of a wide range of plasmids and strains. The protein must then be purified, a process in which the cells are lysed and the protein is extracted using buffers to maintain protein stability. This is followed by the primary purification, which is dependent on any tags attached to the protein.[86] For example, a commonly used expression vector within the Centre for Medicines Discovery (CMD) at the University of Oxford is pNIC28-Bsa4, which generates the protein with a 'His tag' at the N terminus (6 consecutive histidine residues).[80] This can be purified via immobilised metal affinity chromatography (IMAC), in which a column is packed with resin containing metal ions resulting in the reversible binding of a target protein on the

solid support matrix via the His tag. A secondary purification step may then be employed to obtain higher purity, for example size exclusion chromatography (SEC).[87] The protein is then characterised to ensure its identity, purity, and stability; commonly used approaches are using sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE), western blotting, and mass spectrometry (MS).[88]

1.3.2 Protein crystallisation

Various methods are used to crystallise a protein after obtaining the protein sample. The approach of choice in the CMD is a vapor diffusion method, specifically 'sitting drop' (Figure 6).[89-90] In this technique, a small volume (typically 1-2 µL) of the protein solution is mixed with a precipitant that contains salts, solvents, and other agents that promote crystallisation. The resulting mixture is dispensed onto a flat surface, generally a crystallisation plate well, and the precipitant solution is added in a proximal reservoir. The premise of the approach is controlled nucleation, defined as the initial step in the formation of a crystal.[91] Over time, the protein and precipitant solution will reach a state of equilibrium in the drop which can result in several situations (Figure 7).

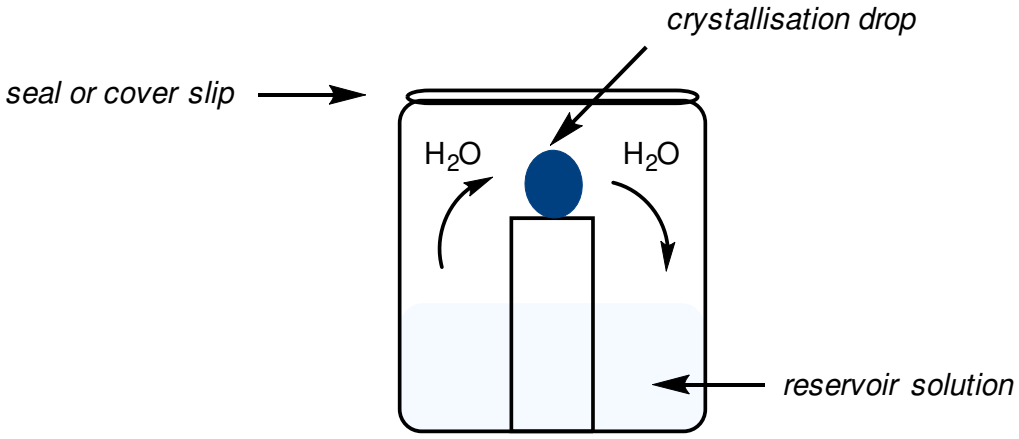


Figure 6 *– Schematic of 'sitting drop' vapour diffusion. A crystallisation drop (dark blue) comprised of a protein sample and buffer solution will be in vapour equilibrium with a reservoir of the buffer solution, encouraging controlled nucleation. A cover slip is applied to minimise evaporation.*

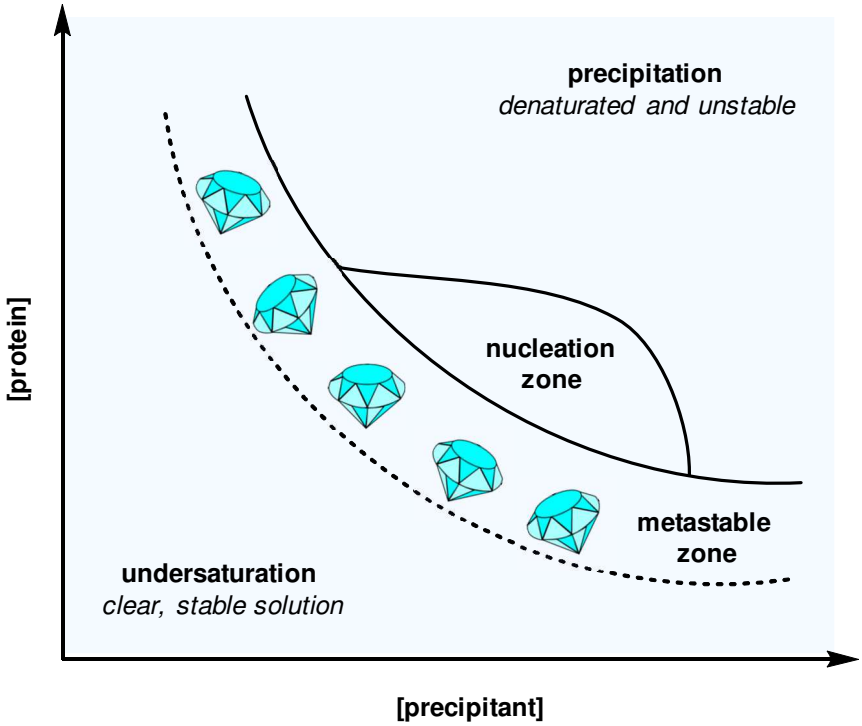


Figure 7 – *Solubility phase diagram for crystallisation.*

Ideally supersaturation occurs, in which the protein concentration exceeds its solubility limit. This should result in nucleation, the initial formation of small clusters of protein molecules. Under favourable conditions, these clusters can continue to grow into larger crystals. They may, however, continue to precipitate in isolation resulting in a mixture of protein solution and precipitate that cannot be used. Alternatively, the mixture won't reach supersaturation, in which a clear, stable protein solution will be the outcome.

Determining the best conditions for crystallisation can be unpredictable and require extensive exploration. This has led to the advancement of crystal screens, which are pre-formulated solutions designed to systematically explore a range of conditions for crystal formation.[92-93] They are diverse in their pH value, salt concentrations, precipitants and additives. Commercial screens are well-established and ultimately increase the likelihood of obtaining a high-quality and reproducible protein crystal for structure determination or fragment screening.[80]

1.3.3 Data acquisition

Once crystal growth is optimised, the crystals are added a cryoprotectant, mounted onto small pins, and stored temporarily in liquid nitrogen to minimise radiation damage.[94-95] The crystals will then be rotated in an X-ray beam.[71] The X-rays interact with the electrons in the crystal lattice causing them to scatter in various directions, creating a diffraction pattern of spots on a detector.[96-97] Mathematical calculations are then applied, including Fourier transformation, to extract the electron density distribution within the crystal.[98-99]

The central processing challenge is known as the 'phase problem'.[100-101] 'Phases' refer to the angles or numerical values that describe the position of each diffracted X-ray wave in its oscillation cycle, indicating the direction of electron density at each point in the crystal.[102-103] These values are crucial for reconstructing the electron density map, but are not directly measured by the detector. Several techniques have been developed to address the phase problem. Molecular replacement (MR) is a popular approach when the structure of a homologous or related molecule with a known structure is available, known as a model.[104-105] The chosen model is then orientated in various positions within the unit cell obtained from the diffraction pattern. The ultimate goal of this step, referred to as 'phasing', is to find the best orientation and position that fits the observed data.[100] The phases are determined based on the phases of the calculated structure factors from the model, providing information to construct the electron density map.[101] The map is used to refine the structure, known as 'refinement', by adjusting atomic positions and other parameters to improve the fit between the model and map.[106]

The quality of the structure will then be assessed through various criteria that provide information about the reliability, accuracy and precision of the atomic model and electron density map.[107-109] The resolution is a fundamental indicator of the quality; it represents the finest level of detail that can be observed in an electron density map, where a higher resolution indicates better quality. There are also statistical measures that compare the calculated electron density map with the experimental diffraction data. The R_{work} measures the agreement between the calculated structure factors and the observed data, in which a lower value indicates a better fit.[110] R_{free} is a specific type of R-factor that is calculated using a subset of the diffraction data that was

excluded during refinement, serving as an independent validation.[107] Again, a lower value is desirable. High 'completeness' is essential for accurate structure determination, which measures the fraction of all possible unique reflections collected.[111] A Ramachandran plot will often be used during refinement to follow progress; this assesses the backbone dihedral angles of the residues, identifying regions that are allowed, disallowed, or outliers.[112] Finally, B-factors, or temperature factors, describe the thermal vibrations of atoms in the crystal and should reflect the expected motion for a given temperature.[113]

Finally, the structure is validated by the protein data bank (PDB) to ensure it accurately represents the experimental data; this usually results in the deposition of the structure to the PDB.[75, 114]

1.3.4 Ligand screening

X-ray crystallography can be applied to FBDD as a powerful tool for identifying hits whilst simultaneously determining the structure of the hit bound to their targets.[70] A variety of commercial libraries are available, including diversity libraries and focused libraries that are designed to target specific protein families or binding sites.[115] Institutions may also develop in-house libraries tailored to their specific interests.[116]

Introducing fragments to protein crystals involves either co-crystallisation or soaking.[117] For co-crystallisation, the fragment will be mixed with the protein solution prior to crystallisation, and then set up for crystallisation experiments. This approach is more suitable for fragments that have relatively high binding affinities for the protein are so are expected to remain bound during the crystallisation process.[118] For

soaking, pre-grown crystals are exposed to a solution containing a higher concentration of the fragment than the original crystallisation mixture.[50, 119] As the protein crystal is typically ~50% solvent, the fragment can diffuse into the crystal and bind to the protein within the crystal lattice. This is particularly useful for weaker binders.

One platform for the effective screening of fragments is the XChem facility at Diamond Light Source, a synchrotron radiation facility in the UK.[120] This is a dedicated platform for FBDD accessible to both academia and industry. Automation is a key feature of the operation which plays a crucial role in maximising efficiency and throughput of experiments. The facility is equipped with robotic systems to handle the loading and retrieval of crystal samples into the X-ray beams, and aspects of sample preparation. This includes mounting crystals, changing sample holders, and positioning them correctly for data collection.

A typical workflow will involve the following: 1) the researcher produces the crystals and transports them to Diamond Light Source; 2) fragments are added to the protein crystals using acoustic dispensing; 3) the soaked crystals are mounted with Shifter technology and data collected; 4) the XChemExplorer pipeline processes thousands of datasets, and the PanDDA software identifies binders by increasing the signal:noise ratio through a background density reduction algorithm; 5) structures with a fragment bound are flagged and can be solved by the researcher.[119, 121-123] The level of optimisation required dictates the number of crystals processed (Figure 8). The automated workflow makes the XChem facility highly suitable for high-throughput crystallography and thus the rapid screening of fragments for FBDD. The platform was

applied in the screening of SARS-CoV-2 targets throughout the COVID-19 pandemic; the turnaround for the main protease was just 3 weeks.[120, 124]

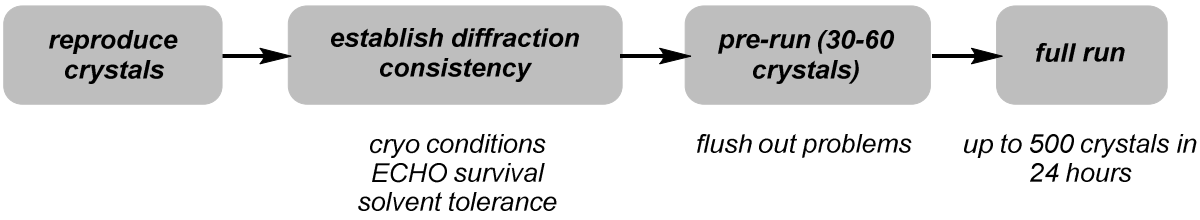


Figure 8 *– Typical XChem workflow as described by Diamond Light Source.*

1.4 Computational docking and visualisation

Following the identification of a fragment hit, the 3D structure of the fragment/target complex can be visualised in chemical simulation software such as Schrödinger and Molsoft ICM.[125-126] The software can also be used for exploring potential binders using docking simulations. This is a multi-stage computational process used in structure-based drug design (SBDD) to identify and optimise molecules that bind to a target protein.[127-129] First, an appropriate fragment library will be selected or enumerated to add functionality to a known binder. The compounds will then be 'docked' individually into the binding site, sampling possible ligand orientations and conformations, and the binding poses and energies are evaluated using a scoring function.[125] Various scoring functions evaluate the predicted binding affinity of each fragment.[130] Researchers will typically focus on fragments with the highest docking scores, as they are more likely to be effective binders. The compounds can then be synthesised and tested to validate the predicted binding.

Whilst computational docking is a valuable technique for prioritising compounds for synthesis and testing, there are limitations.[93, 131] For example, the target structures themselves may be of high quality, but could still contain errors or missing regions that lead to inaccuracies in docking results. Biological targets also exhibit conformational flexibility which can be challenging to account for in simulations.[132] Exhaustive sampling of all possible conformations can also be computationally expensive and time-consuming, and thus a researcher with limited access may miss binding modes.[133] Other considerations are solvent effects and the flexibility of the ligand.[134] Ultimately, scoring functions are approximations, and the hypotheses that arise require experimental validation. The approach is generally used as a source of guidance and

prioritisation, and is a valuable tool for compound optimisation when the limitations are appropriately acknowledged.

1.5 Protein structure prediction

Impressively, ~100,000 unique protein structures have been determined and published.[82] This, however, represents only a fraction of known protein sequences. The bottleneck for structural identification is the extensive time, effort, and cost to determine one, and thus the accurate *in silico* prediction of structures is desirable. One technique is the use of comparative modelling; if a protein's sequence is related to one with a known structure, alignment can be used to build a model based on the template. A more complex approach is Ab Initio (de novo) modelling. Here, the principles of chemistry and physics are used to predict how the atoms in a protein would arrange themselves in a stable conformation. This is computationally intensive and generally less accurate for larger or more complex proteins.

In recent years, the AlphaFold2 server has cemented itself as a noteworthy intelligence system for predicting 3D protein structure. AlphaFold is trained on a vast number of protein structures from the PDB and the neural networks recognise patterns in protein sequences and their corresponding 3D structures.[82] As well as predicting the structure, confidence scores are also provided. These allow researchers to assess the reliability of the structure. Prediction servers such as this reduce the need for costly experimental techniques but are not without their disadvantages. Although AlphaFold has shown remarkable accuracy in predicting structures, it is not infallible, especially for challenging protein structures or specific regions. Disordered regions are also challenging to predict. Whilst this is unsurprising given the dynamic nature of such regions, it can result in the poor prediction of connected domains relative to one another in space. It is important to recognise that AlphaFold is an extremely powerful

tool for structural biology, but should be used in conjunction with other experimental and computational methods.

1.6 Targeting the GEF/GTPases

1.6.1 The Ras superfamily

The small guanosine triphosphatases (GTPases) are a family of proteins that play a crucial role in various cellular processes, particularly in signal transduction, cell division, and intracellular transport.[135] The Ras superfamily of over 150 GTPases includes several subfamilies, each with distinct functions and roles.[136-137]

The Ras subfamily is perhaps the most well-known, and is involved in cell growth, differentiation, and survival.[138-139] Mutations in Ras family members, particularly KRas, are commonly associated with cancer.[140] The Rab family is associated with vesicular trafficking and intracellular transport, and targets vesicles to specific cellular compartments, such as the Golgi apparatus, endosomes, and lysosomes.[141] Ran GTPases control nucleocytoplasmic transport by regulating the movement of proteins and ribonucleic acid (RNA) between the nucleus and cytoplasm, essential for mitosis and RNA export.[142-143] The Arf family regulate vesicle trafficking and membrane dynamics, including the formation of clathrin-coated vesicles during endocytosis and exocytosis.[144-145]

The remaining subfamily is the focus of this thesis. The Rho subfamily comprises 20 members that regulate the dynamics of the actin cytoskeleton, vital for a myriad of cellular processes.[146-148] Rho GTPases are involved in cell shape, cell migration, cell adhesion and other processes that require changes in the organisation of the cytoskeleton. The most highly studied are Ras homolog family member A (RhoA), Ras-related C3 botulinum toxin substrate 1 (Rac1), and Cell division control protein 42

homolog (Cdc42), all of which have been linked to neurodegeneration. It is unsurprising, therefore, that the subfamily are of interest in the medical field.

1.6.2 The GTPase activation cycle

The activation cycle of the GTPases is a fundamental process that governs their molecular switch activity (Figure 9). The GTPases transition between a guanosine triphosphate (GTP) bound active state, and a guanosine diphosphate (GDP) bound inactive state.[135, 149] Both nucleotides are bound with extremely high affinity, typically in the nanomolar to picomolar range[150], and so the switch process is tightly regulated by effectors.[135] Activation of the GTPases begins with nucleotide exchange; when GDP is bound, the guanine nucleotide exchange factors (GEFs) bind to the GTPase and increase the GDP dissociation rate, allowing GTP to bind in the nucleotide binding pocket. The binding of GTP results in a conformational change of the GTPase, which can in turn interact with downstream effectors and initiate signalling cascades or processes. Over time, the intrinsic activity of the GTPase hydrolyses the bound GTP to GDP, thus inactivating the GTPase. This process, however, is very slow. The GTPase-activating proteins (GAPs) significantly accelerate this process, enhancing the ability of the GTPase to return to an inactive state. This prevents prolonged activation, ensuring proper regulation and finely tuned responses to signals.[150-151] Guanine dissociation inhibitors (GDIs) also exist for the Rho and Rab GTPases, which prevent the dissociation of GDP and maintain the GTPase in its inactive state.

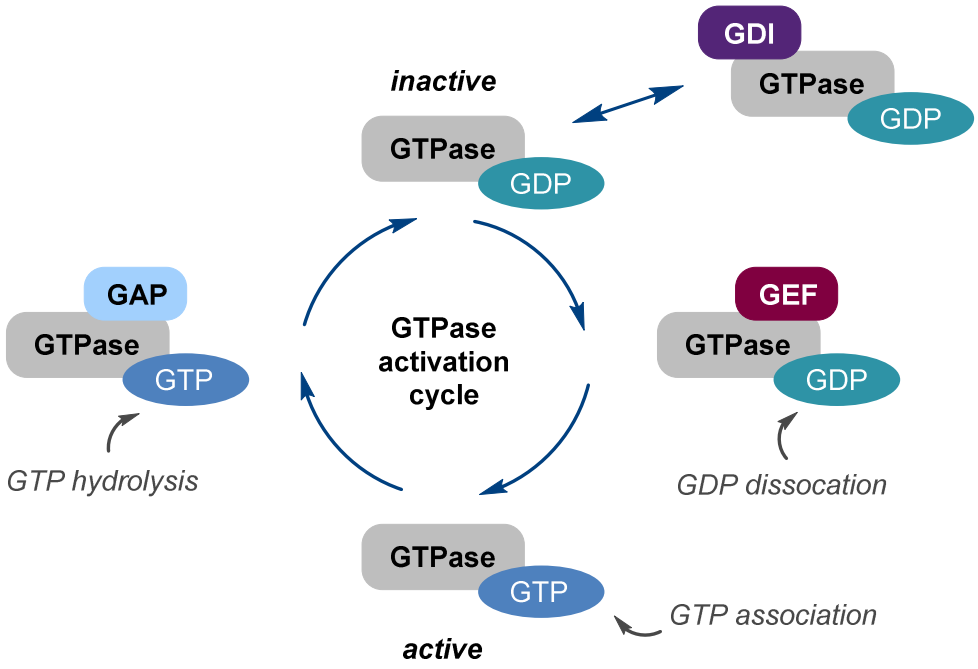


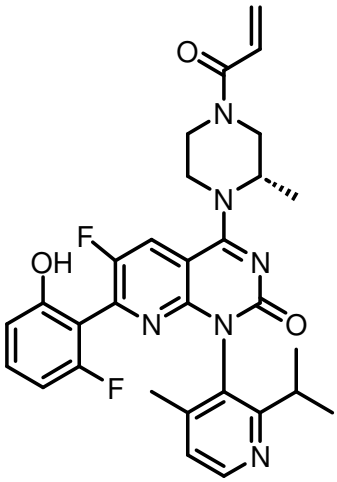
Figure 9 *– The GTPase activation cycle. GTPases (grey) exist in an inactive GDP (teal) bound state and active GTP (blue) bound state. The GEFs (magenta) catalyse the dissociation of GDP and GAPs (light blue) promote the hydrolysis of GTP to GDP. GDIs (purple) exist for the Rho and Rab GTPases which maintain the GDP-inactive form of the GTPase.*

1.6.3 GTPases in drug discovery

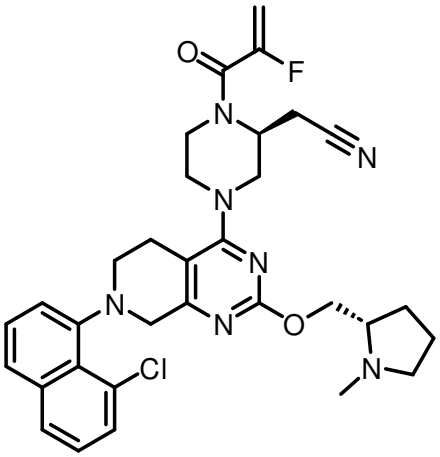
Historically, GTPases have been targeted for drug development primarily through indirect approaches. This is due to the challenges associated with directly inhibiting the nucleotide binding pocket, namely generating a competitive inhibitor against the tightly bound nucleotides. This, in combination with the high concentration of GDP (>30 µM) and GTP (>300 µM) in the cell, has resulted in many unsuccessful attempts to modulate GTPase activity directly.[152] One alternative approach is regulating post-translational modifications.[153] Tipifarnib is being investigated for the treatment of HRAS-mutant head and neck squamous cell carcinoma (HNSCC). The drug blocks the farnesylation of Ras, preventing it from proper membrane localisation and

activation.[154-155] Downstream targets have also been of focus, such as targeting Raf kinases downstream of Ras in cancer therapy.[156-158]

Therapies that directly target the GTPases have only been approved in the last two years and were a breakthrough in GTPase druggability. In 2021, sotorasib was approved by the FDA for the treatment of non-small cell lung cancer (NSCLC) (Figure 10).[159-160] The drug is specific for the proto-oncoprotein KRas G12C.[161] Typically, patients with this mutation have a poor prognosis and have not responded well to conventional treatments. Sotorasib covalently interacts with the G12C mutated residue, whilst the remainder of the drug binds in a cryptic pocket near to the nucleotide binding site. This in turn prevents the binding of GTP, whilst still allowing GDP to bind. KRas G12C is thus engaged in an inactive conformation, preventing the binding to effectors that lead to the effects seen in NSCLC. Adagrasib was approved a year later for the same indication, and targets KRas G12C in a similar mode to sotorasib.[162] Both treatments were found to shrink tumours in 37% and 43% of cases, respectively.[163]



sotorasib



adagrasib

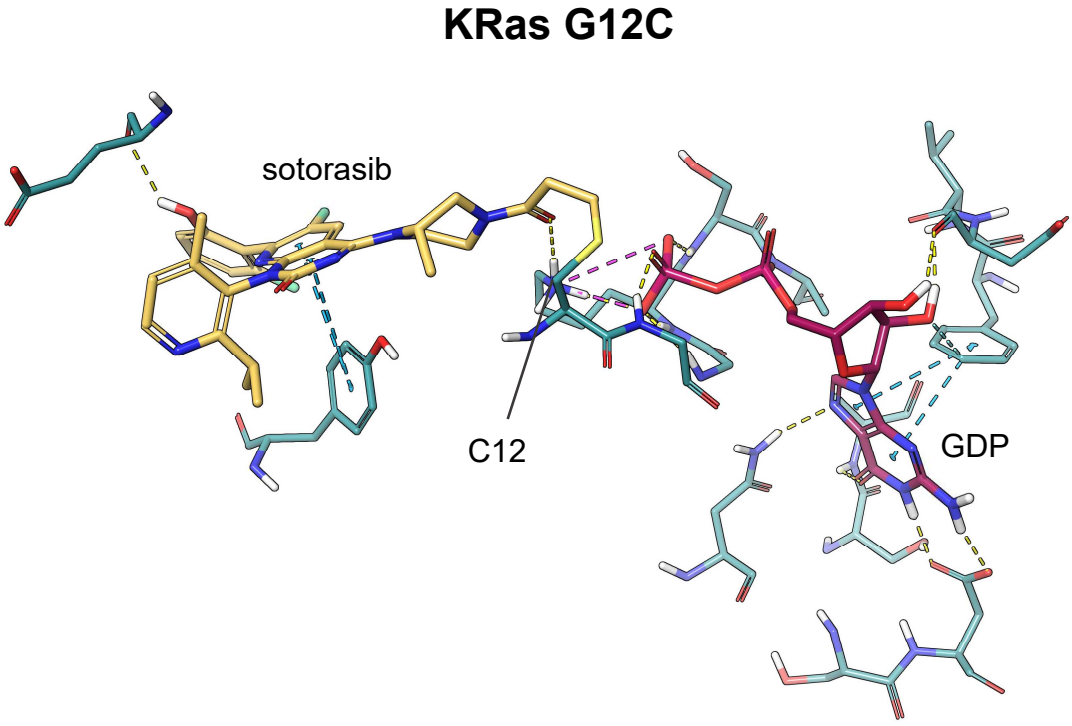


Figure 10 *– FDA-approved inhibitors of the KRas G12C mutant. Above: Structures of sotorasib and adagrasib. Below: Sotorasib (yellow sticks) partially binds at the nucleotide binding site of KRas G12C (blue sticks) through a covalent linkage with C12. The drug binds alongside GDP (purple sticks) but prevents the binding of GTP.*

1.6.4 Targeting Rho GEF/GTPases

There is not yet an approved drug or high quality probe specifically targeting the Rho GTPase subfamily.[164] The availability of a selective probe could significantly progress the understanding of Rho GTPase pathology. However, due to the reasons previously described, directly drugging the GTPases is challenging. An alternative approach is targeting GTPase activity through effectors.[164-165] As described in Section 1.6.2, the GEFs bind to the GDP-bound GTPase and form a low-affinity complex and induce an extensive conformational change of specific regions of the GTPase, and thus reduce the affinity of the bound GDP. Upon nucleotide dissociation, a high affinity GEF-GTPase complex is formed. Then, once GTP associates, the complex is in a low-affinity state and the GEF dissociates to give a high-affinity GTP-GTPase complex.

Targeting the Rho GEF/GTPase complex would have multiple benefits. First, a competitive inhibitor would need to outcompete the nucleotide bound with micromolar affinity, rather than the picomolar affinity of the nucleotide-GTPase complex. Additionally, the Rho GEFs outnumber the Rho GTPases by 3:1, allowing a more rigorous spatial control of activity.[165-166] There is also evidence to show that the binding site would be sufficiently different from the GTPase in isolation, allowing for selectivity between the GEF-GTPase complex and GTPase.[167]

Thus far, attempts to inhibit Rho GTPase activation directly or via the GEF/GTPase complex have largely resulted in compounds with low potency and limited selectivity (Figure 11).[164] The majority of these efforts focus on Rac1, RhoA and Cdc42. A commonly used inhibitor of Rac1 activity is NSC23766 (IC_{50} ~ 50 μM).[168] Whilst NSC23766 was shown to produce a dose-dependent inhibition of Rac1 activity in

cellular assays, off-target effects are possible due to poor potency. This led to multiple campaigns generating similar scaffolds, with no significant improvement in potency.[164] Another inhibitor, MBQ-167, was reported to inhibit Rac1 activity in cells with an IC_{50} of 103 nM; MBQ-167 also inhibited Cdc42 with an IC_{50} of 78 nM, demonstrating the challenges of obtaining selectivity.[169]

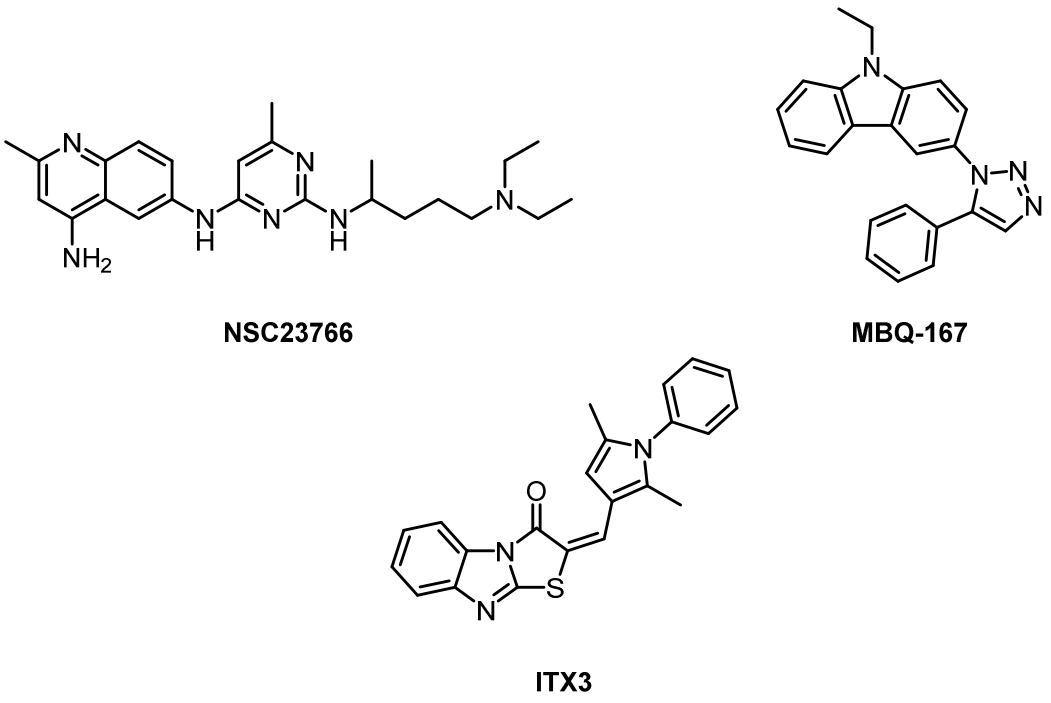


Figure 11 *– Structures of previously reported inhibitors of Rho GTPase activity.*

Inhibitors have also been reported that inhibit GEF activity. ITX3[170] (IC_{50} = 76 µM) is advertised as a specific GEF Trio inhibitor, and yet contains a 1, 2, 3-alkyl pyrrole motif reported to be a PAIN, in addition to a reactive toxicophore. ITX3 was also used to explore the effects of inhibiting Rho GEF activity of GEF Kalirin *in vivo*, but the structural alerts and low potency of the inhibitor question the validity of the results.[171] In summary, potent, selective, and well characterised inhibitors of Rho GTPase activity to prevent the overactivation of GTPase activity are still desired.

1.6.5 Rac1 and RhoA

Rac1 is one of the most highly studied Rho GTPases and is associated with a plethora of disease areas.[172-174] One of its primary functions is to regulate the dynamics of the cytoskeleton.[175] When activated, Rac1 promotes the formation of actin filaments, essential for processes like cell motility, shape changes and formation of membrane protrusions.[176-177] It is also involved in cell migration, intracellular signalling, and endocytosis.[178-179] Dysregulation of Rac1 activity, generally it's overactivation, has been linked to cancer, neurodegenerative diseases, and cardiovascular diseases.[180-182] More specifically, Rac1 has been linked to Alzheimer's disease (AD) and Huntington's disease (HD) based on the implication of Rac1 in synaptic plasticity and neuronal survival.[183-186] The schizophrenia risk factor 'disrupted in schizophrenia 1' (DISC1) was also found to regulate the spines of the glutamate synapse via Rac1.[187] It is unsurprising, therefore, that Rac1 is a target of interest in neuropathology and drug development. Inhibitors could prevent the overactivation of Rac1, thus downregulating GTPase activity in disease.

A structurally related Rho GTPase of interest is RhoA.[188] Like Rac1, RhoA is a central regulator of the actin cytoskeleton.[188-190] However, whilst Rac1 stimulates the polymerisation of branched actin networks, RhoA induces the formation of actomyosin stress fibres and other contractile structures.[177, 191-193] Accordingly, Rac1 has been associated with membrane protrusion, and RhoA more closely with cell contraction.[194] RhoA also has distinct roles in cell adhesion and gene expression.[195-197] The GTPase has similarly been studied in the context of neurodegeneration, as it plays a role in neuronal morphology and axon guidance.[198] Associated diseases include AD, Parkinson's disease (PD), and amyotrophic lateral sclerosis (ALS).[199-202] Ultimately,

both GTPases induce coordinated changes to drive a number of crucial biological responses.

1.6.6 Kalirin

Both Rac1 and RhoA are activated by Kalirin (from gene KALRN), a multi-domain protein that plays a significant role in the regulation of synaptic function and neuronal morphology in the brain.[203-204] It is critical to the formation and maintenance of dendritic spines, which are small protrusions on the dendrites of neurons.[205-207] Changes in this morphology are intimately related to cognitive function, and so the dysregulation of Kalirin has been linked to neurological disorders such as AD, schizophrenia, and bipolar disorder.[208-211]

Kalirin's name, after the many-handed goddess Kali, reflects its multiple interacting properties (Figure 12). At the N-terminus, Kalirin contains a Sec14p domain thought to be involved in membrane interactions, that was recently reported to bind a limited set of phospholipids.[212-213] Following this domain are nine spectrin-like repeats. These generally contribute to the structural stability of a protein, and in Kalirin were found to form a 'nested' structure.[214] They facilitate protein-protein interactions (PPIs), oligomerisation, cytoskeletal interactions and subcellular localisation, contributing heavily to the maintenance of neuronal structure.[215] Kalirin was found to bind to DISC1 via its spectrin repeat domain, suggesting that the domain may regulate Kalirin's catalytic activity.[187]

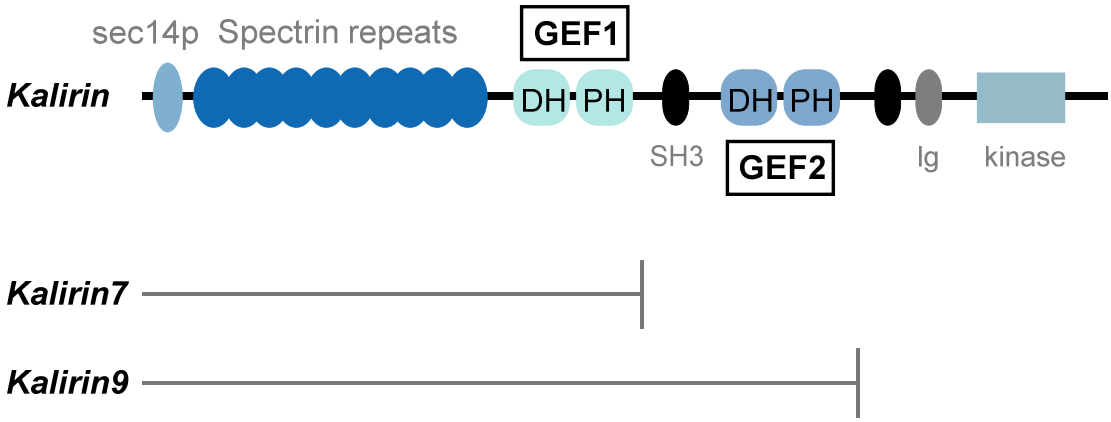


Figure 12 – *Domains and relevant splice variants of protein Kalirin.*

Kalirin also comprises two Scr homology 3 (SH3) domains, which typically bind to proline-rich sequences and mediate interactions with other signalling proteins.[216] At the C-terminal, Kalirin encompasses a immunoglobulin-like (Ig) domain, a fibronectin III (FnIII) domain, and a kinase domain predicted to be a serine-kinase. The function of these domains is largely unknown; the only characterisation of the kinase domain is the autophosphorylation of the skeletal muscle KALRN isoform.[217] A more in-depth study is required to characterise the role of these domains in regulating Kalirin activity.

Relevant to this thesis is the presence of two Rho GEF domains in Kalirin. The first (Kalirin-GEF(1)) is able to bind and activate Rac1, and the second (Kalirin-GEF(2)) RhoA (Figure 12).[218-219] The GEF domains are typical of the Rho Dbl GEFs, comprising a Dbl homology (DH) domain - the core catalytic domain responsible for promoting nucleotide exchange - and a tandem Pleckstrin homology (PH) domain.[165, 220] The PH domain is variable in its function depending on the GEF - it can bind to phospholipids to help target the GEFs to the cell membrane, or variably influence GEF activity.[221-223] The Dbl GEFs are structurally distinct from another group of Rho GEFs,

the dedicators of cytokinesis (DOCKs), which are characterised instead by DOCK homology regions 1 and 2 (DHR1 and DHR2).[224-225]

The better studied isoforms of Kalirin are Kalirin7 and Kalirin9, two adult splice variants.[219] Intriguingly, Kalirin-9 and Kalirin-12 (full length) were found to express predominantly during embryonic neuronal development, whereas Kalirin-7 expression was only observed in the adult cortex.[211, 226] Moreover, expression appeared to shift almost exclusively to Kalirin-7 over other isoforms in the adult cortex, suggesting a shift in the gene function of KALRN in later neuronal development.[211]

Kalirin7, the isoform containing the first GEF domain only, has been associated with anxiety in Kalirin7 knockout mice.[227] Kalirin7 mRNA was also found to be downregulated in the prefrontal cortex of post-mortem schizophrenic brains.[228-229] Contrastingly, Kalirin9 mRNA was found to be upregulated in the superior temporal cortex; this splice variant contains both of the described GEF domains.[206] These studies highlight the important role Kalirin has in neurodevelopment, but do not confirm that the observed phenotypic effects are due to the differing GEF activity of the Kalirin7/9 isoforms. The generation of inhibitors for each of the Kalirin/GTPase complexes could inhibit the activity arising from the active complex, and establish the role of Kalirin in neurological signalling networks.

1.6.7 Previous Kalirin-GEF(1)/Rac1 progress

Dr. Janine Gray, previously of the Brennan Group, solved the novel crystal structure of the DH1 domain of Kalirin (1232-1410) bound to Rac1 (1-177) and GDP (Figure 13).[167] Small GTPases exhibit a similar overall fold, containing a central ß-sheet with

strands interspersed by α-helices. The nucleotide binding site is formed at the juncture of two mobile loops known as the switch I and switch II regions (residues 26–45 and 59–74 respectively for Rac1) and the phosphate binding region (P-loop; residues 10–17 for Rac1).[230] The P-loop contains a GxxxxGK[S/T] sequence that recognises the ß-phosphate of bound nucleotides and crucially coordinates Mg^{2+} to complex the charged pyrophosphate moiety, contributing to high ligand affinity. The switch I region contains a highly conserved threonine residue (T35 in Rac1) that is responsible for forming interactions with the *γ*-phosphate in GTP and the bound Mg^{2+}.[231]

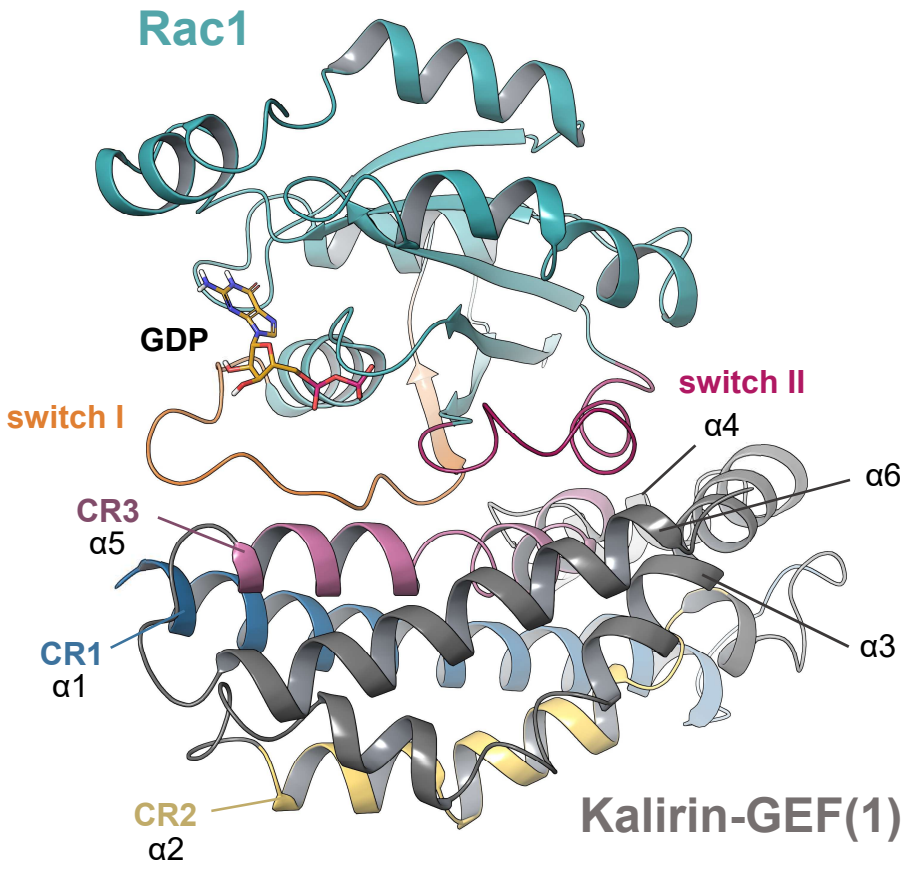


Figure 13 – *The GDP-bound co-crystal structure of Rac1 and the DH domain of Kalirin-GEF(1) (PDB: 5O33). Rac1 (teal ribbons) contains the switch I (orange ribbons) and switch II (magenta ribbons) regions that form the nucleotide binding site for GDP (yellow sticks) and interact with the GEF domain. The DH domain of Kalirin-GEF(1) has a 6 α-helical structure containing conserved regions (CRs) CR1 (blue ribbons), CR2 (yellow ribbons) and CR3 (pink ribbons). CR1 and CR3 interact with Rac1 at the switch I and II regions.*

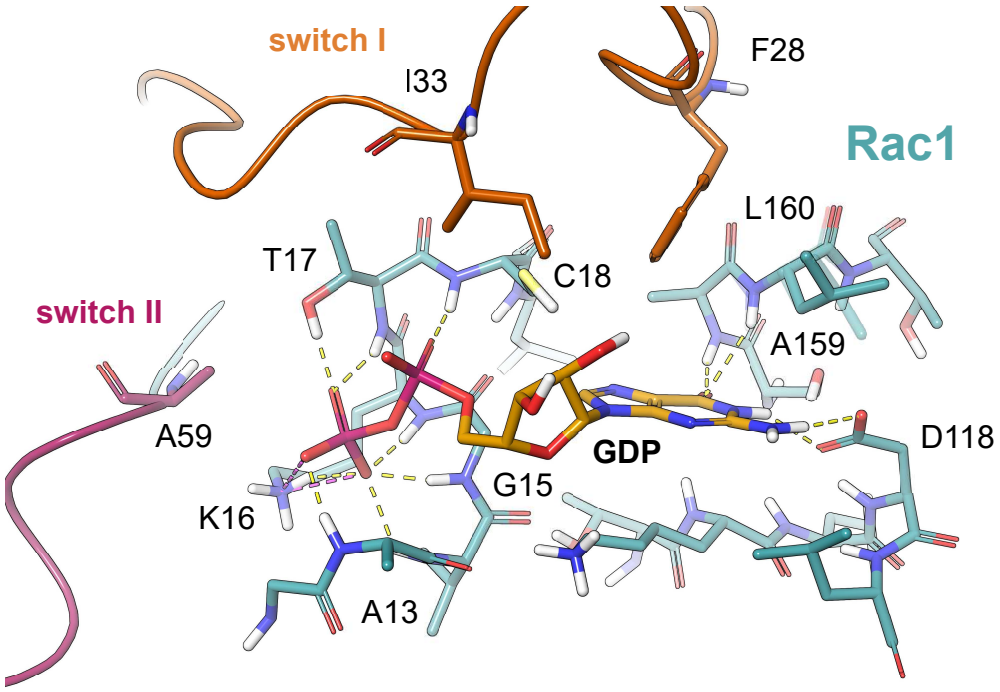


Figure 14 *– The nucleotide binding site of Kalirin-GEF(1)/Rac1. GDP (yellow sticks) binds to Rac1 (teal sticks) through various interactions and is formed at the juncture of the switch I region (orange sticks) and switch II region (magenta sticks). Hydrogen bonding interactions (yellow dashes) and salt bridges (pink dashes) shown.*

GEFs exert their activity by binding to the GTPase and disrupting these key interactions, whilst leaving the remainder of the GTPase unperturbed.[165] A canonical DH domain, and indeed Kalirin-GEF(1), consists of elongated helical bundles described as helices α1–α6 in accordance with previous nomenclature, with three conserved regions (CR1, 2 and 3).[232] CR1 and CR3 predominantly interact with switch I and disrupt key interactions with the bound nucleotide and Mg^{2+} cofactor through steric hindrance. CR3 and part of the α6-helix interact with switch II, which stabilises the GTPase and occludes the Mg^{2+} co-factor. This occlusion is crucially enacted by an alanine residue from switch II (A59 in Rac1).[150, 230] The affinity of the GDP reduces from picomolar to micromolar affinity *via* this mechanism (Figure 14).[164] It was envisaged that Kalirin-GEF(1)/Rac1 could be targeted by a small molecule, and that

targeting the low-affinity binding state of the nucleotides would yield more potent inhibitors. Additionally, a proximal cysteine residue was identified in the pocket (C18) that could be harnessed with covalent inhibitors.

Multiple screening techniques were applied in this effort. Isoxazole **1** was identified from an XChem screen and was found to bind in high occupancy in the nucleotide binding site, showing its capability in displacing GDP (Figure 15). The urea moiety formed two hydrogen bonds with D118, and the isoxazole interacted with the peptide backbone of A159, mimicking the hydrogen bond donor-acceptor pattern observed in the GDP-bound structure. Efforts were made to grow the fragment from accessible vectors, but no new analogues were found to bind with increased affinity.

An *in-silico* screen was subsequently performed to look for compounds with similar pharmacophores but diverse chemical scaffolds to isoxazole **1**, GDP, or fragmented GDP. The top 300 compounds were purchased and screened using XChem, and pyridone **2** was identified as a fragment of interest. This fragment interacted with a similar binding pose to isoxazole **1**, and notably possessed a dual hydrogen bonding interaction with D118.

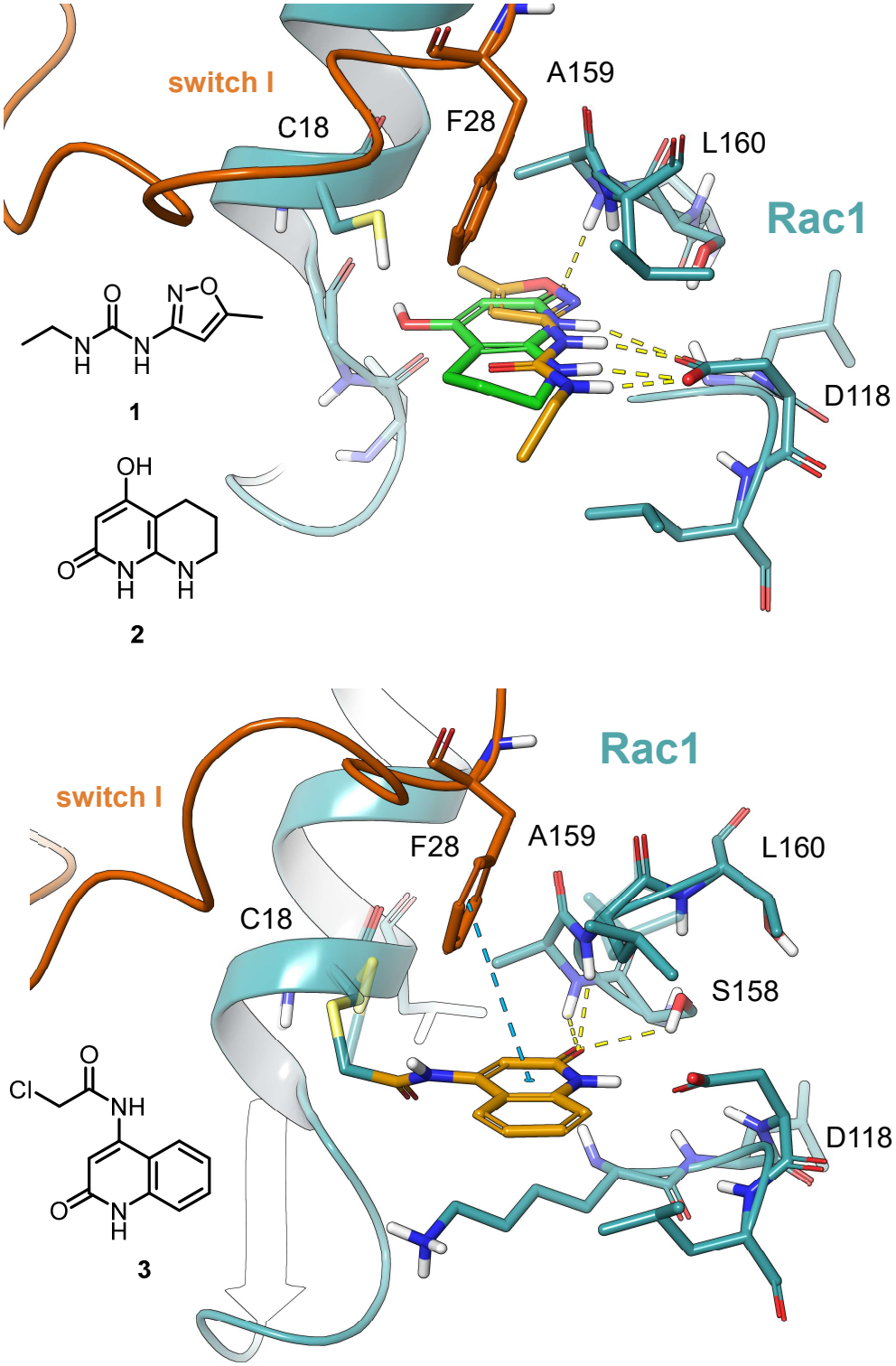
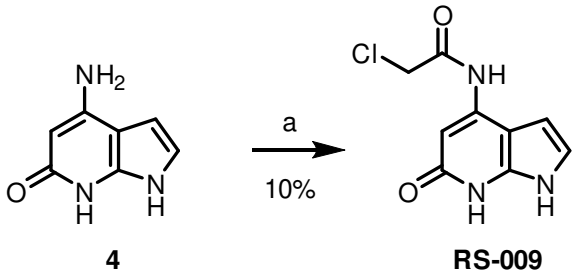


Figure 15 – *Fragment hits identified for Kalirin-GEF(1)/Rac1. A) Fragments 1 (yellow sticks) and 2 (green sticks) interact with D118 through their hydrogen bond donors, and in the case of 2 with the backbone of A159. B) Fragment 3 (yellow sticks) hydrogen bonds to the backbone of A159 and L160 and side chain of S158, and forms a covalent bond with C18 in high occupancy. Hydrogen bonding interactions (yellow dashes) are shown.*

Concurrently, another screen was performed with a focus on covalent fragments to harness C18. Accordingly, 67,127 chloroacetamides were screened using CovDock; the bound fragments were later ranked based on their proposed capability to mimic the binding of fragment **1** and the pharmacophore of GDP. Chloroacetamide **3** was found to have preferential labelling of Kalirin-GEF(1)/Rac1 over Rac1 alone by intact MS, and thus was crystallised with Kalirin-GEF(1)/Rac1. The heterocyclic rings were found to be orientated in a manner comparable with GDP and electron density indicated that the compound was covalently bound to C18 (Figure 15).

Through a fragment merging approach, it was anticipated that the second hydrogen bonding interaction observed in the co-crystal structures of fragments **1** and **2** could be achieved by incorporating a hydrogen bond donor into the phenyl ring of **3**. It was envisioned that the presence of this interaction, as well as the presence of the appropriately situated chloroacetamide, would significantly boost the binding interactions within the pocket. This led to the design of **RS-009** (Scheme 1).

RS-009 was first synthesised by Dr. Raina Seupel from precursor **4**, the synthesis of which was outsourced. Co-crystallisation experiments confirmed that the fragment covalently bound to the nucleotide binding site of Kalirin-GEF(1)/Rac1 at C18 and the core scaffold shifted in comparison to chloroacetamide **3** to generate the second hydrogen bonding interaction with D118.



4 **RS-009**

Scheme 1 – *Synthesis of RS-009 performed by Raina Seupel. a) chloroacetyl chloride (1.1 eq.), TEA (2 eq.), DCM/DMF (0.49 M).*

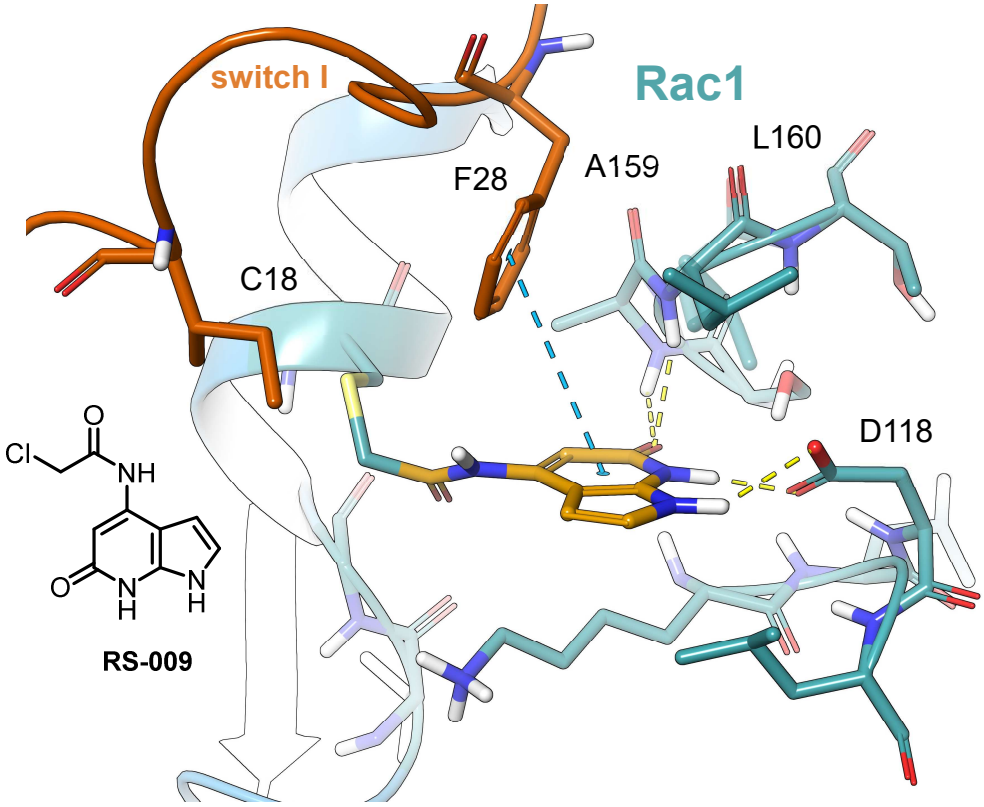


Figure 16 – *Binding of fragment RS-009 to the Kalirin-GEF(1)/Rac1 complex at the nucleotide binding site. RS-009 (yellow sticks) forms hydrogen bonding interactions with D118, A159 and L160 were observed and a pi-stacking interaction with F28. A covalent bond was observed with C18. Hydrogen bonding interactions (yellow dashes) are shown.*

To corroborate the binding ability of the fragments, a nucleotide exchange assay was developed in-house by Dr. Carmen Jimenez-Antunez. The fluorescence intensity (FI) association assay was established to identify the ability of compounds to inhibit the binding of the fluorophore-containing nucleotide analogue BODIPY-GTP. Upon binding to the GTPase, BODIPY-GTP exhibits an increased FI than in solution; inhibitors cause a slower initial increase of FI upon addition of BODIPY-GTP. Fragments **1–3** and **RS-009** were subsequently tested in this assay at 250 µM with a 2 h incubation. Pyridone **2** and **RS-009** were found to significantly inhibit BODIPY-GTP binding compared with the negative control, offering a starting point for an inhibitor of Rac1 activation via Kalirin-GEF(1)/Rac1.

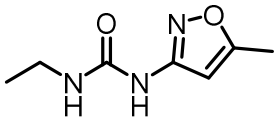
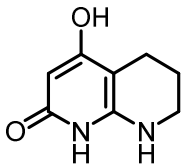
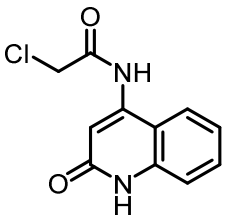
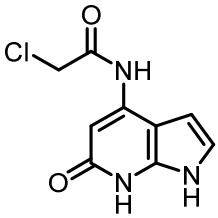
Compound	Structure	Activity at 250 µM
1		**-**
2		**Y**
3		**-**
RS-009		**Y**

Table 2 – *Activity of fragments identified from previous screens and rational design. All fragments tested at 250 µM by Janine Gray.*

1.6.8 Previous Kalirin-GEF(2)/RhoA progress

Compared to the progress for Kalirin-GEF(1)/Rac1, Kalirin-GEF(2)/RhoA was much less explored. RhoA is one of the most highly studied Rho GTPases and the structure is similar in fold to Rac1.[233] There is no known structure of the Kalirin-GEF(2) domain, nor that of the Kalirin-GEF(2)/RhoA co-crystal complex. Additionally, no fragments that bind to the complex have been reported in the literature.

1.6.9 Alternative Rho GEF/GTPase complexes as targets

In addition to Kalirin-based complexes, it was desirable to expand the number of Rho GEF/GTPase targets to be considered for modulation. In this effort the literature was explored for activation data pertaining to the Rho Dbl GEFs. Of the 70 GEF domains that were explored, 37% lacked any *in vitro* nucleotide exchange data to define their GEF activity and 81% lacked structural data to confirm the binding mechanism (Figure 17). In the absence of complete data, it was hypothesised that a comprehensive structure-based analysis of the Rho Dbl GEFs could aid in the prediction of complementary GEF/GTPase pairings and therefore new druggable targets. This would be of particular interest for GEFs that are associated with diseases *in vivo*, to offer an alternative strategy for probing related biological processes.

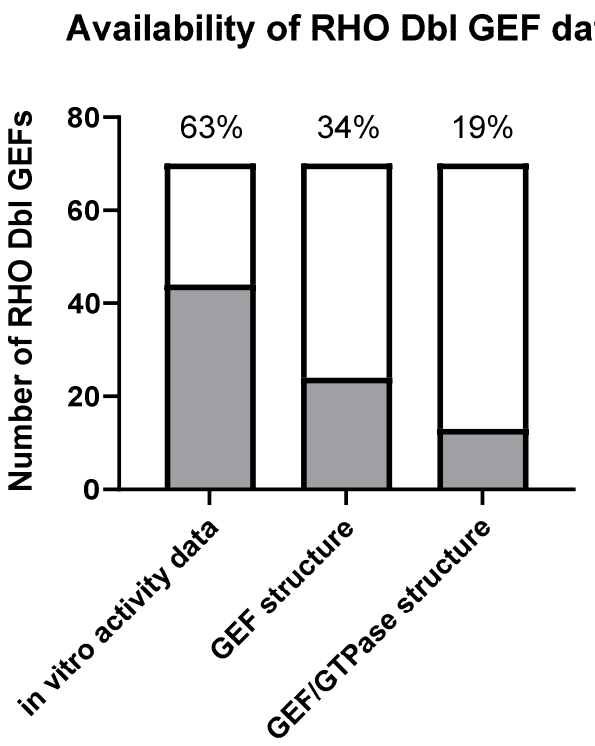


Figure 17 – *Availability of public data for the Rho Dbl GEFs (n = 70).*

1.7 Project Objectives

The aim of the work presented in this thesis was to investigate the approach of modulating Rho GTPase activity via the Rho GEF/GTPase complex. A deeper understanding of downregulating GTPase activity via these targets was necessary for the development of chemical probes for target validation. The focal points were the Kalirin-related complexes Kalirin-GEF(1)/Rac1 and Kalirin-GEF(2)/RhoA based on their associations with synaptic plasticity and their potential to be druggable targets for schizophrenia. The identification of other Rho GEF/GTPases was also explored in an attempt to broaden the number of targets that should be considered for this approach.

1.7.1 Kalirin-GEF(1)/Rac1

Fragment hits were previously identified for Kalirin-GEF(1)/Rac1 by taking a FBDD approach. Exploiting these, the objectives were as follows:

1.1. To quantitatively analyse the binding of previous fragments.

1.2. To probe the binding of fragments by designing, synthesising, and testing new analogues.

1.3. To corroborate the binding mechanism of active fragments.

1.4. To consider the synthetic accessibility of active fragments for the generation of analogues.

1.5. To structurally elaborate active fragments to improve potency and selectivity.

1.6. To explore the selectivity of active fragments.

1.7.2 Kalirin-GEF(2)/RhoA

At the time of writing this thesis, there are no known structures of the Kalirin-GEF(2) domain or the Kalirin-GEF(2)/RhoA complex. Additionally, no fragments that inhibit the complex were reported. Towards obtaining this information, the objectives were as follows:

2.1. To design Kalirin-GEF(2) constructs and perform their expression and purification.

2.2. To investigate the GEF activity of the produced constructs.

2.3. To obtain structural information of the Kalirin-GEF(2) domain and the Kalirin-GEF(2)/RhoA complex.

2.4. To explore the ligandability of the complex.

1.7.3 Predicting new Rho GEF/GTPase targets

Given the GEF activity for many Rho Dbl subfamily members is unknown, it was advantageous to generate a comprehensive understanding of Rho Dbl GEF activity to broadly understand the catalytic mechanisms of the protein family. Such a resource was theorised to aid the prediction of unreported GEF/GTPase complexes that could be targeted for disease. The objectives were as follows:

3.1. To determine an effective method of comparing the Rho Dbl GEFs at the molecular level.

3.2. To investigate the conserved moieties of the Rho Dbl GEFs and highlight retained motifs for activity, binding and Rho selectivity.

3.3. To apply the learnings to predict unknown GEF/GTPase partners.

3.4. To experimentally validate the prediction of GEF/GTPase partners and summarise the applicability of the approach.

2 Kalirin-GEF(1)/Rac1

This chapter will describe efforts to generate potent and selective inhibitors of Kalirin-GEF(1)/Rac1 and thoroughly explore their binding mechanisms. Previous hits reported by Dr. Janine Gray, **RS-009** and pyridone **2**, were exploited as starting points for this objective.

2.1 Quantitative analysis of previous hits

The first aim was to obtain quantitative measures of potency for hits **RS-009** and pyridone **2**, to understand the current potency of the inhibitors and for comparative measures. The compounds were previously tested in the FI nucleotide exchange assay described in Section 1.6.7, but only at a single concentration of 250 µM. It was also desirable to further corroborate the covalent binding mechanism of **RS-009**, in which the covalent bond was observed in the RS-009-Kalirin-GEF(1)/Rac1 crystal structure.

2.1.1 IC_{50}

The half-maximal inhibitory concentration (IC_{50}) of both fragments was calculated first. The dose-response curve for **RS-009** was reproducible and an IC_{50} of 11.9 ± 0.2 µM (incubation time = 2 h) was calculated (Figure 18). Contrastingly, the activity profile for pyridone **2** was poorly reproducible or irregular Hill Slopes were observed. This is often indicative of precipitation or aggregation in the protein system at high compound concentrations. The compound was also observed to precipitate easily from DMSO, and so precipitation under the assay conditions was likely. A conclusive IC_{50} value was not calculated for pyridone **2**, and the fragment was not explored further over the difficulties in obtaining bioactivity data.

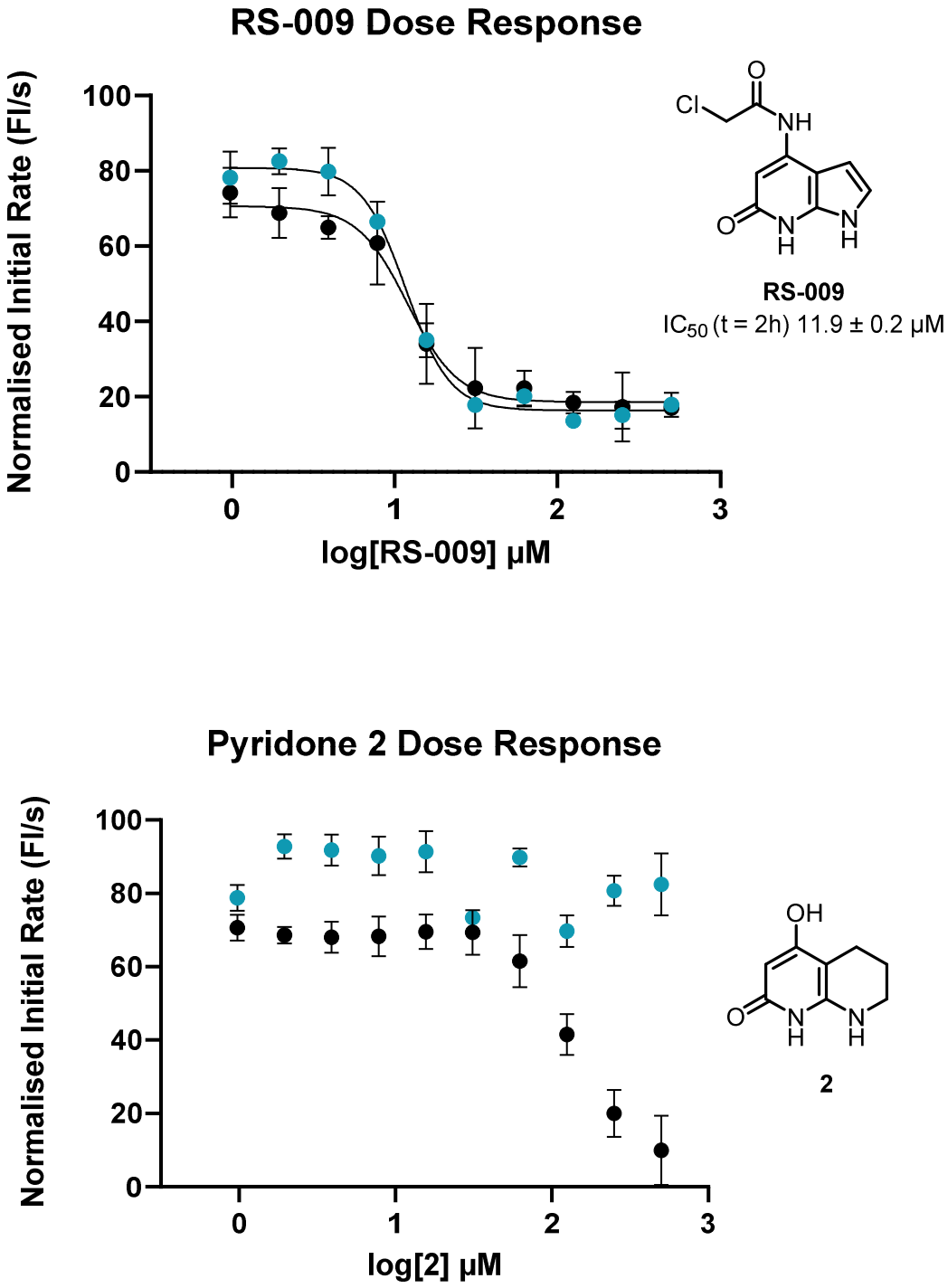


Figure 18 – *Dose response curves of RS-009 and pyridone 2. n = 1 (black) and n = 2 (blue) are shown.*

2.1.2 k_{inact}/K_i

The covalent nature of **RS-009** was subsequently corroborated. Whilst determining IC_{50} values are a quick approach for confirming hits and comparing activity at a specified incubation time, a more appropriate measure of comparison for covalent compounds is k_{inact}/K_i, which describes the rate of bond formation resulting from the potency of the first irreversible step (K_i) and the maximum potential rate of inactivation (k_{inact}).[234] This value is more applicable when comparing covalent compounds as both the reversible and irreversible steps are considered.

Krippendorf *et al* reported the ability to calculate k_{inact}/K_i using IC_{50} values obtained over varying incubation periods.[235] This data was obtained for **RS-009** and time-dependent inhibition was observed, corroborating the covalent binding mode (Figure 19).

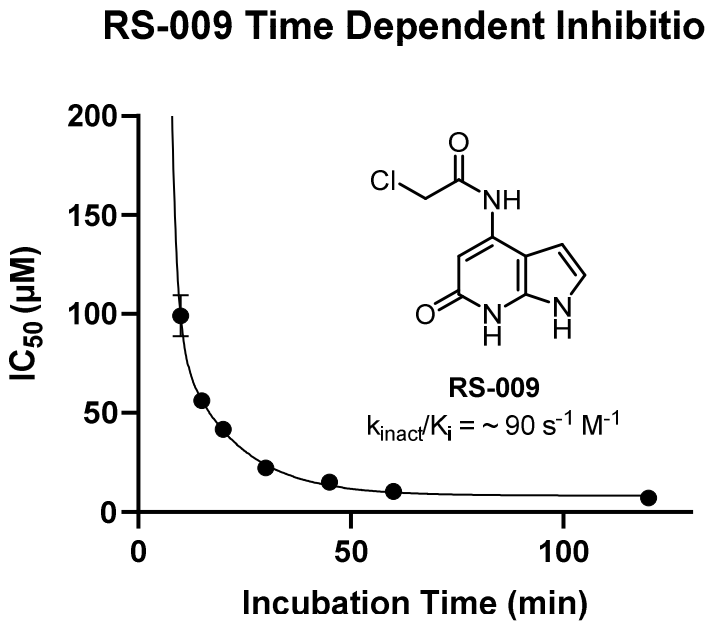


Figure 19 – *Time dependent inhibition of Kalirin-GEF(1)/Rac1 with RS-009. k_{inact}/K_i was calculated as described by Krippendorf et al.[235]*

The k_{inact}/Ki value of **RS-009** for Kalirin-GEF(1)/Rac1 was subsequently calculated with the equation reported by Krippendorf (Equation 1). Parameter S denotes the concentration of the substrate, which in this case is Rac1, and parameter K_M is the Michaelis constant. This value was experimentally determined previously by Carmen Jimenez Antunez.

Equation 1 – *Reported by Krippendorf for the derivation of K_i and k_{inact} directly from time-dependent IC_{50} values.*

$$IC_{50}(t) = K_i\left(1 + \frac{S}{K_M}\right) \bullet \left(\frac{2 - 2e^{-\eta IC_{50} \bullet k_{inact} \bullet t}}{\eta IC_{50} \bullet k_{inact} \bullet t} - 1\right) \text{ with}$$

$$\eta IC_{50} = \frac{IC_{50}(t)}{K_i\left(1 + \frac{S}{K_M}\right) + IC_{50}(t)}$$

where:
K_i = potency of the first reversible step
S = substrate concentration
K_M = Michaelis constant
k_{inact} = maximum potential rate of inactivation
t = time

Applying these values, the k_{inact}/Ki value was found to be approximately 10^6 fold lower than ibrutinib for BTK.[236] This was unsurprising given the small size of the fragment. In summary, the binding of **RS-009** was corroborated and quantitatively described by both IC_{50} and k_{inact}/K_i, which could be used to compare the activity of future analogues.

2.2 Core SAR analysis

2.2.1 Design of core analogues

The next objective was to improve the affinity of **RS-009** by first understanding the key binding interactions it makes within the nucleotide binding pocket. This was done by devising new target molecules to iteratively improve the binding through an understanding of the SAR of the series. An analysis of the co-crystal structure revealed that **RS-009** binds to Kalirin-GEF(1)/Rac1 by hydrogen bonding to the acidic side chain of D118 and the backbone of A159 and L160 and covalently bonding to C18 (Figure 20). Analogues were designed to modulate these interactions by changing the geometry and electronics of the fused ring system (Table 3).

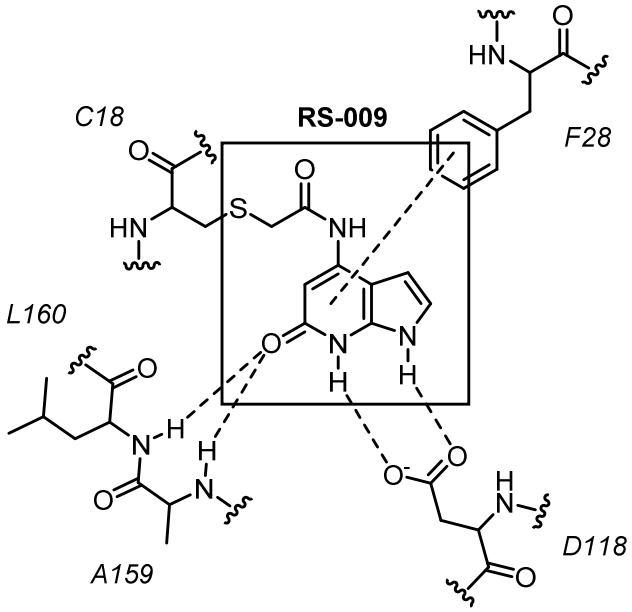


Figure 20 – *Binding interactions of RS-009 within the nucleotide binding pocket of Kalirin-GEF(1)/Rac1.*

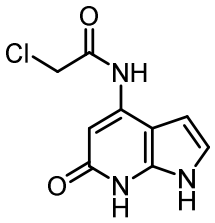
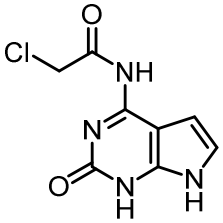
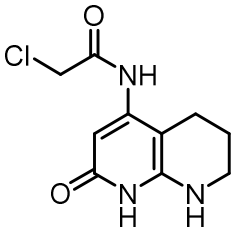
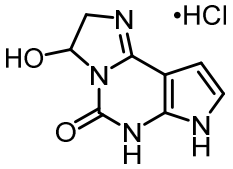
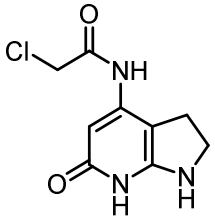
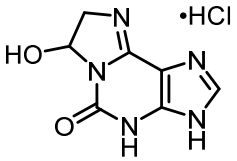
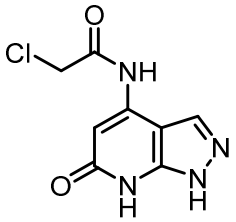
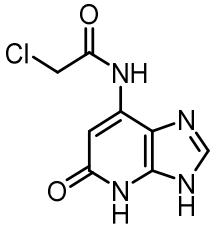
Compound	Structure	Compound	Structure
RS-009		**8**	
MC-278		**9**	
5		**10**	
6			
7			

Table 3 – *Designed analogues of RS-009.*

The piperidinopyridone analogue, **MC-278**, was inspired by previous hits pyridone **2** and **RS-009** (Figure 21). A fragment merging approach was expected to enhance potency by tuning the D118 interaction and engaging C18 through a covalent bond.

Pyrrolidine **5**, pyrazolopyridone **6**, and imidazolopyridone **7** were also designed with this purpose.

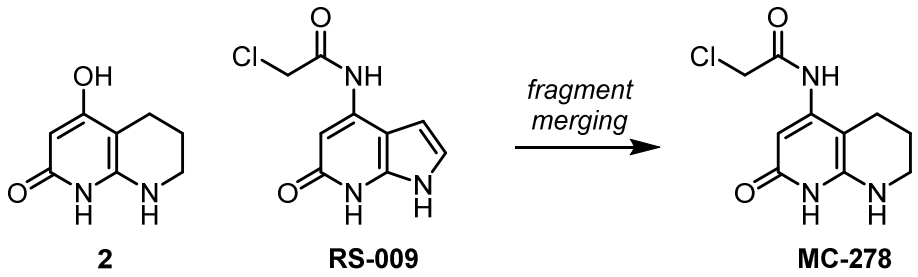
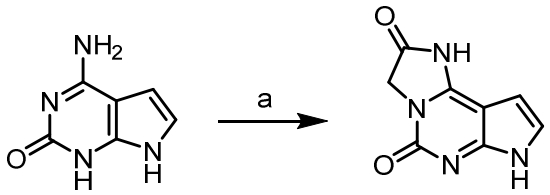


Figure 21 – *Design of MC-278 using a fragment merging approach from hits identified by Dr. Janine Gray.*

To probe the interactions with residues C18, A159 and L160, analogues of the pyridone core were considered. Dr. Mary Wilson, a rotation student, previously attempted to synthesise pyrrolopyrimidone **8** to explore the addition of a nitrogen atom at C5; this resulted in the intramolecular cyclised product, in which the pyrimidone reacted with the chloroacetamide warhead (Scheme 2).



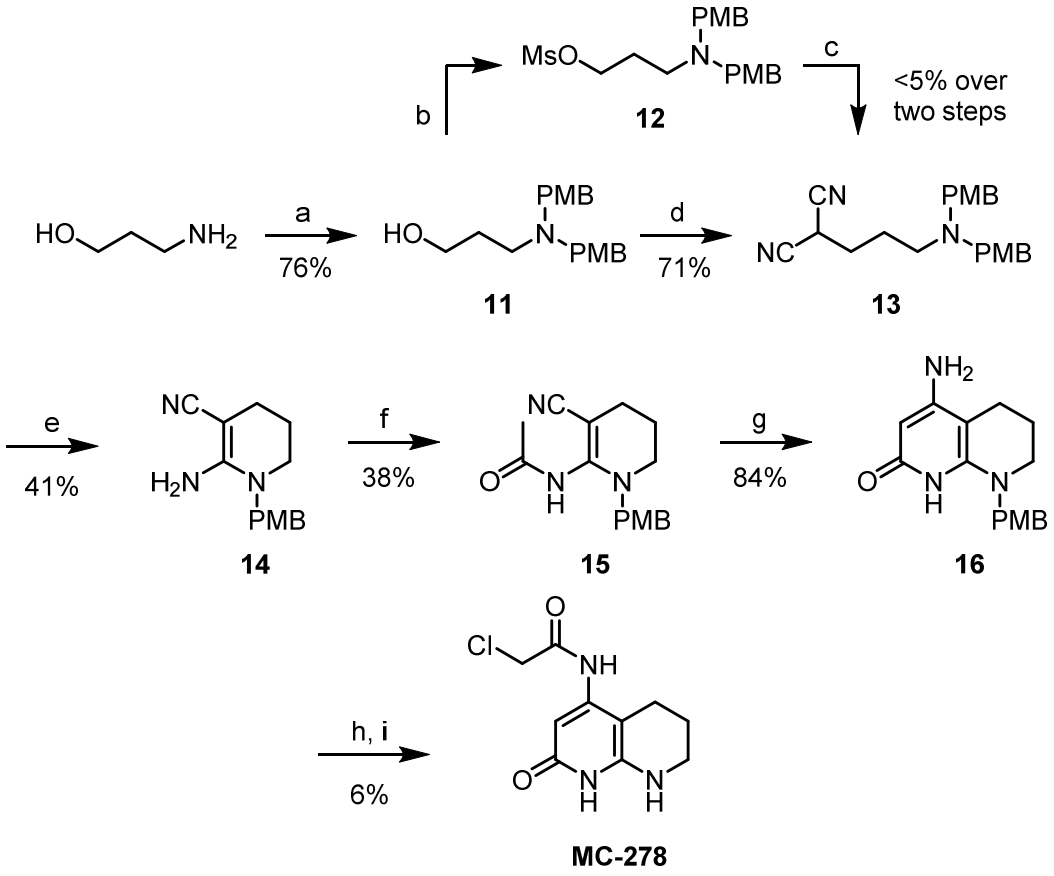
Scheme 2 – *Side product observed in the attempted synthesis of pyrrolopyrimidine **8**, reported by Dr. Mary Wilson. a) chloroacetyl chloride, DIPEA, DCM, 0 °C, 1 h.*

To circumvent this issue, hemi-aminal compounds **9** and **10** were designed. The cyclised warheads were reported to be transformed to the corresponding aldehyde in pH 7.5 assay buffer, and were subsequently cysteine reactive.[237] In total, six new

fragments were designed. The analogues were predicted to have analogous binding to **RS-009** by molecular docking.

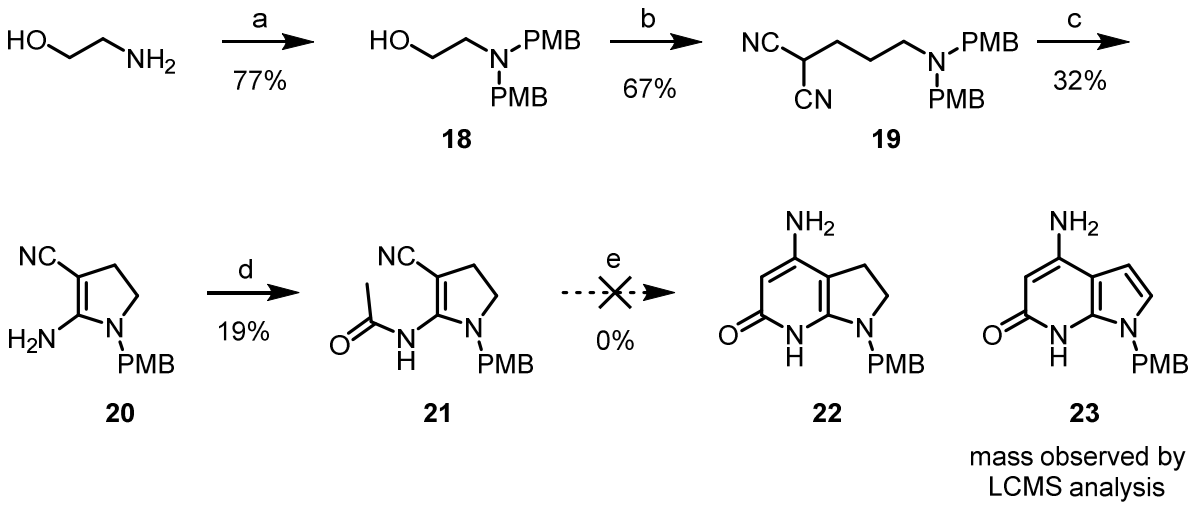
2.2.2 Synthesis of core analogues

Accessing **MC-278** proved to be a multi-step synthesis despite the small size of the molecule (Scheme 3). The novel piperidinopyridone core was built from simple linear starting material 3-aminopropanol. The amino group was subsequently protected with a PMB group to yield alcohol **11**. A malononitrile group was initially introduced through mesylate **12** followed by alkylation to access nitrile **13**. However, the yield over two steps was limiting. Accordingly a Mitsunobu reaction was implemented, which removed a step in the synthesis and significantly increased the yield. The piperidine ring was then constructed using 1-chloroethyl chloroformate (ACE-Cl), which allowed for a mono-selective PMB-deprotection followed by intramolecular cyclisation to form amino piperidine **14**. The product was acetylated to give amide **15**, and finally ring closure under basic conditions generated piperidinopyridone **16**. The covalent warhead was introduced with chloroacetyl chloride followed by a PMB deprotection to yield **MC-278**.



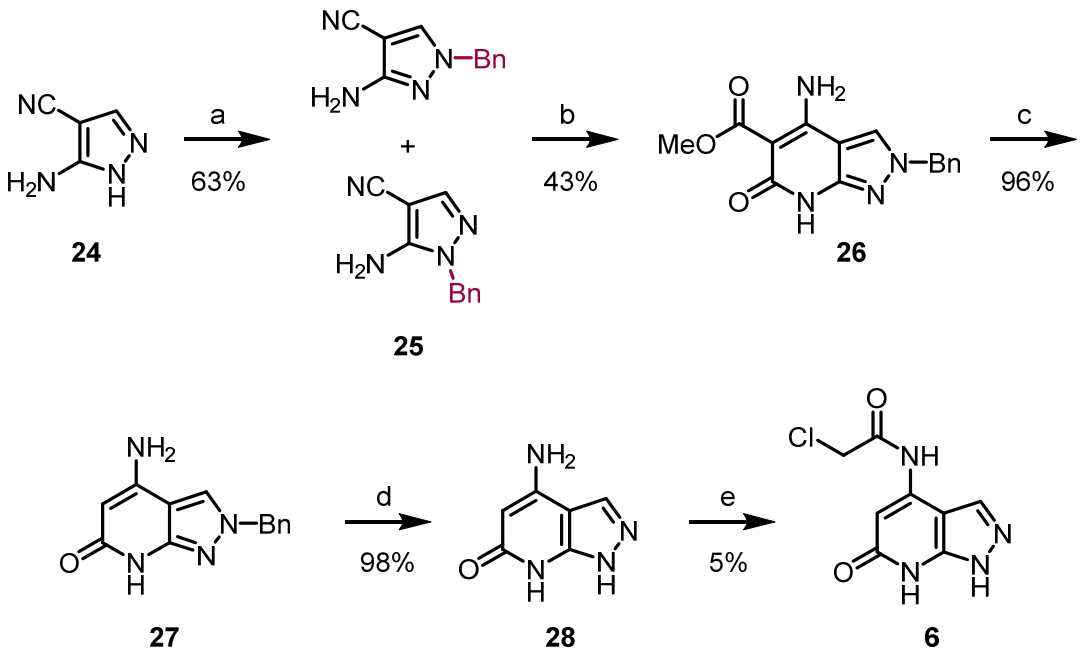
Scheme 3 – *Synthesis of MC-278. a) PMBCl (2 eq.), K_2CO_3 (4 eq.), MeCN (0.67 M), 50 °C, 16 h. b) MsCl (2 eq.), DIPEA (5 eq.), DMAP (0.1 eq.), DCM (0.44 M), 0 °C–r.t., 16 h. c) malononitrile (1.5 eq.), DIPEA (2 eq.), MeCN (0.2 M), 60 °C, 3 days. d) DIAD (1 eq.), PPh_3 (1 eq.), malononitrile (0.8 eq.), THF (0.22 M), 0 °C–r.t., 18 h. e) ACE-Cl (4 eq.), DCM (0.07 M), reflux, 16 h, then MeOH (0.07 M), reflux, 16 h. f) Ac_2O (1.5 eq.), MeCN (1.09 M), 45 °C, 1 h. g) LiHMDS (6 eq.), THF (0.13 M), 65 °C, 5 h. h) TEA (2 eq.), chloroacetyl chloride (4 eq.) DCM (0.13 M), r.t., 1 h. i) TFA (0.04 M), r.t., 2 h.*

The same strategy was used in the attempted synthesis of pyrrolidinopyridone **5** albeit beginning with ethanolamine to produce intermediates **18-21** (Scheme 4). However, the attempted ring closure to form pyrrolidinopyridone **22** resulted in a complex mixture of products. This included the unsaturated product **23** by liquid chromatography mass spectrometry (LCMS) analysis, suggesting the desired heterocycle was not stable under the conditions required for ring closure. The synthesis of pyrrolidinopyridone **5** was ultimately abandoned due to stability issues.



Scheme 4 – *Attempted synthesis of 5. a) PMBCl (2 eq.), K_2CO_3 (4 eq.), MeCN (0.67 M), 50 °C, 16 h. b) DIAD (1 eq.), PPh_3 (1 eq.), malononitrile (0.8 eq.), THF (0.22 M), 0 °C–r.t., 18 h. c) ACE-Cl (4 eq.), DCM (0.07 M), reflux, 16 h, then MeOH (0.07 M), reflux, 16 h. d) Ac_2O (1.5 eq.), MeCN (1.09 M), 45 °C, 1 h. e) LiHMDS (6 eq.), THF (0.13 M), 65 °C, 5 h.*

Pyrazolopyridone **6** was synthesised in five steps from readily available starting materials (Scheme 5). Aminopyrazole **24** was first benzyl protected to yield **25** as a mixture of benzylated products that was not purified further. Condensation with dimethyl malonate successfully produced methyl ester **26**, followed by a one-pot hydrolysis/decarboxylation to give pyrazolopyridone **27**. A benzyl deprotection was then performed under acidic conditions to form pyrazole **28**, and the covalent warhead was incorporated to access the final desired pyrazolopyridone **6**.



Scheme 5 – *Synthesis of 6. a) BnBr (1.2 eq.), K_2CO_3 (1.5 eq.), DMF (0.93 M), r.t., 21 h. b) methyl malonate (1.3 eq.), NaOMe (1.3 eq.), MeOH (1.11 M), reflux, 2 days. c) NaOH (15% aq), reflux, 19 h. d) H_2, Pd/C (50% wetted, 0.8 eq.), HCl (4 M in dioxane, 0.21 M), MeOH (0.04 M), r.t., 3 days. e) TEA (2 eq.), chloroacetyl chloride (4 eq.), DCM/DMF (1:1, 0.56 M), r.t., 1 h.*

Imidazolopyridone **7** was expected to be synthesised employing the equivalent route to pyrazolopyridone **6**. However, the synthesis was abandoned during attempts. It was hypothesised that the imidazolo N2 was likely to react with the final chloroacetamide warhead to give the cyclised product (Figure 22). Due to these concerns, the synthesis of imidazolopyridone **7** was deprioritised.

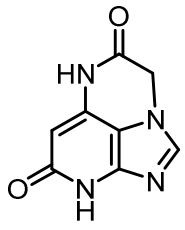
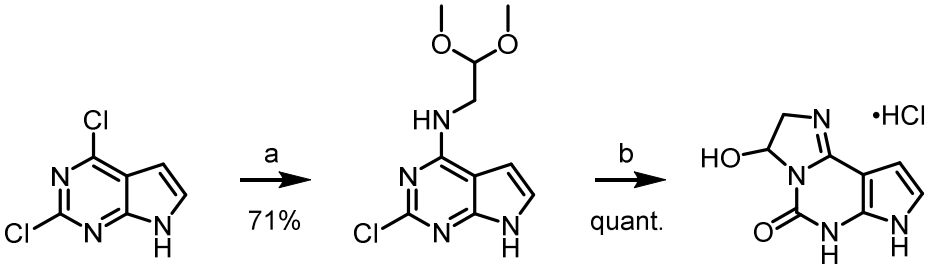


Figure 22 – *Side product predicted to form during the attempted synthesis of imidazolopyridone 7.*

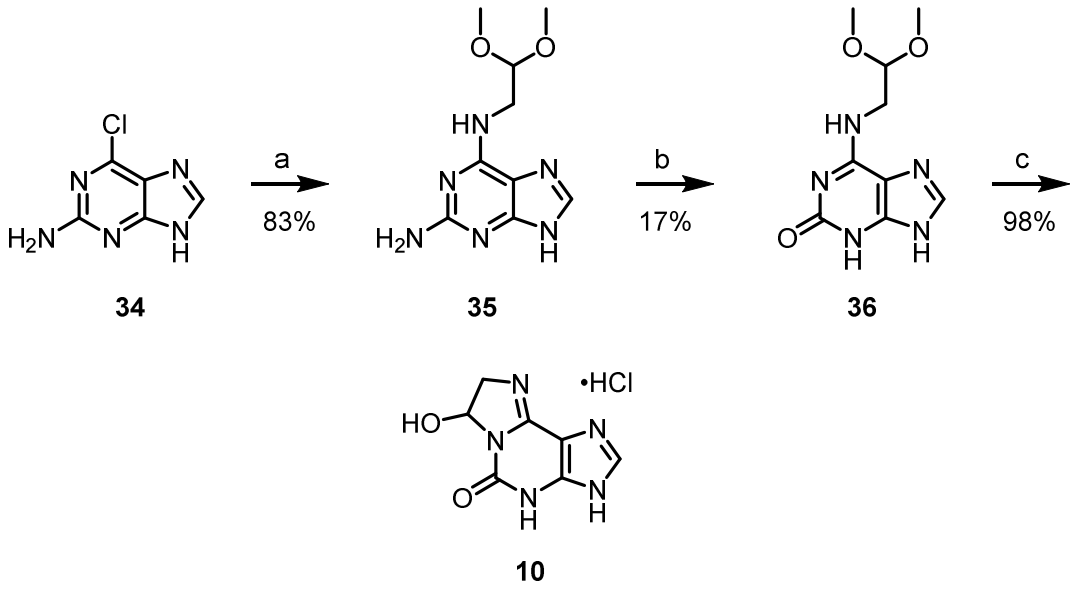
Hemi-aminals **9** and **10** were more easily accessed in 2–3 steps. Dichloropyrimidine **32** first underwent an S_NAr reaction to yield acetal **33**, and upon exposure to an aqueous solution of HCl at high temperature, a one-pot hydrolysis, deprotection and cyclisation sequence occurred to give the desired product **9** (Scheme 6).



32 **33** **9**

Scheme 6 – *Synthesis of 9. a) aminoacetaldehyde dimethyl acetal (3 eq.), TEA (1.2 eq.), n-BuOH (0.23 M), 70 °C, 4 h. b) HCl (1.0 M), reflux, 24 h.*

A similar approach was taken for hemi-aminal **10**, beginning with the analogous S_NAr approach from aminochloropyrimidine **34** to yield acetal **35**. An extra step was necessary to introduce the pyrimidone moiety through a diazotisation/hydrolysis reaction to yield pyrimidone **36**, and finally the intermediate was exposed to acid to form the desired hemi-aminal **10** (Scheme 7). In summary, four of the six designed fragments were successfully synthesised.



Scheme 7 – *Synthesis of 10. a) aminoacetylaldehyde dimethyl acetal (3 eq.), TEA (1.2 eq.), n-BuOH (0.23 M), 70 °C, 24 h. b) $NaNO_2$ (3 eq.), AcOH (3.6 M), water (0.74 M), 0 °C, 1 h. c) HCl (1 M), reflux, 24 h.*

2.2.3 Activity of core analogues

The synthesised fragments were tested for activity in the FI nucleotide exchange assay at 500 µM (Table 4). **MC-278** was found to inhibit BODIPY-GTP binding at this concentration, suggesting the piperidine ring system retained the hydrogen bonding interaction with D118. The calculated IC_{50} (incubation time = 2 h) was 10.8 ± 1.2 µM, a comparable value to **RS-009** (Figure 23).

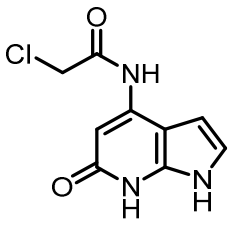
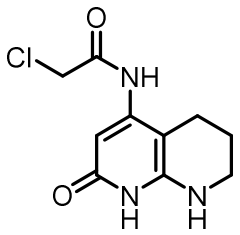
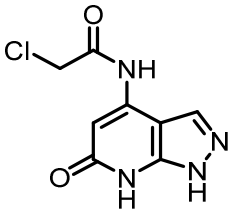
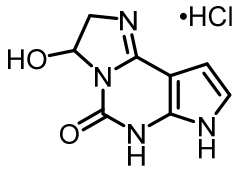
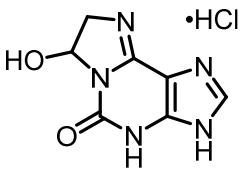
Compound	Structure	IC_{50} (µM)
RS-009		11.9 ± 0.2
MC-278		10.8 ± 1.2
7		> 500
9		> 500
10		> 500

Table 4 – *Inhibition of BODIPY-GTP binding, using the FI nucleotide exchange assay.*

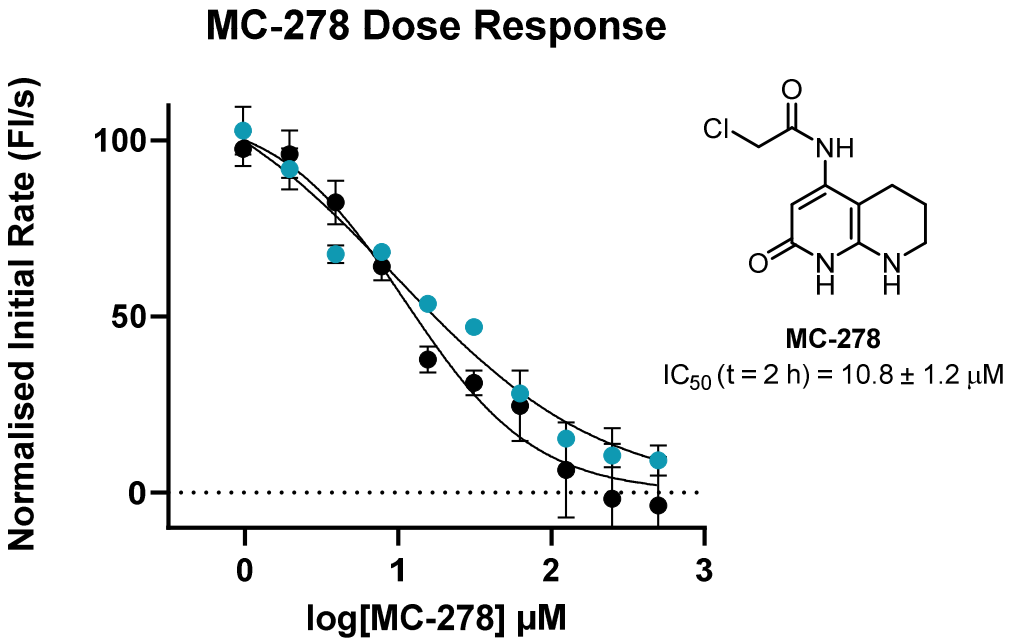
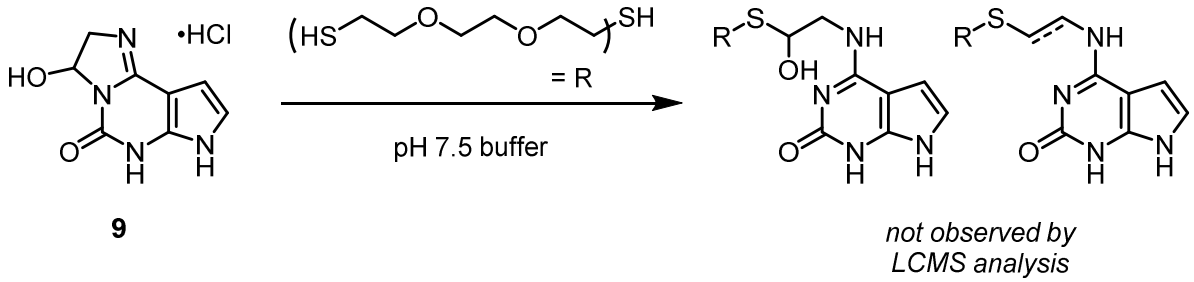


Figure 23 – *Dose response of MC-278 against Kalirin-GEF(1)/Rac1. n = 1 (black) and n = 2 (blue) are shown.*

The remaining fragments showed no significant activity at 500 µM. The inactivity of pyrazolopyridone **7** was surprising, as it was expected that the pyrazole would be an appropriate hydrogen bond donor to interact with D118. The lack of activity indicates the importance of tuning this specific interaction; changes in tautomerisation and the distribution of electron density of the heteroaromatic ring system should be considered.

Hemi-aminals **9** and **10** were also inactive. Three hypotheses were made for this result: the fragments were not cysteine reactive under the assay conditions, the warhead was poorly situated from C18 to covalently bind, or the cores themselves were poorly binding. To test the former, hemi-aminal **9** was exposed to thiol 2,2′-(ethylenedioxy)diethanethiol (EDT) under the assay conditions. By LCMS analysis no reaction was observed, even at prolonged stir times. It was concluded that the

hemiaminal was poorly active towards thiols, and thus may not have been able to form the covalent bond with C18 of Rac1. This did not, however, deem the alternative theories as false.



Scheme 8 – *Test reaction between hemi-aminal 9 and EDT under the assay conditions.*

2.2.4 Binding mode of MC-278

The binding mechanism of **MC-278** was first confirmed by calculating the k_{inact}/K_i value (Figure 24). The IC_{50} was found to be time-dependent, which supports the notion that the fragment covalently interacts with Kalirin-GEF(1)/Rac1. A crystal structure of the complex was also obtained at a resolution of 1.48 Å (Figure 24). As expected, the heterocyclic core of **MC-278** overlaid well with that of **RS-009** and the dual hydrogen bonding interaction with D118 was retained. Electron density was also observed between the acetamide moiety of **MC-278** and C18 in high occupancy, a characteristic of covalent bond formation. The binding mode of **MC-278** was confidently confirmed with these experiments.

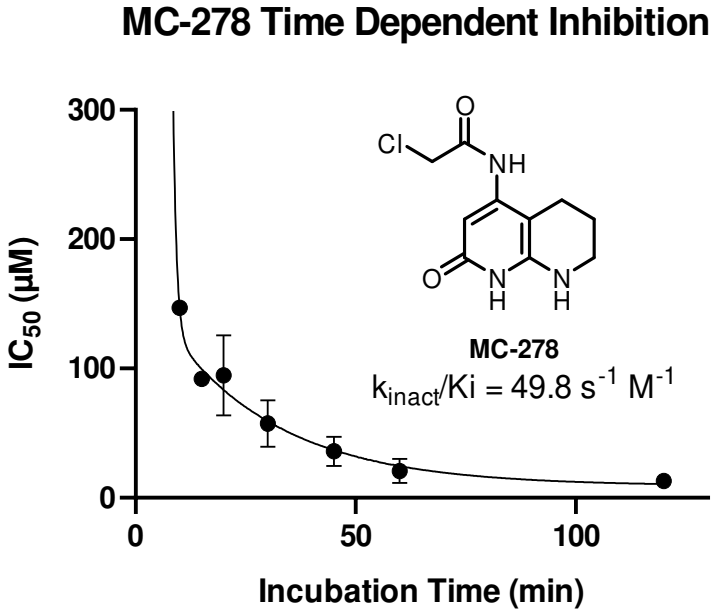


Figure 24 – *Time dependent inhibition of Kalirin-GEF(1)/Rac1 by MC-278.*

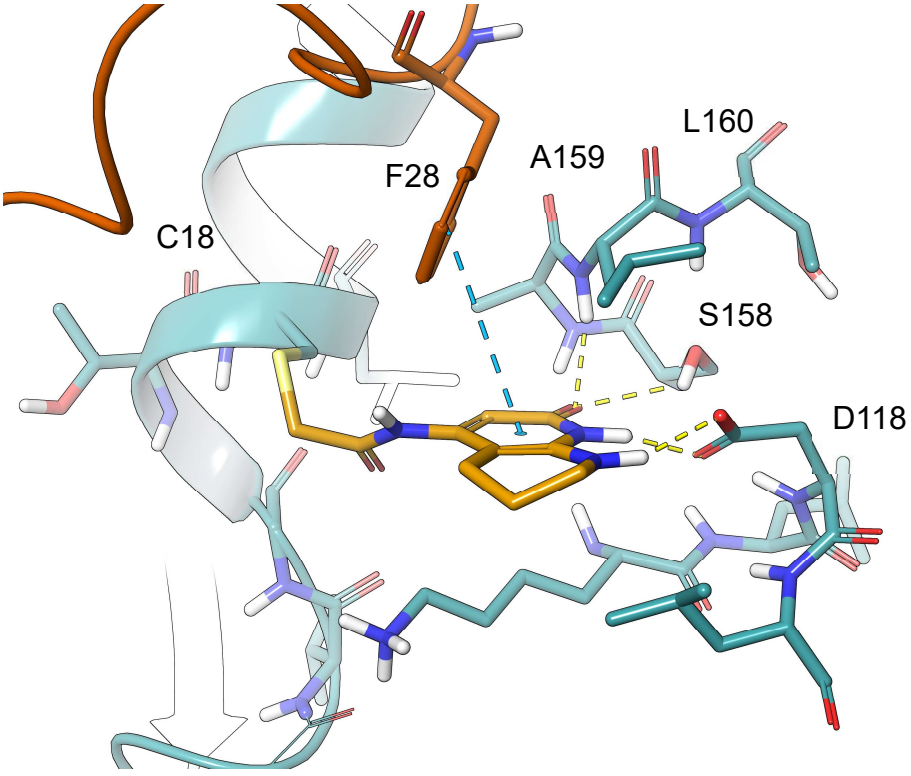


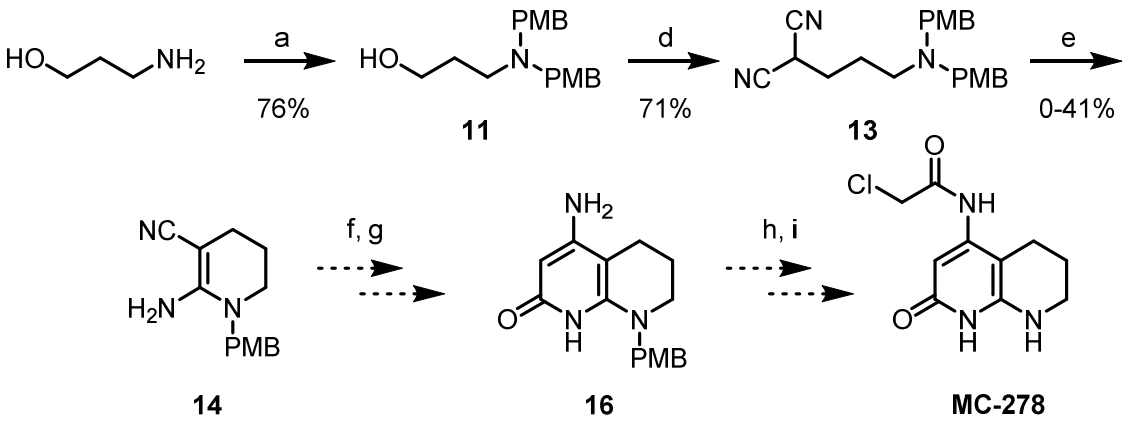
Figure 25 – *Crystal structure of MC-278 (yellow sticks) bound to the nucleotide binding pocket of Kalirin-GEF(1)/Rac1. The switch I region (orange sticks) is shown.*

2.3 Warhead SAR analysis

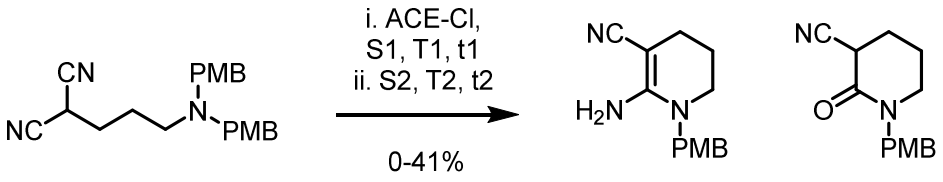
The lack of activity of hemi-aminals **9** and **10** led to a further investigation into the influence of the covalent interaction with C18. It was expected that a variety of non-covalent and covalent appendages at the chloroacetamide vector could be synthesised and the SAR of the series analysed. To synthesise these analogues a larger quantity of advanced intermediates, such as intermediate **16**, was required. This led to an investigation into scalable in-house syntheses for **MC-278** and **RS-009**.

2.3.1 Scalable synthesis of MC-278

The original route toward **MC-278** was initially repeated on a larger scale to access precursors for analogue generation (Scheme 9). Whilst the steps a–d were repeatable in good yield, problems quickly arose with the reproducibility of the mono-selective deprotection of **13** using ACE-Cl, in which the reaction would unpredictably fail and produce lactam **38** in varying yield (Table 5).



Scheme 9 – *Attempted scale up of the original synthesis of MC-278. a) PMBCl (2 eq.), K_2CO_3 (4 eq.), MeCN (0.67 M), 50 °C, 16 h. d) DIAD (1 eq.), PPh_3 (1 eq.), malononitrile (0.8 eq.), THF (0.22 M), 0 °C–r.t., 18 h. e) ACE-Cl (4 eq.), DCM (0.07 M), reflux, 16 h, then MeOH (0.07 M), reflux, 16 h. f) Ac_2O (1.5 eq.), MeCN (1.09 M), 45 °C, 1 h. g) LiHMDS (6 eq.), THF (0.13 M), 65 °C, 5 h. h) TEA (2 eq.), chloroacetyl chloride (4 eq.) DCM (0.13 M), r.t., 1 h. i) TFA (0.04 M), r.t., 1 h.*



13 **14** **38**

Entry	S1	T1 (°C)	t1 (h)	S2	T2 (°C)	t2 (h)	38	Yield
1	DCE	reflux	12	MeOH	reflux	20 h	Y	0-41%
2	DCE	reflux	12	MeOH	r.t.	20 h	Y	6%
3	DCM	r.t.	12	MeOH	r.t.	2 h	Y	13%
4	DCM	r.t.	3	iPrOH	r.t.	12 h	N	25%
4 (n = 2)	DCM	r.t.	3	iPrOH	r.t.	12 h	N	10%
4 (n = 3)	DCM	r.t.	3	iPrOH	r.t.	12 h	N	38%
4 (n = 4)	DCM	r.t.	3	iPrOH	r.t.	12 h	N	0%

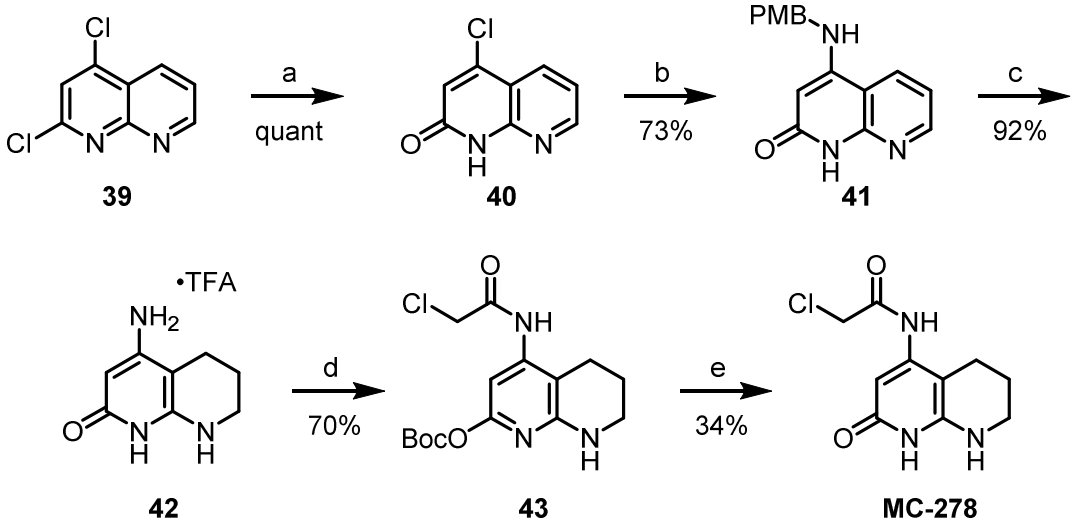
Table 5 – *Selected entries for the optimisation of step e performed by Anya Robinson. S = solvent, T = temp, t = time.*

At this time an undergraduate student, Anya Robinson, explored the poor reproducibility by making iterative changes to the original conditions (Entry 1). The reaction involves exposing intermediate **13** to ACE-Cl in DCE (stage 1), followed by a solvent swap to MeOH to encourage intramolecular cyclisation to form **14** (stage 2). It was first postulated that exposure to moisture increased the formation of undesired lactam **38** via a hydrolysis of enamine **14**. To reduce this exposure, the temperature was lowered of stage 2 (Entry 2), and the yield was improved after changing the solvent of stage 1 to DCM and lowering the temperature concurrently (Entry 3). Despite these changes, lactam **38** remained a major side product. Eventually, a

solvent swap from MeOH to iPrOH was found to completely abolish the formation of this side product (Entry 4). Despite this improvement, the reaction continued to behave unpredictably, and upwards of 20 g of intermediate **13** was lost when attempting to proceed through the synthesis.

Whilst the synthesis did produce a limited amount of intermediate **16**, a robust route was still required. The second route toward **MC-278** took advantage of a related fused heterocycle, naphthyridine **39**, which was first exposed to acidic conditions to form naphthyridone **40** (Scheme 10). An S_NAr reaction was then performed with 4-methoxybenzylamine to yield aminonaphthyridone **41**. The key step in the synthesis was a high-pressure flow hydrogenation that selectively reduced the pyridine ring to access the novel piperidinopyridone **42**; this reaction did not occur under standard hydrogenation conditions. The PMB group was only partially cleaved during this step, requiring the addition of TFA to remove the protecting group completely. The warhead was then introduced via an *in-situ* Boc protection of the pyridone moiety followed by an acylation to produce Boc-protected intermediate **43**, which underwent a Boc deprotection to yield **MC-278**.

This route was a vast improvement on the previous; there were a reduced number of steps, a greater overall yield, and flow chemistry was utilised, all of which are appealing for scale up. Additionally, column chromatography was not required for the steps a–c, that too improving the scalability. The route was used to generate a large batch of the original fragment and intermediates for analogue synthesis. Based on this success, efforts to demonstrate a robust synthesis toward the **RS-009** core were deprioritised.



Scheme 10 – *Improved synthesis of MC-278. a) 6M HCl (0.5 M), reflux, 21 h. b) 4-methoxybenzylamine (16 eq.), DMSO (0.22 M), 150 °C, 1 h. c) H-Cube, H_2, Pd/C, MeOH (0.18 M), formic acid (0.07 M), 40 °C, 20 bar, 0.5 mL/min, then TFA (0.29 M), 60 °C, 24 h. d) TEA (4 eq.), Boc_2O (1 eq.), chloroacetyl chloride (1 eq.), DMF (0.2 M), 0 °C–r.t., 5 h. e) 1-methyl piperidine (1.2 eq.), DCM (0.2 M), r.t., 1.5 h.*

2.3.2 Design of non-covalent analogues

Non-covalent analogues of **MC-278** were explored first. In this effort, a library was enumerated by reaction in the Maestro suite (Schrödinger). Two potential linkers were explored, one retaining the amide moiety, and another exploiting a more flexible N-alkyl chain. Over 30,000 compounds were subsequently docked and scored, and for both series, it appeared that a wide range of groups would be tolerated with respect to steric hindrance; this was attractive for fragment elaboration, given the space available in the binding pocket at this vector (Figure 26). Non-covalent compounds **45–58** were eventually chosen based on their diversity and potential to make new interactions within the binding pocket (Figure 27). For example, ether **56** was predicted to form a hydrogen bond with K116, and smaller hydrophobic fragments such as **53** and **54** were expected to interact favourably with I33. Acetamide **46** represented the closest analogue to **MC-278**, albeit with the removal of the electrophile.

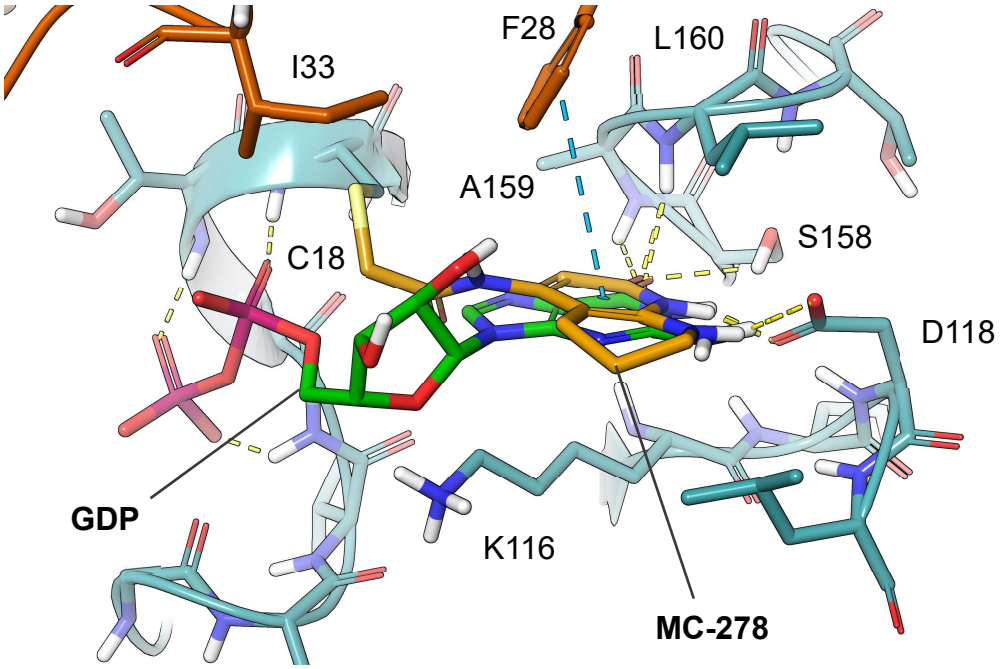
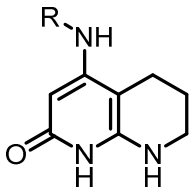


Figure 26 – *Crystal structure of MC-278 (yellow sticks) and GDP (green) bound to Kalirin-GEF(1)/Rac1 (teal ribbons). The switch I region is shown (orange sticks).*



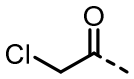
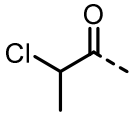
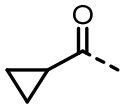
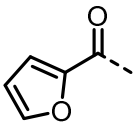
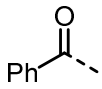
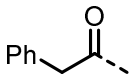
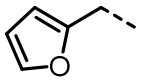
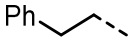
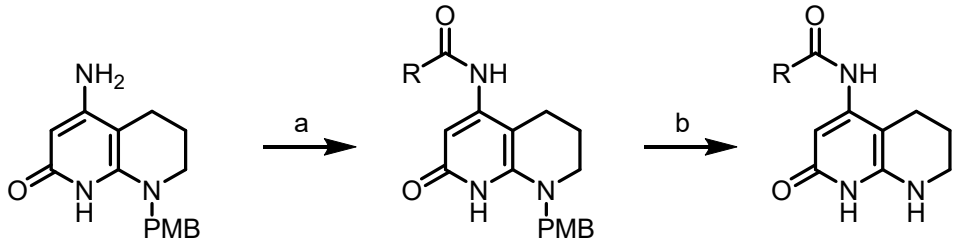
Compound	Structure	Compound	Structure
MC-278		**51**	
45		**52**	
46		**53**	
47		**54**	
48		**55**	
49		**56**	
50		**57**	
		58	

Figure 27 – *MC-278 analogues differing at the chloroacetamide vector enumerated by reaction in Maestro, Schrödinger.*

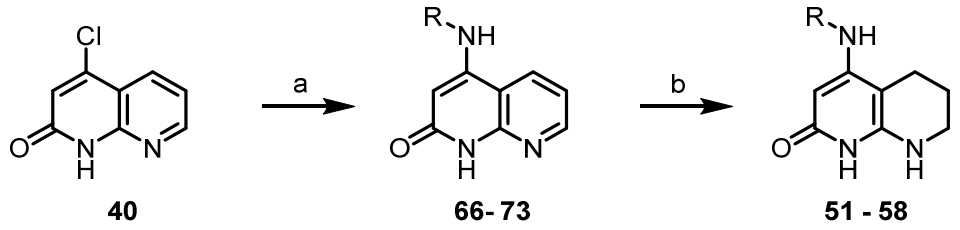
2.3.3 Synthesis of non-covalent analogues

Amide fragments **45–50** were synthesised by exposing piperidinopyridone **16** to the corresponding acid chloride to synthesise the PMB protected compounds **59–64**, followed by PMB cleavage in TFA to give the desired products (Scheme 11).



Scheme 11 – *Synthesis of amides 45–50. a) RCOCl (4 eq.), TEA (4 eq.), DCM (0.2 M), r.t., 6–24 h. b) TFA (0.1 M), reflux, 3–5 h.*

N-alkyl substituents **51**–**58** were synthesised by Anya Robinson using the second route toward the **MC-278** core (Scheme 12). S_NAr reactions of naphthyridone **40** were performed in the presence of the desired amine to yield naphthyridones **66–73**, followed by selective reduction under the flow hydrogenation conditions to access the desired products **51–58**.



Scheme 12 – *Synthesis of N-alkyls 51–58. a) NH_2R, DMSO (0.2 M), 50–150 °C, 4–6 h. b) H-Cube, H_2, Pd/C, MeOH (0.18 M), formic acid (0.07 M), 40 °C, 20 bar, 0.5 mL/min*

2.3.4 Activity of non-covalent analogues

The synthesised fragments were subsequently tested in the FI nucleotide exchange assay (Figure 28). In quick summary, none of the fragments showed significant activity at 500 µM. The most likely hypothesis for this result is that BODIPY-GTP was able to outcompete the bound fragments readily, which does not occur with a covalent bond between the fragment and the complex. The newly appended functional groups were

unable to compensate for the loss of activity arising from this covalent interaction. Whilst no new active fragments were identified, the influence of the chloroacetamide warhead was clearly demonstrated.

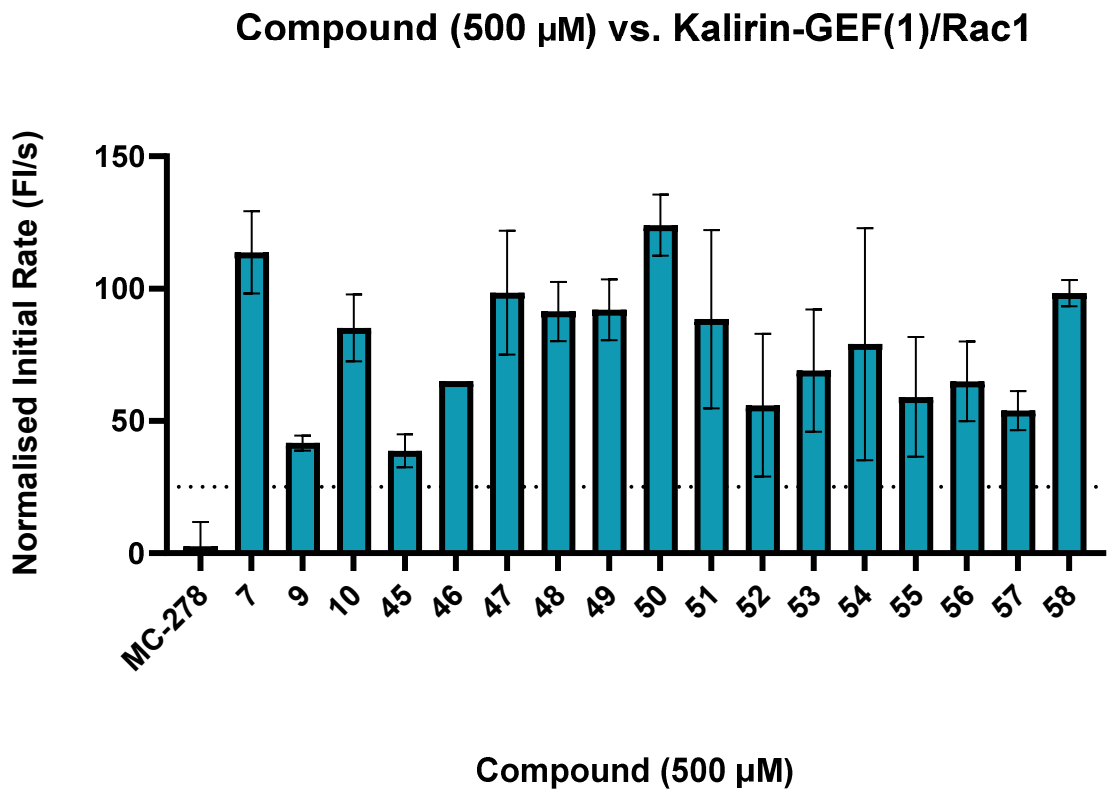
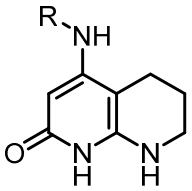


Figure 28 *– Normalised initial rates of BODIPY-GTP binding to Kalirin-GEF(1)/Rac1 in the presence of compounds (500 μM). Compounds were tested n = 2 and are shown as the mean and standard deviation. Inhibition was considered significant at < 25 FI/s.*

2.3.5 Design of covalent analogues

To explore alternative covalent interactions, analogues **74–78** were designed using cysteine reactive warheads (Table 6).[21] The generation of acrylamides was particularly interesting, as the electrophilic centre of attack was spatially translated compared with the chloroacetamide. Once again, all analogues were predicted to have comparative

binding to **MC-278** by molecular docking when comparing the location and orientation of the core scaffold.



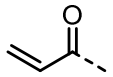
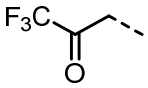
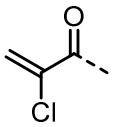
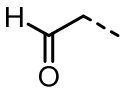
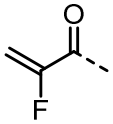
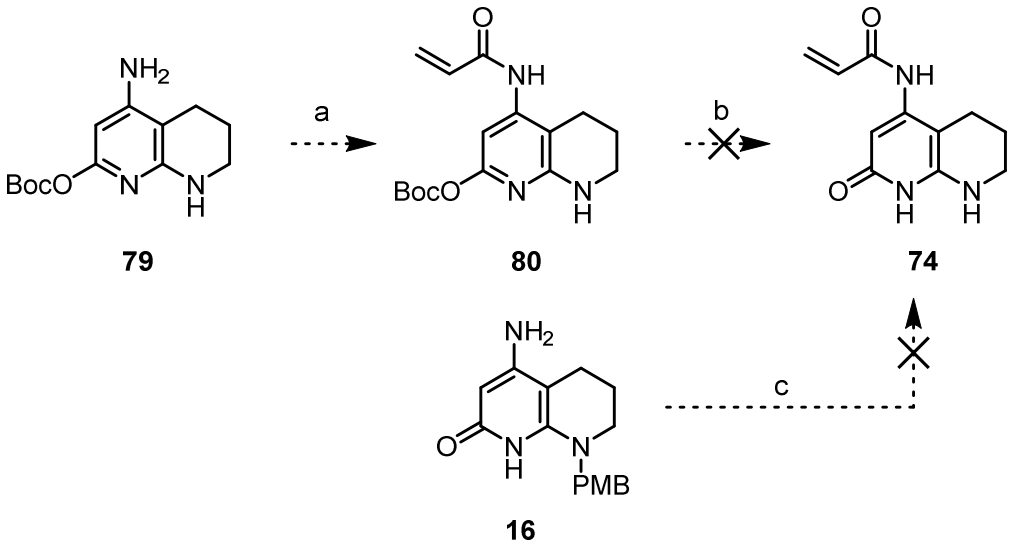
Compound	Structure	Compound	Structure
74		**77**	
75		**78**	
76			

Table 6 – *Covalent analogues of MC-278.*

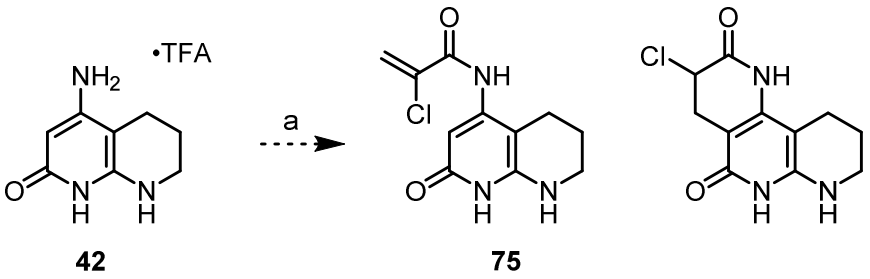
2.3.6 Synthesis of covalent analogues

In an attempt to access acrylamide **74**, Boc protected intermediate **79** was added acryloyl chloride which resulted in a mixture of **80** and other regioisomers that could not be separated; this mixture was subjected to Boc deprotection conditions and separation was attempted once again (Scheme 13). The result was a mixture of products not limited to the expected regioisomers. A complex mixture of products was also observed when attempting the access acrylamide **74** through intermediate **16**, despite attempting to selectively protect the pyridone prior to the introduction of the warhead.



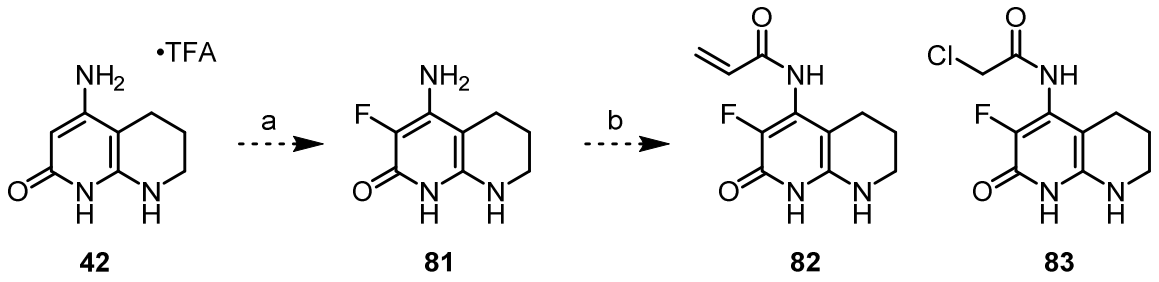
Scheme 13 – *Synthesis of 74. a) acryloyl chloride (1 eq.), TEA (2 eq.), DMF (0.2 M), 0 °C–r.t. b) 1-methylpiperazine (2 eq.), DCM (0.2 M), r.t. c) acryloyl chloride (3 eq.), TEA (2 eq.), DCM (0.2 M), r.t. then TFA (0.04 M), r.t.*

This side product formation was persistent when attempting to access covalent fragments. For example, an amide coupling approach was used in an attempt to access chloroacrylamide **75**, which resulted in products that had undergone intramolecular cyclisation by NMR analysis (Scheme 14). Attempts to synthesise fluoroacrylamide **76** were abandoned due to these accessibility issues.



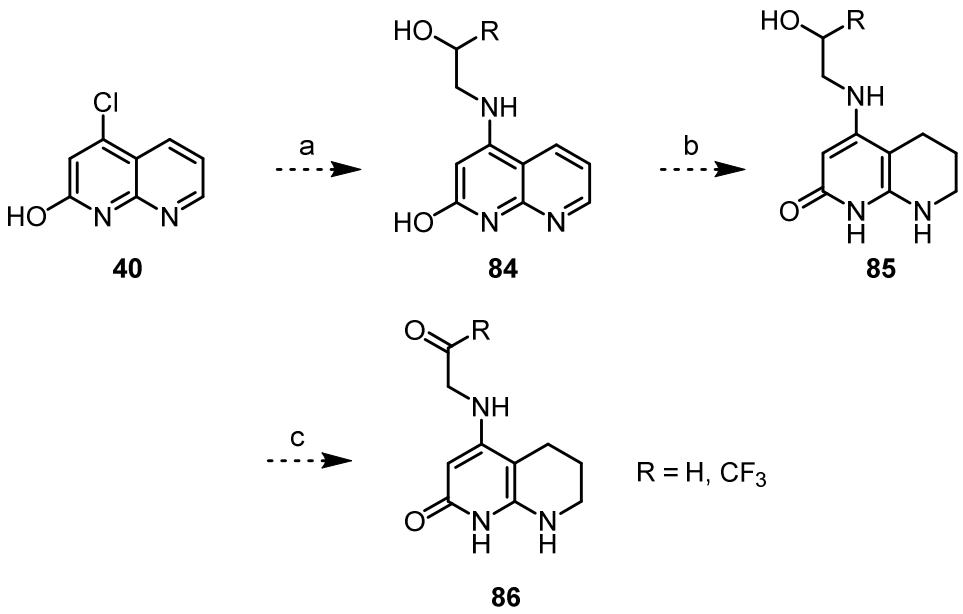
Scheme 14 – *Synthesis of 75. a) 2-Chloroacrylic acid (2 eq.), HATU (2 eq.), DIPEA (3 eq.), 1-methylpiperazine (2 eq.), THF (0.12 M), 0 °C–r.t.*

To prevent this intramolecular cyclisation, it was hypothesised that the pyridone C3 position could be blocked by a substituent. However, the addition of a methyl group was predicted by molecular docking to force the scaffold into a conformation that would prevent the correct binding of the fragment; only substitution by fluorine was likely to be suitable. Fluorination was not attempted as part of this thesis, but future work could include the fluorination of piperidinopyridone **42** to access fluoropyridone **81**, followed by the optimisation of a protecting group strategy and acylation to access chloroacetamide **82** and acrylamide **83** (Scheme 15). The ability of fluorine to attenuate the basicity of amines would also be an interesting effect to explore with respect to the activity of the final products and synthetic accessibility.



Scheme 15 – *Synthesis of 82 and 83. a) Selectfluor, DCM, r.t. b) chloroacetyl chloride or aryloyl chloride, DCM, TEA, r.t.*

Attempts to synthesise carbonyls **77** and **78** were not performed as part of this thesis due to the prioritisation of other aspects of the project. Future work should include the attempted synthesis of these compounds via an S_NAr reaction with naphthyridone **40** to access amino intermediates **84**, followed by the selective reduction of the ring to access piperidinopyridones **85**, and finally an oxidation to generate the trifluoroketone or aldehydic fragments **86** (Scheme 16).



Scheme 16 – *Synthesis of intermediates* ***86****. a) 3-chloro-1,1,1-trifluoroacetone or 2-aminopropanol, DMSO (0.22 M), 150 °C, 1 h. b) H-Cube, H_2, Pd/C, MeOH (0.18 M), formic acid (0.07 M), 40 °C, 20 bar, 0.5 mL/mi. c) [O].*

In summary, no covalent analogues were successfully synthesised to be tested. This was due to the production of numerous undesired side products in the reaction scope that was explored. Ultimately, this objective was deprioritised for efforts to improve the affinity of the fragments through fragment growing, described in Section 2.6.

2.4 Selectivity screening

Following the SAR analyses, the original fragments **MC-278** and **RS-009** remained optimal with respect to potency. The next aim was to determine the current selectivity of the fragments. The desired outcome was a proof-of-concept that the specific GEF/GTPase pair could be selectively inhibited. Based on the availability of material, **MC-278** was used to explore this objective.

2.4.1 Limited Rho GEF/GTPase panel

The first screening attempt utilised the FI nucleotide exchange assay and a limited number of GEF/GTPases. The covalent nature of **MC-278** was an important consideration; GTPases not bearing the analogous C18 or alternative nucleophilic side chain cannot form a covalent bond with the fragment. Accordingly, only GTPases retaining the equivalent C18 residue were selected. All were members of the Rho subfamily, except for c9orf72/RAB39B of the Rab family (Figure 29).

Gratifyingly, **MC-278** showed varying potency across the GEF/GTPases used in the assay (Figure 29). The activity against three GEF/GTPases (Vav1/Rac1, Kalirin-GEF(2)/RhoA, Dbs/Cdc42) was comparable to that of Kalirin-GEF(1)/Rac1, but for others (Tiam1/Rac1, Vav1/Cdc42, c9orf72/Rab39B) no activity was observed at 500 µM. The data suggests that the GTPase or GEF do not alone account for selectivity; it is dictated by the specific GEF/GTPase pairing. Moreover, **MC-278** was found to be inactive against c9orf72/Rab39B, demonstrating that selectivity outside of the subfamily is possible. This was a very promising for the suggested approach of selectively targeting Rho GEF/GTPase complexes.

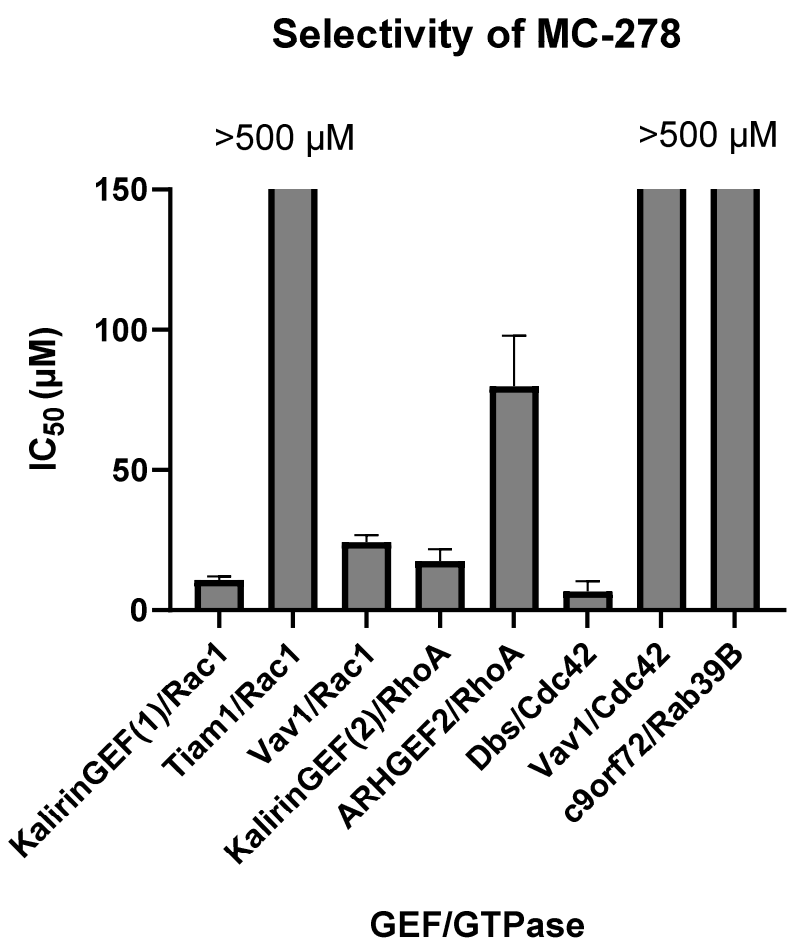


Figure 29 – *Selectivity of MC-278 against Rho and Rab GEF-GTPase pairs, measured by FI nucleotide exchange.*

Understanding the mechanisms behind the observed activity could be used to improve the selectivity of fragments during the development of probe-like compounds. This elucidation remains a key part of future work. Two hypotheses were suggested from the outset: a) the binding site between the complexes is distinct enough to influence the binding of the fragment, or b) the varying kinetics between the formation of the GEF-GTPase and MC-278-GEF-GTPase complexes influence the ability of MC-278 to covalently bind. For example, one complex may be more stable in the apo GEF-GTPase form than the MC-278-GEF-GTPase form, reducing the time for MC-278 to

form the covalent linkage. The former would require an analysis of the binding sites of the complexes, and the latter could be explored using binding analyses and the kinetics of nucleotide exchange, and correlating the data with the observed activity.

2.4.2 Activity-based proteomic profiling

The selectivity of **MC-278** was also explored on a proteome-wide level, specifically using activity-based protein profiling (ABPP). This is a chemical proteomic method for the functional interrogation of enzymes. The probe is comprised of three components; an inhibitor or site-directed active group that covalently binds a subset or family of related enzymes - a tag, for the quantification and identification of the labelled enzymes - and a linker joining them.[238] Due to many large ABPP probes having permeability issues, the inhibitor is often incubated with the proteome containing a latent handle in place of the full reporter tag. This is usually an alkyne or azide, which orthogonally reacts with the corresponding partner on the reporter tag via a click reaction, generating the entire ABPP probe *in situ*. After the proteins have been labelled, they are often enriched with immobilised streptavidin and subjected to on-bead proteolytic digestion. The resulting peptides are analysed by MS-MS, and the peptides are identified using sources such as Uniprot and data processing platforms like MaxQuant or MSFragger.[239-241]

In this case, isotopic Tandem Orthogonal Proteolysis–Activity-Based Protein Profiling (isoTOP-ABPP) was used.[242] This approach allows for the quantitative analysis of native cysteine reactivity; an electrophilic iodoacetamide containing an alkyne handle (IA-alkyne) labels cysteine residues and undergoes click chemistry for the attachment of a tag for purification or visualisation. A variation of this experiment, competitive

isoTOP-ABPP, was employed to measure the selectivity of **MC-278** given the fragment lacks any latent handle to incorporate a tag for visualisation.[243] In this instance, two experiments were performed in which the cell lysates were incubated with **MC-278** or DMSO, added IA-alkyne, and 'clicked' with isotopic heavy and light tags; the cysteines reacting with **MC-278** were then identified by comparing the redundancies of the heavy and light tags (Figure 30).

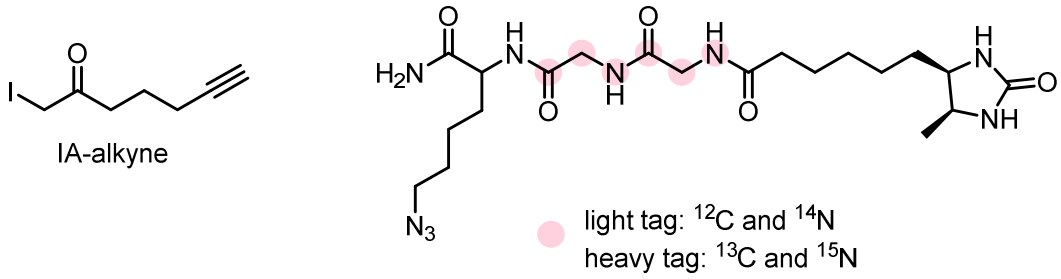


Figure 30 – *Components of the isoTOP-ABPP experiment.*

After performing the experiment using a K-562 cell line pellet obtained from Dr. Elisabeth Mira Rothweiler (two biological repeats, each with three technical repeats), only a limited number of peptides were enriched with the IA-alkyne tags. Whilst many published isoTOP-ABPP experiments report upwards of 5000 peptides from lysates, the experiments yielded only 700-1500 peptides, indicating that a limited proportion of the proteome was accessed. Moreover, no GTPases at any cysteine were among the identified. Due to this result, the selectivity and enrichment of **MC-278** across the proteome was ambiguous.

There are many possibilities for the limited enrichment of proteins. One could speculate the quality of the cell pellet or the buffers and conditions used in the experiment. However Amber Truepenny, a DPhil student of the Brennan and Kessler

groups, performed the experiment with the same pellets and experimental set up, albeit using an alternative probe to **MC-278**. In this case, 2000-6000 peptides were identified. Another possibility is the quality of the digesting enzyme; the batch of chymotrypsin used was fresh, and no issues were reported in the department previously. The most likely cause of the poor enrichment is the degradation of proteins during the experiment, or the disruption of binding of the tagged proteins to the streptavidin-binding beads. In both cases, the incompatibility of the **MC-278** probe is the likely cause, through mechanisms that are not understood.

In summary, the enrichment of Rho GTPases across the proteome by **MC-278** was not recognised. In future experiments, an alternative probe should be applied, such as **RS-009**, or one incorporating a latent handle for the direct interaction with the isotopic tags. This future project is outlined in detail in Section 2.6.3.

2.5 Potency at cellular nucleotide concentrations

It is important to note that the selectivity experiments described previously fail to take account of the whole cellular environment, most notably the high concentration of the endogenous nucleotides. Exploring this effect was desirable, as it would highlight the affinity required of a probe to outcompete the endogenous ligands. The following section will describe the strategies investigated to obtain this data using **MC-278**.

2.5.1 Fluorescence intensity assay

In the first attempt, the original FI nucleotide exchange assay was applied under the pre-set conditions with the addition of GppNHp (300 µM) (Figure 31). However, excess GppNHp resulted in low signal:background due to the suppression of the fluorescence arising from BODIPY-GTP binding.

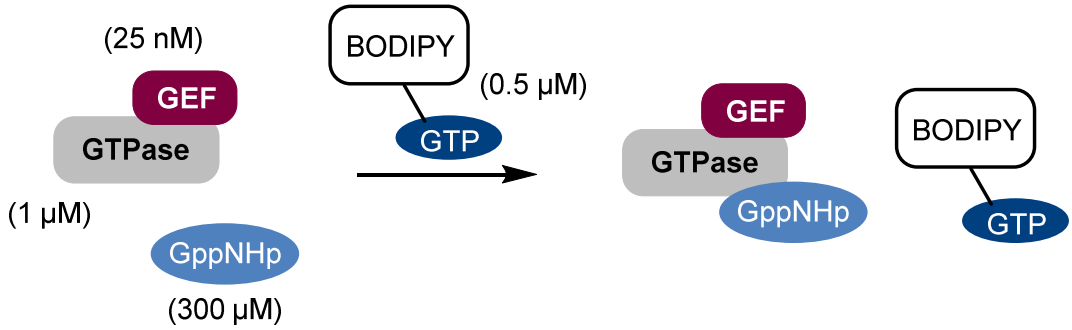


Figure 31 – *Competition assay using FI nucleotide exchange. The presence of 300 µM GppNHp reduced the signal of BODIPY-GTP binding, thus reducing the signal:noise.*

To optimise the signal, BODIPY-GTP was titrated to measure the change in the FI value and initial rate of BODIPY-GTP binding (Figure 32). A concentration of 2.5 µM

was considered optimal based on the maximum signal read by the fluorescence reader (260,000 FI) and the largest assay window. The assay was performed with the optimised concentration of BODIPY-GTP, but the readout was poorly reproducible and large errors occurred (Figure 33).

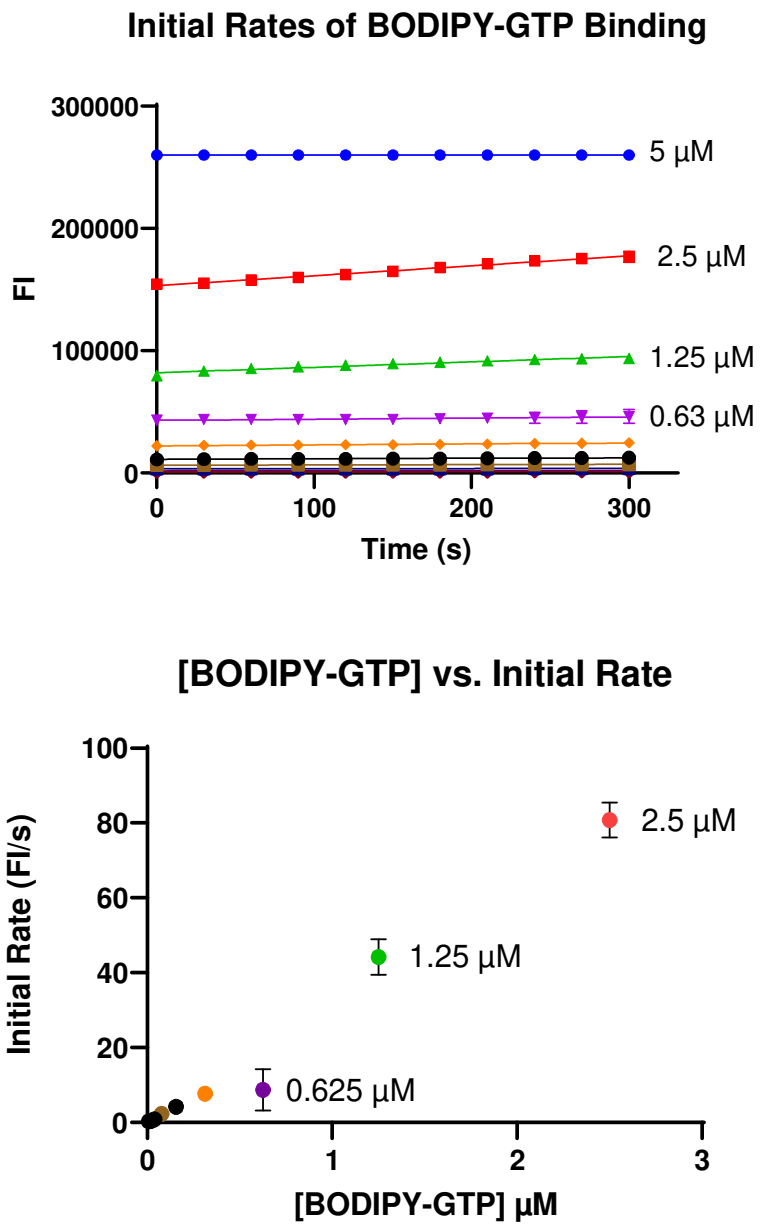


Figure 32 – *Optimisation of [BODIPY-GTP] in the presence of 300 µM GppNHp.*

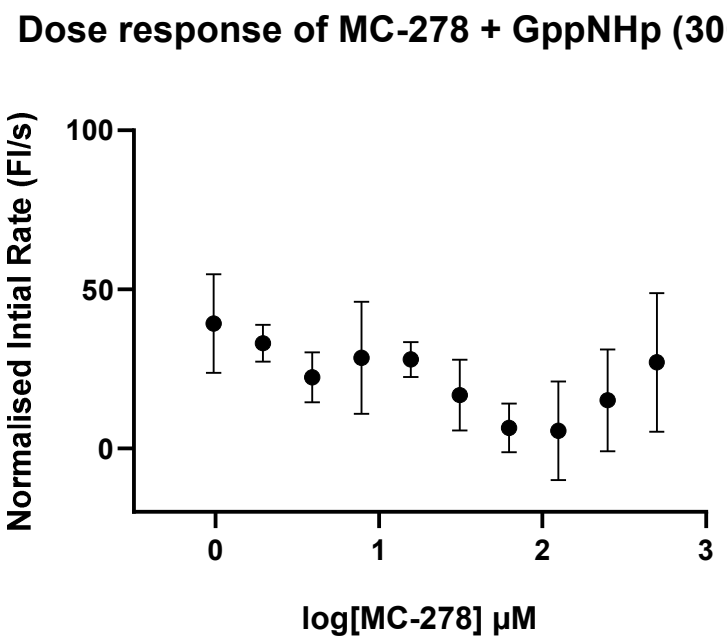


Figure 33 – *Dose response of MC-278 to Kalirin-GEF(1)/Rac1 (incubation time = 2 h) followed by the addition of GppNHp (300 µM).*

This readout was suspected to be due to small differences in mixing of the BODIPY-GTP and GppNHp upon addition. It was also probable that a molecule of BODIPY-GTP could bind at the nucleotide binding site, only to be displaced by a molecule of GppNHp, leading to uncertainty in the readout. Accordingly, alternative assays were considered.

2.5.2 Mass spectrometry assay

MS was explored as a direct approach of measuring the abundance of wild type (WT) Rac1 and **MC-278** covalently modified Rac1 (MC-278-Rac1) in the presence of Kalirin-GEF(1) and GppNHp. The assay was initially set up using the concentrations and conditions inherited from the original FI nucleotide exchange assay (Table 7, Entry 1). By MS analysis no covalent modification was observed, contrary to the prior evidence supporting the covalent interaction.

Entry	Rac1 (μM)	Kalirin-GEF(1) (μM)	Buffer additions	Temp (°C)	Time (h)	MC-278-Rac1 (%)
1	1.25	0.03	BSA, DTT	r.t.	2	0
2	1.25	0.03	-	r.t.	2	0
3	1.25	1.25	-	r.t.	2	0
4	1.25	1.25	-	r.t.	24	0
5	1.25	0.03	-	4	24	0
6	1.25	1.25	-	4	24	79
7	1.25	0	-	4	24	0

Table 7 – *Optimisation of MS assay measuring the covalent modification of Rac1 by MC-278 (200 μM).*

Multiple changes were made to the procedure to encourage covalent binding. Altering the buffer composition offered no change in signal (Entry 2), nor did increasing the ratio of Kalirin-GEF(1)/Rac1 1:1 (Entry 3). Extending the incubation time was unproductive (Entry 4), even at a lower incubation temperature (Entry 5). However, the combination of lowering the temperature, extending the incubation time, and increasing the stoichiometry of Kalirin-GEF(1):Rac1 resulted in a 79% abundance of MC-278-Rac1 by MS analysis (Entry 6). No MC-278-Rac1 was observed in the absence of Kalirin-GEF(1) under the same conditions (Entry 7), demonstrating the selectivity the compound has for the GEF/GTPase complex vs. the GTPase alone.

The optimised conditions were then applied in the presence of GppNHp. In one experiment, **MC-278** was incubated with Kalirin-GEF(1)/Rac1 followed by the addition of GppNHp after 24 h. In this instance, the abundance of MC-278-Rac1 was unchanged; GppNHp was unable to disrupt the binding of **MC-278** once covalently bound as expected (Figure 34).

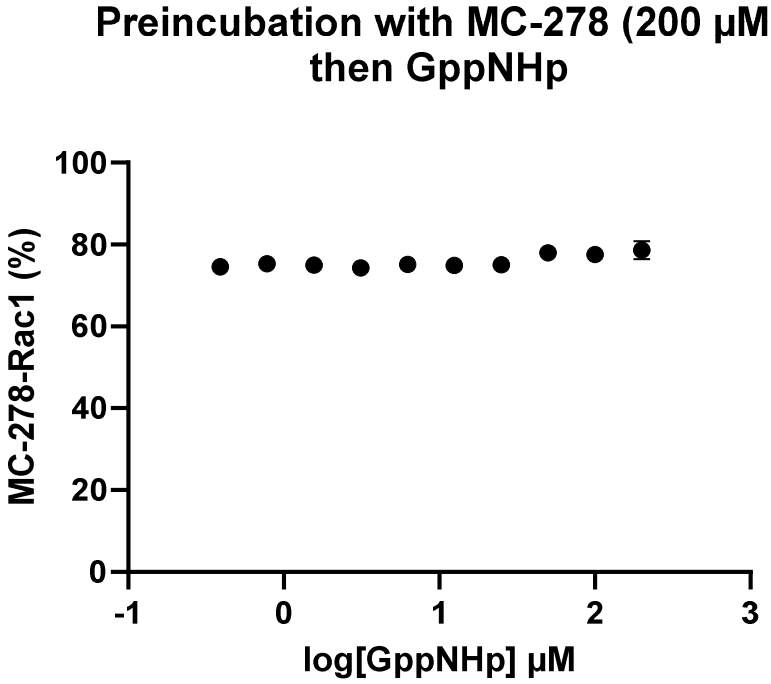


Figure 34 – *Percentage of MC-278 modified Rac1 after incubation of Kalirin-GEF(1) and Rac1 with MC-278 (200 µM), followed by addition of GppNHp (n = 2).*

When incubating **MC-278** (200 µM) and GppNHp with Kalirin-GEF(1)/Rac1 simultaneously at the selected time point, measurable MC-278-Rac1 was only observed when [GppNHp] < 100 µM (Figure 35A). This was further emphasised by the change in IC_{50} of **MC-278** in the presence of just 10 µM GppNHp (Figure 35B). Given the concentration of GTP in the cell is 300 µM, it is likely a negligible quantity of **MC-278** would be covalently bound to Rac1 under cellular conditions at this time point.

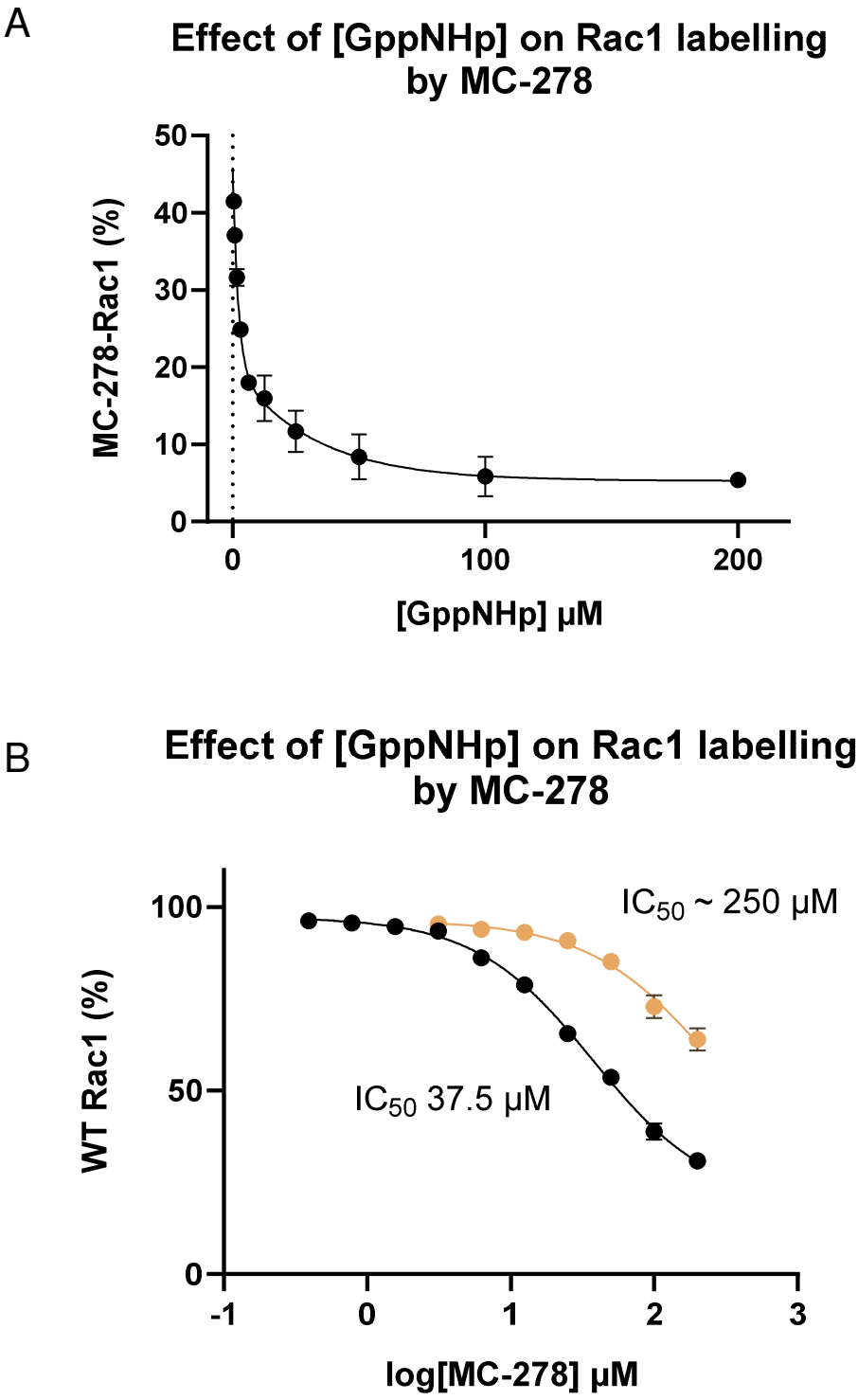


Figure 35 – *(A) Percentage of MC-278 modified Rac1 following the incubation of Kalirin-GEF(1) and Rac1 with MC-278 (200 µM) and GppNHp. (B) Dose-response of MC-278 ± 10 µM GppNHp. black = no GppNHp, orange = 10 µM GppNHp.*

The experiment highlights that a potent binder is required for the first reversible binding step to outcompete the endogenous ligands, even in the presence of a reactive covalent warhead. To improve this affinity, fragment elaboration efforts where explored.

2.6 Fragment elaboration

The next aim was to improve the potency of **MC-278** and **RS-009** through fragment elaboration, whilst considering the important binding mechanisms identified from the previous SAR analyses. A number of approaches were applied to this end.

2.6.1 XChem screen

The XChem screening platform was one strategy considered to assist fragment elaboration. The protein crystals of MC-278-Kalirin-GEF(1)/Rac1 that were previously obtained (Section 2.2.4) could be acquired reproducibly; with this in-hand, it was envisaged that complementary ligands binding in the nucleotide binding pocket could be identified by XChem and a fragment linking approach could be applied. As such, crystals of Kalirin-GEF(1)/Rac1 were grown and pre-soaked with **MC-278**, followed by the addition of fragments from XChem libraries. This XChem campaign was performed by Lizbe Koekemoer of the protein crystallography group at the CMD.

In total, 359 soaks were performed, 243 were successful, and 19 structures were flagged by the PanDDA software as having unmodelled electron density in the nucleotide binding pocket, suggesting the presence of a new binder. However, the structures of the corresponding fragments fit the electron density poorly. This was surprising, given that the resolution of the obtained structures was good (~ 2 Å). Upon further visual analysis, it was found that the nucleotide binding site was occupied by GDP, rather than **MC-278** and a new fragment. The presence of GDP arises during protein expression, and was not completely displaced by **MC-278** during pre-soaking. With this understanding, it became clear that the XChem campaign resulted in no new binders at the nucleotide binding pocket. A future XChem campaign could be explored

with a higher concentration of **MC-278** to completely displace GDP and applying alternative fragment libraries. Such a campaign was deprioritised for alternative methods of fragment elaboration, such as probing various vectors of **MC-278** through chemical synthesis.

2.6.2 Elaboration of MC-278

The second strategy was identifying positions on the fragment that could be functionalised. One area of the fragment that was unexplored and could be fit for this purpose was the piperidine ring, which contained three positions for modification (Figure 36). Whilst C2 was observed to be mostly solvent exposed, hydrophobic groups at C1 or C3 had the potential to interact with L160, F28, or I33.

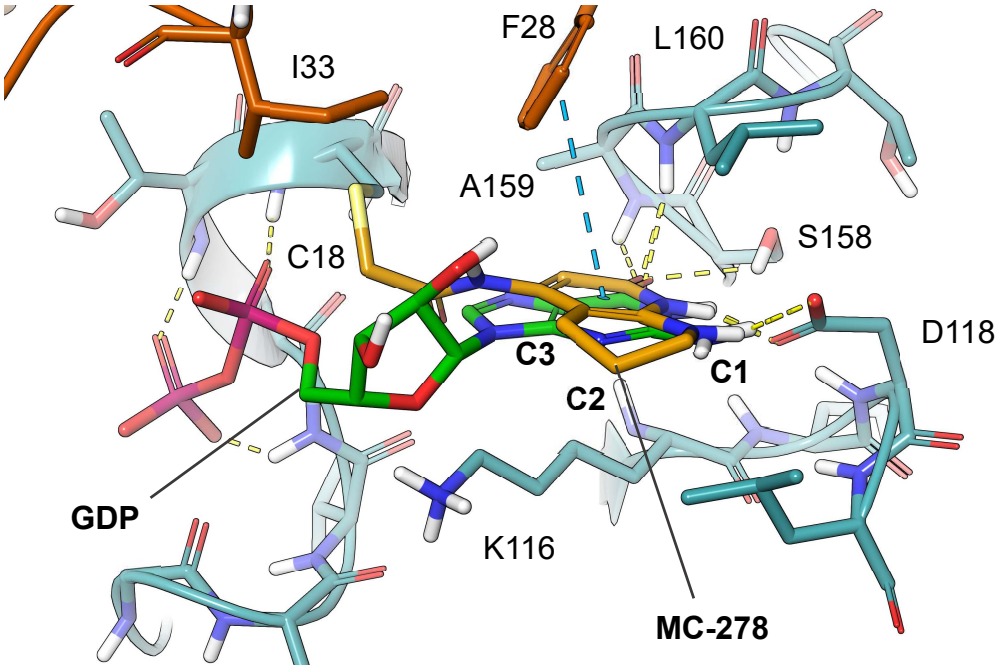
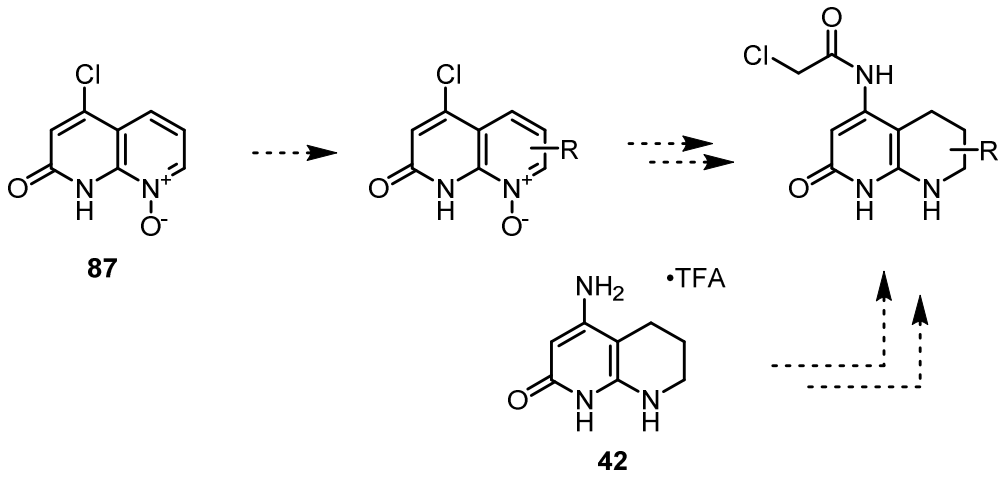


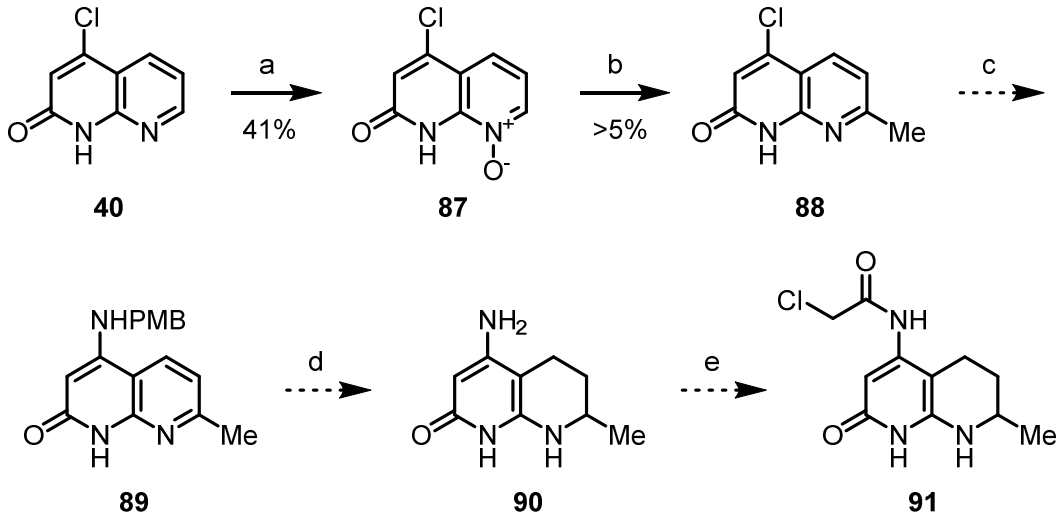
Figure 36 - *Crystal structure of MC-278 (yellow sticks) and GDP (green) bound to Kalirin-GEF(1)/Rac1 (teal ribbons). The switch I region is shown (orange sticks).*

Two strategies were proposed (Scheme 17); introducing functionality through N-oxide intermediate **87**, or directly from piperidinopyridone **42** using electrochemistry or photochemistry.[244-247] Preliminary experiments were performed toward this objective to provide a starting point for future projects.



Scheme 17 – *Possible routes to access piperidine analogues of MC-278.*

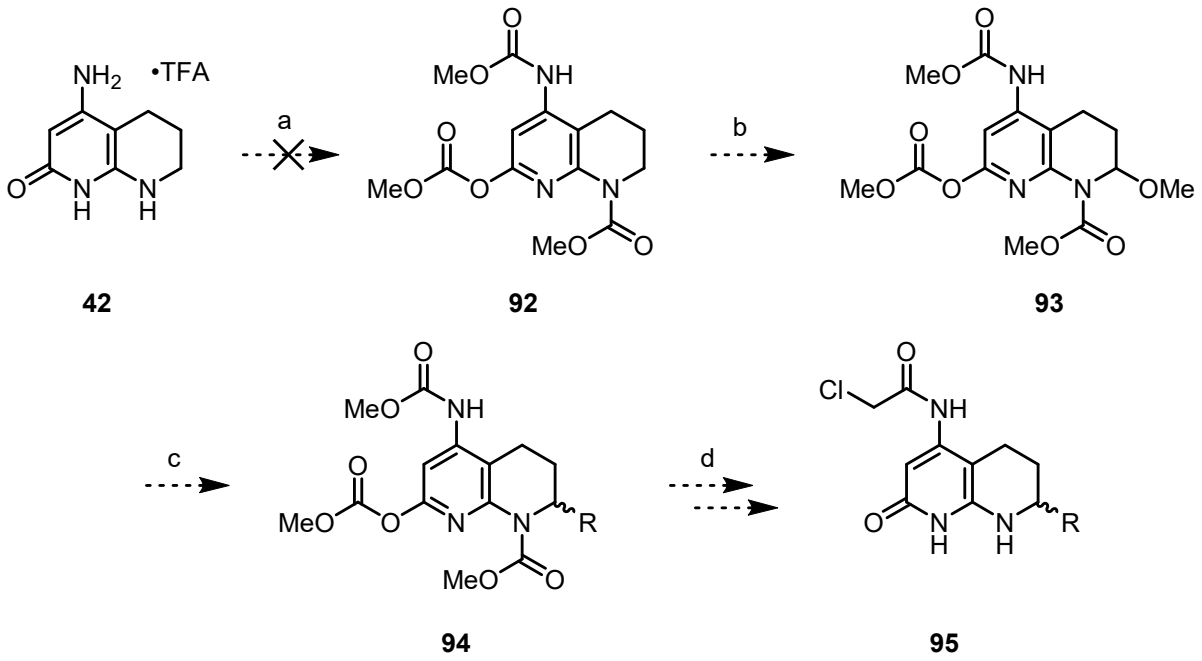
The N-oxide route was first attempted with a simple methyl substituent (Scheme 18). Naphthyridine **40** was successfully oxidised to give N-oxide **87**, which was subsequently alkylated under Grignard conditions to give the methylated product **88**.[248-249] A low yield was observed in this step; this was postulated to be due to the number of products forming by LCMS analysis, such as the ring opened product and the non-deoxygenated product. Optimisation of this step, including changing the solvent and temperature to prevent the formation of side products, would be required to access intermediates **89** and **90**, and finally the methylated product **91**. Following optimisation, the route could be repeated with diverse Grignard reagents. There are also a plethora of alternative strategies for pyridine N-oxide functionalisation that could be investigated simultaneously.[247]



Scheme 18 – *Attempted synthesis of 91. a) mCPBA (1.5 eq.), DCM (0.2 M), 0 °C–r.t., 15 h. b) MeMgBr (2 eq.), $MgCl_2$ (1.5 eq.), CuCl (0.1 eq.), diethyl ether (0.1 M), 0 °C–r.t., 30 min. c) 4-methoxybenzylamine (16 eq.), DMSO (0.22 M), 150 °C, 1 h d) H-Cube, H_2, Pd/C, MeOH (0.18 M), formic acid (0.07 M), 40 °C, 20 bar, 0.5 mL/min, then TFA (0.29 M), 60 °C, 24 h. e) TEA (4 eq.), Boc_2O (1 eq.), chloroacetyl chloride (1 eq.), DMF (0.2 M), 0 °C–r.t., 5 h. then 1-methyl piperidine (1.2 eq.), DCM (0.2 M), r.t., 1.5 h.*

Preliminary experiments were also performed for the direct functionalisation of the piperidine ring. Shono reported the ability to oxidise carbamates under anodic conditions which has since been used to functionalise the α-position of N-heterocycles.[246, 250] This was considered as a functionalisation approach for **MC-278**. As such, piperidinopyridone **42** was added an excess of methyl chloroformate to generate carbamate **92**. Despite this large excess, the carbamate only formed on the pyridone moiety despite multiple additions and attempts. Optimisation of this step would need to be performed to explore this electrochemical approach. Carbamate **92** could then be exposed to anodic oxidation in MeOH to form hemi aminal **93**, followed by the addition of a nucleophile to functionalise the *ortho* position of the piperidine ring to give intermediates **94**. Finally, global cleavage of the carbamate groups followed by

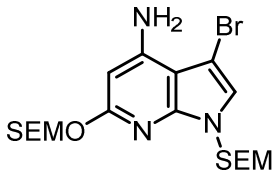
selective acylation could yield the final products **95**. Following the demonstration of a robust synthesis, diverse substituents should be investigated and the SAR explored, to ultimately aid in the iterative improvement of the potency and selectivity of **MC-278**.



Scheme 19 – *Functionalisation of the piperidine ring of MC-278. a) methyl chloroformate (3 eq.), TEA (4 eq.), DCM (0.2 M), 0 °C–r.t. b) anodic oxidation, MeOH, c) Nu. d) THF/MeOH/2 M LiOH (1:1:1), 120 °C then, TEA (4 eq.), Boc_2O (1 eq.), chloroacetyl chloride (1 eq.), DMF (0.2 M), 0 °C–r.t., 5 h. then 1-methyl piperidine (1.2 eq.), DCM (0.2 M), r.t., 1.5 h.*

2.6.3 Elaboration of RS-009

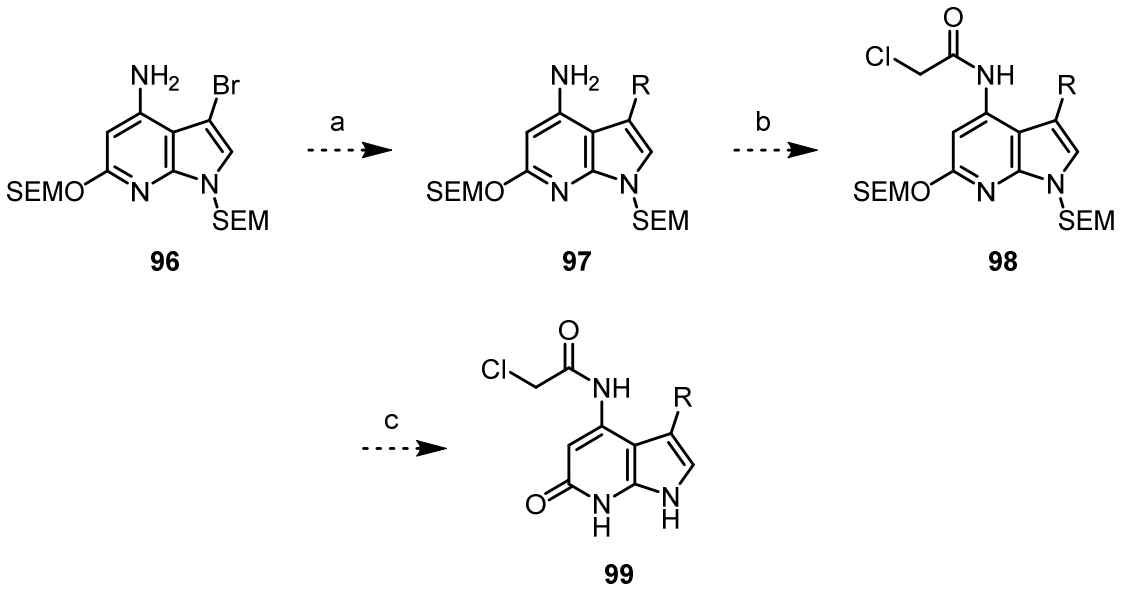
The elaboration of fragment **RS-009** was also of interest, albeit of less focus than **MC-278**. This was due to the multiple attempts previously made by Dr. Janine Gray to synthesise the core in-house which proved challenging. Therefore, advanced intermediate **96** was purchased from Cortex Organics to expedite progress (Figure 37).



96

Figure 37 – *Advanced intermediate toward RS-009 and related analogues.*

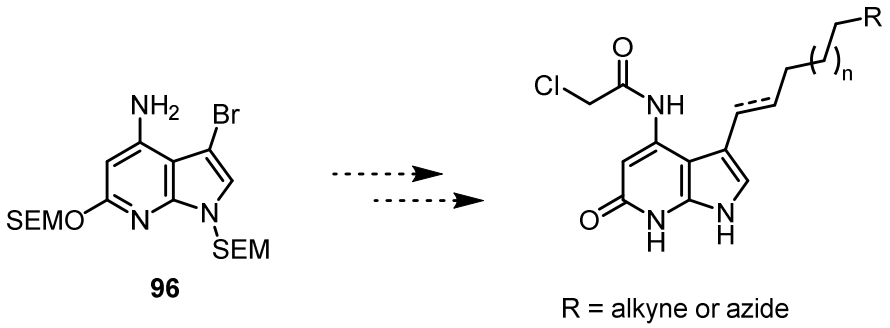
The intermediate will serve multiple purposes for future projects. First, and most relevant to the aims of this thesis, the bromide handle could be used to append various groups at C3. The SAR of the series could then be explored to improve the potency and selectivity of **RS-009**. A robust route should be designed to access these analogues, followed by activity testing in the FI nucleotide exchange assay. This could include a cross-coupling to introduce functionality at C3 and access functionalised intermediates **97**, followed by an introduction of the warhead to access chloroacetamide **98**, and finally a global SEM deprotection to access the functionalised RS-009 core **99** (Scheme 20).



Scheme 20 – *Synthesis of functionalised RS-009 analogues. a) metal cross coupling reactions e.g. Suzuki. b) TEA (2 eq.), chloroacetyl chloride (4 eq.) DCM (0.13 M), r.t., 1 h. c) TFA (0.04 M), r.t., 1 h.*

Intermediate **96** could also be used in the development of a Rho GTPase probe for ABPP, given the issues observed using **MC-278** in the competitive isoTOP-ABPP experiment described in Section 2.4.2. It is expected that an alkyne or azide handle could be appended to intermediate **96** via a linker to covalently bind Rho GTPases in the proteome (Scheme 21). A promiscuous Rho GTPase probe would yield many benefits; competitive-ABPP could be used to explore the potency and selectivity of new compounds against multiple targets simultaneously, in which the binding of the promiscuous probe will be reduced in the presence of more potent compounds.[251] The probe could also be used to identify the presence of Rho GTPases in various cell lines, or explore the Rho GEF/GTPase proximitome. A variety of linker types and lengths with alkyne or azide handles should be explored with molecular docking, and the most promising fragments synthesised and tested in the FI nucleotide exchange assay. This will confirm compounds with appended linkers and handles are still able to bind in the

nucleotide binding pocket. Following this optimisation, the probe should be applied in ABPP and the selectivity investigated. This project will be undertaken by Zhihe Lei, a final year undergraduate chemistry student.



Scheme 21 – *Overview of ABPP probe structure for Rho GTPases.*

Finally, intermediate **96** will be used to explore the possibility of degrading GTPases or GEFs via the GEF/GTPase complex using a proteolysis targeting chimera (PROTAC).[252-254] PROTACs are quickly emerging as potential therapeutic agents for the degradation of targets via ubiquitination, and none have been reported for GEFs or GTPases. They consist of an inhibitor of the target protein or complex for degradation, and a linker connecting the inhibitor to ligand capable of recruiting an E3 ubiquitin ligase. The result is the ubiquitination of the protein of interest and ultimately degradation. A linker could be incorporated to intermediate **96** via the bromide handle, and optimised for activity through changing the linker or the E3-ligase binding moiety. This project will be led by Amelia Henley, a DPhil student in the Brennan Group.

2.7 Conclusions

To conclude, the work described in this chapter offers a greater insight into the mechanisms surrounding the inhibition of Rho GEF/GTPase complexes by small molecules, focusing on Kalirin-GEF(1)/Rac1. Much more is understood about the binding and selectivity of inhibitors **RS-009** and **MC-278**, as well as the obstacles that must be overcome in order to progress the fragments toward usable chemical probes. This developed insight has also led to new projects focussed on targeting the Rho GEF/GTPase family.

The binding of a previously identified hit, **RS-009**, was first quantitatively measured using IC_{50} and k_{inact}/Ki, and analogues were designed to probe the binding interactions. In total four alternative cores were synthesised and tested, and tuning the dual hydrogen bonding interaction with D118 and interaction with C18 was found to be crucial for activity. A new active fragment, **MC-278**, was also identified. The binding mechanism of **MC-278** was successfully corroborated through time-dependent kinetic studies and high-resolution crystallography.

The covalent interaction with C18 was further explored by designing a diverse set of non-covalent and covalent analogues. Prior to their synthesis, an improved route toward the **MC-278** core was required. A route with fewer steps, superior overall yield and improved scalability was successfully demonstrated and used to produce useful intermediates. This resulted in the synthesis and testing of 17 non-covalent analogues, and the lack of activity of the series emphasised the influence of the chloroacetamide warhead.

The selectivity of **MC-278** was later explored. It was demonstrated in a limited biochemical screen that the specific GEF/GTPase pairing could dictate the binding of fragments. This supports the approach of targeting specific GEF/GTPases for disease. **MC-278** was also applied in proteome-wide isoTOP-ABPP experiment, in which it appeared the probe was incompatible with the experiment, and should be performed with an alternative in the future.

As the previous selectivity studies were absent of the competing endogenous ligands, a mass spectrometry assay was designed and optimised to measure the covalent modification of Rac1 in their presence. The data confirmed that there was selectivity for the GEF/GTPase complex over the GTPase alone, but also highlighted the negligible binding the fragments were likely to have at cellular concentrations of GDP and GTP. This emphasises a central piece of future work, to improve the potency of the fragments to outcompete the endogenous ligands.

Finally, preliminary experiments were performed to elaborate **MC-278** and **RS-009** for this objective. Initially, an XChem campaign was initiated with MC-278-Kalirin-GEF(1)-Rac1 soaked crystals that yielded no new ligands for fragment linking. Instead, reactions were performed that offer starting points for the synthetic elaboration of **MC-278**. A core focus should be the optimisation of these routes to grow the fragments, and explore the SAR with diverse functionality. An advanced intermediate toward **RS-009** was also designed and purchased, which will be exploited in fragment elaboration, and two new projects to develop probes for ABPP and PROTAC-induced degradation of GEF/GTPases.

3 Kalirin-GEF(2)/RhoA

The second target of interest was Kalirin-GEF(2)/RhoA. The key objective was to design and produce active Kalirin-GEF(2) to generate material for screening and crystallography studies; neither the structure of Kalirin-GEF(2) or the Kalirin-GEF(2)/RhoA complex was previously known, and no inhibitors had been reported for the complex. As such, the potential to inhibit nucleotide exchange on Kalirin-GEF(2)/RhoA was also explored.

3.1 Production of active Kalirin-GEF(2)

3.1.1 Construct design

Multiple factors were considered to improve the likelihood of expression and crystallisation of Kalirin-GEF(2) constructs as described in detail in Section 1.3.1. It was important to include constructs that contained the DH domain (core catalytic domain) and the DH-PH tandem, as the PH domain is reported to have a variety of influences depending on the specific GEF. After employing these factors, 22 constructs were designed and purchased from Twist Biosciences (Figure 38).

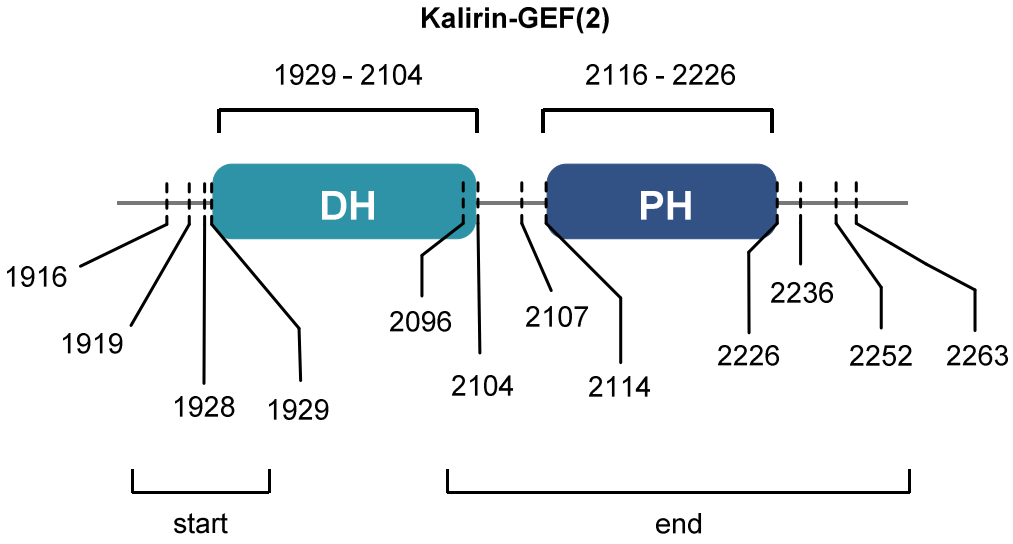


Figure 38 – *Start and end residues used in the design of 22 Kalirin-GEF(2) constructs. Residue numbers as dictated by UniProt for Kalirin: O60229.*

3.1.2 Expression and purification

Prior to large scale expression, a test expression of the 22 purchased constructs was performed by Leela Shrestha of the Biotechnology small research facility (Biotech SRF) in the CMD (Figure 39). Of these, 12 expressed successfully. Constructs with the start residue G1929 all expressed poorly, and thus was avoided in future designs.

The 12 constructs were subsequently scaled up as part of this thesis. A common approach used in the CMD is the parallel expression and purification of constructs known as 'PrepX' which was applied in this work. Following bacterial expression and IMAC purification, 8 of the constructs were obtained in significant yield (Figure 40A). The purity of the constructs was significantly improved following reverse IMAC purification (Figure 40B).

Construct	Start	End
A1	K1916	P2263
A2	E1919	P2263
A3	K1916	R2252
A4	G1929	P2263
A5	G1929	R2252
A6	G1929	R2252
A7	K1916	Q2236
A8	E1919	Q2236
A9	K1916	E2226
A10	E1919	E2226
A11	G1929	Q2236
A12	G1929	E2226

Construct	Start	End
B1	K1916	E2114
B2	E1919	E2114
B3	K1916	M2104
B4	E1919	M2104
B5	G1929	E2114
B6	K1916	L2096
B7	R1928	L2107
B8	E1919	L2096
B9	G1929	M2104
B10	G1929	L2096

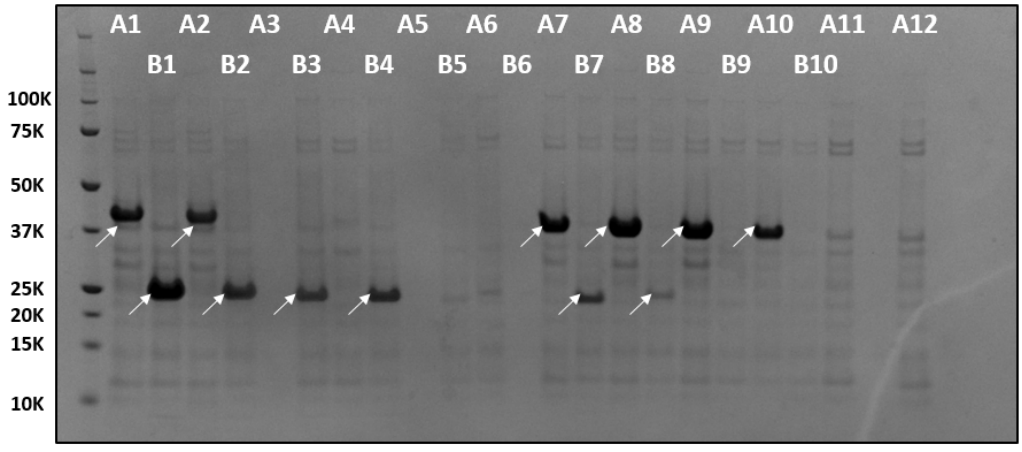


Figure 39 – *Test expression of Kalirin-GEF(2) constructs. Well expressed constructs are highlighted in green. White arrows indicate expected mass.*

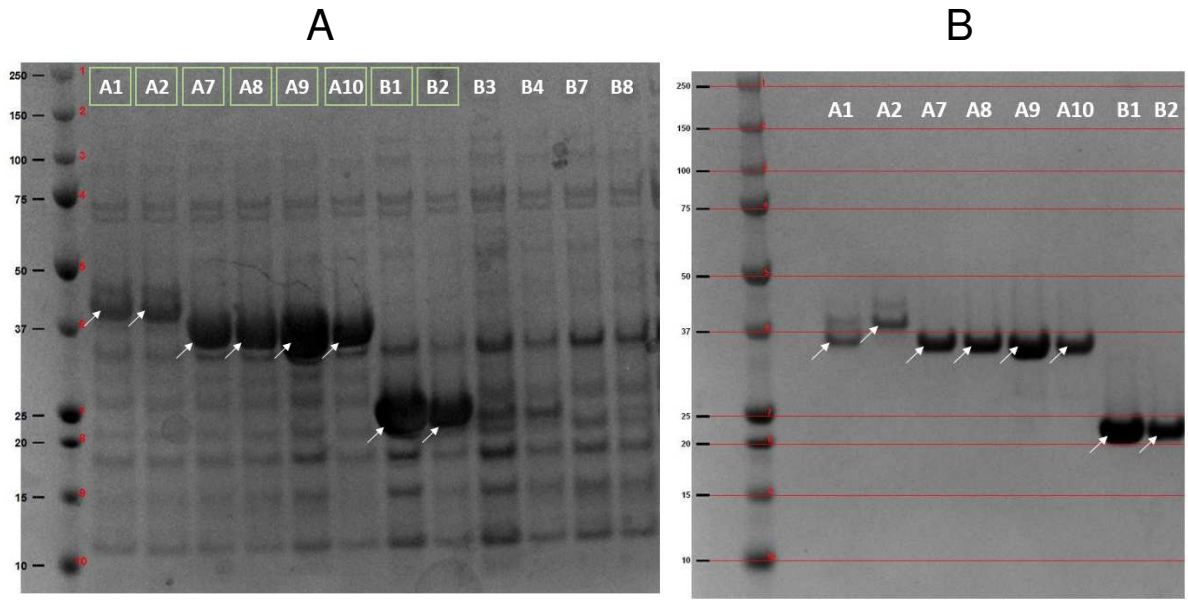


Figure 40 – *Primary purification of constructs. (A) Coomassie brilliant blue stain following IMAC purification. Constructs highlighted in green were taken forward. (B) Coomassie brilliant blue stain following reverse IMAC purification. White arrow indicates expected mass.*

Intact mass spectrometry was used to confirm the expected masses of the constructs (Supplementary Table 2). Constructs **A1** and **A2** were found to have undergone complete and partial proteolysis, respectively, and thus were not purified further. After further purification by size exclusion chromatography (SEC), constructs **A9**, **B1** and **B2** were obtained at a practical concentration to use in crystallography or activity experiments.

3.1.3 Kalirin-GEF(2) activity analysis

The GEF activity of the Kalirin-GEF(2) constructs was subsequently explored with the FI nucleotide exchange assay. The most significant finding was the difference in activity observed for the DH domain (**B1**) and the DH-PH tandem (**A9**), where the DH-PH domain was inferior in GEF activity (Figure 41). This data alludes to the PH2 domain of Kalirin being autoinhibitory in nature; this is a contrast to Kalirin-GEF(1), in

which Dr. Janine Gray observed the PH1 domain to enhance nucleotide exchange on Rac1. Such a phenomenon is well studied for Rho GEFs, and is a useful consideration for future assays.[166, 221, 255-256] For example, recombinant full length Kalirin may be observed to activate Rac1 to a greater extent than RhoA in the absence of an effector to regulate the PH2 domain. To fully elucidate this mechanism, the **A9** DH-PH construct could be crystallised and the structure solved. The PH2 domain may be observed to reside in front of the GTPase binding surface, which would partially prevent GTPase binding. This objective was not explored as part of this thesis.

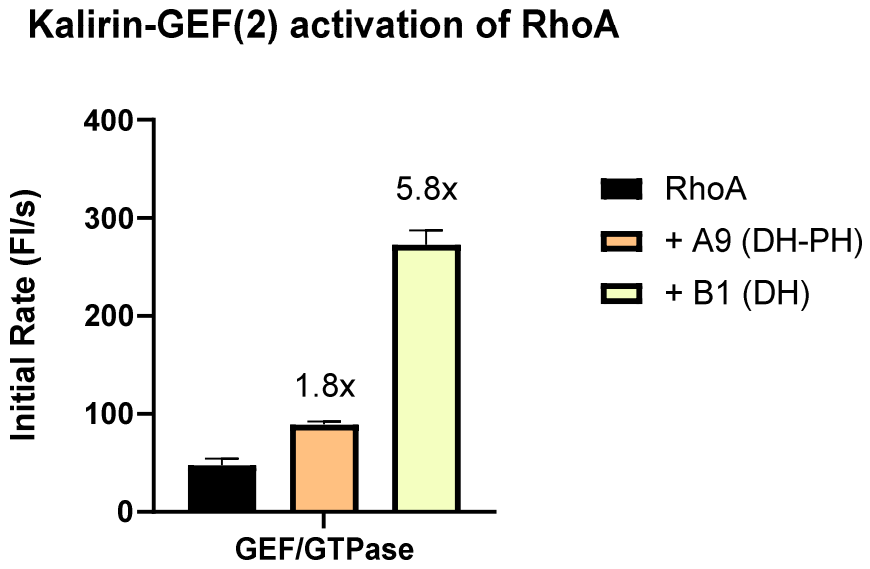


Figure 41 – *Kalirin-GEF(2) activation of RhoA.*

3.2 Crystallisation of Kalirin-GEF(2)/RhoA

After confirming the GEF activity of the produced constructs, the next objective was to obtain novel structural data for the Kalirin-GEF(2) domain and the Kalirin-GEF(2)/RhoA complex. Optimisation of the crystallisation process would also be valuable for XChem screening, to identify a starting point for chemical probe development.

3.2.1 Kalirin-GEF(2)/RhoA complex purification

The Kalirin-GEF(2) constructs were first combined with RhoA (supplied by Dr. Vittorio Katis) and purified by SEC following incubation, as it was expected that the GEF/GTPase complex would be observable by UV trace analysis. For the **A9**/RhoA mixture, two distinct peaks were observed; this suggested the Kalirin-GEF(2) DH-PH/RhoA complex had not formed under the purification conditions (Figure 42A). This separation was confirmed by the concentrations observed in the resulting SEC fractions (Figure 42B).

The SEC trace of **B1**/RhoA was ambiguous (Figure 43A). Whilst the trace contained just one peak, it was possible that the two components had co-eluted due to similarities in size and mass. The resulting SEC fractions appeared to show the components in similar stoichiometry but did not support any conclusion with respect to complex formation (Figure 43B). The SEC trace for **B2**/RhoA was blank due to an instrument error, but the resulting fractions were also found to contain both **B2** and RhoA in similar concentration (Figure 43B). The mixtures were concentrated and taken forward for crystallography screening, as it was hypothesised the crystallisation conditions could encourage complex formation.

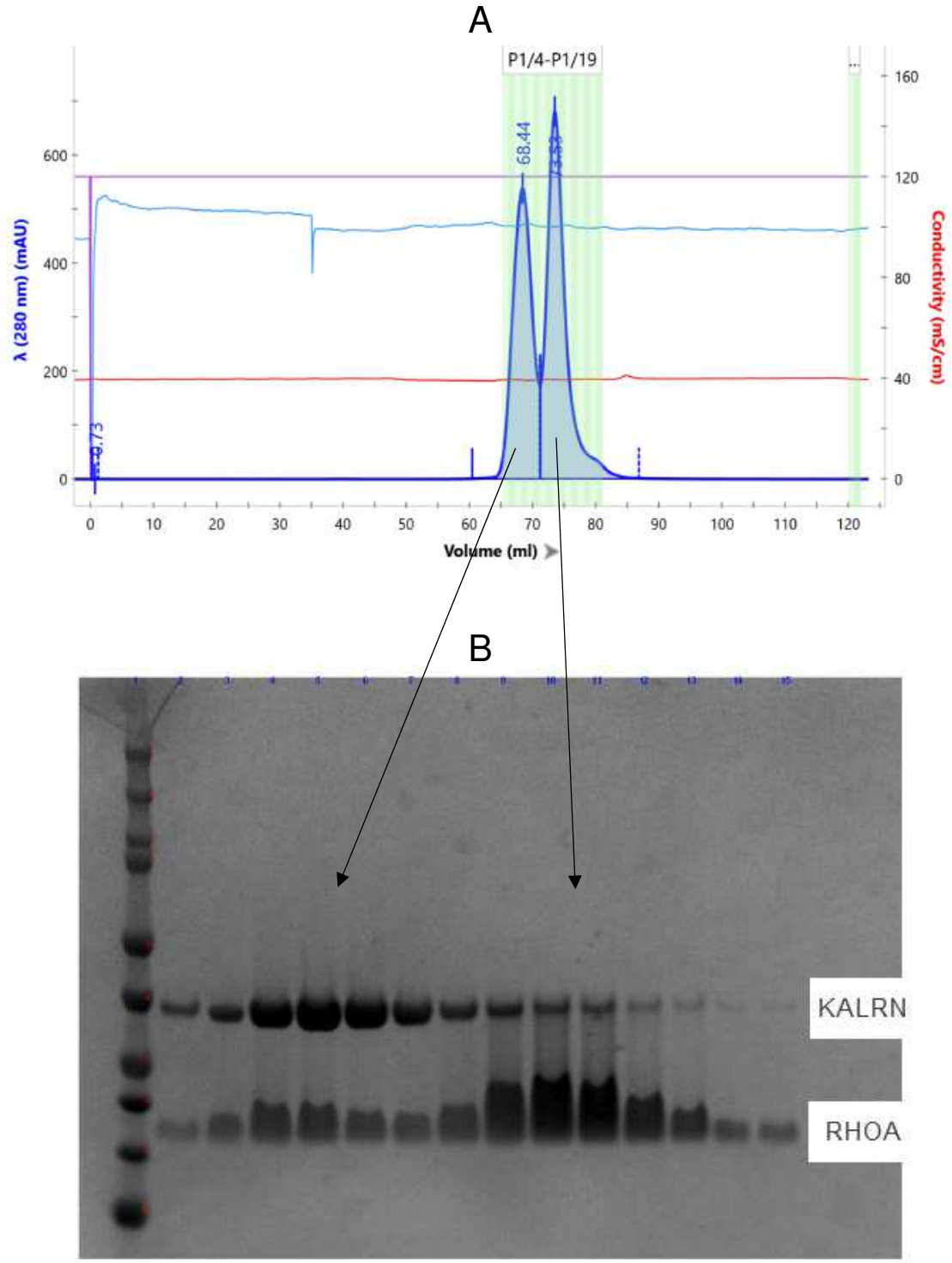


Figure 42 – *SEC purification of A9/RhoA. (A) The SEC UV purification trace. (B) Coomassie brilliant blue stain of the resulting SEC fractions.*

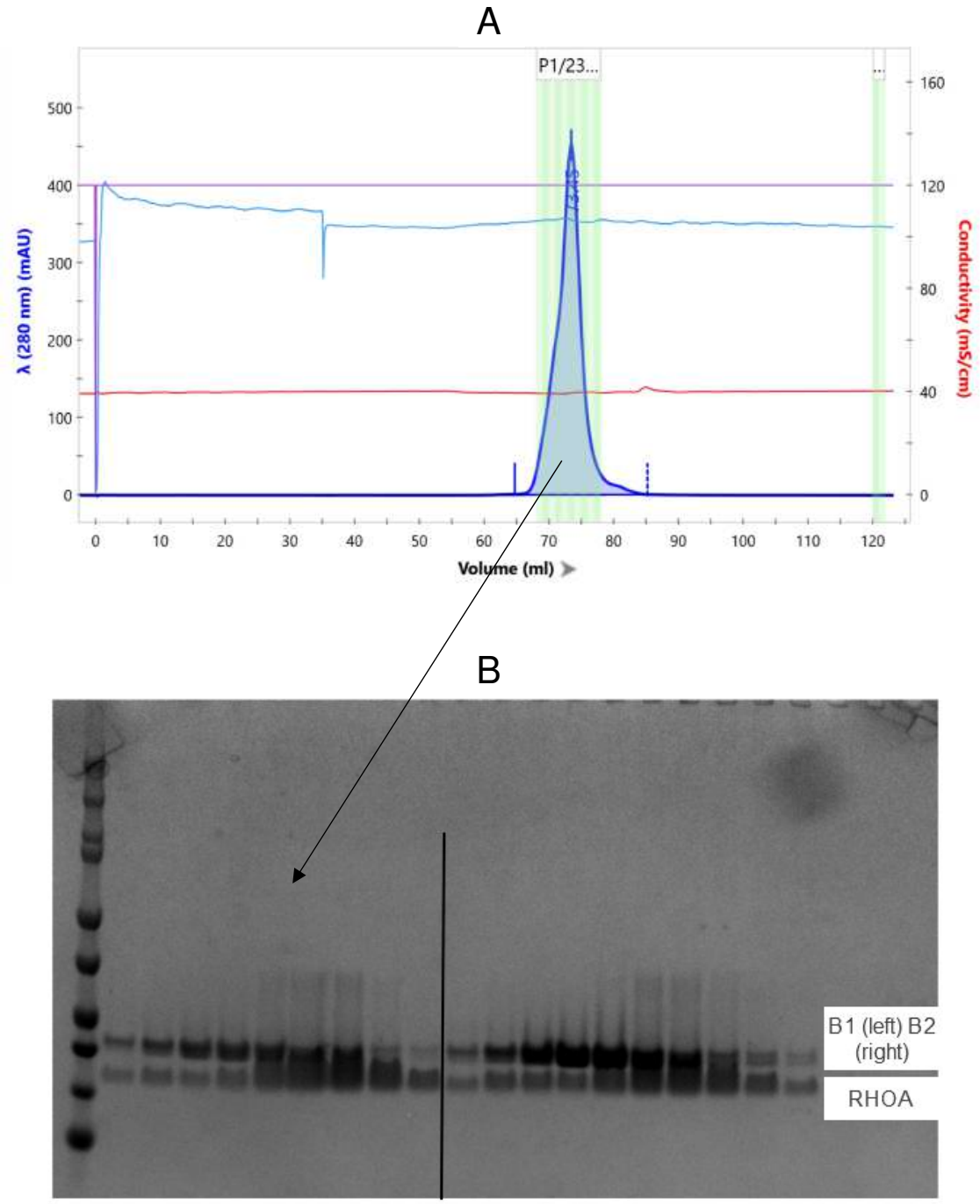


Figure 43 – *SEC purification of B1/RhoA and B2/RhoA (A) SEC UV purification trace of B1/RhoA. (B) Coomassie brilliant blue stain of fractions from SEC purification of B1/RhoA and B2/RhoA.*

3.2.2 Crystallisation screening

Conditions to crystallise the GEF/GTPase complex were explored next, with a stoichiometry of 1:1 and a concentration of 10–20 mg/mL (Table 8). Toward **A9**/RhoA, two plates were set up at 10 mg/mL using the coarse screen JCSG7 from Hampton Research, but very limited precipitation was observed (Entries 1 and 2). This prompted an increase in protein concentration and an alternative coarse screen (Entries 3 and 4). Whilst the level of precipitation increased, no crystallisation was observed. Later, microcrystals obtained from construct **B1** were seeded to encourage crystallisation, which still yielded no crystals (Entries 5). Crystallisation attempts for this construct were subsequently abandoned.

The results for the remaining mixtures were promising. After employing the coarse JCSG7 screen for **B1**/RhoA, a crystal was observed in 20% PEG3350/0.2 M ammonium formate (Entry 6). However, the crystal was extremely thin and needle-like and diffracted poorly (Figure 44A). This prompted the design of follow up screens: one varied the concentrations of PEG3350 and ammonium formate (Entry 7), and one seeded the original JCSG7 screen with microcrystals collected from the original screen (Entry 8). Good crystal growth was observed across both screens and multiple crystals were successfully mounted for diffraction (Figure 44B and C). For construct **B2**, two crystals from the JCSG7 screen were observed and mounted; the diffraction data was poor and follow up screens were not attempted due to the **B1** and **B2** constructs only differing by three amino acids (Entry 9).

Entry	Construct	Concentration (mg/mL)	Screen	Temperature (°C)	Crystals
1		10	JCSG7	20	
2		10	JCSG7	4	
3	A9 (DH-PH)	20	JCSG7	20	
4		20	Morpheus	20	
5		20	JCSG7 (seeded)	20	
6		23	JCSG7	20	
7	B1 (DH)	18	Custom Screen **1**	20	
8		18	JCSG7 (seeded)	20	
9	B2 (DH)	21	JCSG7	20	

Table 8 – *Crystallisation screens for Kalirin-GEF(2)/RhoA. Red = no crystals, orange = < 2 crystals, green = > 20 crystals. Screens acquired from Hampton Research.*

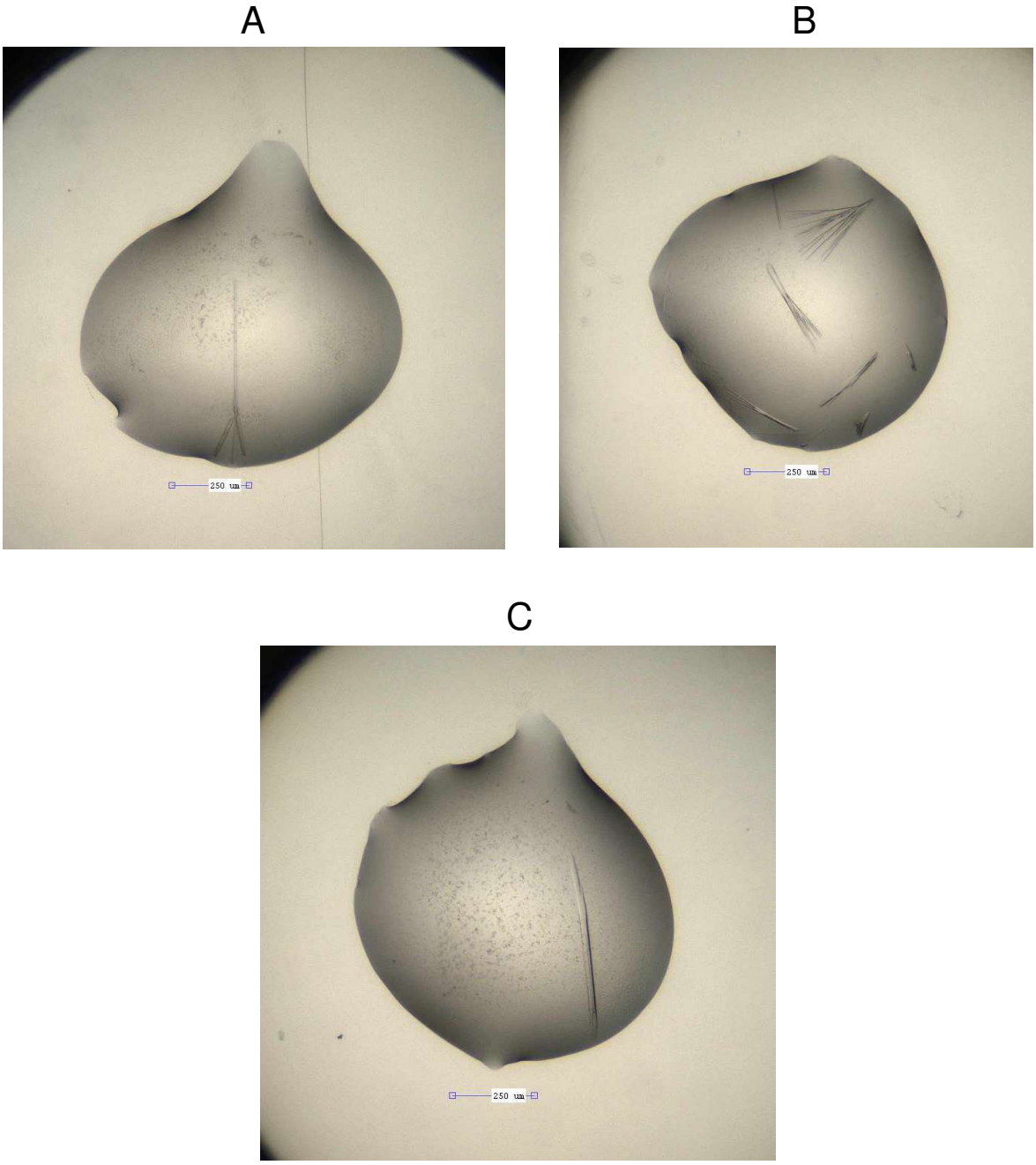


Figure 44 – *Optimisation of crystal growth of the B1/RhoA mixture. (A) First crystal identified in PEG3350 – ammonium formate. (B) Representative example of crystals obtained from a follow-up PEG3350/ammonium formate screen. (C) Crystal obtained in jeffamine ED-2003 / HEPES pH 7.0 with seeding.*

Many of the mounted crystals for the B1/RhoA mixture resulted in good diffraction (< 2 Å). The collected data was initially phased using MR with the known structure of RhoA and a model of Kalirin-GEF(2) generated by Phyre2.[257] Surprisingly, the phasing was unsuccessful. MR was instead performed with the separate components; using

the Kalirin-GEF(2) model alone resulted in a significant improvement in the phasing step, whereas the RhoA model phased in poor confidence. Upon refinement, the protein crystals were revealed to be the Kalirin DH2 domain only.

3.2.3 Novel structure of the Kalirin DH2 domain

The novel crystal structure of the Kalirin DH2 domain was successfully solved to 1.8 Å (PDB: 8C7D). The domain was found to have a characteristic Rho Dbl GEF fold consisting of a highly conserved six α-helical bundle with the three conserved regions intact (Figure 45). As the crystallisation of the domain was highly reproducible, using the XChem pipeline to identify binders was considered. Fragments binding near the GTPase binding surface could be explored as PPI inhibitors to prevent the GEF/GTPase interaction and thus GTPase activation. Such a campaign could be pursued as an alternative approach if no fragments are identified for the Kalirin-GEF(2)/RhoA complex.

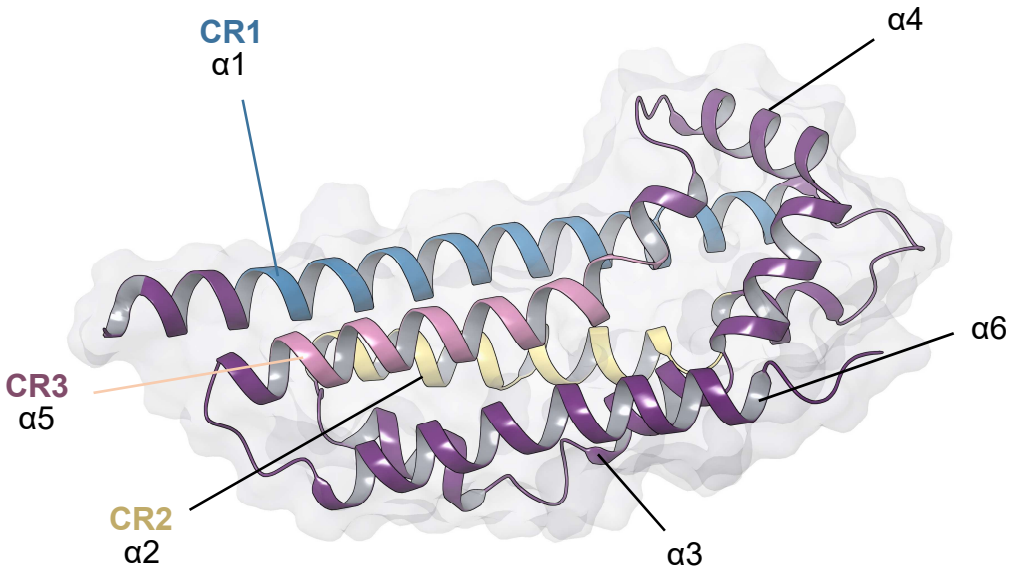


Figure 45 – *Novel crystal structure of the Kalirin-GEF(2) DH domain (PDB: 8C7D). Conserved regions (CR) 1 (blue), 2 (yellow) and 3 (pink) are highlighted.*

3.2.4 Encouraging Kalirin-GEF(2)/RhoA crystallisation

The co-crystallisation of the Kalirin-GEF(2)/RhoA complex was still required. As Kalirin DH2 construct **B1** was shown to active RhoA in nucleotide exchange (Section 3.1.3), there was confidence that the components were able to bind sufficiently in solution. Instead of attempting to improve this binding, the superior crystallisation of the DH2 domain was investigated. After visually analysing the packing contacts in the structure, several contacts were found to be in proximity of the N-terminal of the domain. When limiting the distance to 10 Å, six **B1** molecules were proximal to the starting residue E1916 (Figure 46). It was hypothesised that the N-terminal could be extended to disrupt these cooperative interactions and therefore discourage the preferential crystallisation of **B1**. Four new constructs were designed for this aim (Table 9).

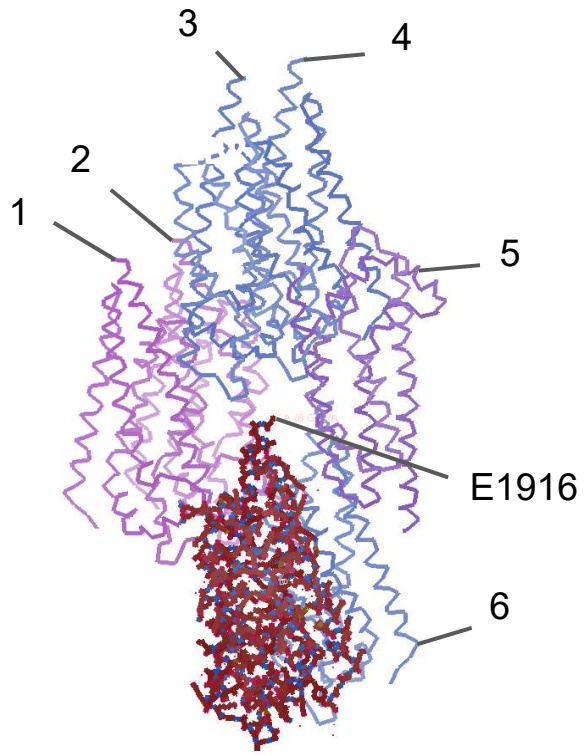


Figure 46 – *Close packing contacts (10 Å) to E1916 in the crystal structure of construct B1.*

Table 9 – *New Kalirin-GEF(2) DH constructs designed to discourage N-terminal packing. Residue positions as per UniProt: O60229. Original construct B1 is also shown.*

Entry	Start	End
B1	1916	2114
C1	1907	2114
C2	1903	2114
C3	1907	2122
C4	1903	2122

The constructs were ordered from Twist Biosciences but were not expressed and purified as part of this thesis. The production of the new constructs, followed by crystallisation trials, should be performed as part of future work. Crystallisation of the complex should also be optimised for XChem screening, with the aim of obtaining hits to use as starting points for chemical probe development.

3.3 Fragment screening

The next objective was to provide a proof-of-concept that the Kalirin-GEF(2)/RhoA complex was amenable to small molecule inhibition. Based on the suspected autoinhibitory mechanism of the PH2 domain, the DH2 domain was used exclusively in this study (construct **B1**). The FI nucleotide exchange assay was utilised for inhibitor screening in the absence of a Kalirin-GEF(2)/RhoA crystallisation system.

3.3.1 Limited fragment screen

A limited library of previously synthesised compounds for Kalirin-GEF(1)/Rac1 was used first. Given that the endogenous ligands are the same for both targets, it was predicted that these rationally designed fragments were more likely to be active vs. a non-specific purchased library. In total, 21 fragments were screened against Kalirin-GEF(2)/RhoA at 500 µM, three of which were identified as having significant activity (Figure 47). The hits were those previously identified for Kalirin-GEF(1)/Rac1; **MC-278**, **RS-009** and piperidinopyridone **2**.

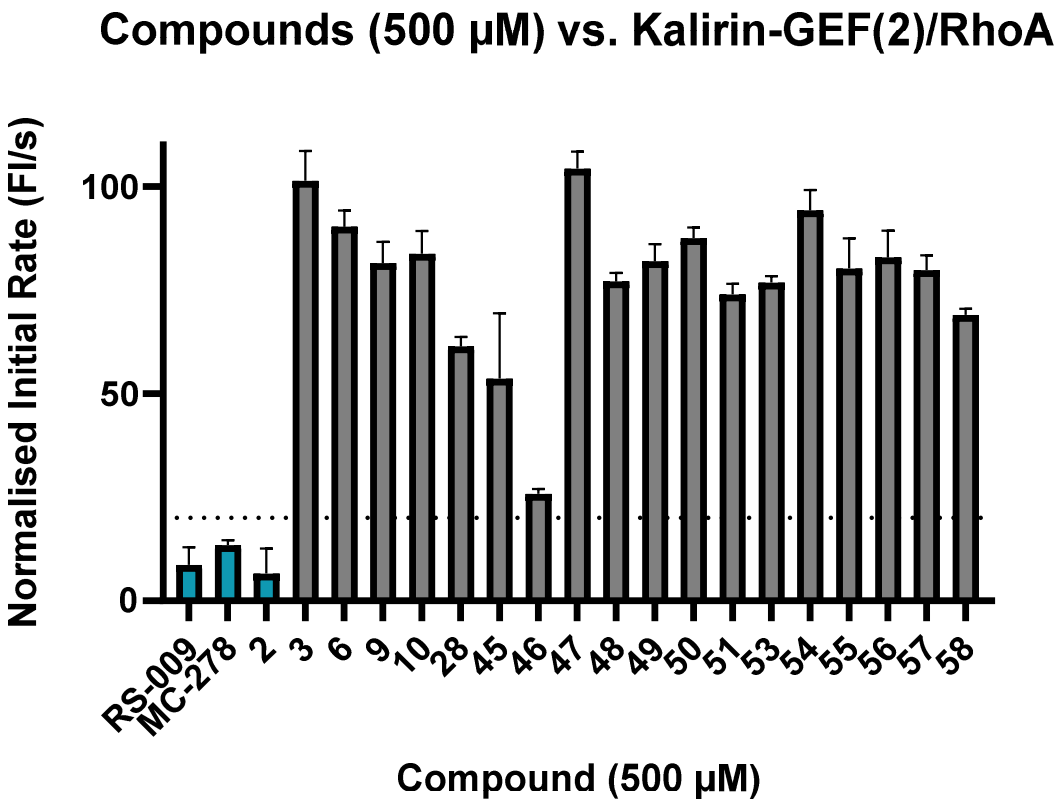


Figure 47 – *Limited fragment screen against Kalirin-GEF(2)/RhoA.*

Dose response curves were subsequently obtained. **MC-278** and **RS-009** were found to inhibit Kalirin-GEF(2)/RhoA within the µM range, much like their inhibition of Kalirin-GEF(1)/Rac1 (Figure 48). The binding modes were theorised to be comparative to those observed for Kalirin-GEF(1)/Rac1. Pyridone **2**, the only non-covalent hit, was initially calculated to have IC_{50} of 39 µM (Figure 49). However, the dose response curves had irregularly large Hill Slopes that were typical of precipitation or aggregation, as observed for Kalirin-GEF(1)/Rac1. The binding of the fragment could not be quantitatively described based on the ambiguity of this bioactivity data.

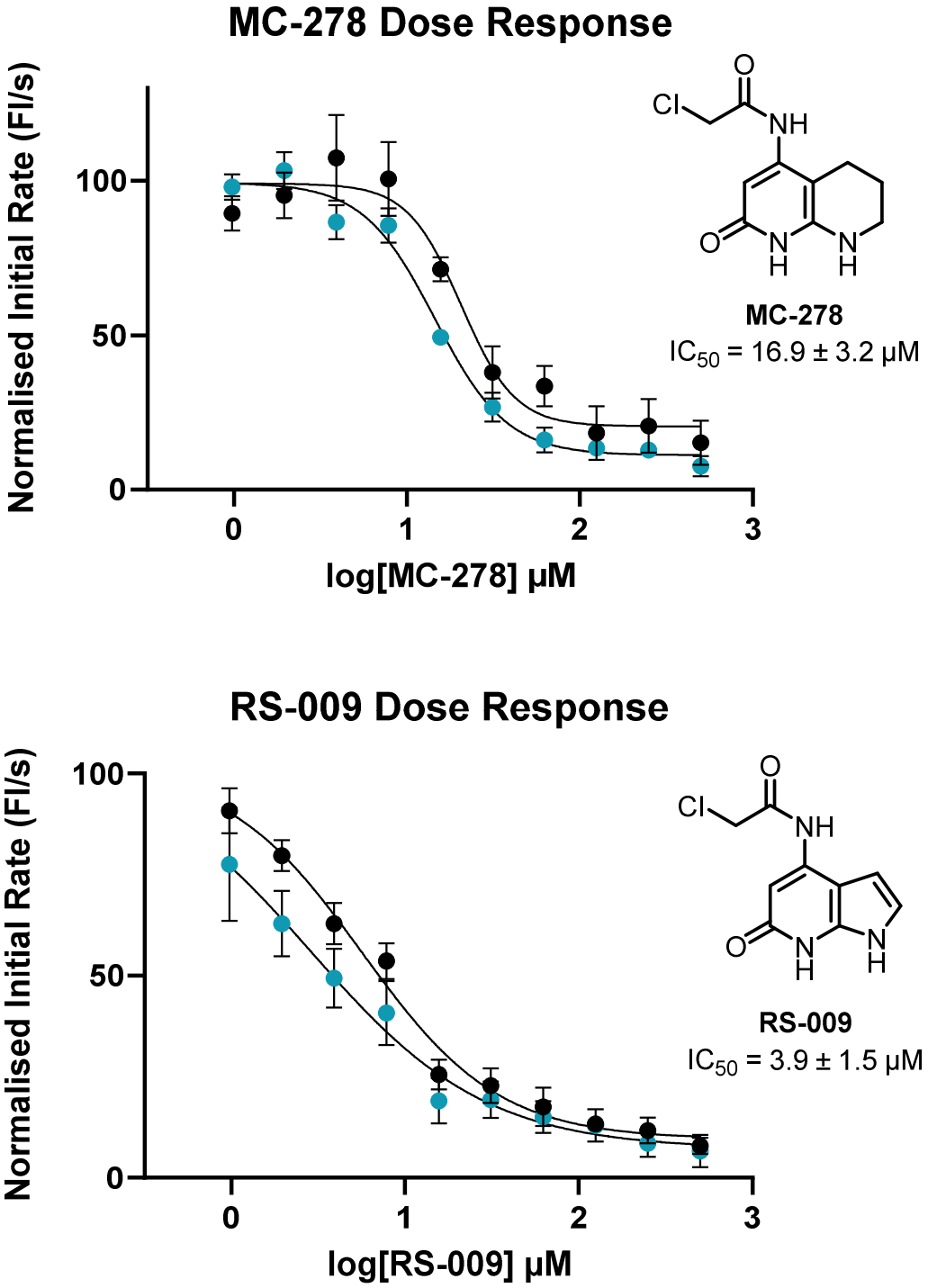


Figure 48 – *Dose response curves of MC-278 and RS-009 against Kalirin-GEF(2)/RhoA. n = 1 (black) and n = 2 (blue) are shown.*

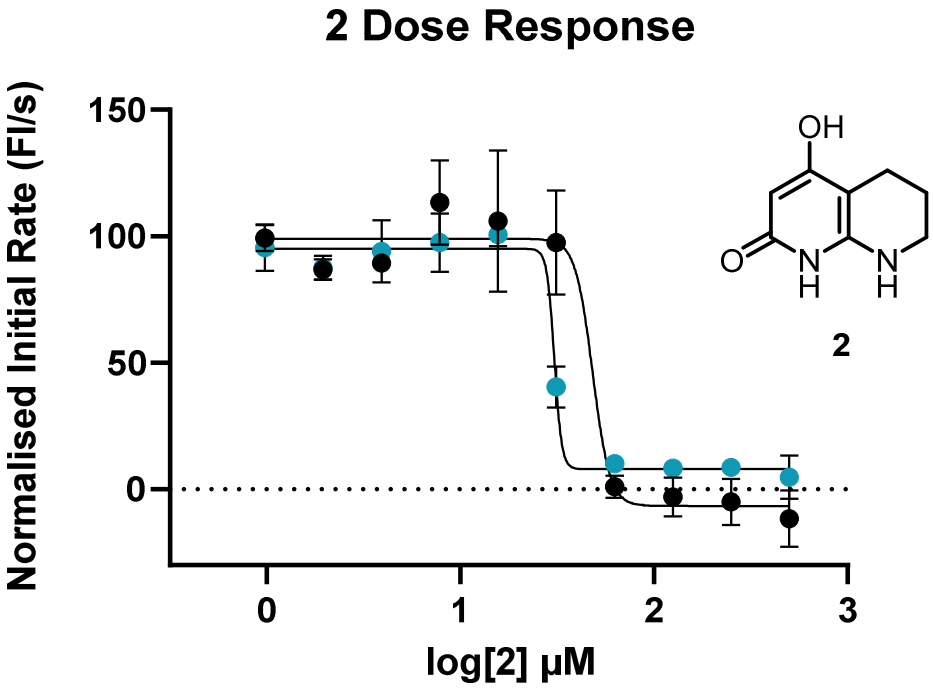


Figure 49 – *Dose response curve of piperidinopyridone 2 against Kalirin-GEF(2)/RhoA. n = 1 (black) and n = 2 (blue) are shown.*

The data provides a proof-of-concept that the complex has the potential to be inhibited by small molecules. Based on the data obtained for Kalirin-GEF(1), it was postulated that **MC-278** and **RS-009** also covalently bind in the nucleotide binding pocket of the Kalirin-GEF(2)/RhoA complex, thus preventing nucleotide exchange. The fragments could be used as starting points for the development of a chemical probe for Kalirin-GEF(2)/RhoA. A binding analysis should be performed to confirm the postulated mechanism, and new target molecules should be devised to explore fragment affinity by obtaining an SAR analysis. With active Kalirin-GEF(2) and RhoA in-house, a library screen could also be performed using the FI nucleotide exchange assay. Such screens may reveal diverse inhibitors to **MC-278** and **RS-009** that could be developed to selectivity bind Kalirin-GEF(2)/RhoA over Kalirin-GEF(1)/Rac1.

3.4 Conclusions

In summary, this chapter describes the early progress made towards probing the Kalirin-GEF(2)/RhoA complex. The production of active Kalirin-GEF(2) was successfully demonstrated, which is necessary for future crystallography studies or screening campaigns. Moreover, preliminary studies were performed in both of these areas.

To provide active Kalirin-GEF(2) a parallel expression-purification approach was applied, which led to the successful production of Kalirin-GEF(2) constructs. The GEF activity of the DH and DH-PH domains were explored in the FI nucleotide exchange assay, and it was postulated that the PH2 domain of Kalirin is autoinhibitory in nature. This is an important consideration for construct selection for future screening campaigns.

Attempts were subsequently made to crystallise the novel Kalirin-GEF(2)/RhoA complex for XChem fragment screening. This resulted in the novel crystal structure of the Kalirin DH2 domain which exposed its characteristic Rho Dbl GEF fold. Although the GEF/GTPase co-crystal structure was not obtained, new Kalirin-GEF(2) constructs were designed to discourage the preferential crystallisation of the DH2 domain; production of these constructs remains a priority in future work, which could ultimately lead to solving the GEF/GTPase structure and fragment screening by XChem.

Finally, the inhibition of the Kalirin-GEF(2)/RhoA complex was explored using the FI nucleotide exchange assay. Both **MC-278** and **RS-009** were found to inhibit BODIPY-GTP binding in the μM region, supporting the notion that small molecules can prevent

nucleotide exchange on Kalirin-GEF(2)/RhoA. Alternative screens should be applied in future work that may reveal diverse inhibitors to **MC-278** and **RS-009** that could be developed to selectivity bind Kalirin-GEF(2)/RhoA over Kalirin-GEF(1)/Rac1. These fragments could be used as starting points for chemical probe development, or as control compounds in larger library screens.

4 Predicting new GEF/GTPase targets

The final objective of this thesis was to widen the scope of GEF/GTPase targets that could be considered for inhibition by small molecules, given the successes described in Chapters 2 and 3 showing multiple complexes may be amenable to small molecules. It was anticipated that a comprehensive resource on Rho Dbl GEF activity at the molecular level would offer an improved understanding of GEF/GTPase complementary pairs and was hypothesised to aid the prediction of unknown GEF/GTPase complexes, and thus new targets for GEF-associated diseases.

4.1 Rho Dbl GEF structure analysis

The first aim was to select an efficient method of comparing the structures of Rho Dbl GEFs. It is well documented that the GEFs have a characteristic fold at the DH domain comprising a 6-α helical bundle, described in detail in Section 1.6.7. The suggested strategy, therefore, was the 3D structural alignment of Rho Dbl GEFs to analyse the spatial conservation of residues. Highly conserved residues at the GTPase binding surface were expected to play a crucial and ubiquitous role in the GEF activity of DH domain containing proteins. Furthermore, conserved residues in GTPase-specific GEFs were hypothesised to reveal selectivity markers for future predictions. With the release of the EMBL-EBI AlphaFold protein structure database described in Section 1.5, predictive models of the entire human genome are readily accessible and thus were used in the absence of crystal structures.

4.1.1 Selection criteria and alignment protocols

Rho Dbl GEFs were selected for this study based on three criteria: (i) the availability of either a crystal structure or AlphaFold prediction of the Rho Dbl GEF DH domain, and if using the AlphaFold prediction, (ii) a per-residue confidence score (pLDDT) of the secondary structure greater than 70 and (iii) the presence of a complete DH domain.[82] This resulted in 70 GEF DH domain structures that were truncated as per the DH domain boundaries dictated in UniProt. The structures were aligned utilising the Protein Structure Alignment tool in the Maestro suite (Schrödinger). As the Kalirin-GEF(1)/Rac1 structure and mechanism was described in detail in Section 1.6.7, the structure was used as the alignment reference and model system (PDB: 5O33, Chain B).[125] Residue side chains on the helices overlapped well (RMSD 0.711–3.940 Å for alpha and beta carbons), which encouraged the suitability of the alignment approach.

A spatially conserved residue was flagged when >80% of GEFs comprised the same type of residue at the same position: acidic, basic, polar, non-polar, and aromatic. All residues were referenced to the corresponding position on Kalirin-GEF(1) and denoted by that position (Supplementary Table 4).

4.1.2 Highly conserved residues

Overall, nine positions on the GTPase binding surface were identified as being highly conserved, three residing on the α1 helix (CR1) and six on the α5 helix (CR3) (Figure 50, Supplementary Table 5 and 6). Using the crystal structure of Kalirin-GEF(1)/Rac1 as a model system, the conserved residues on CR1 were shown to interact exclusively with switch I of the GTPase, and those on CR3 with either switch I or II (Figure 51).

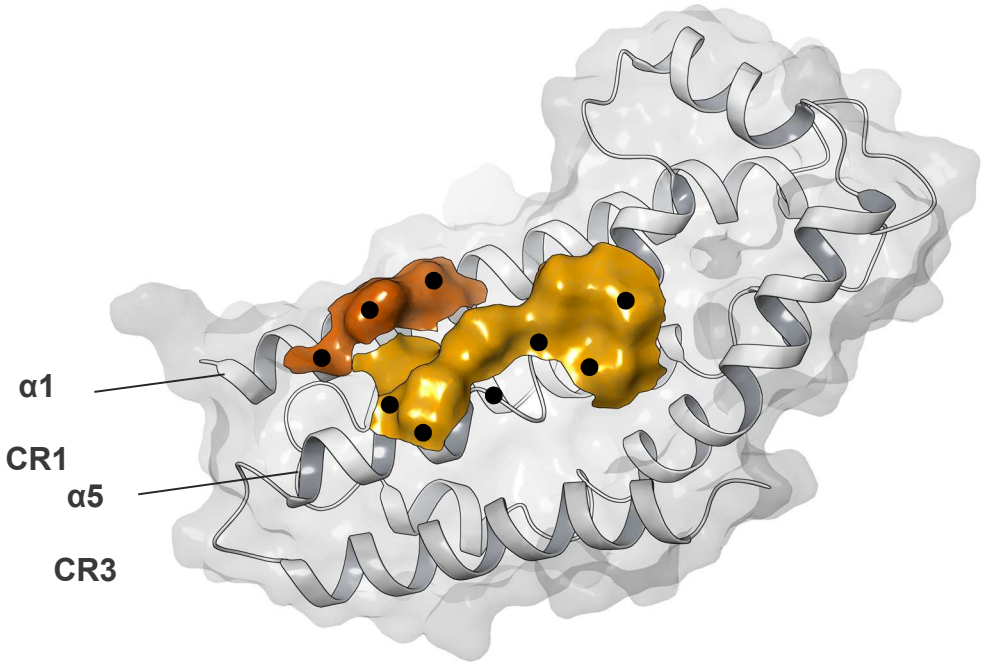


Figure 50 – *Locations of highly conserved residues of the Rho Dbl GEFs identified by 3D alignment. The crystal structure of Kalirin-GEF(1) is shown as the model system (grey ribbons and surface). Conserved residues were found at CR1 (orange surface) and CR3 (yellow surface). Only GTPase binding surfaces were considered.*

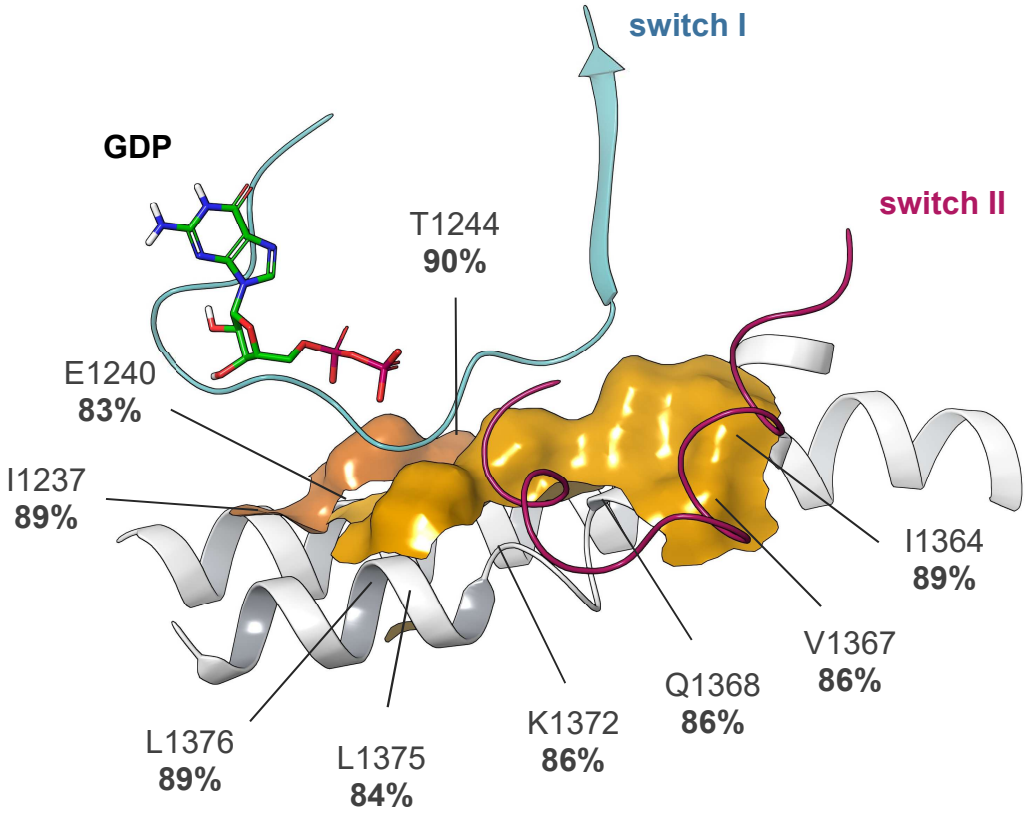


Figure 51 – *Conserved residues of the Rho Dbl GEFs and proximity of the switch I and switch II regions of the GTPase. Highly conserved residues on CR1 (orange surface) and CR2 (yellow surface) are shown with the % of GEFs containing the same type of residue at the same position. Switch I (blue ribbons) and switch II (magenta ribbons) are shown. Kalirin-GEF(1)/Rac1 was used as both a model system and alignment reference.*

The redundancy of these residues across the Rho Dbl GEFs suggests their contribution to GEF activity during the nucleotide exchange process, and their importance for Rho GTPase selectivity. To further corroborate this hypothesis, the interactions of the residues with the GTPase were explored with the Kalirin-GEF(1)/Rac1 model system. Mutational examples from the literature were also used as further evidence.

The three conserved positions on CR1 (I1237, E1240, T1244) interact with Y32, T35 and V36 of the switch I region of Rac1 (Y34, T37, V38 in RhoA and Y32, T35, V36 in

Cdc42) (Figure 52). The movement of these GTPase residues is critical for disrupting interactions that contribute to the tight binding of GDP and therefore nucleotide exchange; most notable is spatial removal of T35 from the nucleotide binding site upon GEF binding, which usually interacts with the bound Mg^{2+}. One can conclude that the combined contribution of these GEF residues is imperative for the movement of the switch I region during nucleotide exchange. This influence is especially pronounced in site-directed mutational studies; alanine substitutions of the equivalent E1240 residue in GEFs Trio1[258] (E1240A) and PDZ-RHOGEF[259] (E741A) and of the equivalent T1244 residue in Trio1[258] (T1244A) all decreased GDP exchange by up to 20-fold.

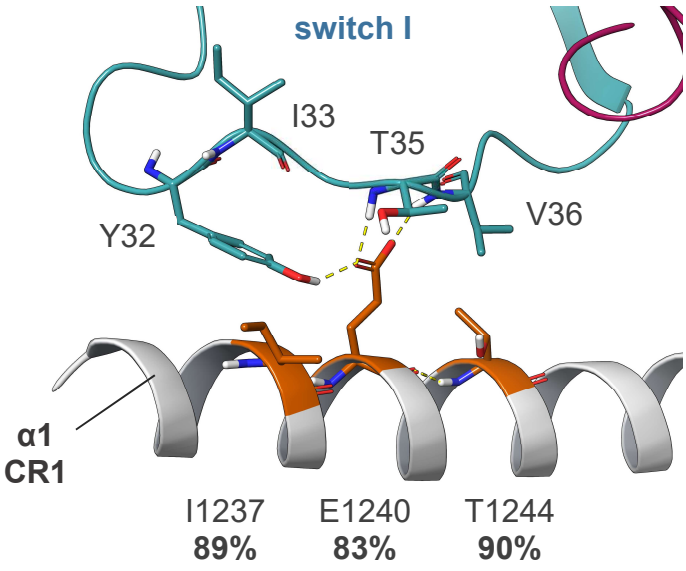


Figure 52 – *Conserved positions of Rho Dbl GEFs on CR1 (dark orange ribbons and sticks, denoted by the corresponding position on Kalirin-GEF(1)) and interactions with the switch I region of the GTPase (teal ribbons and sticks, denoted by Rac1).*

Residues at positions Q1368, L1375 and L1376 on CR3 were also found to interact with the switch I region in the Kalirin-GEF(1)/Rac1 model (Figure 53). Q1368 was observed to hydrogen bond to N39 of Rac1, which remains relatively unperturbed during GEF binding. This implies the interaction contributes to overall GEF-GTPase

affinity rather than having a catalytic role. The corresponding alanine mutation in Trio1 (Q1368A) significantly reduced the GEF activity toward Rac1, demonstrating the influence of the conserved residue.[258] Residues L1375 and L1376 form a hydrophobic accommodating patch that stabilises the positioning of T35 of Rac1 during remodelling. The corresponding alanine substitutions of L1375 in GEFs Trio1 (L1375A)[258] and Tiam1 (L1198A)[260] resulted in a notable decrease in GEF activity toward Rac1, supporting the importance of this reoccurring motif.

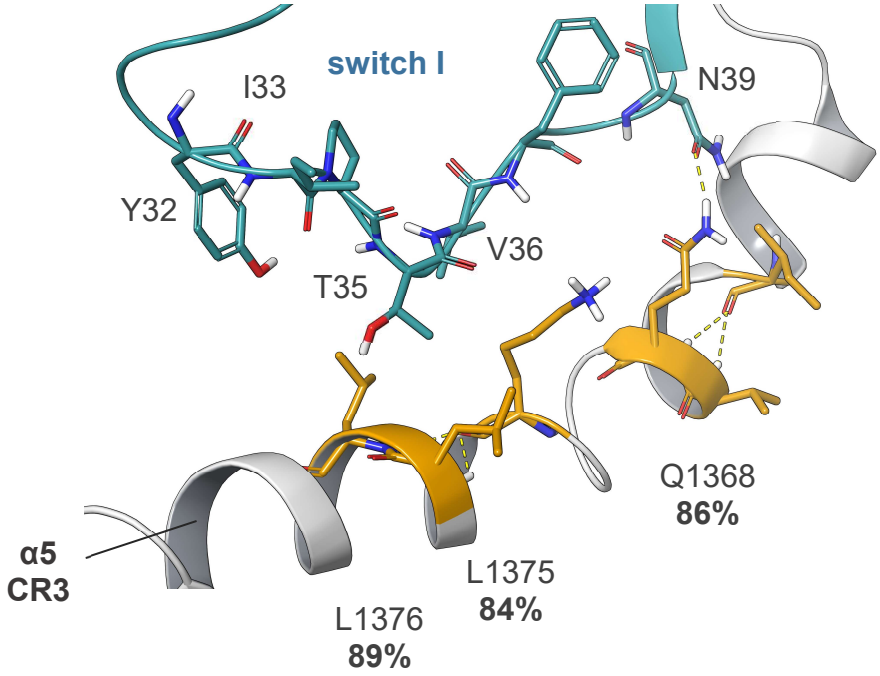


Figure 53 – *Conserved positions of Rho Dbl GEFs on CR3 (yellow ribbons and sticks, denoted by the corresponding position of Kalirin-GEF(1)) interacting with switch I of the GTPase (blue ribbons and sticks denoted by the corresponding position on Rac1).*

Finally, conserved residues I1364, V1367 and K1372 were found to interact with the switch II region of the GTPase from the reference model (Figure 54). Non-polar residues I1364 and V1367 form a hydrophobic cleft that accommodates L67 and L70 of Rac1; this region of switch II remains unchanged during nucleotide exchange,

highlighting its role in anchoring and orienting critical GEF residues and stabilising the overall interaction. The corresponding alanine mutation in Trio1 (V1367A) reduced the activity of the GEF to <20%, demonstrating its influence.[258] Furthermore, mutations to alternative hydrophobic groups of the equivalent I1364 residue in PDZ-RHOGEF (I876) significantly impacted the GEF activity toward RhoA.[261]

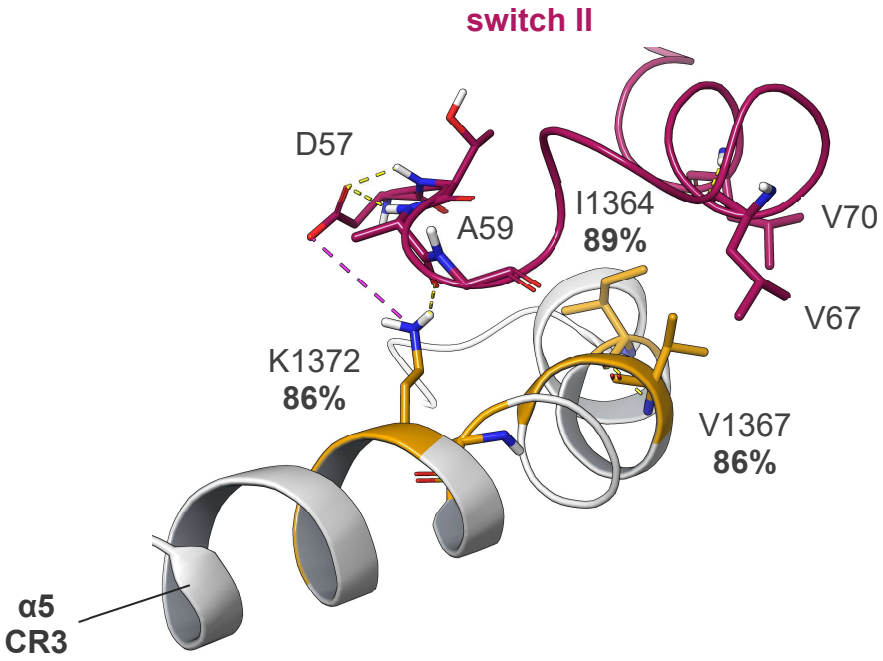


Figure 54 – *Conserved positions on Rho Dbl GEFs on CR3 (yellow ribbons and sticks, denoted by the corresponding position on Kalirin-GEF(1)) interacting with switch II of the GTPase (magenta ribbons and sticks, denoted by the corresponding position on Rac1).*

The conserved basic side chain at K1372 forms a backbone interaction with A59 and a salt bridge with the side chain of D57 in Rac1. The remodelling of this region of switch II is crucial, where the transposition of A59 ultimately occludes the Mg^{2+} co-factor during nucleotide exchange. Corresponding mutations of the equivalent basic residue of Trio1 (K1372A)[258] and Tiam1 (K1195A)[260] were deleterious to GEF activity, only further confirming its role.

The α6-helix was also explored for spatial conservation, which has been reported to interact with the switch II region (Supplementary Table 7). This analysis identified no position at the GTPase binding surface with a redundancy of >80%. This suggests a more versatile role of the α6-helix during nucleotide exchange.

4.1.3 Substrate specificity markers

The next objective was the identification of regions or residue positions that influence GTPase specificity. Rho GTPases display high amino acid sequence and structural homology within the family, leaving only a few surface residues for the GEFs to exploit for selectivity. Previous work has highlighted the importance of the β1, 2 and 3 sheets of the GTPases in this respect, where five positions are notably different in Rac1, RhoA and Cdc42 (Figure 55).[222, 232, 260, 262-266]

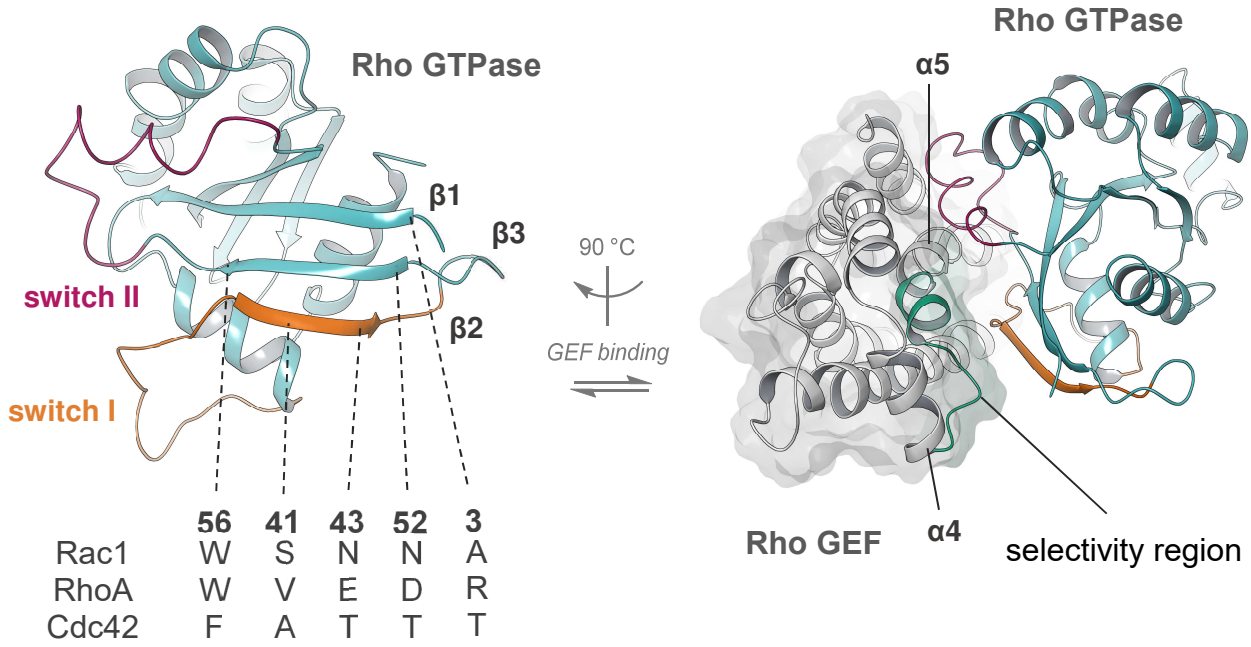


Figure 55 – *Structural differences between Rac1, RhoA and Cdc42. Kalirin-GEF(1) (grey ribbons and surface) and Rac1 (teal ribbons) are used as a reference model. Relevant regions β1–3 are shown with switch I (orange ribbons) and switch II (magenta ribbons). The 'selectivity region' of the GEF is highlighted (green ribbons and surface).*

This region was explored using the Kalirin-GEF(1)/Rac1 model. The region of the GEF that binds to the β1–3 sheets was found between the α4 and α5 helices and partially included the α5 helix; this will be referred to as the selectivity region henceforth. The aim was to analyse the aligned GEF models at the selectivity region and identify indicators of substrate specificity.

First, the literature was explored to identify reported GEF/GTPase partners focusing solely on the most highly studied Rho GTPases Rac1, RhoA and Cdc42 (Figure 56, Supplementary Table 8). This restriction was applied due to the limited activation data available for other Rho GTPases. The agreement of GEF/GTPase activation across studies was more consistent when limiting the analysis to nucleotide exchange-based assays with purified protein. Consequently, a GEF was only considered to activate Rac1, RhoA or Cdc42 with this *in vitro* data; 44 Rho Dbl GEFs were assigned GTPase partners with these restrictions (Supplementary Table 8). RhoA was found to be activated by the most GEFs exclusively and Cdc42 the least. It is also important to note that a GEF was not necessarily reported to active one GTPase only, demonstrating the versatility of these proteins.

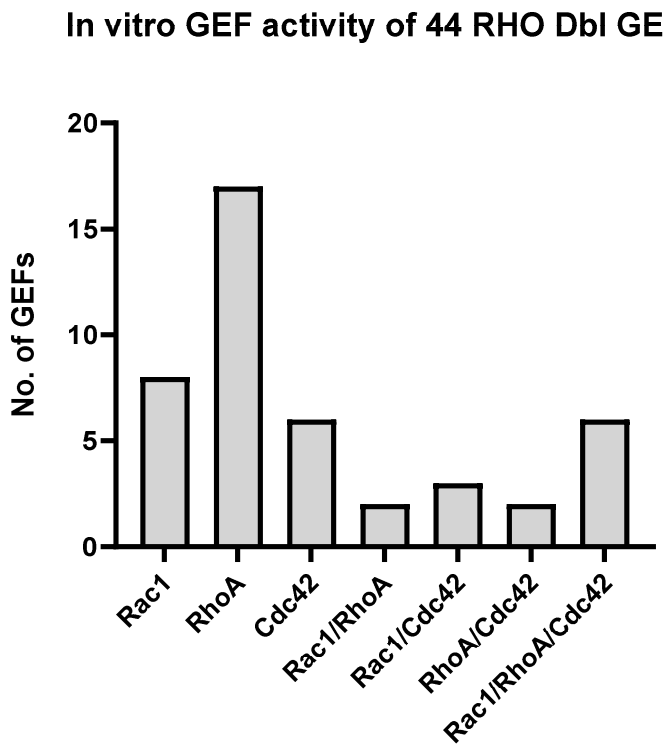


Figure 56 – *Rho Dbl GEF activation identified from the literature considering in vitro nucleotide exchange data with purified protein.*

4.1.4 Rac1 selectivity

The selectivity regions of the Rac1 activating GEFs were visually analysed for reoccurring motifs, specifically those on the GTPase binding surface. None were identified, and rather the selectivity regions varied in both sequence and architecture across the GEF models (Figure 57). The only commonality between the structures was the occurrence of a hydrophobic residue at I1364 in 84% of cases. This residue was previously highlighted as being highly conserved across all Rho GEFs, specifically in 89% of the 70 that were analysed. This suggests that this residue alone cannot account for Rac1 specificity. Other conserved residues observed from the alignment were not present at the GTPase binding surface, and likely contribute to the overall fold of the GEF.

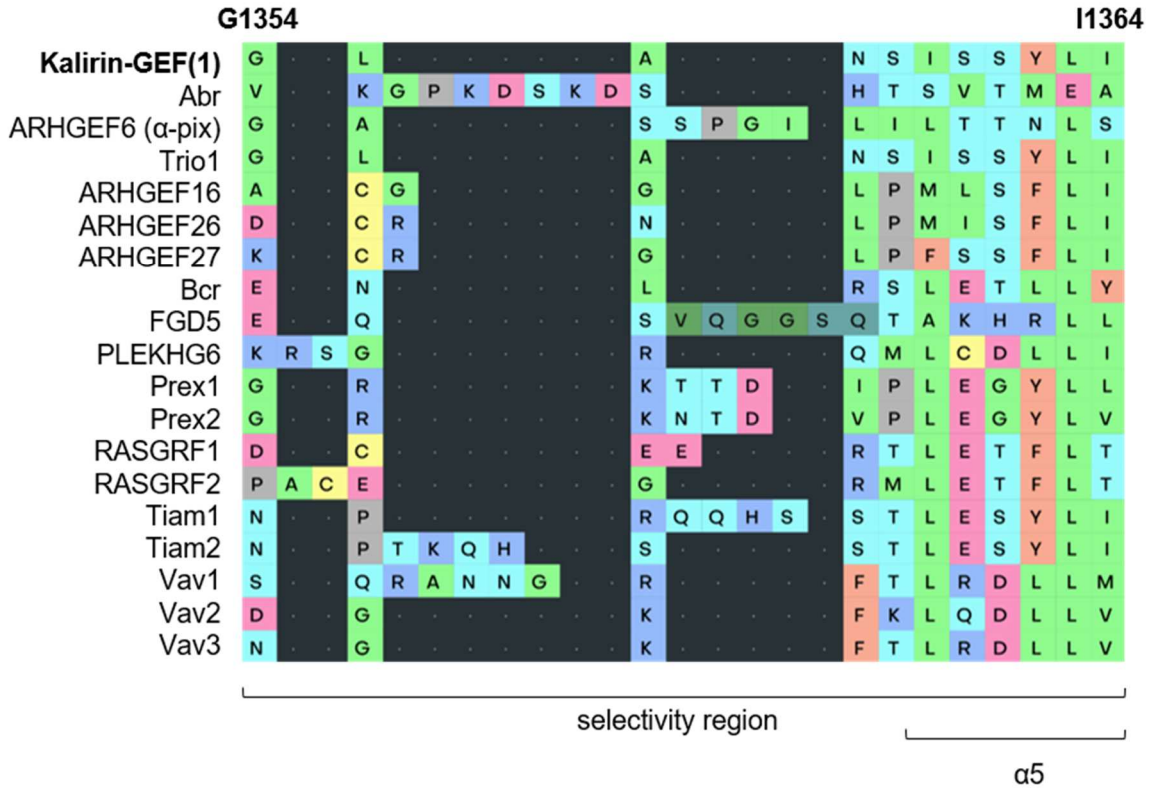


Figure 57 *– Selectivity region analysis of Rac1 activating GEFs. The models were structurally aligned, and then the sequences reported based on the 'current structure superposition' tool in the Multiple Sequence Viewer in Maestro, Schrodinger. Reference positions G1354 and I1364 are of Kalirin-GEF(1).*

Based on the alignment, it was postulated that Rac1 activating GEFs are able to enact their specificity through diverse mechanisms. This was supported by an analysis of available GEF/Rac1 crystal structures: Kalirin-GEF(1) (PDB: 5O33), Trio1 (PDB: 2NZ8), Tiam1 (PDB: 1FOE), Prex1 (PDB: 5FI0) and Vav1 (PDB: 3BJI). Tiam1/Rac1 was the first Rho GEF/GTPase crystal structure to be reported, which exposed an extensive interaction of the selectivity region of Tiam1 and β1-3 sheets of Rac1 (Figure 58).[260, 266] This interaction network was noticeably absent in the remaining Rac1 co-crystal structures. As represented by Trio1, the selectivity region is comparably spatially removed from the β1–3 sheets of the GTPase (Figure 59).

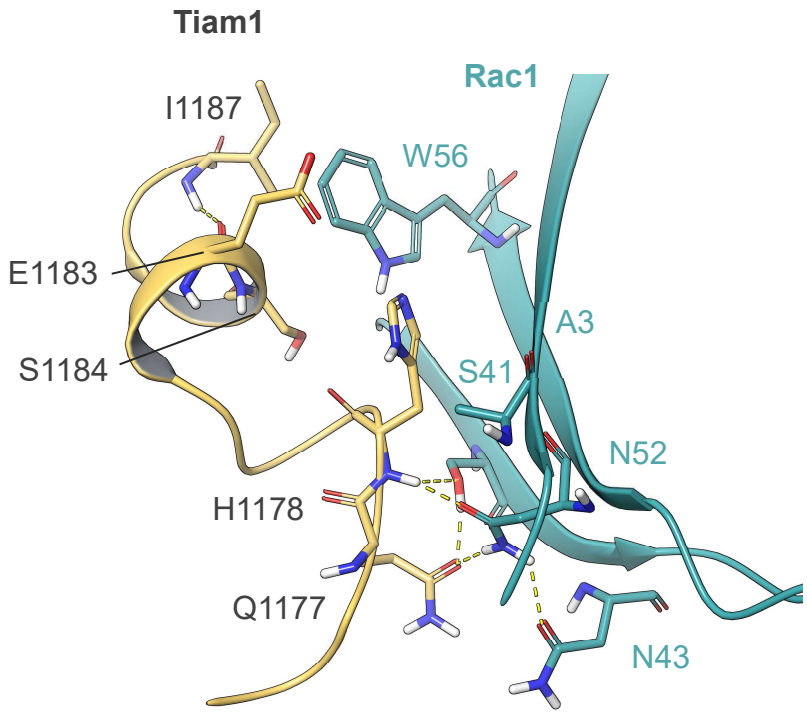


Figure 58 – *Selectivity region of Tiam1 (yellow ribbons and sticks) interacting with the β1–3 sheets of Rac1 (teal ribbons and sticks).*

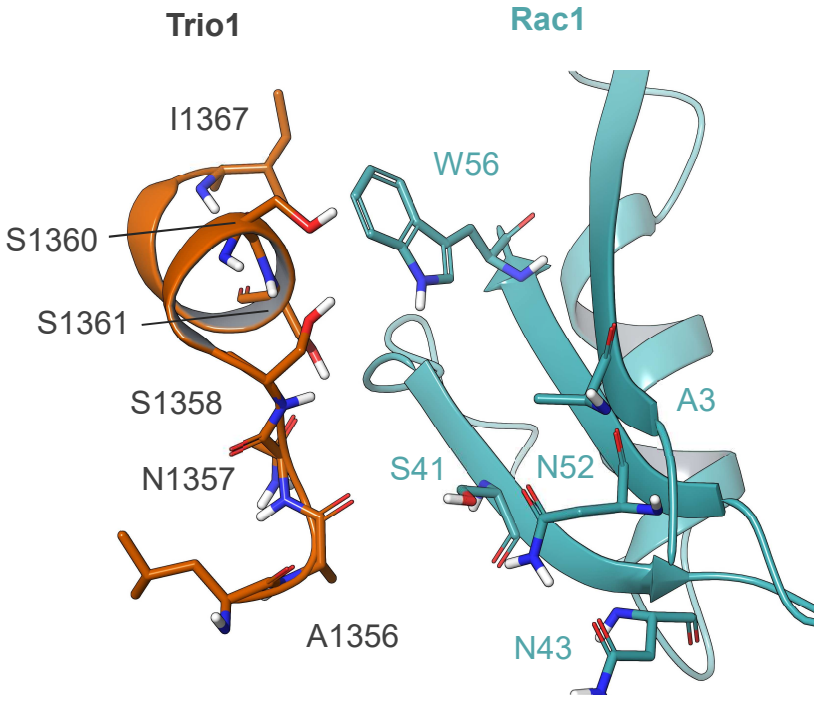


Figure 59 – *Selectivity region of Trio1 (orange sticks and ribbons) interacting with the β1–3 sheets of Rac1 (teal ribbons and sticks).*

The only retained interaction is with W56, which is sufficiently stabilised by the selectivity region in each case.

The importance of W56 on Rac1 activation was previously explored in mutational studies: a W56F mutation was enough to completely abolish Tiam1 binding and rendered the mutant capable of binding to Intersectin1, a Cdc42 specific GEF.[266] W56 was also found to be the determining residue for Trio1 recognition.[265] This accumulation of data suggests favourable interaction with W56 is key for Rac1 recognition, although it appears from this study that many GEF residues and topologies can make this interaction. Due to this diversity, predicting new GEF/Rac1 complexes could not be performed based on a statistical analysis.

4.1.5 RhoA selectivity

The selectivity region of RhoA activating GEFs was explored next. A common motif was clearly identified: 96% contained at least one basic residue protruding from the selectivity region toward the GTPase binding surface, with 85% containing the basic residue specifically at position R923 as referenced to the Leukemia-associated Rho GEF (ARHGEF12, Figure 60). The importance of this selective motif was further emphasised by the fact all GEFs with the basic side chain were reported to be activators of RhoA (where *in vitro* data was available). Adjacent to the conserved basic residue, at position L924, was also a hydrophobic residue in 74% of cases.

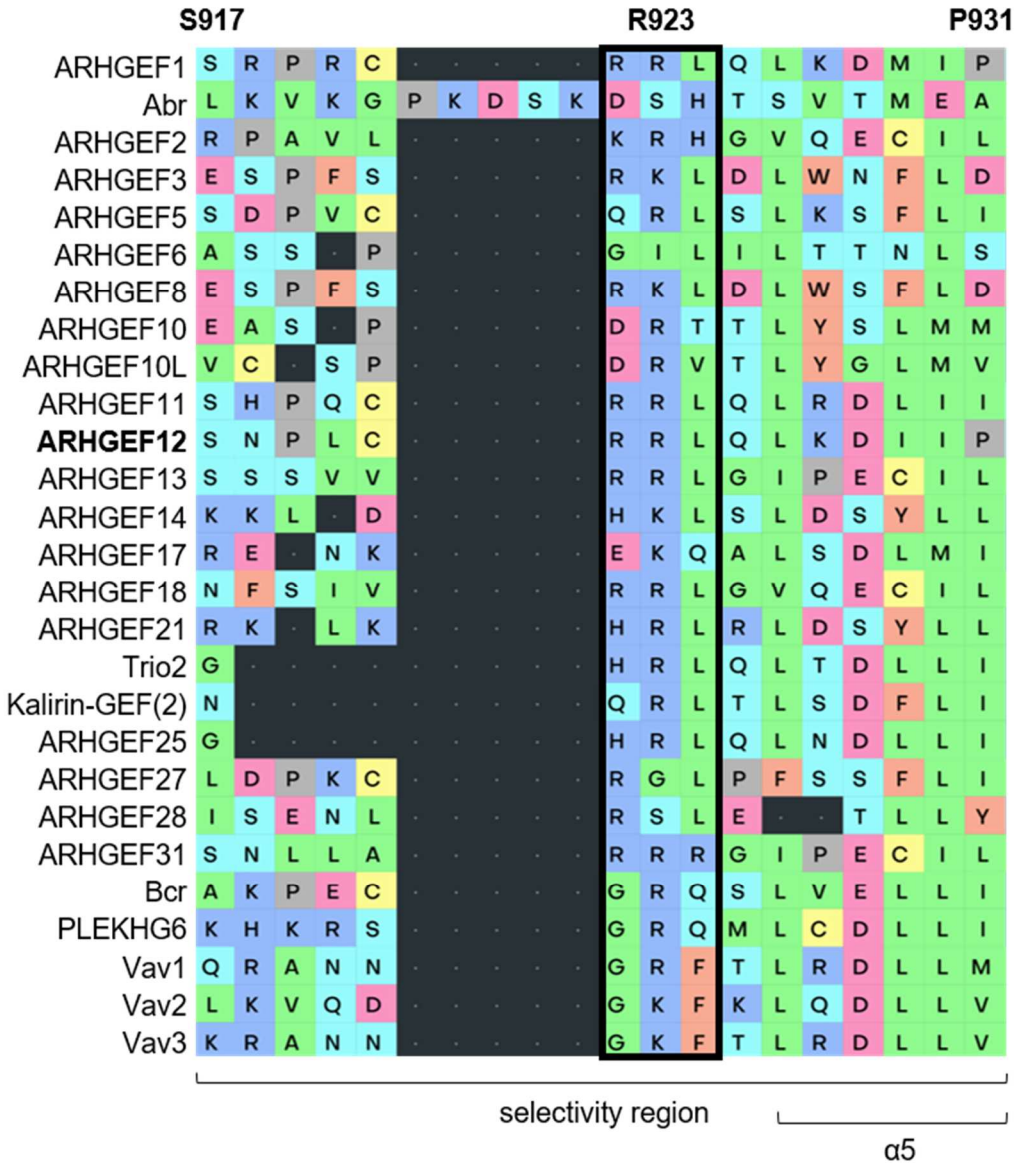


Figure 60 *– Selectivity region analysis of RhoA activating GEFs. Reference positions S917, R923 and P931 are of LARG, also known as ARHGEF12.*

The significance of this motif was explored with the known GEF-RhoA structures from the PDB: ARHGEF8 (NET1, PDB: 4XH9), ARHGEF11 (PDZ-RHOGEF, PDB: 1XCG), ARHGEF12 (LARG, PDB: 1X86), ARHGEF13 (AKAP13, PDB: 4D0N), ARHGEF14 (Dbs, PDB: 1LB1) and ARHGEF25 (GEFT, PDB: 2RGN). Each contained the identified basic side chain, which was found to interact with D45 and E54 of RhoA through a highly complementary dual salt bridge interaction. This is exemplified by R923 of

LARG (Figure 61). The adjacent hydrophobic residue – illustrated by L924 of LARG – forms part of a hydrophobic cradle that interacts favourably with V43 and W58 of RhoA.[267] V43 is substituted for polar residues S41 and T41 in Rac1 and Cdc42 respectively, highlighting the additional selectivity handle.

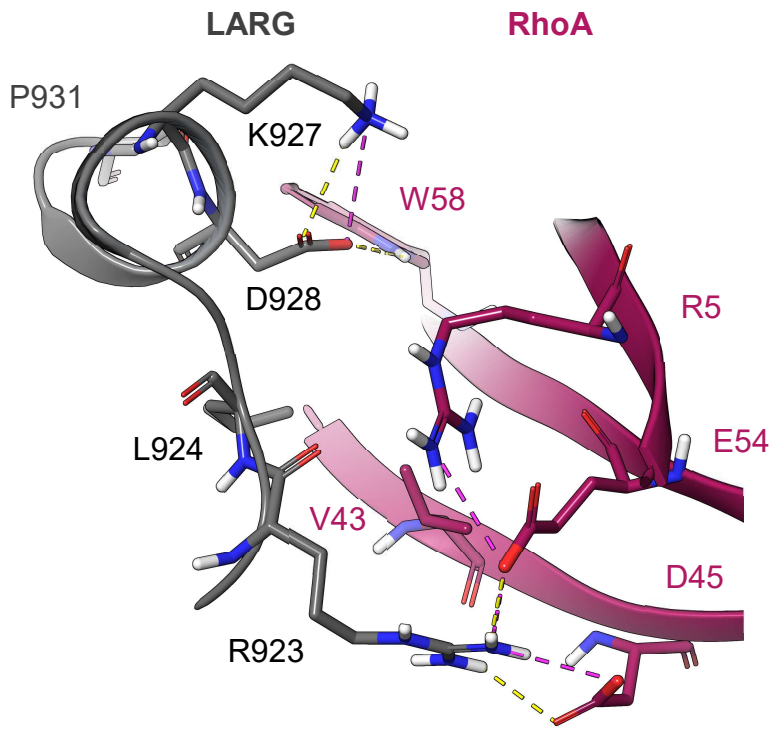


Figure 61 – *Selectivity region of LARG (grey ribbons and sticks) interacting with the β1–3 of RhoA (magenta ribbons and sticks). Hydrogen bonding interactions (yellow dashes) and salt bridge interactions (pink dashes) are shown.*

Multiple mutational studies support the importance of this selectivity motif. The removal of the corresponding basic residue in PDZ-RHOGEF (R868G)[261] was deleterious to GEF activity, and the mutation of the equivalent Lys residue (K758A) in Dbs[262] significantly reduced exchange. The influence of V43, and therefore the contribution of the non-polar side chain in RhoA activating GEFs, is perhaps most pronounced in the study of ARHGEF3. The GEF is known to activate RhoA and RhoB,

but not the very closely related RhoC, which has an I43 residue in the place of V43. An I43V mutation of RhoC allowed activation by ARHGEF3, demonstrating the sensitive influence of this single residue.[268]

Fine-tuning both of the identified positions was also found to induce RhoA recognition: the G1368K/M1369L double mutation of ITSN1 was successful in altering wild-type Cdc42 nucleotide exchange to favour RhoA.[262] This introduced a basic side chain, whilst simultaneously removing the steric encumbrance of the original M1369 with V43 of RhoA.

These studies, along with our broader structural analysis, support that DH domains with the combination of a basic and adjacent hydrophobic side chain are likely to be activators of RhoA. This was applied in the analysis of uncharacterised GEFs, revealing 4 GEFs that were predicted to be activators of RhoA: ARHGEF15, ARHGEF19, PLEKHG4B and PLEKHG5 (Figure 62). PLEKHG5 is likely to be of interest for researchers exploring neuropathy, given the reported associations with Charcot-Marie-Tooth neuropathy and lower motor neuron disease.[269-270]

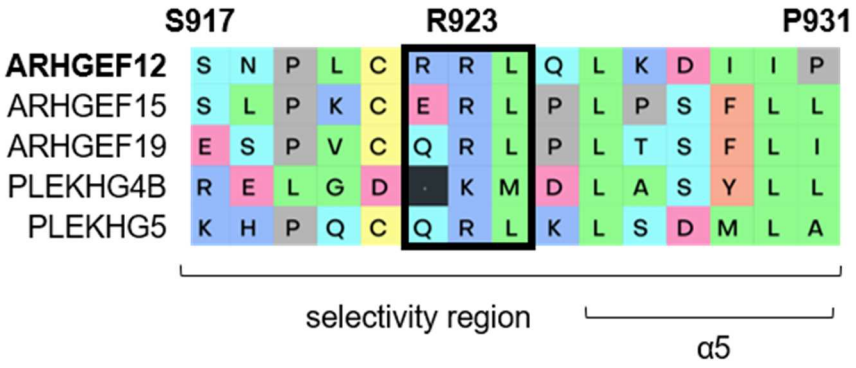


Figure 62 *– Alignment of unexplored GEFs predicted to be RhoA activators. LARG (ARHGEF12) is shown as a reference.*

4.1.6 Cdc42 selectivity

Of the Cdc42 activating GEFs, a pattern was also identified on the GTPase binding surface – 88% contained a large hydrophobic residue at position L759 as referenced to Cdc42 activating GEF Dbs (Figure 63). This interaction was investigated with the GEF-CDC42 crystal structures: Dbs (PBD: 1KZ7), Prex1 (PDB: 5GI1), ARHGEF9 (Collybistin II, PDB: 2DFK) and ITSN1 (PDB: 1KI1). Each GEF contained the same hydrophobic residue identified from the structural analysis, and in combination with another large hydrophobic residue at position L766 of Dbs, was found to form a hydrophobic cleft around F56 of Cdc42 (Figure 64).

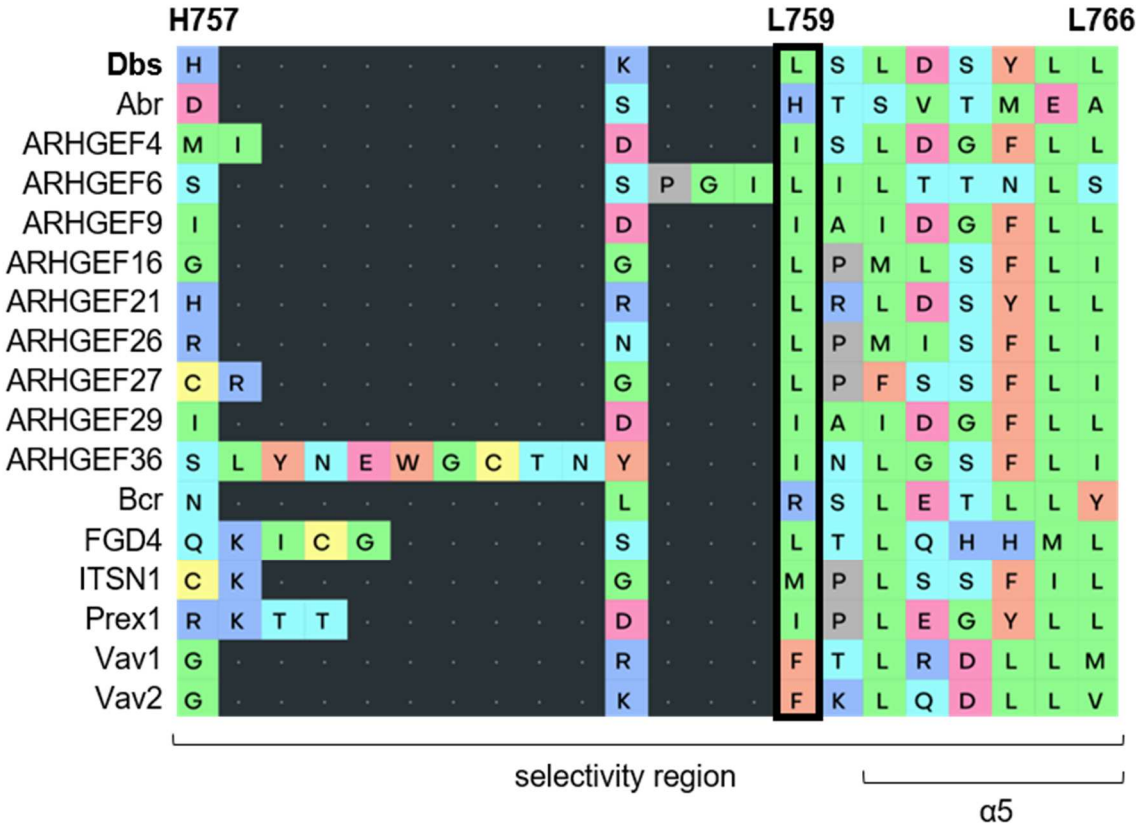


Figure 63 *– Selectivity region analysis of Cdc42 activating GEFs. Reference positions H757, L759 and L766 are of GEF Dbs.*

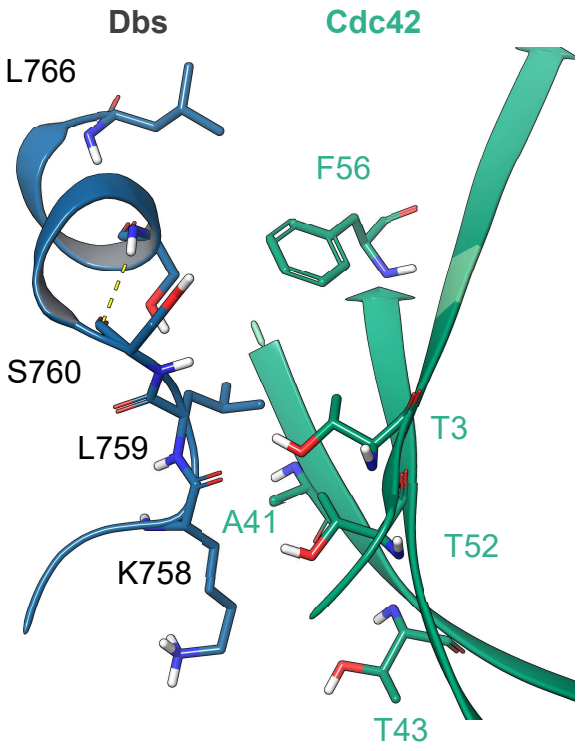


Figure 64 – *Interaction of the selectivity region of GEF Dbs (blue ribbons and sticks) with Cdc42 (green ribbons and sticks).*

The presence of F26 on Cdc42 is the most notable variance compared with Rac1 (W56) and RhoA (W58), and so it was postulated that GEFs with larger, more hydrophobic groups will preferentially bind to Cdc42 through F56. Mutational studies have proven the importance of these hydrophobic GEF residues for Cdc42 recognition. The M1369L/L1376I double mutant of ITSN1 (L759 and L766 of Dbs) reduced the GEF activity for Cdc42, suggesting that larger hydrophobic residues are highly favoured for recognising Cdc42 and prevent interaction with Rac1 and RhoA.[262] This, along with the alignment study, supports the notion that such a hydrophobic motif could result in favourable Cdc42 binding. Accordingly, 16 of the unexplored GEFs were predicted to be activators of Cdc42: ARHGEF15, ARHGEF19, ARHGEF32, ARHGEF38, ARHGEF41, ARHGEF42, ARHGEF43, ARHGEF44, FARP1, FARP2,

FGD1, FGD2, FGD3, FGD6, ITSN2, and PLEKHG4B (Figure 65). This abundance suggests a gap in Cdc42 activation studies, which may be significant for understanding related biological processes. For example, FARP1 has been linked to multiple cancers, including a correlation with poor prognosis of advanced gastric cancers. [271-273]

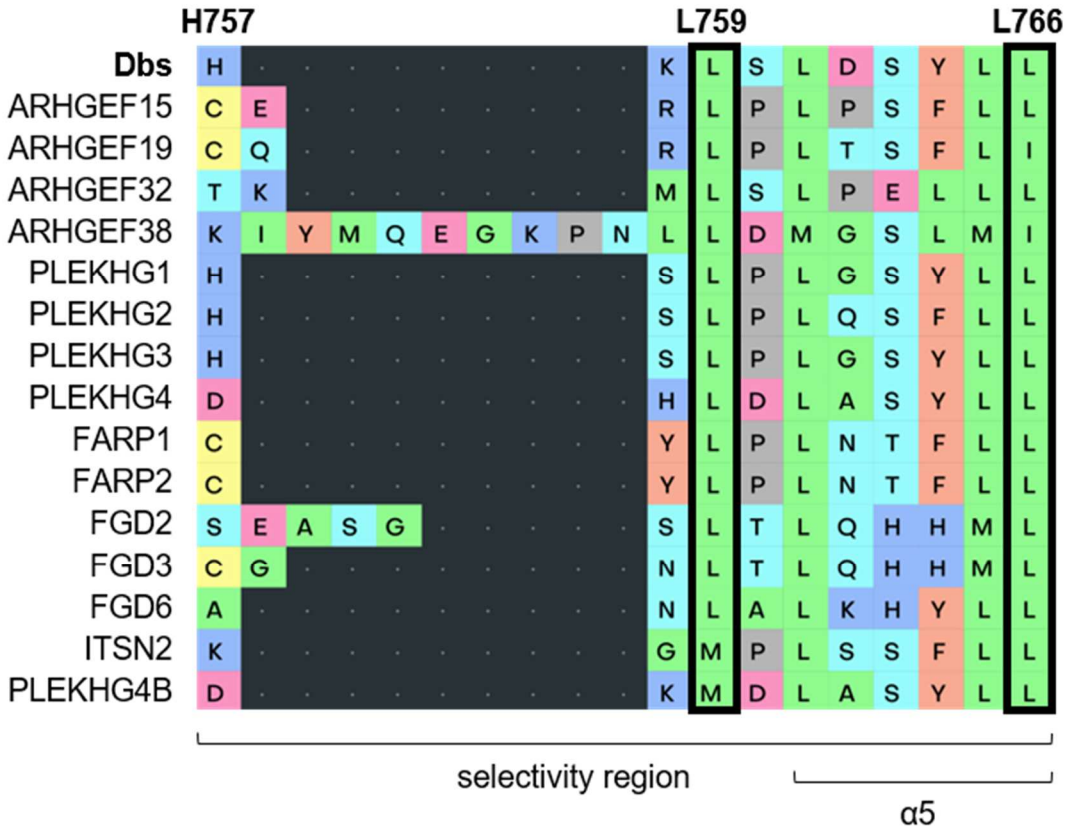


Figure 65 *– Alignment of unexplored GEFs predicted to be Cdc42 activators. Dbs is shown as a reference.*

4.1.7 Activation of multiple GEFs

Although many of the analysed GEFs solely activate one GTPase, there are a significant number that can act on multiple. In some instances, this lack of specificity could be rationalised by applying the knowledge of the discussed selectivity motifs.

For example, GEF Dbs may be able to activate both RhoA and Cdc42 as it comprises a basic and adjacent hydrophobic residue capable of interacting favourably with RhoA, in addition to a hydrophobic cleft that is seemingly inherent of Cdc42 recognition (Figure 64). Others are more unusual in the architecture of their selectivity region and discerning their selectivity is more challenging. This is true of the breakpoint cluster region protein (Bcr), which is predicted to have a selectivity region with an unusual β-sheet between the α4 and α5 helices (AlphaFold). This GEF is reported to activate Rac1, RhoA and Cdc42 and may use the distinctive structure of the selectivity region to recognise all three GEFs. Structurally atypical selectivity regions were also observed in many of the remaining GEFs lacking *in vitro* data: ALS2, β-pix, ARHGEF32, ARHGEF33, ARHGEF37, ARHGEF39, ARHGEF40, SOS1 and SOS2. Due to this structural disparity, the listed GEFs could not be assigned to GTPase partners using the recognition motifs identified in the study.

4.2 Validation of GEF/GTPase predictions

4.2.1 GEF activity of selected GEFs

The final aim was to test the GEF activity of a subset of poorly explored GEFs and compare the data to the predictions made based on the structural alignment study. GEFs ARHGEF19, ARHGEF38, ARHGEF39, PLEKHG1, PLEKHG4 and PLEKHG5 were selected based on their diverse predictions or associations with disease *in vivo* (Table 10). Constructs were designed based on the structure of the active Kalirin DH2 domain obtained in Section 3.1.3 and the proteins were obtained from the Biotech SRF; only ARHGEF19, ARHGEF38 and ARHEF39 were successfully produced.

Table 10 – *Designed GEF DH constructs and alignment predictions. Only ARHGEF19, ARHGEF38 and ARHGEF39 were successfully obtained.*

GEF	Alignment study predictions	Start	End
ARHGEF19	RhoA, Cdc42	S364	E568
ARHGEF38	Cdc42	E83	K296
ARHGEF39	N/A	R9	Q199
PLEKHG1	Cdc42	D145	Q307
PLEKHG4	Cdc42	R732	D912
PLEKHG5	RhoA	R332	S543

The obtained GEFs were tested in the FI nucleotide exchange assay as part of this work. As predicted, ARHGEF19 was found to activate RhoA and Cdc42 in a concentration-dependent manner. However, ARHGEF38 failed to activate Cdc42 as

expected. ARHGEF39, which didn't comprise any clear selectivity markers, was found to lack GEF activity for any of the three GTPases.

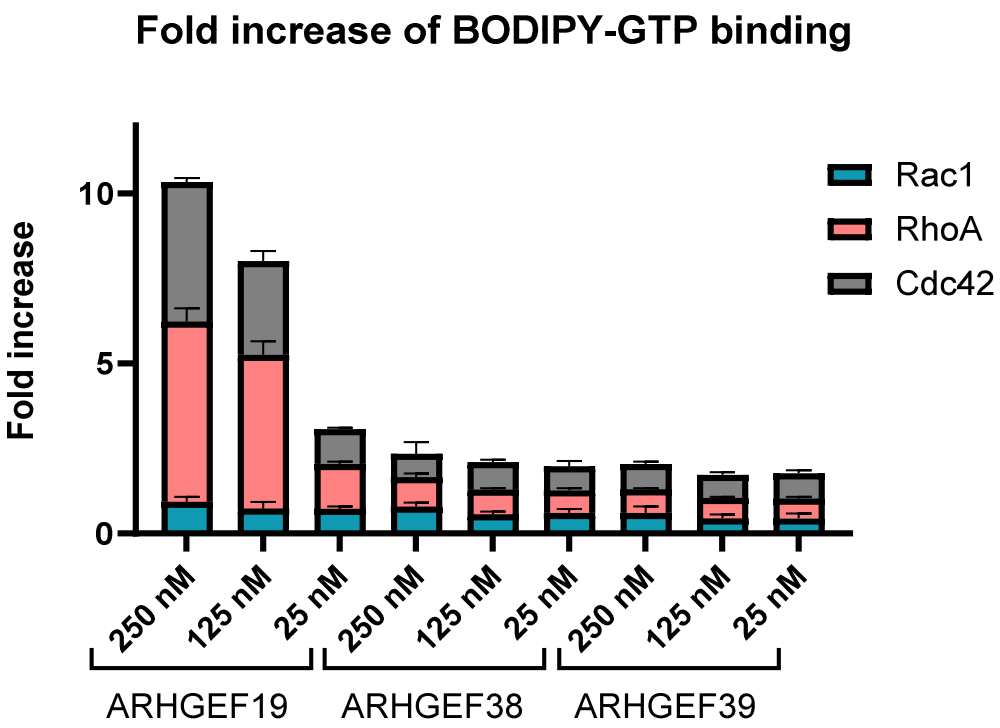


Figure 66 – *Fold increase of BODIPY-GTP binding to GTPases in the presence of ARHGEF19, ARHEF38 and ARHGEF39.*

Multiple conclusions were made from this data. The confirmation that ARHGEF19 can act as a GEF on RhoA and Cdc42, as predicted from the alignment study, reveals two potential druggable targets that could be inhibited by small molecules as described in Chapters 2 and 3. This may be especially appealing to researchers interested in the progression of lung or breast cancers, of which ARHGEF19 has been linked to.[274-276] The lack of GEF activity of the ARHGEF38 and ARHGEF39 constructs highlights the limitations of predicting pairings based on this alignment study. The first is that only Rac1, RhoA and Cdc42 were considered based on the availability of literature information, when 20 mammalian Rho GTPase members exist. It is therefore probable that some of the unexplored GEFs will activate GTPases outside of those considered

in this study. It is also possible that proteins predicted to be GEFs by sequence similarity may be incorrect. An alternative assay to measure binding instead of activation, such as surface plasmon resonance (SPR), would also offer an insight into whether the GEFs are able to bind to the GTPases, but not activate them.

The second is the contribution of domains outside of the DH domain. The major aim of this study was to highlight the conversed mechanisms of GEF activation specifically within the DH domain, given it is at the catalytic centre. It should be noted, however, that nucleotide exchange is often influenced by other domains within the protein. The adjacent PH domain has been cited numerous times to directly influence GEF activity, both in binding affinity, exchange potential and auto-inhibition. [222, 232, 256, 277-278] This was even pronounced in Chapter 2 of this thesis, in which the PH2 domain of Kalirin demonstrated signs of autoinhibition. Other domains have also shown their aptitude in influencing the exchange potential: FARP1 and FARP2 contain a DH-PH-PH tandem that results in a complex autoinhibitory mechanism.[255] Vav1 contains a DH-PH-CRD tandem, in which the cysteine rich domain (CRD) resides next to the DH domain and stabilises GTPase binding.[264] Each of these examples highlight the multi-faceted nature of the larger proteins, which should be considered when investigating full length GEF/GTPase complexes.

Finally, despite a residue or motif being considered highly conserved when occurring in over 80% of GEFs, there was no position in which a residue was conserved in 100% of the alignments. This infers that no residue is vital for activity, but accumulates to form part of a wider interaction network of affinity or catalytic activity, and thus should be considered in the context of one another.

4.3 Conclusions

In conclusion, the largest structure-based analysis of the Rho Dbl GEF family was successfully performed. The accessibility of the EMBL-EBI AlphaFold2 server allowed for facile analysis of this large family through accurate protein structure prediction, and thus was used where crystal structures were absent. In total, 70 Rho Dbl GEFs were structurally aligned and analysed. Nine residues were found to be highly conserved on the GTPase binding surface and their functions were explored with known crystal structures and mutation studies reported in the literature. Moreover, selectivity motifs were identified for RhoA and Cdc42 activating GEFs, and thus used to understand complementary GEF/GTPase pairings.

The selectivity motifs were subsequently used to predict GEF/GTPase pairings with GEFs that lacked *in vitro* data. This ultimately led to the discovery of the ARHGEF19/RhoA and ARHGEF19/Cdc42 complexes, which could be explored for target validation. The limitations of the study were also emphasised, highlighting the multi-faceted nature of DH domain containing proteins. Moreover, much less is known about the Rho GTPases outside of Rac1, RhoA and Cdc42, which could be explored further. It is expected that this broad structural study will aid researchers who wish to explore the Rho Dbl GEF family for biology or drug discovery.

5 Conclusions and Future Work

The Rho GTPases are pleotropic molecular switches that govern a multitude of cellular processes. Although there have been attempts to regulate GTPase activity across many disease areas, direct modulation remains a challenge. Targeting the Rho GEF/GTPase complex was hypothesised to be more feasible with respect to competitive inhibition at the nucleotide binding pocket and would benefit from selective GEF/GTPase signalling. Two Kalirin-based targets were of focus based on their associations with synaptic plasticity and dendritic spine formation. This morphology is crucial to regular cognitive function and may influence diseases such as schizophrenia, AD and bipolar disorder. A greater understanding of probing these targets was required, with the long-term goal of generating high-quality chemical probes for target validation. This thesis highlights the progress made toward developing inhibitors of Kalirin-GEF(1)/Rac1 and Kalirin-GEF(2)/RhoA as well as a broader investigation into Rho Dbl GEF activity to aid future researchers.

RS-009 was a promising covalent fragment previously found to inhibit Kalirin-GEF(1)/Rac1 at the nucleotide binding pocket. This hit inspired the synthesis of new analogues that varied the central pyrrolopyridone core and the SAR of the series was analysed. **MC-278** was identified as a second inhibitor of the complex and the binding mode was corroborated with crystallography and kinetics. The selectivity of **MC-278** was then tested in a limited GEF/GTPase screen, which provided evidence that GEF/GTPase complexes could be selectivity probed. An understanding of the mechanisms behind this could be used to further improve selectivity, and an investigation should be considered for future work. This could include obtaining the

(ligand)-GEF/GTPase crystal structures to explore conformational differences or kinetic data on the rates of the GEF-GTPase interaction.

A mass spectrometry assay was subsequently optimised to investigate the ability of the fragments to outcompete the endogenous ligands; Rac1 was negligibly modified by **MC-278** in the presence of high nucleotide concentrations, highlighting the requirement of a high-affinity compound. To this end, the elaboration of **MC-278** was attempted. An XChem screen performed on **MC-278**-soaked Kalirin-GEF(1)/Rac1 crystals yielded no new ligands due to incomplete occupation of the binding pocket, and instead the fragment was probed through the design and synthesis of novel analogues. The lack of activity observed for the non-covalent counterparts of **MC-278** demonstrates the influence of the chloroacetamide warhead. The elaboration of alternative vectors of **MC-278** is a major piece of future work - analogues should be synthesised to generate a deeper SAR analysis to iteratively enhance potency and selectivity. Generating analogues of **RS-009** is similarly significant.

The availability of data for Kalirin-GEF(2)/RhoA was much more limited. Accordingly, active Kalirin-GEF(2) constructs were designed and successfully produced for screening and crystallography purposes. The DH-PH tandem was found to have inferior GEF activity compared with the DH domain only, suggesting the PH2 domain is autoinhibitory in nature. This mechanism could be corroborated by obtaining a crystal structure of the DH-PH tandem, an objective that should be considered for future work. Crystallisation screens were exploited to crystallise the Kalirin-GEF(2)/RhoA complex and optimise crystal growth for XChem screening. Only crystals of the Kalirin-GEF(2) DH domain were obtained, and the novel structure was

solved to 1.8 Å. The packing of the DH domain molecules was highly favourable at the N-terminal; to discourage these interactions and improve the likelihood of complex crystallisation, new constructs were designed with extended termini. Future work should involve the expression and purification of the new constructs, followed by crystal trials to obtain the crystallised Kalirin-GEF(2)/RhoA complex for XChem screening.

In the absence of a Kalirin-GEF(2)/RhoA screening system for XChem, fragments originally synthesised for Kalirin-GEF(1)/Rac1 were screened against Kalirin-GEF(2)/RhoA in the nucleotide exchange assay. **RS-009** and **MC-278** were discovered to inhibit BODIPY-GTP binding, providing evidence that Kalirin-GEF(2)/RhoA can be probed with small molecules. Future work should involve orthogonal confirmation of the binding mode of the fragments should they be used as starting points for chemical probe development, such as by crystallography and kinetic studies.

Finally, a broad structure-based alignment study was performed with the crystal structures and AlphaFold predictions of the Rho Dbl GEFs. Spatially conserved positions were identified and their roles were understood using known co-crystal structures and reported mutational studies. Moreover, motifs for selectivity were investigated, which resulted in the prediction of new GEF/GTPase pairings. These predictions were explored in a FI nucleotide exchange assay which highlighted new potential targets and also the limitations of the approach. The study offered a comprehensive structural understanding of the GEF family, and highlighted potential GEF/GTPase targets that could be probed to generate a better understanding of

related diseases. Additional work could be performed to measure the GEF activity of the remaining predictions; this would involve the design and production of the protein constructs, followed by testing their GEF activity in the nucleotide exchange assay. Preventing nucleotide exchange on newly discovered GEF/GTPases could also be explored with **MC-278** and **RS-009**, as a proof-of-concept that other Rho GEF/GTPases have the potential to be targeted with this approach.

6 Appendices

Supplementary Table 1 – *Crystallography data collection and refinement statistics for MC-278-Kalirin-GEF(1)/Rac1.*

Overall	
Ligand/complex	MC-278-Kalirin-GEF(1)/Rac1
Beamline	I03
Wavelength	0.9763
Space group	P65 2 2
a, b, c (Å)	62.05, 62.05, 339.44
a, b, g (°)	90, 90, 120

Data collection	
Resolution	56.57–1.48
R_{merge}	0.11
I/σ	17.2
Completeness (%)	99.8
Redundancy	30.9
No. of reflections	65984

Refinement	
Resolution (Å)	1.48
No. reflections	65303
Rwork/Rfree	0.21/0.23
Average B factor (Å^2)	23.6
RMSD: bonds (Å^2)	0.005
RMSD: angles (°)	0.79
Ramachandran favoured (%)	98.5
Ramachandran outliers (%)	0.0
Clashscore	1.31
Protein molecules per unit cell	1
Protein atoms	2704
Solvent atoms	294
Ligand atoms	15
Average B, all atoms (Å^2)	23.0

Supplementary Table 2 – *Intact mass spectrometry of purified Kalirin-GEF(2) constructs.*

Construct	Expected (no tag)	Regions Uniprot O60229	Observed Mass
A1	41205.7	K1916 to P2263	37599.97 (-3605.73)
A2	40866.3	E1919 to P2263	40866.92 (0.62)
A7	38139.2	K1916 to Q2236	38139.69 (0.49)
A8	37799.8	E1919 to Q2236	37800.25 (0.45)
A9	36951.9	K1916 to E2226	36952.3 (0.4)
B1	24023.0	K1916 to E2226	24022.9 (0.25)
B2	23683.6	K1919 to E2114	23682.5 (-0.15)

Supplementary Table 3 – *Crystallography data refinement and statistics for the DH2 domain of protein Kalirin.*

Overall	
Ligand/complex	Kalirin-GEF(2) (PDB: 8C7D)
Beamline	I03
Wavelength	0.9763
Space group	P21 21 21
a, b, c (Å)	27.32, 74.48, 100.12
a, b, g (°)	90, 90, 90
Data collection	
Resolution	59.76–1.86
R_{merge}	0.249
I/σ	7.1
Completeness (%)	100
Redundancy	11.7
No. of reflections	17994
Refinement	
Resolution (Å)	1.86
No. reflections	17831
Rwork/Rfree	0.28/0.23
Average B factor (Å^2)	33.9
RMSD: bonds (Å^2)	0.008
RMSD: angles (°)	0.89
Ramachandran favoured (%)	98
Ramachandran outliers (%)	0
Clashscore	5
Protein molecules per unit cell	2
Protein atoms	1512
Solvent atoms	72
Average B, all atoms (Å^2)	33.0

Supplementary Table 4 - *Rho-Dbl GEFs used in the alignment study. Alternative names, PDB codes used or the use of AlphaFold (AF) is shown. Start and end boundaries were taken from UniProt (www.uniprot.org). Protein structure alignment was performed on Maestro, Schrödinger[125] using PDB: 5O33 Chain B as a reference. The RMSD (Å) is quoted as aligned to PDB: 5O33 Chain B.*

GEF	Names	PDB/AF	With GTPase?	Start	End	RMSD (Å)	Comment
Kalirin-GEF(1)	KalirinN	5O33	Rac1	1281	1456	-	
Abr		AF	N	91	284	3.406	
ALS2	Alsin	AF	N	690	885	3.128	
ARHGEF1	p115-RHOGEF, Lsc	3ODO	N	416	605	2.226	
ARHGEF2	Lfc	AF	N	235	432	2.089	
ARHGEF3	XPLN	AF	N	122	304	1.887	
ARHGEF4	Asef	2DX1	N	285	468	1.971	
ARHGEF5	Tim, Ephexin3	AF	N	1174	1358	1.831	
ARHGEF6	α-pix	AF	N	421	421	2.067	
ARHGEF7	β-pix	1BY1	N	271	451	3.940	
ARHGEF8	NET1	4XH9	RhoA	174	356	2.201	
ARHGEF9	Collybistin II	2DFK	CDC42	103	287	1.770	Rat (100% homology to human)
ARHGEF10		AF	N	421	608	1.759	
ARHGEF10L	GrinchGEF	AF	N	317	502	1.773	
ARHGEF11	PDZ-RHOGEF	IXCG	RhoA	734	923	1.982	
ARHGEF12	LARG	1X86	RhoA	787	977	1.899	

ARHGEF13	Lbc, AKAP13	4D0N	RhoA	1994	2191	2.282	
ARHGEF14	Dbs	1KZ7	CDC42	632	812	1.316	With RhoA PDB: 1LB1
ARHGEF15	Ephexin5	AF	N	417	601	1.631	
ARHGEF16	Ephexin4	AF	N	284	468	1.715	
ARHGEF17	p164-RHOGEF	AF	N	1066	1354	1.950	
ARHGEF18	p114-RHOGEF, SA-RHOGEF	AF	N	447	644	2.030	
ARHGEF19	WGEF, Ephexin2	AF	N	376	560	1.789	
ARHGEF21	Dbl, MCF2L	AF	N	495	675	1.333	
ARHGEF22	MCF2L2	AF	N	619	822	1.551	
Trio1	ARHGEF23, TrioN	2NZ8	Rac1	1292	1467	0.711	
Trio2	ARHGEF23, TrioC	6D8Z	N	1969	2145	1.443	
Kalirin-GEF(2)	KalirinC	8C7D	N	1929	2104	1.419	
ARHGEF25	p63-RHOGEF, GEFT	2RGN	RhoA	160	336	1.320	
ARHGEF26	SGEF	AF	N	439	623	1.717	
ARHGEF27	NGEF, Ephexin	AF	N	273	457	1.675	
ARHGEF28	p190-RHOGEF	AF	N	849	1044	2.006	

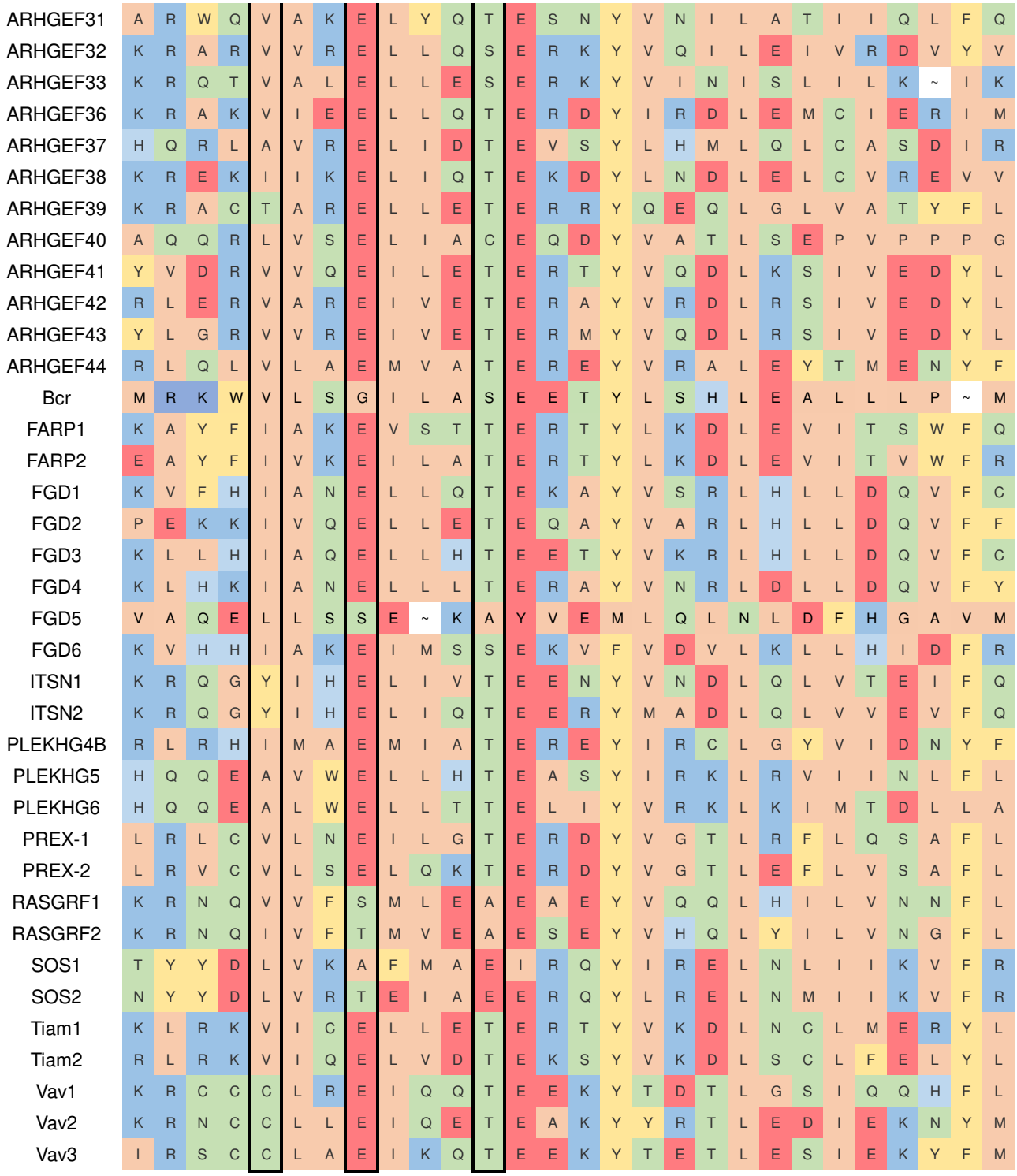
ARHGEF29	Asef2, SPATA13	AF	N	240	424	1.668
ARHGEF30	OBSCN	No AF structure				
ARHGEF31	ECT2	6L30	N	452	641	2.587
ARHGEF32	ECT2L	AF	N	566	754	3.022
ARHGEF33		AF	N	265	440	1.911
ARHGEF36	DNMBP, Tuba	AF	N	784	967	1.642
ARHGEF37		AF	N	30	213	1.729
ARHGEF38		AF	N	94	285	1.804
ARHGEF39	C9orf100	AF	N	22	197	2.246
ARHGEF40	solo	AF	N	1085	1253	2.386
ARHGEF41	PLEKHG1	AF	N	113	293	1.347
ARHGEF42	PLEKHG2	AF	N	102	283	1.573
ARHGEF43	PLEKHG3	AF	N	93	272	1.547
ARHGEF44	PLEKHG4	AF	N	732	908	1.300
Bcr		5N7E – incomplete structure so AF used	N	498	691	3.262
FARP1	FERM, CDEP	4H6Y	N	540	730	2.252
FARP2	FIR	AF	N	535	726	2.794
FGD1		AF	N	373	561	2.287
FGD2		AF	N	102	290	2.111
FGD3		AF	N	157	341	2.180
FGD4	Frabin	AF	N	206	393	2.397
FGD5		3MPX– incomplete structure	N	892	1084	3.135

		so AF used					
FGD6		AF	N	871	1060	2.093	
ITSN1		3QBV	CDC42	1237	1423	1.776	
ITSN2		3GF9	N	1178	1367	1.612	
PLEKHG4B		AF	N	805	984	1.252	
PLEKHG5	Tech	AF	N	336	528	1.945	
PLEKHG6	MyoGEF	AF	N	161	353	1.990	
PLEKHG7		AF	N	313	488	2.441	
PREX-1		5FI0	Rac1	49	240	1.773	With Cdc42 PDB: 5FI1
PREX-2		AF	N	23	214	1.870	
RASGRF1		AF	N	240	426	2.116	
RASGRF2		AF	N	243	429	2.061	
SOS1		1DBH	N	200	390	3.250	
SOS2		AF	N	198	388	3.112	
Tiam1		1FOE	Rac1	1040	1234	1.896	Mouse (99% homology to human)
Tiam2	STEF	AF	N	1099	1293	1.837	
Vav1		3BJI	Rac1	194	373	2.008	
Vav2		AF	N	198	376	1.771	
Vav3		AF	N	192	371	1.811	

Supplementary Table 5 - *Structure-based alignment of GEF residues on conserved region 1 (CR1) of Rho-Dbl GEFs. Residue positions are aligned as referenced to the corresponding position on Kalirin-GEF(1) (1233-1260, PDB: 5O33, Chain B). Solvent-exposed residues that were considered highly conserved (>80% same residue type) are highlighted (black outline). Polar residues (S, T, C, N, Q) are coloured green, non-polar (G, A, V, L, I, P, M) in orange, basic (K, R, H) in blue, acidic (D, E) in red and aromatic (Y, F, W) in yellow. ~ indicates poor overlap.*

CR1

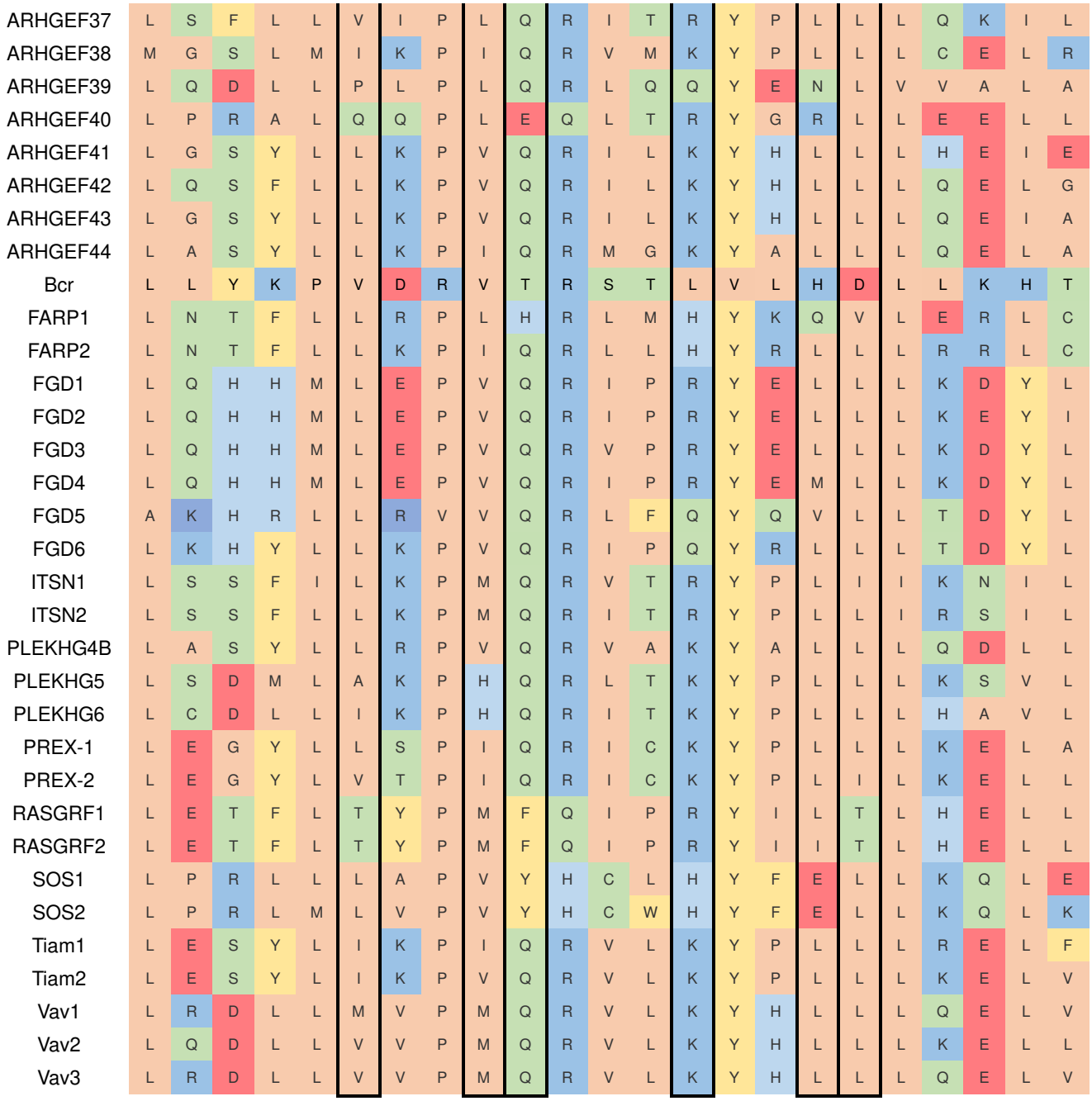
	1233	1234	1235	1236	1237	1238	1239	1240	1241	1242	1243	1244	1245	1246	1247	1248	1249	1250	1251	1252	1253	1254	1255	1256	1257	1258	1259	1260
Kalirin-GEF(1)	K	K	E	F	I	M	A	E	L	L	Q	T	E	K	A	Y	V	R	D	L	H	E	C	L	E	T	Y	L
Abr	M	R	K	L	V	L	S	G	F	L	A	S	E	E	I	Y	I	N	Q	L	E	A	L	L	L	P	~	M
ALS2	G	Y	I	A	S	L	H	E	L	A	T	T	E	R	R	F	Y	S	K	L	S	D	I	K	S	Q	I	L
ARHGEF1	K	R	Q	E	V	I	S	E	L	L	V	T	E	A	A	H	V	R	M	L	R	V	L	H	D	L	F	F
ARHGEF2	K	Q	Q	D	V	I	Y	E	L	I	Q	T	E	L	H	H	V	R	T	L	K	I	M	T	R	L	F	R
ARHGEF3	K	R	Q	E	A	I	F	E	L	S	Q	G	E	E	D	L	I	E	D	L	K	L	A	K	K	A	Y	H
ARHGEF4	~	R	T	N	V	I	N	E	I	L	S	T	E	R	D	Y	I	K	H	L	R	D	I	C	E	G	Y	V
ARHGEF5	K	L	Q	E	V	K	F	E	L	I	V	S	E	A	S	Y	L	R	S	L	N	I	A	V	D	H	F	Q
ARHGEF6	Y	Y	T	V	V	L	Q	N	I	L	D	T	E	K	E	Y	A	K	E	L	Q	S	L	L	V	T	Y	L
ARHGEF7	Y	Y	T	V	V	L	Q	N	I	L	D	T	E	K	E	Y	A	K	E	L	Q	S	L	L	V	T	Y	L
ARHGEF8	R	R	Q	E	A	I	Y	E	M	S	R	G	E	Q	D	L	I	E	D	L	K	L	A	R	K	A	Y	H
ARHGEF9	M	R	A	N	V	I	N	E	I	M	S	T	E	R	H	Y	I	K	H	L	K	D	I	C	E	G	Y	L
ARHGEF10	V	R	R	Y	I	L	G	S	V	V	D	S	E	K	N	Y	V	D	A	L	K	R	I	L	E	Q	Y	E
ARHGEF10L	W	R	R	H	I	L	G	S	I	V	Q	S	E	G	S	Y	V	E	S	L	K	R	I	L	Q	D	Y	R
ARHGEF11	D	R	Q	E	V	I	N	E	L	F	V	T	E	A	S	H	L	R	T	L	R	V	L	D	L	I	F	Y
ARHGEF12	K	R	Q	E	V	I	N	E	L	F	Y	T	E	R	A	H	V	R	T	L	K	V	L	D	Q	V	F	Y
ARHGEF13	K	R	Q	E	V	I	Y	E	L	M	Q	T	E	F	H	H	V	R	T	L	K	I	M	S	G	V	Y	S
ARHGEF14	L	R	R	H	V	M	N	E	L	L	D	T	E	R	A	Y	V	E	E	L	L	C	V	L	E	G	Y	A
ARHGEF15	R	M	Q	E	S	L	F	E	V	V	T	S	E	A	S	Y	L	R	S	L	R	L	L	T	D	T	F	V
ARHGEF16	K	R	Q	E	A	M	F	E	I	L	T	S	E	F	S	Y	Q	H	S	L	S	I	L	V	E	E	F	L
ARHGEF17	M	R	K	H	V	A	M	T	L	L	D	T	E	Q	S	Y	V	E	S	L	R	T	L	M	Q	G	Y	M
ARHGEF18	K	R	Q	D	V	L	Y	E	L	M	Q	T	E	V	H	H	V	R	T	L	K	I	M	L	K	V	Y	S
ARHGEF19	K	L	Q	E	A	K	F	E	L	I	T	S	E	A	S	Y	I	H	S	L	S	V	A	V	G	H	F	L
ARHGEF21	L	K	N	H	V	L	N	E	L	I	Q	T	E	R	V	Y	V	R	E	L	Y	T	V	L	L	G	Y	R
ARHGEF22	P	R	R	R	I	I	R	D	L	L	E	T	E	E	I	Y	I	K	E	I	K	S	I	I	D	G	Y	I
Trio1	R	K	E	F	I	M	A	E	L	I	Q	T	E	K	A	Y	V	R	D	L	R	E	C	M	D	T	Y	L
Trio2	R	R	H	Y	V	L	Q	E	L	V	E	T	E	R	D	Y	V	R	D	L	G	Y	V	V	E	G	Y	M
Kalirin-GEF(2)	G	R	M	F	V	L	N	E	L	V	Q	T	E	K	D	Y	V	K	D	L	G	I	V	V	E	G	F	M
ARHGEF25	R	S	M	Y	V	L	S	E	L	V	E	T	E	K	M	Y	V	D	D	L	G	Q	I	V	E	G	Y	M
ARHGEF26	K	R	Q	E	A	I	F	E	V	I	S	S	E	H	S	Y	L	L	S	L	E	I	L	I	R	M	F	K
ARHGEF27	K	L	Q	E	A	M	F	E	L	V	T	S	E	A	S	Y	Y	K	S	L	N	L	L	V	S	H	F	M
ARHGEF28	K	R	Q	D	V	I	F	E	L	M	Q	T	E	M	H	H	I	Q	T	L	F	I	M	S	E	I	F	R
ARHGEF29	M	R	T	N	V	I	R	E	I	M	D	T	E	R	V	Y	I	K	H	L	R	D	I	C	E	G	Y	I


ARHGEF31 A R W Q V A K E L Y Q T E S N Y V N I L A T I I Q L F Q
ARHGEF32 K R A R V V R E L L Q S E R K Y V Q I L E I V R D V Y V
ARHGEF33 K R Q T V A L E L L E S E R K Y V I N I S L I L K ~ I K
ARHGEF36 K R A K V I E E L L Q T E R D Y I R D L E M C I E R I M
ARHGEF37 H Q R L A V R E L I D T E V S Y L H M L Q L C A S D I R
ARHGEF38 K R E K I I K E L I Q T E K D Y L N D L E L C V R E V V
ARHGEF39 K R A C T A R E L L E T E R R Y Q E Q L G L V A T Y F L
ARHGEF40 A Q Q R L V S E L I A C E Q D Y V A T L S E P V P P P G
ARHGEF41 Y V D R V V Q E I L E T E R T Y V Q D L K S I V E D Y L
ARHGEF42 R L E R V A R E I V E T E R A Y V R D L R S I V E D Y L
ARHGEF43 Y L G R V V R E I V E T E R M Y V Q D L R S I V E D Y L
ARHGEF44 R L Q L V L A E M V A T E R E Y V R A L E Y T M E N Y F
Bcr M R K W V L S G I L A S E E T Y L S H L E A L L L P ~ M
FARP1 K A Y F I A K E V S T T E R T Y L K D L E V I T S W F Q
FARP2 E A Y F I V K E I L A T E R T Y L K D L E V I T V W F R
FGD1 K V F H I A N E L L Q T E K A Y V S R L H L L D Q V F C
FGD2 P E K K I V Q E L L E T E Q A Y V A R L H L L D Q V F F
FGD3 K L L H I A Q E L L H T E E T Y V K R L H L L D Q V F C
FGD4 K L H K I A N E L L L T E R A Y V N R L D L L D Q V F Y
FGD5 V A Q E L L S S E ~ K A Y V E M L Q L N L D F H G A V M
FGD6 K V H H I A K E I M S S E K V F V D V L K L L H I D F R
ITSN1 K R Q G Y I H E L I V T E E N Y V N D L Q L V T E I F Q
ITSN2 K R Q G Y I H E L I Q T E E R Y M A D L Q L V V E V F Q
PLEKHG4B R L R H I M A E M I A T E R E Y I R C L G Y V I D N Y F
PLEKHG5 H Q Q E A V W E L L H T E A S Y I R K L R V I I N L F L
PLEKHG6 H Q Q E A L W E L L T T E L I Y V R K L K I M T D L L A
PREX-1 L R L C V L N E I L G T E R D Y V G T L R F L Q S A F L
PREX-2 L R V C V L S E L Q K T E R D Y V G T L E F L V S A F L
RASGRF1 K R N Q V V F S M L E A E A E Y V Q Q L H I L V N N F L
RASGRF2 K R N Q I V F T M V E A E S E Y V H Q L Y I L V N G F L
SOS1 T Y Y D L V K A F M A E I R Q Y I R E L N L I I K V F R
SOS2 N Y Y D L V R T E I A E E R Q Y L R E L N M I I K V F R
Tiam1 K L R K V I C E L L E T E R T Y V K D L N C L M E R Y L
Tiam2 R L R K V I Q E L V D T E K S Y V K D L S C L F E L Y L
Vav1 K R C C C L R E I Q Q T E E K Y T D T L G S I Q Q H F L
Vav2 K R N C C L L E I Q E T E A K Y Y R T L E D I E K N Y M
Vav3 I R S C C L A E I K Q T E E K Y T E T L E S I E K Y F M

Supplementary Table 6 - *Structure-based alignment of GEF residues on conserved region 3 (CR3) of Rho-BDL GEFs. Positions are referenced as per Kalirin-GEF(1) (1359–1381, PDB: 5O33, Chain B). Solvent-exposed residues that were considered highly conserved (>80% same residue type) are highlighted (black outline). Colour annotations are as described in Table S2.*

CR3

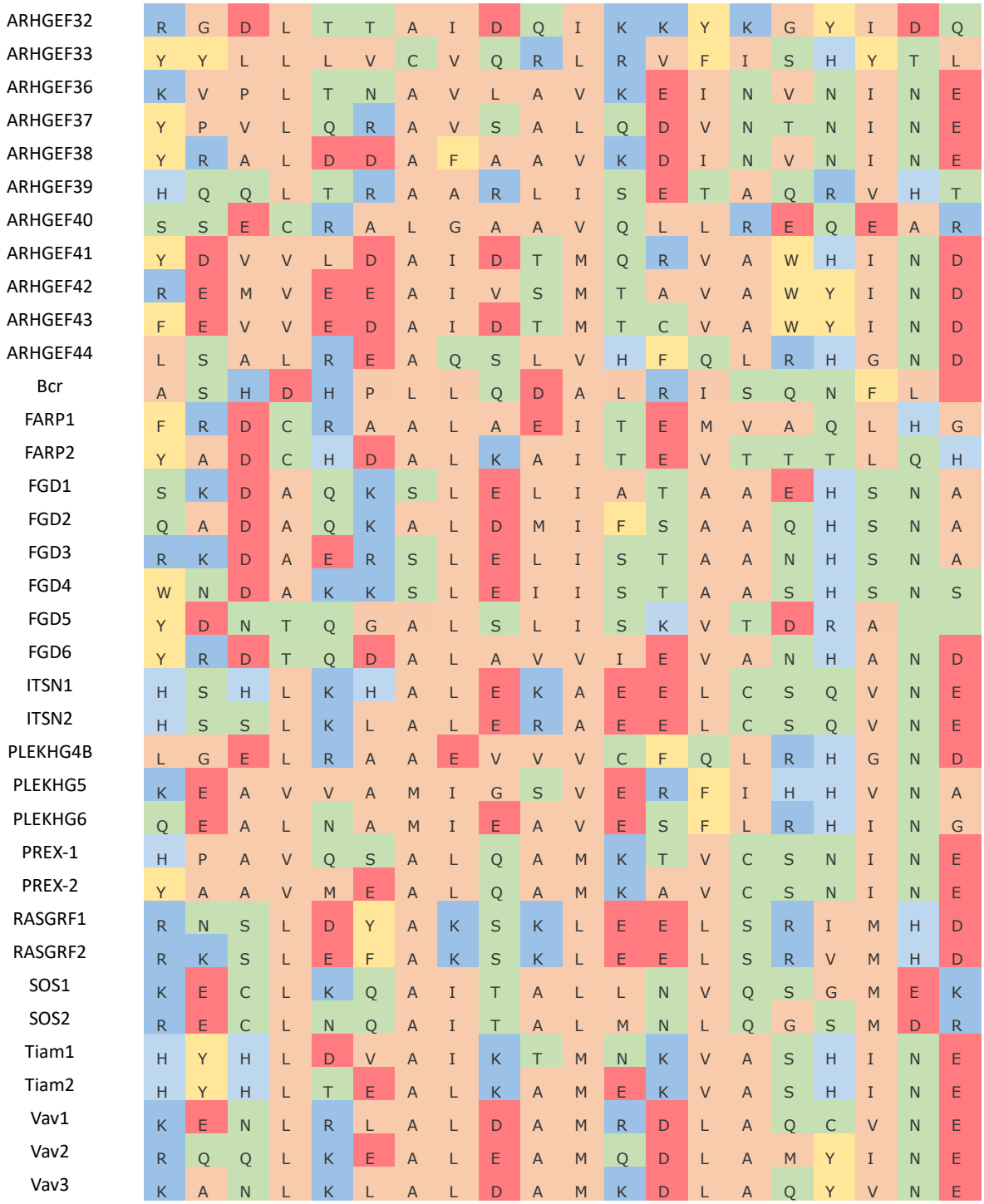
	1359	1360	1361	1362	1363	1364	1365	1366	1367	1368	1369	1370	1371	1372	1373	1374	1375	1376	1377	1378	1379	1380	1381
Kalirin-GEF(1)	I	S	S	Y	L	I	K	P	V	Q	R	V	T	K	Y	Q	L	L	L	K	E	L	L
Abr	L	L	Y	K	P	I	D	R	V	T	R	S	T	L	V	L	H	D	L	L	K	H	T
ALS2	L	N	T	L	F	F	L	P	I	R	R	L	H	N	Y	A	K	V	L	L	K	L	A
ARHGEF1	L	K	D	M	I	P	T	E	M	Q	R	L	T	K	Y	P	L	L	L	Q	S	I	G
ARHGEF2	V	Q	E	C	I	L	L	V	T	Q	R	I	T	K	Y	P	L	L	I	S	R	I	L
ARHGEF3	L	W	N	F	L	D	I	P	R	S	R	L	V	K	Y	P	L	L	L	R	E	I	L
ARHGEF4	L	D	G	F	L	L	T	P	V	Q	K	I	C	K	Y	P	L	Q	L	A	E	L	L
ARHGEF5	L	K	S	F	L	I	L	P	F	Q	R	I	T	R	L	K	L	L	L	Q	N	I	L
ARHGEF6	L	T	T	N	L	S	K	P	F	M	R	L	E	K	Y	V	T	L	L	Q	E	L	E
ARHGEF7	L	T	T	N	L	S	K	P	F	M	R	L	E	K	Y	V	T	L	L	Q	E	L	E
ARHGEF8	L	W	S	F	L	D	I	P	R	S	R	L	V	K	Y	P	L	L	L	K	E	I	L
ARHGEF9	I	D	G	F	L	L	T	P	V	Q	K	I	C	K	Y	P	L	Q	L	A	E	L	L
ARHGEF10	L	Y	S	L	M	M	K	P	I	Q	R	F	P	Q	F	I	L	L	L	Q	D	M	L
ARHGEF10L	L	Y	G	L	M	V	K	P	I	Q	R	F	P	Q	F	I	L	L	L	Q	D	M	L
ARHGEF11	L	R	D	L	I	I	S	E	M	Q	R	L	T	K	Y	P	L	L	L	E	S	I	I
ARHGEF12	L	K	D	I	I	P	T	Q	M	Q	R	L	T	K	Y	P	L	L	L	D	N	I	A
ARHGEF13	I	P	E	C	I	L	L	V	T	Q	R	I	T	K	Y	P	V	L	F	Q	R	I	L
ARHGEF14	L	D	S	Y	L	L	K	P	V	Q	R	I	T	K	Y	Q	L	L	L	K	E	M	L
ARHGEF15	L	P	S	F	L	L	L	P	F	Q	R	I	T	R	L	R	M	L	L	Q	N	I	L
ARHGEF16	M	L	S	F	L	I	L	P	M	Q	R	V	T	R	L	P	L	L	M	D	T	L	C
ARHGEF17	L	S	D	L	M	I	K	P	V	Q	R	I	P	R	Y	E	L	L	V	K	D	L	L
ARHGEF18	V	Q	E	C	I	L	L	V	T	Q	R	I	T	K	Y	P	V	L	V	E	R	I	I
ARHGEF19	L	T	S	F	L	I	L	P	F	Q	R	I	T	R	L	K	M	L	V	E	N	I	L
ARHGEF21	L	D	S	Y	L	L	K	P	V	Q	R	I	T	K	Y	Q	L	L	L	K	E	L	L
ARHGEF22	L	F	K	Y	L	K	G	P	S	Q	R	L	I	K	Y	Q	M	L	L	K	G	L	L
Trio1	I	S	S	Y	L	I	K	P	V	Q	R	I	T	K	Y	Q	L	L	L	K	E	L	L
Trio2	L	T	D	L	L	I	K	P	V	Q	R	I	M	K	Y	Q	L	L	L	K	D	F	L
Kalirin-GEF(2)	L	S	D	F	L	I	K	P	I	Q	R	I	T	K	Y	Q	L	L	L	K	D	F	L
ARHGEF25	L	N	D	L	L	I	K	P	V	Q	R	I	M	K	Y	Q	L	L	L	K	D	F	L
ARHGEF26	M	I	S	F	L	I	L	P	M	Q	R	V	T	R	L	P	L	L	M	D	T	I	C
ARHGEF27	F	S	S	F	L	I	L	P	F	Q	R	I	T	R	L	K	L	L	V	Q	N	I	L
ARHGEF28	I	P	E	C	I	L	L	V	T	Q	R	I	T	K	Y	P	V	L	V	E	R	I	L
ARHGEF29	I	D	G	F	L	L	T	P	V	Q	K	I	C	K	Y	P	L	Q	L	A	E	L	L
ARHGEF31	L	V	E	L	L	I	R	P	V	Q	R	L	P	S	V	A	L	L	L	N	D	L	K
ARHGEF32	L	P	E	L	L	L	Y	P	S	R	R	F	E	E	Y	L	N	L	L	Y	A	V	R
ARHGEF33	I	Y	T	L	F	F	H	I	V	Q	R	I	P	E	Y	L	I	H	L	Q	N	V	L
ARHGEF36	L	G	S	F	L	I	K	P	V	Q	R	V	M	R	Y	P	L	L	L	M	E	L	L


ARHGEF37 L S F L L V I P L Q R I T R Y P L L L Q K I L
ARHGEF38 M G S L M I K P I Q R V M K Y P L L L C E L R
ARHGEF39 L Q D L L P L P L Q R L Q Q Y E N L V V A L A
ARHGEF40 L P R A L Q Q P L E Q L T R Y G R L L E E L L
ARHGEF41 L G S Y L L K P V Q R I L K Y H L L L H E I E
ARHGEF42 L Q S F L L K P V Q R I L K Y H L L L Q E L G
ARHGEF43 L G S Y L L K P V Q R I L K Y H L L L Q E I A
ARHGEF44 L A S Y L L K P I Q R M G K Y A L L L Q E L A
Bcr L L Y K P V D R V T R S T L V L H D L L K H T
FARP1 L N T F L L R P L H R L M H Y K Q V L E R L C
FARP2 L N T F L L K P I Q R L L H Y R L L L R R L C
FGD1 L Q H H M L E P V Q R I P R Y E L L L K D Y L
FGD2 L Q H H M L E P V Q R I P R Y E L L L K E Y I
FGD3 L Q H H M L E P V Q R V P R Y E L L L K D Y L
FGD4 L Q H H M L E P V Q R I P R Y E M L L K D Y L
FGD5 A K H R L L R V V Q R L F Q Y Q V L L T D Y L
FGD6 L K H Y L L K P V Q R I P Q Y R L L L T D Y L
ITSN1 L S S F I L K P M Q R V T R Y P L I I K N I L
ITSN2 L S S F L L K P M Q R I T R Y P L L I R S I L
PLEKHG4B L A S Y L L R P V Q R V A K Y A L L L Q D L L
PLEKHG5 L S D M L A K P H Q R L T K Y P L L L K S V L
PLEKHG6 L C D L L I K P H Q R I T K Y P L L L H A V L
PREX-1 L E G Y L L S P I Q R I C K Y P L L L K E L A
PREX-2 L E G Y L V T P I Q R I C K Y P L I L K E L L
RASGRF1 L E T F L T Y P M F Q I P R Y I L T L H E L L
RASGRF2 L E T F L T Y P M F Q I P R Y I I T L H E L L
SOS1 L P R L L L A P V Y H C L H Y F E L L K Q L E
SOS2 L P R L M L V P V Y H C W H Y F E L L K Q L K
Tiam1 L E S Y L I K P I Q R V L K Y P L L L R E L F
Tiam2 L E S Y L I K P V Q R V L K Y P L L L K E L V
Vav1 L R D L L M V P M Q R V L K Y H L L L Q E L V
Vav2 L Q D L L V V P M Q R V L K Y H L L L K E L L
Vav3 L R D L L V V P M Q R V L K Y H L L L Q E L V

Supplementary Table 7 - *Structure-based alignment of GEF residues on the α6 helix of Rho-BDL GEFs. Positions are referenced as per Kalirin-GEF(1) (1388–1410, PDB: 5O33, Chain B). Solvent-exposed residues that were considered highly conserved (>80% same residue type) are highlighted (black outline). Colour annotations are as described in Table S2. In some instances, helices are shorter where the DH domain boundaries were dictated by UniProt. *Abr poorly overlapped in this region, and so corresponding residues were not identified. ~ indicates poor overlap.*

α6 helix

	1388	1389	1390	1391	1392	1393	1394	1395	1396	1397	1398	1399	1400	1401	1402	1403	1404	1405	1406	1407
Kalirin-GEF(1)	K	G	E	L	K	D	G	L	E	V	M	L	S	V	P	K	K	A	N	D
Abr*																				
ALS2	E	Y	Q	K	L	Q	D	S	S	S	C	Y	E	C	L	A	L	H	L	G
ARHGEF1	R	E	K	V	E	L	A	A	E	C	C	R	E	I	L	H	H	V	N	Q
ARHGEF2	R	Q	D	L	T	T	A	L	G	L	V	K	E	L	L	S	N	V	D	E
ARHGEF3	Q	Q	H	L	E	E	A	I	N	I	I	Q	G	I	V	A	E	I	N	T
ARHGEF4	F	K	D	V	E	A	A	L	H	A	M	K	N	V	A	Q	L	I	N	E
ARHGEF5	E	A	E	A	T	K	A	H	H	A	L	E	Q	L	I	R	D	C	N	N
ARHGEF6	H	Q	D	I	L	K	A	I	V	A	F	K	T	L	M	G	Q	C	Q	D
ARHGEF7	R	Q	D	I	Q	K	S	M	A	A	F	K	N	L	S	A	Q	C	Q	E
ARHGEF8	V	Q	L	L	E	D	A	I	L	I	I	Q	G	V	L	S	D	I	N	L
ARHGEF9	Y	R	Y	V	A	A	A	L	A	V	M	R	N	V	T	Q	Q	I	N	E
ARHGEF10	R	L	P	L	Q	M	A	L	T	E	L	E	T	L	A	E	K	L	N	E
ARHGEF10L	R	L	S	L	Q	L	A	L	T	E	L	E	T	L	A	E	K	L	N	E
ARHGEF11	H	E	K	L	C	R	A	R	D	Q	C	R	E	I	L	K	Y	V	N	E
ARHGEF12	R	E	K	V	K	K	A	A	D	H	C	R	Q	I	L	N	F	V	N	Q
ARHGEF13	Q	E	D	L	A	Q	S	L	S	L	V	K	D	V	I	G	A			
ARHGEF14	A	E	D	L	Q	E	A	L	S	S	I	L	G	I	L	K	A	V	N	D
ARHGEF15	Q	E	N	A	Q	K	A	L	G	A	V	S	K	I	I	E	R	C	S	A
ARHGEF16	Y	K	A	A	S	R	A	L	K	A	I	S	K	L	V	R	Q	C	N	E
ARHGEF17	H	P	L	L	L	E	A	Q	R	N	I	K	Q	V	A	E	R	I	N	K
ARHGEF18	Y	E	D	L	T	Q	A	L	N	L	I	K	D	I	I	S	Q	V	D	A
ARHGEF19	E	D	M	A	T	K	A	F	N	A	L	K	E	L	V	Q	E	C	N	A
ARHGEF21	S	A	L	L	K	K	A	L	D	A	M	L	D	L	L	K	S	V	N	D
ARHGEF22	S	T	E	L	Q	Q	A	L	A	V	I	E	D	L	I	K	S	C	E	L
Trio1	K	G	E	I	K	D	G	L	E	V	M	L	S	V	P	K	R	A	N	D
Trio2	D	T	S	E	L	E	R	A	V	E	V	M	C	I	V	P	R	R	C	N
Kalirin-GEF(2)	C	S	D	I	E	K	A	V	E	L	M	C	L	V	P	K	R	C	N	D
ARHGEF25	T	A	D	L	E	Q	A	V	E	V	M	C	F	V	P	K	R	C	N	D
ARHGEF26	Y	E	V	C	K	R	A	L	K	E	V	S	K	L	V	R	L	C	N	E
ARHGEF27	E	C	T	A	L	D	A	H	K	E	L	E	M	V	V	K	A	C	N	E
ARHGEF28	H	K	D	L	R	K	A	L	C	L	I	K	D	M	I	A	T	V	D	L
ARHGEF29	Y	S	N	I	K	A	A	Y	E	A	M	K	N	V	A	C	L	I	N	E
ARHGEF31	K	S	T	L	E	K	A	I	G	S	L	K	E	V	M	T	H	I	N	E


ARHGEF32 R G D L T T A I D Q I K K Y K G Y I D Q
ARHGEF33 Y Y L L L V C V Q R L R V F I S H Y T L
ARHGEF36 K V P L T N A V L A V K E I N V N I N E
ARHGEF37 Y P V L Q R A V S A L Q D V N T N I N E
ARHGEF38 Y R A L D D A F A A V K D I N V N I N E
ARHGEF39 H Q Q L T R A A R L I S E T A Q R V H T
ARHGEF40 S S E C R A L G A A V Q L L R E Q E A R
ARHGEF41 Y D V V L D A I D T M Q R V A W H I N D
ARHGEF42 R E M V E E A I V S M T A V A W Y I N D
ARHGEF43 F E V V E D A I D T M T C V A W Y I N D
ARHGEF44 L S A L R E A Q S L V H F Q L R H G N D
Bcr A S H D H P L L Q D A L R I S Q N F L
FARP1 F R D C R A A L A E I T E M V A Q L H G
FARP2 Y A D C H D A L K A I T E V T T T L Q H
FGD1 S K D A Q K S L E L I A T A A E H S N A
FGD2 Q A D A Q K A L D M I F S A A Q H S N A
FGD3 R K D A E R S L E L I S T A A N H S N A
FGD4 W N D A K K S L E I I S T A A S H S N S
FGD5 Y D N T Q G A L S L I S K V T D R A
FGD6 Y R D T Q D A L A V V I E V A N H A N D
ITSN1 H S H L K H A L E K A E E L C S Q V N E
ITSN2 H S S L K L A L E R A E E L C S Q V N E
PLEKHG4B L G E L R A A E V V V C F Q L R H G N D
PLEKHG5 K E A V V A M I G S V E R F I H H V N A
PLEKHG6 Q E A L N A M I E A V E S F L R H I N G
PREX-1 H P A V Q S A L Q A M K T V C S N I N E
PREX-2 Y A A V M E A L Q A M K A V C S N I N E
RASGRF1 R N S L D Y A K S K L E E L S R I M H D
RASGRF2 R K S L E F A K S K L E E L S R V M H D
SOS1 K E C L K Q A I T A L L N V Q S G M E K
SOS2 R E C L N Q A I T A L M N L Q G S M D R
Tiam1 H Y H L D V A I K T M N K V A S H I N E
Tiam2 H Y H L T E A L K A M E K V A S H I N E
Vav1 K E N L R L A L D A M R D L A Q C V N E
Vav2 R Q Q L K E A L E A M Q D L A M Y I N E
Vav3 K A N L K L A L D A M K D L A Q Y V N E

Supplementary Table 8 - *GTPases activated by Rho-Dbl GEFs. Limited only to RhoA, Rac1 and Cdc42. Allocations were exclusively based on in vitro data with purified proteins. *Indicates conflicting data in the literature, where the activation of the GTPases is controversial across the referenced studies.*

RHO-Dbl GEFs	RhoA	Rac1	CDC42	Reference
Kalirin-GEF(1)	N	Y	N	unpublished data
Abr	Y*	Y*	Y*	279-280
ARHGEF1	Y	N	N	281
ARHGEF2	Y	N	N	unpublished data
ARHGEF3	Y	N	N	268, 280
ARHGEF4	N	N	Y	280
ARHGEF5	Y	N	N	282-283
ARHGEF6	Y*	Y*	Y*	280, 284
ARHGEF8	Y	N	N	285
ARHGEF9	N	N	Y	280
ARHGEF10	Y	N	N	286
ARHGEF10L	Y	N	N	287
ARHGEF11	Y	N	N	261
ARHGEF12	Y	N	N	267
ARHGEF13	Y	N	N	288
ARHGEF14	Y	N	Y	232, 280
ARHGEF16	N	Y	Y	289-290
ARHGEF17	Y	N	N	291
ARHGEF18	Y	N	N	292
ARHGEF21	Y	N	Y	280, 293
Trio1	N	Y	N	294
Trio2	Y	N	N	294
Kalirin-GEF(2)	Y	N	N	unpublished data
ARHGEF25	Y	N	N	295
ARHGEF26	N	N	Y	296
ARHGEF27	Y	Y	Y	297
ARHGEF28	Y	N	N	298
ARHGEF29	N	N	Y	280, 299

ARHGEF31	Y	N	N	300
ARHGEF36	N	N	Y	280
Bcr	Y	Y	Y	279
FGD4	N*	N*	Y*	279-280, 301
FGD5	N	Y	N	302
ITSN1	N	N	Y	280
PLEKHG6	Y	Y	N	303
PREX-1	N	Y	N	280
PREX-2	N	Y	N	304
RASGRF1	N	Y	N	305
RASGRF2	N	Y	N	306
Tiam1	N	Y	N	280
Tiam2	N	Y	N	307
Vav1	Y	Y	Y	264, 308
Vav2	Y	Y	Y	308-309
Vav3	Y	Y	N	310

7 Experimental

7.1 General chemistry experimental

7.1.1 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. All reactions involving moisture sensitive reagents were carried out under a nitrogen atmosphere. Standard practices were employed when handling moisture- and air-sensitive compounds. Room temperature (r.t.) refers to ambient temperature. Temperatures of 0 °C were maintained using an ice-water bath.

7.1.2 Purification and chromatography

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 F254 silica gel. Plates were visualised using UV light (254 nm). 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 % WATER, 5 % MeCN, 2 % of 0.5 M NH4OAc pH 6.0) in solvent B (18% WATER, 80% MeCN, 2% of 0.5 M NH4OAc 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). Purification by HPLC was performed using a Waters SFO with 515 HPLC pump and Waters Binary

Gradient 2545 device (5-95% solvent A [18% Water, 80% MeCN, 2% NH4OAc pH 6.0] in solvent B [93% Water, 5% MeCN, 2% NH4OAc pH 6.0]).

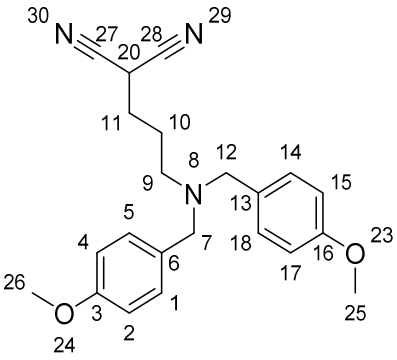
7.1.3 Characterisation

Melting points (Mpt) were recorded using a Stuart SMP40 apparatus and are reported in °C. Infrared (IR) spectra were recorded neat on a Thermo Scientific Nicolet iS5 with an iD7 ATR module. Selected absorption maxima (nmax) 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 (d) 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 QTOF; data acquisition and processing were performed using Agilent MassHunter Workstation software. Compound names were generated using ChemDraw v20.1

systematic naming. Atom numbering in structures is purely for the purposes of assignment and does not reflect IUPAC numbering conventions.

7.1.4 Synthetic procedures

2-(3-(bis(4-methoxybenzyl)amino)propyl)malononitrile, 13

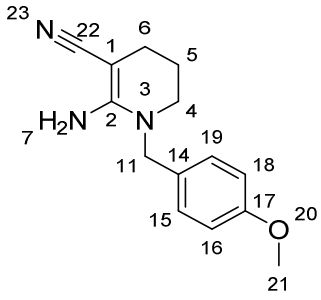


To a flask containing K_2CO_3 (36.8 g, 266 mmol) and 3-aminopropan-1-ol (5.1 mL, 66.6 mmol) was added MeCN (100 mL) and PMB-Cl (9.0 mL, 66.4 mmol) and stirred at 50 °C for 22 h. The resulting mixture was cooled, filtered, and evaporated to dryness. The residue was partitioned between DCM and water, the organic washed with brine, dried (Na_2SO_4) and evaporated to dryness and the resulting residue was used without further purification in the next step.

In a separate oven-dried flask, PPh_3 (13.3 g, 50.8 mmol) in THF (100 mL) was cooled to 0 °C and DIAD (10.0 mL, 50.8 mmol) was added dropwise and stirred for 30 min. The residue was then added in one portion and stirred. After an additional 30 min, malononitrile (2.80 g, 42.3 mmol) was added and the reaction stirred at room temperature for 18 h. The mixture was evaporated to dryness and the product purified by flash column chromatography (0-70% EtOAc in Cy) to give **13** as a clear oil (13.1 g, 54%).

V_{max} (cm^{-1}) 2973, 2884, 1651, 1380, 1086, 1046, 879; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 7.27–7.19 (m, 5H, *H1, 5, 14, 18*), 6.93–6.85 (m, 5H, *2, 4, 15, 17*), 4.70 (t, *J* = 6.7 Hz, *H20*), 3.73 (s, 6H, *H25, 26*), 3.43 (s, 4H, *H7, 12*), 2.39 (t, *J* = 6.6 Hz, 2H, *H9*), 1.96–1.86 (m, 2H, *H11*), 1.63 (dq, *J* = 6.7, 14.3 Hz, 2H, *H10*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 158.7 (*C3, 16*), 131.4 (*C6, 13*), 130.2 (*C1, 5, 14, 18*), 114.9 (*C27, 28*), 114.1 (*C2, 4, 15, 17*), 57.0 (*C7, 12*), 55.4 (*C25, 26*), 51.4 (*C9*), 28.1 (*C11*), 24.0 (*C10*), 22.4 (*C20*); **LR-ESI-MS** $C_{22}H_{25}N_3O_2$ $[M+H]^+$ *m/z* found 364.6 cald 364.5; **HR-ESI-MS** $C_{22}H_{25}N_3O_2$ $[M+Na]^+$ *m/z* found 386.1766 cald 386.1845.

2-amino-1-(4-methoxybenzyl)-1,4,5,6-tetrahydropyridine-3-carbonitrile, 14

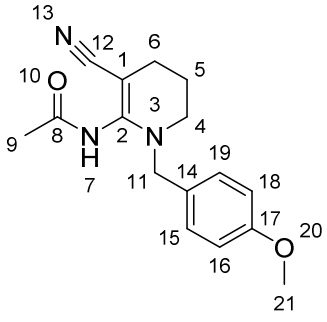


Malononitrile **13** (0.546g, 1.50 mmol) was added DCM (11.7 mL) and ACE-Cl (0.654 mL, 6.04 mmol). The mixture was stirred at r.t. for 3 h and the solvent evaporated. The resulting residue was added IPA (11.7 mL) and stirred at r.t. for 30 h. The reaction was quenched with saturated Na_2CO_3 and added water and DCM. The aqueous was extracted with DCM (3x) and the organic washed with brine, dried (Na_2SO_4) and dry loaded onto silica. The crude product was purified via flash column chromatography (0-10% MeOH in DCM) to give **14** (0.091 g, 0.375 mmol, 25%) as an orange gum.

V_{max}(cm^{-1}) 3301, 1446, 1023, 619; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 7.16 (d, *J* = 8.8 Hz, 2H, *H12, H16*), 6.91 (d, *J* = 8.8 Hz, 2H, *H13, H15*), 5.42 (s, 2H, *H7*), 4.34 (s, 2H, *H10*), 3.73 (s, 3H, *H18*), 3.02 (t, *J* = 5.4 Hz, 2H, *H4*), 2.09 (t, *J* = 6.4 Hz, 2H, *H6*), 1.58

(dt, *J* = 6.0, 5.8 Hz, 2H, *H5*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 158.9 (*C14*), 157.5 (*C2*), 130.1 (*C11*), 129.1 (*C12, C16*), 125.8 (*C8*), 114.3 (*C13, C15*), 55.5 (*C18*), 52.2 (*C10*), 49.1 (*C1*), 47.9 (*C4*), 23.3 (*C6*), 22.3 (*C5*); **LR-ESI-MS** $C_{14}H_{17}N_3O$ $[M+H]^+$ *m/z* found 244.5 cald 244.3; **HR-ESI-MS** $C_{14}H_{17}N_3O$ $[M+H]^+$ *m/z* found 244.1425 cald 244.1450.

N-(3-cyano-1-(4-methoxybenzyl)-1,4,5,6-tetrahydropyridin-2-yl)acetamide, 15

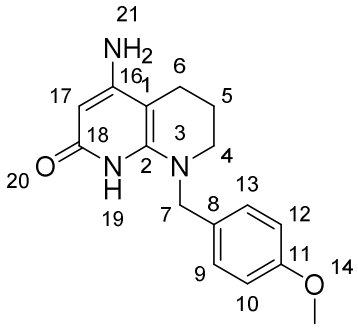


14 (60.3 mg, 0.25 mmol) was dissolved in MeCN (0.5 mL) and acetic anhydride (37 µL, 0.4 mmol) was added. The reaction was stirred at 50 °C for 2 h. The reaction was cooled, partitioned between DCM and water, and the organic washed with brine, dried (Na_2SO_4) and evaporated to dryness. The product was purified by flash column chromatography (0-10% MeOH in DCM) to give the product **15** as an orange gum (21.9 mg, 31%).

V_{max} (cm^{-1}) 2973, 2874, 1927, 2242, 1677, 1266, 1044; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 9.78 (s, 1H, *H7*), 7.22–7.15 (m, 2H, *H15, 19*), 6.94–6.84 (m, 2H, *H16, 18*), 4.19 (s, 2H, *H11*), 3.74 (s, 3H, *H21*), 3.06 (t, *J* = 5.4 Hz, 2H, *H4*), 2.19 (t, *J* = 6.3 Hz, 2H, *H6*), 1.98 (s, 3H, *H11*), 1.55 (p, *J* = 6.0 Hz, 2H, *H5*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 169.4 (*C8*), 159.0 (*C17*), 148.7 (*C2*), 129.6 (*C15, 19*), 129.5 (*C14*), 122.4 (*C12*), 114.3 (*C16, 18*), 72.1 (*C1*), 55.5 (*C21*), 52.5 (*C11*), 46.6 (*C4*), 24.0 (*C6*), 23.1 (*C9*), 21.0 (*C5*); **LR-**

ESI-MS $C_{16}H_{19}N_3O_2$ $[M+H]^+$ *m/z* found 286.5 cald 286.4; **HR-ESI-MS** $C_{16}H_{19}N_3O_2$ $[M+H]^+$ *m/z* found 286.1550 cald 286.1556.

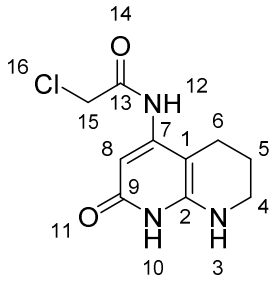
4-amino-8-(4-methoxybenzyl)-5,6,7,8-tetrahydro-1,8-naphthyridin-2(1H)-one, 16



15 (445 mg, 1.56 mmol) was dissolved in anhydrous THF (12 mL) and LiHMDS (1.0 M in THF, 9.4 mL, 9.36 mmol) was added. The reaction was stirred at 65 ºC for 6 h. The reaction was cooled, poured onto water and DCM was added. The aqueous layer was extracted with DCM (3x), and the combined extracts were washed with brine, dried (Na_2SO_4) and evaporated to dryness to give the crude product **16** as a red waxy solid (373 mg, 84%).

V_{max} (cm^{-1}) 2971, 2871, 1990, 1922, 1569, 1053, 1092; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 7.13–7.05 (m, 2H, *H9, 13*), 6.85–6.79 (m, 2H, *H10, 12*), 5.28 (s, 2H, *H21*), 5.26 (s, 1H, *H17*), 4.62 (s, 2H, *H7*), 3.70 (s, 3H, *H15*), 3.06 (t, *J* = 5.6 Hz, 2H, *H4*), 2.29 (t, *J* = 6.4 Hz, 2H, *H6*), 1.80–1.71 (m, 2H, *H5*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 161.5 (*C18*), 158.6 (*C11*), 156.0 (*C16*), 151.8 (*C2)*, 131.5 (*C8*), 129.2 (*C9, 13*), 114.1 (*C10, 12*), 85.8 (*C17*), 83.4 (*C1*), 55.4 (*C15*), 51.1 (*C7*), 47.3 (*C4*), 21.5 (*C5*), 21.1 (*C6*); **LR-ESI-MS** $C_{16}H_{19}N_3O_2$ $[M+H]^+$ *m/z* found 286.5 cald 286.4; **HR-ESI-MS** $C_{16}H_{19}N_3O_2$ $[M+Na]^+$ *m/z* found 308.0523 cald 308.1375.

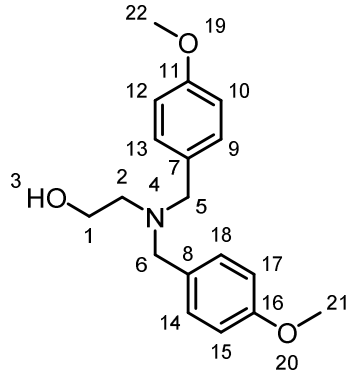
MC-278 (via route 1)



16 (100 mg, 0.35 mmol) was dissolved in DCM (1.8 mL) and added TEA (98 µL, 0.70 mmol) and chloroacetyl chloride (56 µL, 0.70 mmol) consecutively. The reaction was stirred for 1 h at room temperature and partitioned between DCM and water. The organic was washed with water, brine, dried (Na_2SO_4) and evaporated to dryness. The crude was dissolved in TFA (2.2 mL) and stirred at reflux for 2 h. The TFA was evaporated and the crude purified by preparative HPLC to yield **MC-278** (5.1 mg, 6%) as a green waxy solid.

V_{max} (cm^{-1}) 3300, 2974, 2883, 1651, 1380, 1086, 1045; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 10.31 (s, 1H, *H10*), 9.19 (s, 1H, *H12*), 5.97 (s, 1H, *H3*), 5.88 (t, J = 2.7 Hz, 1H, *H3*), 4.32 (s, 2H, *H15*), 3.19 (dq, J = 6.0, 2.8 Hz, 2H, *H4*), 2.36 (t, J = 6.4 Hz, 2H, *H6*), 1.72 (p, J = 6.1 Hz, 2H, *H5*). **^{13}C NMR** (101 MHz, DMSO-*d*$_6$) δ 165.5 (*C13*), 161.8 (*C9*), 149.6 (*C2*), 147.7 (*C7*), 94.0 (*C8*), 89.1 (*C1*), 43.9 (*C4*), 21.3 (*C5*), 20.6 (*C6*); **LR-ESI-MS** $C_{10}H_{12}ClN_3O_2$ $[M+H]^+$ *m/z* found 242.4/244.4 cald 242.1/244.1; **HR-ESI-MS** $C_{10}H_{12}ClN_3O_2$ $[M+H]^+$ *m/z* found 242.0024 cald 242.0696.

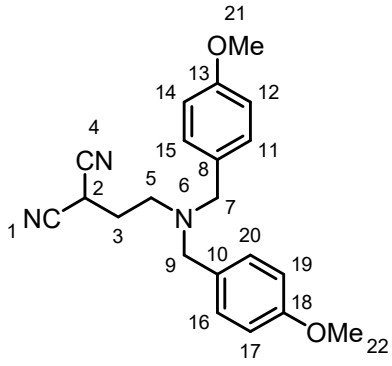
2-(bis(4-methoxybenzyl)amino)ethan-1-ol, 18



Ethanolamine (3.0 g, 0.05 mmol) was added MeCN (74 mL), K_2CO_3 (27.2 g, 0.2 mmol) and PMB-Cl (13.3 mL, 0.1 mmol) and heated to 50 °C for 16 h. The reaction was cooled, filtered and partitioned between DCM and water. The organic was washed with water (2x), brine (1x), dried (Na_2SO_4) and filtered. The solvent was evaporated to give crude **18** as an orange oil (11.4 g, 77%).

V_{max} (cm^{-1}) 2833, 1609, 1583, 1508, 1242, 1177; **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.18–7.10 (m, 4H, *H9, 13, 14, 18*), 6.84–6.73 (m, 4H, *H10, 12, 15, 17*), 3.72 (s, 6H, *H21, 22*), 3.50–3.46 (m, 6H, *H1, 5, 6*), 2.56 (t, *J* = 5.4 Hz, 2H, *H2*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 158.8 (*C11, 16*), 130.2 (*C9, 13, 14, 18*), 130.1 (*C7, 8*), 113.8 (*C10, 12, 15, 17*), 58.4 (*C1*), 57.3 (*C5, 6*), 55.3 (*C21, 22*), 54.3 (*C2*); **LR-ESI-MS** $C_{18}H_{23}NO_3$ $[M+H]^+$ *m/z* found 302.5 cald 302.4; **HR-ESI-MS** $C_{18}H_{23}NO_3$ $[M+H]^+$ *m/z* found 302.1877 cald. 302.1751.

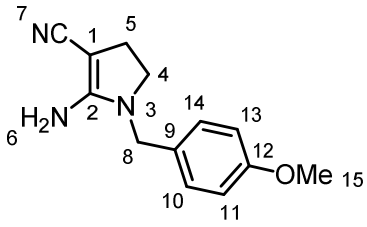
2-(2-(bis(4-methoxybenzyl)amino)ethyl)malononitrile, 19



To an oven-dried flask was added PPh_3 (4.4 g, 16.6 mmol) and *n*-BuOH (75 mL).The mixture was cooled over ice and DIAD (3.4 mL, 16.6 mmol) added and stirred for 30 min, then **18** (5.0 g, 16.6 mmol) added. After 30 min malononitrile (0.9 g, 13.8 mmol) was added and the reaction warmed to r.t. and stirred for a further 18 h. The solvent was evaporated and the crude purified by flash column chromatography (0-50% EtOAc in Cy) to give **19** as a yellow gum (3.9 g, 67%).

V_{max} (cm^{-1}) 1610, 1509, 1453, 1243, 1137; **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.23–7.15 (m, 4H, *H11, 15, 16, 20*), 6.92–6.85 (m, 4H, *H12, 14, 17, 18*), 3.96 (t, *J* = 7.2 Hz, 1H, *C2*), 3.81 (s, 6H, *H21, 22*), 3.52 (s, 4H, *H7, 9*), 2.73–2.65 (m, 2H, *H5*), 1.98 (td, *J* = 7.3, 5.9 Hz, 2H, *H3*). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 159.1 (*C13, 18*), 130.2 (*C8, 10*), 130.1 (*C11, 15, 16, 20*), 114.1 (*C12, 14, 17, 18*), 112.9 (*C1, 4*), 58.5 (*C7, 9*), 55.3 (*C21, 22*), 49.6 (*C5*), 30.2 (*C3*), 19.5 (*C2*); **LR-ESI-MS** $C_{21}H_{23}N_3O_2$ $[M+H]^+$ *m/z* found 350.5 cald 350.4; **HR-ESI-MS** $C_{21}H_{23}N_3O_2$ $[M+H]^+$ *m/z* found 350.1958 cald. 350.1863.

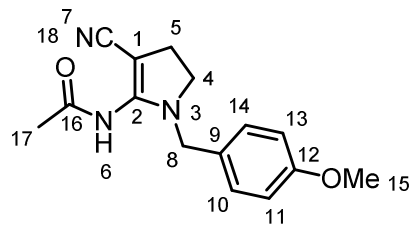
2-amino-1-(4-methoxybenzyl)-4,5-dihydro-1H-pyrrole-3-carbonitrile, 20



20 (1.3 g, 3.7 mmol) was added DCE (57 mL) and ACE-Cl (1.6 mL, 15.0 mmol) and stirred at reflux for 24 h. The solvent was evaporated, the residue added MeOH (57 mL) and stirred at 50 °C for 24 h. The reaction was cooled, quenched with Na_2CO_3, added water and DCM. The organic was washed with brine (1x), dried (Na_2SO_4), and evaporated to dryness. The crude was purified by flash column chromatography (0-5% MeOH in DCM) to yield **20** as a red gum (0.28 g, 32%).

V_{max} (cm^{-1}) 3439, 2102, 1887, 1294; **^{1}H NMR** (400 MHz, DMSO-*d*$_6$) δ 7.21–7.13 (m, 2H, *H10, 14*), 6.95–6.87 (m, 2H, *C11, 12*), 6.44 (s, 2H, *NH*$_2$), 4.20 (s, 2H, *H8*), 3.74 (s, 3H, *H15*), 3.06 (t, *J* = 8.7 Hz, 2H, *H4*), 2.43 (t, *J* = 8.8 Hz, 2H, *H5*); **^{13}C NMR** (101 MHz, DMSO-*d*$_6$) δ 163.8 (*C2*), 158.9 (*C12*), 129.7 (*C10, 14*), 129.4 (*C9*), 123.9 (*C7*), 114.3 (*C11, 13*), 55.5 (*C15*), 49.4 (*C4*), 48.9 (*C1*), 47.2 (*C8*), 25.3 (*C5*). **LR-ESI-MS** $C_{21}H_{23}N_3O_2$ $[M+H]^+$ *m/z* found 230.4 cald 230.3; **HR-ESI-MS** $C_{13}H_{15}N_3O$ $[M+H]^+$ *m/z* found 230.1297 cald. 230.1288

N-(3-cyano-1-(4-methoxybenzyl)-4,5-dihydro-1H-pyrrol-2-yl)acetamide, 21

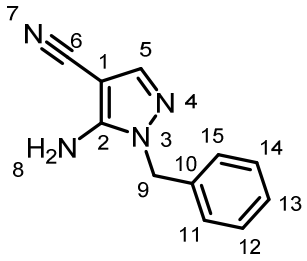


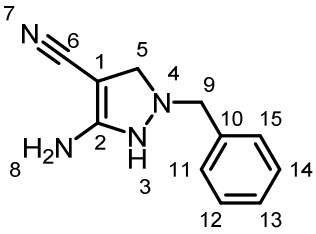
20 (270 mg, 1.2 mmol) was added MeCN (4.5 mL) and acetic anhydride (176 µL, 1.9 mmol) and stirred at 65 °C for 1 h. The reaction was cooled, partitioned between DCM and water, and the organic washed with water (1x)

and brine (1x) and dried (Na_2SO_4). The solvent was evaporated to yield crude **21** as a red gum that was used without further purification.

LR-ESI-MS $C_{15}H_{17}N_3O_2$ $[M+H]^+$ *m/z* found 272.5 cald 272.3

5-amino-1-benzyl-1H-pyrazole-4-carbonitrile, 25



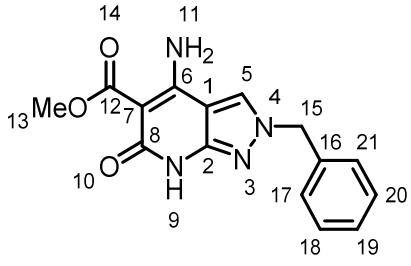


5-amino-1*H*-pyrazole-4-carbonitrile **15** (500 mg, 4.5 mmol), K_2CO_3 (961 mg, 7.0 mmol), and anhydrous DMF (5 mL) were charged into a N_2 flushed flask, followed by the dropwise addition of BnBr (650 µL, 5.44 mmol). The reaction was stirred at room temperature for 21 h. The resulting mixture was diluted in water, extracted with EtOAc, and the organic layer washed with brine. The organic was dried (Na_2SO_4) and evaporated to dryness, and the final product was purified by flash column chromatography (0-50% EtOAc in cyclohexane, Rf = 0.36) to give **16** as a white solid and a mixture of regioisomers (2*N*-benzyl : 1*N*-benzyl 3:1) (568 mg, 63%).

V_{max} (cm^{-1}) 3394, 3031, 2947, 2220, 1509; <u>*N1-Bn*</u> **^{1}H NMR** (400 MHz, DMSO- d_6) 7.59 (s, 1H, *H5*), 7.39–7.13 (m, 5H, *H10-15*), 6.72 (s, 2H, *H8*), 5.16 (s, 2H, *H9*); **^{13}C NMR** (101 MHz, DMSO- d_6) δ 50.4 (*C9*), 72.7 (*C1*), 115.7 (*C6*), 127.9 (*C13*), 128.3 (*C11, 15*), 128.9 (*C12, 14*), 137.2 (*C10*), 141.0 (*C5*), 152.1 (*C2*); <u>*N2-Bn*</u> δ **^{1}H NMR** (400 MHz, DMSO- d_6) 8.24 (s, 1H, *H5*), 7.39–7.13 (m, 5H, *H10-15*), 5.57 (s, 2H, *H8*), 5.08 (s, 2H, *H9*); **^{13}C NMR** (101 MHz, DMSO- d_6) δ 55.3 (*C9*), 77.2 (*C1*), 115.1 (*C6*), 127.7 (*C12*),

128.2 (*C10, 14*), 130.0 (*C11, 15*), 136.3 (*C5*), 137.0 (*C9*), 158.3 (*C2*) NB: *NMR assignments ambiguous due to mixture*; **LR-ESI-MS** $C_{11}H_{10}N_4$ $[M+H]^+$ *m/z* found 199.3 cald 199.2; **HR-ESI-MS** $C_{11}H_{10}N_4$ $[M+Na]^+$ *m/z* 221.0811 cald 221.0804.

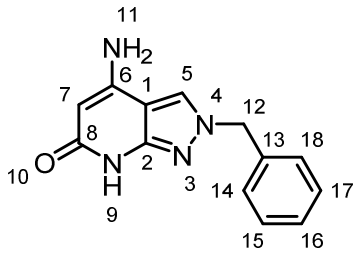
methyl 4-amino-2-benzyl-6-oxo-6,7-dihydro-1H-pyrazolo[3,4-b]pyridine-5-carboxylate, 26



To a nitrogen filled flask was charged MeOH (50.0 mL), NaOMe (25 wt. % in MeOH, 19.9 mL, 74.5 mmol) and dimethyl malonate (8.1 mL, 69.4 mmol) and stirred at room temperature for 15 min. **15** (11.00 g, 55.5 mmol) was charged and stirred at reflux for 2 days. The reaction was cooled, solid filtered, and the filtrate evaporated. To the filtrate was added MeOH and the solid filtered. The filtered solids were combined to give **17** as a white solid that was used without further purification (7.12 g, 43%).

Mpt (ºC) 233–239; **V_{max} (cm^{-1})** 3030, 2943, 2221, 1677, 1602, 1250, 996; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 10.73 (s, 1H, *H9*), 8.35 (s, 1H, *H5*), 8.28 (s, 2H, *H11*), 7.43–7.21 (m, 5H, *H17-H21*), 5.34 (s, 2H, *H22*), 3.66 (s, 3H, *H13*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 170.0 (*C12*), 162.4 (*C8*), 155.6 (*C6*), 149.6 (*C2*), 136.9 (*C16*), 129.1 (*C18, C20*), 128.5 (*C19*), 128.5 (*C17, C20*), 128.2 (*C5*), 100.4 (*C1*), 89.9 (*C7*), 56.0 (*C22*), 51.0 (*C13*); **LR-ESI-MS** $C_{15}H_{14}N_4O_3$ $[M+H]^+$ *m/z* found 299.3 cald 299.3; **HR-ESI-MS** $C_{15}H_{14}N_4O_3$ $[M+Na]^+$ *m/z* found 321.0078 cald 321.0964.

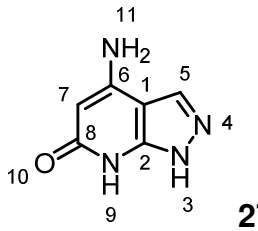
4-amino-2-benzyl-2,7-dihydro-6H-pyrazolo[3,4-b]pyridin-6-one, 27



26 (500 mg, 1.68 mmol) and 15% wt. NaOH solution (10 mL) were charged and stirred at reflux for 18 h. The reaction was cooled, acidified to pH 5, and the resulting solid was filtered and dried on high vacuum to give **18** as an off-white that was used without further purification (385 mg, 96%).

Mpt (°C) 218–220; **V_{max} (cm^{-1})** 3029, 2792, 1634, 1597, 1449, 1072, 814; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 10.53 (s, 1H, *H9*), 8.04 (s, 1H, *H5*), 7.41–7.24 (m, 5H, 14-18*H*), 6.42 (s, 2H, *H11*), 5.32 (s, 2H, *H12*), 4.96 (s, 1H, *H7*). **^{13}C NMR** (101 MHz, DMSO-d_6) δ 165.5 (*C8*), 150.7 (*C6*), 150.4 (*C2*), 137.4 (*C13*), 129.1 (*C15, C17*), 128.6 (*C16*), 128.3 (*C14, C18*), 125.4 (*C5*), 101.6 (*C1*), 89.1 (*C7*), 55.8 (*C12*); **LR-ESI-MS** $C_{13}H_{12}N_4O$ $[M+H]^+$ *m/z* found 241.3 cald 241.3; **HR-ESI-MS** $C_{13}H_{12}N_4O$ $[M+H]^+$ *m/z* found 241.0556 cald 241.1090.

4-amino-1,7-dihydro-6H-pyrazolo[3,4-b]pyridin-6-one, 28

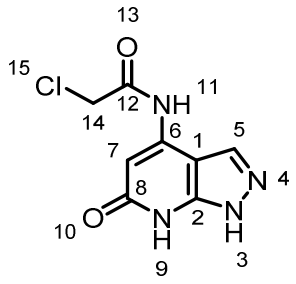


27 (1.00 g, 4.15 mmol) and Pd/C (50% wetted, 800 mg) were charged and the flask was evacuated and filled with nitrogen three times. MeOH (100 mL) and 4M HCl (in dioxane, 20 mL) were then added consecutively by needle under nitrogen. The flask was evacuated and filled with nitrogen three times. The flask was evacuated and the H_2 balloon inserted and tap opened. The flask was then evacuated and back

filled with H_2 twice before finally leaving the flask under H_2 for 3 days at room temperature. The reaction was filtered through celite and evaporated to dryness to give the crude product **28** as a green solid that was used without further purification (611 mg, 98%).

Mpt (°C) 288–290; **V_{max} (cm^{-1})** 3001, 2768, 1644, 1605, 1434, 1247, 799; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 8.09 (s, 1H, *H5*), 6.57 (s, 2H, *H11*), 4.98 (s, 1H, *H7*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 162.7 (*C8*), 157.0 (*C6*), 147.5 (*C2*), 127.4 (*C5*), 102.1 (*C1*), 82.5 (*C7*); **LR-ESI-MS** $C_6H_6N_4O$ $[M+H]^+$ *m/z* found 151.3 cald 151.1; **HR-ESI-MS** $C_6H_6N_4O$ $[M+H]^+$ *m/z* found 151.0626 cald 151.0612.

2-chloro-N-(6-oxo-6,7-dihydro-1H-pyrazolo[3,4-b]pyridin-4-yl)acetamide, 6

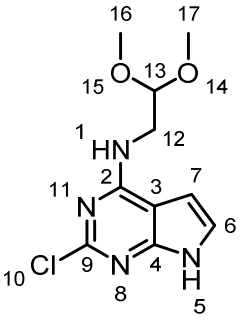


28 (50 mg, 0.33 mmol) and TEA (93 µL, 0.67 mmol) in DCM/DMF (1:1, 0.6 mL) was charged chloroacetyl chloride (106 µL, 1.33 mmol) and stirred at room temperature for 1 h. The reaction was evaporated, added MeOH (1.0 mL) and sat. Na_2CO_3 (0.5 mL) and stirred for 10 min. The reaction was partitioned between DCM and water, and the organic washed with brine, dried (Na_2SO_4), and evaporated to dryness. The crude was purified by preparative HPLC to give **6** as an off-white solid (3.6 mg, 5%).

Mpt (°C) 230–234; **V_{max} (cm^{-1})** 2973, 2886, 1380, 1087, 1044, 879; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 13.10 (s, 1H, *H3*) 11.40 (s, 1H, *H9*), 10.46 (s, 1H, *H11*), 8.26 (s, 1H, *H5*),

6.83 (s, 1H, *H7*), 4.44 (s, 2H, *H14*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 172.9 (*C12*), 166.7 (*C8*), 165.1 (*C2*), 141.4 (*C6*), 125.7 (*C5*), 101.7 (*C7*), 100.5 (*C1*), 43.9 (*C14*) (101 MHz, DMSO-d_6); **LR-ESI-MS** $C_8H_7ClN_4O_2$ $[M+H]^+$ *m/z* found 227.3/229.3 cald 227.0/229.0; **HR-ESI-MS** $C_8H_7ClN_4O_2$ $[M+Na]^+$ *m/z* found 249.0102 cald 249.0155.

2-chloro-N-(2,2-dimethoxyethyl)-7H-pyrrolo[2,3-d]pyrimidin-4-amine, 33

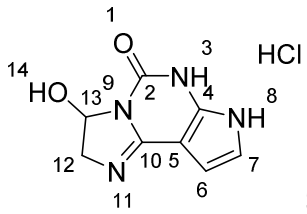


To 2,4-dichloro-7H-pyrrolo[2,3-d]pyrimidine (3.11 g, 16.5 mmol) was added a solution of aminoacetaldehyde dimethyl acetal (5.4 mL, 49.6 mmol) and TEA (2.80 mL, 19.8 mmol) in *n*-BuOH (75 mL) and stirred at 70 °C for 3.5 h. The resulting mixture was cooled, added water and extracted with DCM. The organic layer was washed with brine, dried (Na_2SO_4) and evaporated to dryness. The product was purified by flash column chromatography (0-70% EtOAc in Cy) to give **33** as a white solid (3.01 g, 71%).

Mpt (°C) 167–170; **V_{max} (cm^{-1})** 3329, 2971, 2887, 1582, 1279, 1046, 901; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 11.68 (s, 1H, *H5*), 7.97 (t, *J* = 5.7 Hz, 1H, *H1*), 7.09 (dd, *J* = 3.5, 2.2 Hz, 1H, *H6*), 6.61 (dd, *J* = 3.5, 1.7 Hz, 1H, *H7*), 4.56 (t, *J* = 5.3 Hz, 1H, *H13*), 3.53 (t, *J* = 5.7 Hz, 2H, *H12*), 3.32 (s, 6H, *H16, H17*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 157.3 (*C2)*, 152.9 (*C9*), 151.2 (*C4)*, 122.0 (*C6)*, 102.4 (*C13*), 101.6 (*C3*), 99.7 (*C7*), 53.8 (*C16, C17*), 42.3 (*C12*); **LR-ESI-MS** $C_{10}H_{13}ClN_4O_2$ $[M+H]^+$ *m/z* found 257.3/259.7

cald 257.1/259.1; **HR-ESI-MS** $C_{10}H_{13}ClN_4O_2$ $[M+H]^+$ *m/z* found 257.0836 cald 257.0805.

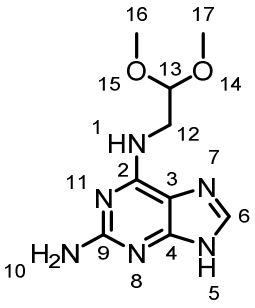
3-hydroxy-2,3,6,7-tetrahydro-5H-imidazo[1,2-c]pyrrolo[3,2-e]pyrimidin-5-one hydrochloride, 9



33 (25.0 mg, 0.097 mmol) was added 1M HCl (0.5 mL) and stirred at reflux for 3 days. The mixture was cooled and evaporated under reduced pressure to give **9** as an off-white solid (22.2 mg, quant).

Mpt (°C) 274–279; **V_{max} (cm^{-1})** 2962, 2831, 1708, 1602, 1311, 1077, 735; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 12.94 (s, 1H, *H3*), 12.00 (s, 1H, *N7, H8*), 11.52 (s, 1H, *H11*), 6.89 (dd, *J* = 3.5, 2.0 Hz, 1H, *H6*), 6.65 (dd, *J* = 3.5, 2.0 Hz, 1H, *H7*), 6.07 (dd, *J* = 7.2, 2.0 Hz, 1H, *H13*), 4.07 (dd, *J* = 12.6, 7.2 Hz, 1H, *H12'*), 3.67 (dd, *J* = 12.6, 2.0 Hz, 1H, *H12"*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 152.8 (*C10*), 146.2 (*C2*), 142.3 (*C4*), 121.3 (*C7*), 103.4 (*C6*), 89.9 (*C5*), 80.2 (*C13*), 52.4 (*C12*); **LR-ESI-MS** $C_8H_8N_4O_2$ $[M+H]^+$ *m/z* found 193.3 cald 193.2; **HR-ESI-MS** $C_8H_8N_4O_2$ $[M+H]^+$ *m/z* found 193.0749 cald 193.0726.

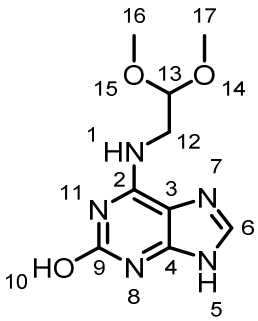
N^6-(2,2-dimethoxyethyl)-9H-purine-2,6-diamine, 35



To 2-amino-6-chloropurine (500 mg, 2.95 mmol) was added a solution of aminoacetaldehyde dimethyl acetal (963 µL, 8.84 mmol) and TEA (491 µL, 3.52 mmol) in *n*-BuOH (13 mL) and stirred at 70 °C for 24 h. The resulting mixture was evaporated to dryness and the product purified by flash column chromatography (0-10% MeOH in DCM) to give **35** as a white solid (585 mg, 83%).

Mpt (°C) 267–269; **V_{max} (cm^{-1})** 3469, 3357, 3230, 2989, 2830, 1600, 1391, 1069, 842; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 7.72 (s, 1H, *H1*), 7.01 (s, 1H, *H6*), 5.77 (s, 2H, *H10*), 4.67 (t, *J* = 5.5 Hz, 1H, *H13*), 3.68–3.55 (m, 2H, *H12*), 3.34 (s, 6H, *H16, 17*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 160.5 (*C9*), 154.8 (*C4*), 152.9 (*C2*), 136.1 (*C6*), 113.0 (*C3*), 102.3 (*C13*), 53.4 (*C16, 17*), 41.5 (*C12*);); **LR-ESI-MS** $C_9H_{14}N_6O_2$ $[M+H]^+$ *m/z* found 239.4 cald 239.3; **HR-ESI-MS** $C_9H_{14}N_6O_2$ $[M+Na]^+$ *m/z* found 261.1122 cald 261.1076.

6-((2,2-dimethoxyethyl)amino)-9H-purin-2-ol, 36

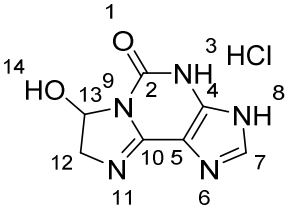


35 (50 mg, 0.21 mmol) was added AcOH (1.0 mL) and water (0.2 mL) at 0 °C. Sodium nitrite (44 mg, 0.64 mmol) in water (0.4 mL) was added dropwise and the solution stirred on ice for 2 h. The mixture was warmed to room temperature,

evaporated to dryness and added MeOH (1.0 mL) and sat. Na_2CO_3 (0.5 mL). The mixture was subsequently filtered and purifed by preperative HPLC collecting to give **36** as a white solid (3.5 mg, 7%).

Mpt (°C) 243–246; **V_{max} (cm^{-1})** 3691, 3321, 2972, 2883, 1380, 1088, 1047, 880; **^{1}H NMR** (400 MHz, DMSO-d_6) 7.14 (s, 1H, *H6*), 6.60 (s, 1H, *H1*), 4.58 (t, *J* = 5.9 Hz, 1H, *H13*), 3.54–3.45 (m, 2H, *H12*), 3.30 (s, 6H, *H16, 17*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 151.8 (*C4*), 145.9 (*C6*), 112.5 (*C3*), 102.4 (*C13*), 53.2 (*C16, 17*), 40.9 (*C12*); **LR-ESI-MS** $C_9H_{13}N_5O_3$ $[M+H]^+$ *m/z* found 240.4 cald 240.2; **HR-ESI-MS** $C_9H_{13}N_5O_3$ $[M+Na]^+$ *m/z* found 262.0223 cald 261.0916.

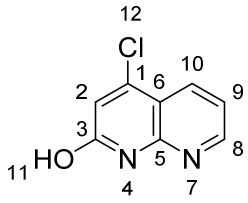
7-hydroxy-3,4,7,8-tetrahydro-5H-imidazo[2,1-i]purin-5-one hydrochloride, 10



36 (3.5 mg, 0.02 mmol) was dissolved in 1M HCl (1 mL) and stirred at reflux for 24 h. The reaction was subsequently evaporated to give **10** as a brown solid (2.8 mg, 98%).

Mpt (°C) 192–194 (degradation); **V_{max} (cm^{-1})** 2972, 2869, 1714, 1668, 1305, 1081, 707; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 12.96 (s, 1H, *H3*), 11.18 (s, 1H, *H8*), 8.24 (s, 1H, *H7*), 6.10 (dd, *J* = 7.2, 2.2 Hz, 1H, *H13*), 4.13 (dd, *J* = 12.9, 7.2 Hz, 1H. *H12'*), 3.75 (dd, *J* = 12.9, 2.2 Hz, 1H, *H12"*). **^{13}C NMR** (101 MHz, DMSO-d_6) δ 146.7 (*C2*), 80.7 (*C13*), 53.1 (*C12*); **LR-ESI-MS** $C_7H_7N_5O_2$ $[M+H]^+$ *m/z* found 194.3 cald 194.1; **HR-ESI-MS** $C_7H_7N_5O_2$ $[M+H]^+$ *m/z* found 194.0624 cald 194.0678.

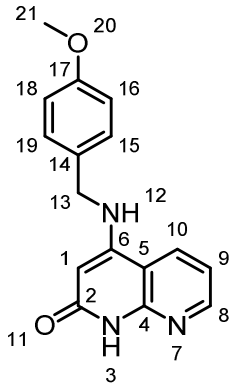
4-chloro-1,8-naphthyridin-2-ol, 40



To a solution of 2,4-dichloroquinoline (500 mg, 2.51 mmol) in dioxane (5 mL) was added HCl (6N, 7.5 mL) dropwise at 0 °C. The mixture stirred at reflux for 17 h, cooled and water added. The resulting precipitate was filtered and dried under vacuum to give **40** as a white solid (455 mg, quant).

Mpt (°C) 264–267; **V_{max}(cm^{-1})** 3306, 2945, 2830, 1667, 1426, 1023, 660; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 12.45 (s, 1H, *H11*), 8.63 (dd, *J* = 4.7, 1.7 Hz, 1H, *H10*), 8.26 (dd, *J* = 8.0, 1.7 Hz, 1H, *H8*), 7.38 (dd, *J* = 8.0, 4.7 Hz, 1H, *H9*), 6.90 (s, 1H, *H2*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 161.6 (*C3*), 152.5 (*C1*), 149.7 (*C10*), 143.4 (*C5*), 134.3 (*C8*), 122.8 (*C2*), 119.4 (*C9*), 113.4 (*C6*); **LR-ESI-MS** $C_8H_5ClN_2O$ $[M+H]^+$ *m/z* found 181.2 cald 181.1; **HR-ESI-MS** $C_8H_5ClN_2O$ $[M+H]^+$ *m/z* found 181.0161 cald 181.0169.

4-((4-methoxybenzyl)amino)-1,8-naphthyridin-2(1H)-one, 41

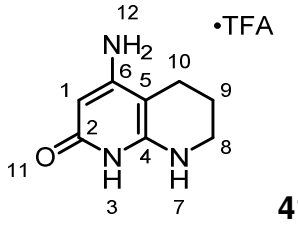


40 (1.04 g, 5.6 mmol) was added 4-methoxybenzylamine (12.3 mL) and DMSO (26.1 mL) and stirred at 150 °C for 1 h. The reaction was cooled, water

added, and the precipitate filtered. The resulting solid was washed with DCM/MeOH to give **41** as a white solid (1.2 g, 73%).

Mpt (°C) 201-202 **V_{max} (cm^{-1})** 3234, 2835, 1628, 1595, 1459, 1334, 1221; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 11.07 (s, 1H, *H3*), 8.49–8.40 (m, 2H, *H8, 10*), 7.71 (t, *J* = 5.9 Hz, 1H, *H12*), 7.33–7.25 (m, 2H, *H15, 19*), 7.20 (dd, *J* = 8.0, 4.7 Hz, 1H, *H9*), 6.95–6.87 (m, 2H, *H16, 18*), 5.19 (d, *J* = 1.6 Hz, 1H, *H1*), 4.37 (d, *J* = 5.7 Hz, 2H, *H13*), 3.73 (s, 3H, *H21*). **^{13}C NMR** (101 MHz, DMSO-d_6) δ 164.4 (*C2*), 159.2 (*C14*), 151.2 (*C8*), 150.9 (*C4*), 131.9 (*C10*), 131.0 (*C14*), 129.1 (*C15, 19*), 117.8 (*C9*), 114.8 (*C16, 18*), 110.1 (*C5*), 92.7 (*C1*), 55.9 (*C21*), 46.1 (*C13*); **LR-ESI-MS** $C_{16}H_{15}N_3O_2$ $[M+H]^+$ *m/z* found 282.4 cald 282.1; **HR-ESI-MS** $C_{16}H_{15}N_3O_2$ $[M+H]^+$ *m/z* found 282.1245 calc. 282.1237.

4-amino-5,6,7,8-tetrahydro-1,8-naphthyridin-2(1H)-one, 42

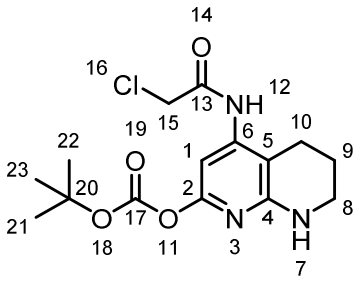


41 (991 mg, 3.52 mmol) was dissolved in formic acid (50 mL) and MeOH (20 mL) and the mixture passed through a Pd/C CatCart (40 °C, 20 bar, 0.5 mL/min, H-cube midi). The resulting mixture was evaporated to dryness, TFA (12 mL) added, and stirred for 13 h at 60 °C. The resulting mixture was added DCM, washed with water, and the aqueous evaporated to dryness to give **42** as an orange solid (900 mg, 92%).

Mpt (°C) 190–195 **V_{max} (cm^{-1})** **^{1}H NMR** (400 MHz, DMSO-d_6) 11.92 (s, 1H, *H3*), 6.80 (s, 2H, *H12*), 6.70 (s, 1H, *H7*), 5.45 (s, 1H, *H1*), 3.23 (t, *J* = 5.5 Hz, 2H, *H8*), 2.27 (t, *J*

= 6.3 Hz, 2H, *H10*), 1.81–1.72 (m, 2H, *H9*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 158.5 (*C6*), 155.6 (*C2*), 148.0 (*C4*), 88.5 (*C5*), 81.9 (*C1*), 40.1 (*C8*), 20.3 (*C9*), 19.5 (*C10*). **LR-ESI-MS** $C_8H_{11}N_3O$ $[M+H]^+$ *m/z* found 166.3 cald 166.1; **HR-ESI-MS** $C_8H_{11}N_3O$ $[M+H]^+$ *m/z* found 166.0981 cald.166.0975.

***tert*-butyl (4-(2-chloroacetamido)-5,6,7,8-tetrahydro-1,8-naphthyridin-2-yl) carbonate, 43**

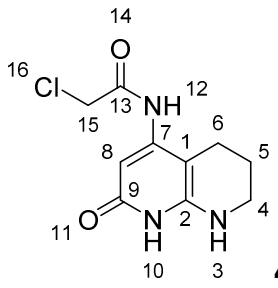


42 (150 mg, 0.54 mmol) was added DMF (2.1 mL), TEA (300 µL, 2.15 mmol), and Boc_2O (117 mg, 0.54 mmol) and stirred at r.t. for 16 h. The reaction was cooled to 0 °C and chloroacetyl chloride (43 µL, 0.54 mmol) in DMF (0.6 mL) was added dropwise. The reaction was warmed to rt, partitioned between DCM and water, and the organic extracted with water (3x). The organic was washed with brine, filtered through a phase separator and the solvent evaporated. The crude was purified by preparative HPLC to yield **43** as an off-white solid (128 mg, 70%).

Mpt (°C) 213–226 **V_{max} (cm^{-1})** 3314, 2942, 2830, 1448, 1022; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 9.58 (s, 1H, *H12*), 6.78 (d, *J* = 2.6 Hz, 1H, *H7*), 6.59 (s, 1H, *H1*), 4.34 (s, 2H, *H15*), 3.26–3.18 (m, 2H, *H8*), 2.52 (d, *J* = 8.7 Hz, 2H, *H10*), 1.76 (d, *J* = 5.9 Hz, 2H, *H9*), 1.46 (s, 9H, *H21, 22, 23*). **^{13}C NMR** (101 MHz, DMSO-d_6) δ 165.7 (*C13*), 156.2 (*C2*), 155.2 (*C6*), 151.3 (*C4*), 145.2 (*C17*), 104.1 (*C5*), 96.2 (*C1*), 83.3 (*C20*), 43.8 (*C15*), 40.4 (*C8*), 27.7 (*C21, 22, 23*), 22.0 (*C10*), 20.6 (*C9*). **LR-ESI-MS** $C_8H_{11}N_3O$

$[M+H]^+$ *m/z* found 342.5/344.4 cald 342.1/344.1; **HR-ESI-MS** $C_{15}H_{20}ClN_3O_4$ $[M+H]^+$ found 342.1301 cald. 342.1215.

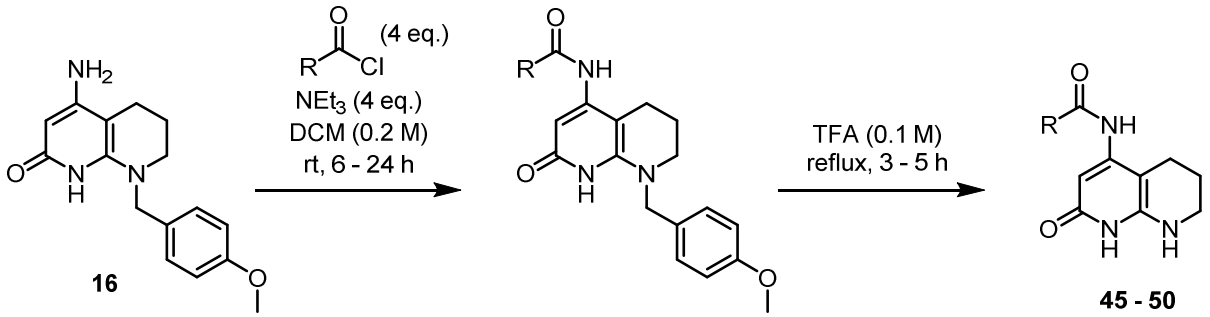
MC-278 (via route 2)



43 (128 mg, 0.37 mmol) was added DCM (1.9 mL) and 1-methylpiperazine (50 µL, 0.45 mmol) and stirred for 2 h. The DCM was blown off with nitrogen and the residue purified by preparative HPLC to give an off-white waxy solid (30.3 mg, 34%).

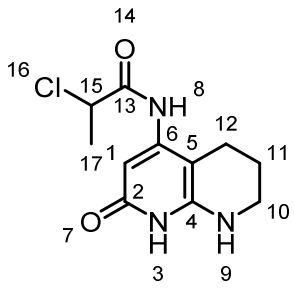
V_{max} (cm^{-1}) 3300, 2974, 2883, 1651, 1380, 1086, 1045; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 10.31 (s, 1H, *H10*), 9.19 (s, 1H, *H12*), 5.97 (s, 1H, *H8*), 5.88 (t, J = 2.7 Hz, 1H, *H3*), 4.32 (s, 2H, *H15*), 3.19 (dq, J = 6.0, 2.8 Hz, 2H, *H4*), 2.36 (t, J = 6.4 Hz, 2H, *H6*), 1.72 (p, J = 6.1 Hz, 2H, *H5*). **^{13}C NMR** (101 MHz, DMSO-d_6) δ 165.5 (*C13*), 161.8 (*C9*), 149.6 (*C2*), 147.7 (*C7*), 94.0 (*C8*), 89.1 (*C1*), 43.9 (*C4*), 21.3 (*C5*), 20.6 (*C6*); **LR-ESI-MS** $C_{10}H_{12}ClN_3O_2$ $[M+H]^+$ *m/z* found 242.4/244.4 cald 242.1/244.1; **HR-ESI-MS** $C_{10}H_{12}ClN_3O_2$ $[M+H]^+$ *m/z* found 242.0024 cald 242.0696.

General procedure for synthesis of compounds 45–50



25 was dissolved in DCM (0.2 M) and added TEA (4 eq.) and the corresponding acid chloride (4 eq.) consecutively. The reaction was stirred for 6-24 h at room temperature and partitioned between DCM and water. The organic was washed with water, brine, dried (Na_2SO_4) and evaporated to dryness. The crude was dissolved in TFA (0.1 M) and stirred at reflux for 3-5 h. The TFA was evaporated under reduced pressure and the products **45–50** were purified by preparative HPLC. All yields are reported over two steps.

2-chloro-N-(2-oxo-1,2,5,6,7,8-hexahydro-1,8-naphthyridin-4-yl)propanamide, 45

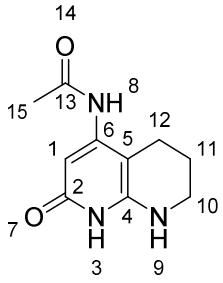


Synthesised as per the general procedure to give **45** as a purple gum (1.3 mg, 14%)

V_{max} (cm^{-1}) 3345, 2979, 2301, 1265, 1099, 956; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 10.42 (s, 1H, *H3*), 9.24 (s, 1H, *H8*), 5.90 (s, 1H, *H1*), 5.89 (s, 1H, *H9*), 4.83 (q, *J* = 6.7 Hz, 1H, *H15*), 3.23–3.15 (m, 2H, *H10*), 2.36 (dt, *J* = 7.9, 4.7 Hz, 2H, *H12*), 1.78–1.67 (m, 2H, *H11*), 1.58 (d, *J* = 6.7 Hz, 3H, *H17*). **^{13}C NMR** (101 MHz, DMSO-d_6) δ 168.2

(*C13*), 161.76 (*C2*), 54.9 (*C15*), 40.4 (*C10*), 21.5 (*C17*), 21.3 (*C11*), 20.7 (*C12*); **LR-ESI-MS** $C_{11}H_{14}ClN_3O_2$ $[M+H]^+$ *m/z* found 256.3/258.3 cald 256.1/258.1; **HR-ESI-MS** $C_{11}H_{14}ClN_3O_2$ $[M+H]^+$ found 256.0349 cald. 256.0847.

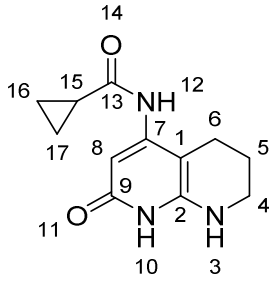
N-(2-oxo-1,2,5,6,7,8-hexahydro-1,8-naphthyridin-4-yl)acetamide, 46



Synthesised as per the general procedure to give **46** as a green waxy solid (1.3 mg, 10%)

V_{max} (cm^{-1}) 3256, 2972, 2206, 1674, 1377, 1044, 878; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 8.83 (s, 1H, *H8*), 6.02 (s, 1H, *H1*), 5.80 (s, 1H, *H9*), 3.19–3.15 (m, 2H, *H10*), 2.36 (t, *J* = 6.3 Hz, 2H, *H12*), 2.05 (s, 3H, *H15*), 1.81–1.66 (m, 2H, *H11*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 170.1 (C2), 169.1 (*C13*), 94.0 (*C1*), 40.3 (*C10*), 24.3 (*C15*), 21.1 (*C11),* 20.6 (*C12*); **LR-ESI-MS** $C_{10}H_{13}N_3O_2$ $[M+H]^+$ *m/z* found 208.5 cald 208.1; **HR-ESI-MS** $C_{10}H_{13}N_3O_2$ $[M+H]^+$ *m/z* found 208.0505 cald 208.1086.

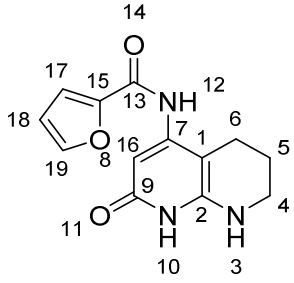
N-(2-oxo-1,2,5,6,7,8-hexahydro-1,8-naphthyridin4yl)cyclopropanecarboxamide, 47



Synthesised as per the general procedure to give **47** as a green waxy solid (5.1 mg, 13%).

V_{max} (cm^{-1}) 3326, 2974, 2890, 1654, 1381, 1086, 1044; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 9.05 (s, 1H, *H12*), 6.01 (s, 1H, *H8*), 5.86 (br s, 1H, *H3*), 3.16 (dt, *J* = 2.9, 6.3 Hz, 2H, *H4*), 2.38 (t, *J* = 6.3 Hz, 2H, *H6*), 2.03–1.92 (m, 1H, *H15*), 1.73 (p, *J* = 6.3 Hz, 2H, *H5*), 0.80–0.71 (m, 4H, *H16, 17*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 172.6 (*C13*), 161.9 (*C9*), 149.4 (*C2*), 148.5 (*C7*), 93.8 (*C8*), 88.4 (*C1*), 40.4 (*C4*), 21.4 (*C5*), 20.6 (*C6*), 14.8 (*C15*), 8.0 (*C16, 17*); **LR-ESI-MS** $C_{12}H_{15}N_3O_2$ $[M-H]^-$ *m/z* found 232.1 cald 232.1; **HR-ESI-MS** $C_{12}H_{15}N_3O_2$ $[M+H]^+$ *m/z* found 234.0588 cald 234.1243.

N-(2-oxo-1,2,5,6,7,8-hexahydro-1,8-naphthyridin-4-yl)furan-2-carboxamide, 48

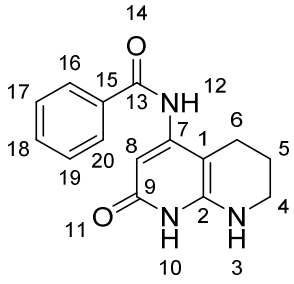


Synthesised as per the general procedure to give **48** as a pink solid (2.2 mg, 12%)

Mpt (°C) 247–249; **V_{max} (cm^{-1})** 3278, 2973, 2889, 1654, 1086, 1045, 879; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 10.29 (br s, 1H, *H10*), 9.21 (s, 1H, *H12*), 7.92 (d, *J* = 1.7 Hz,

1H, *H19*), 7.30 (d, *J* = 3.5 Hz, 1H, *H17*), 6.70 (dd, *J* = 3.5, 1.7 Hz, 1H, *H18*), 5.95 (br s, 1H, *H3*), 5.86 (s, 1H, *H8*), 3.20 (dt, *J* = 2.9, 6.3 Hz, 2H, *H4*), 2.39 (t, *J* = 6.3 Hz, 2H, *H6*), 1.72 (p, *J* = 6.3 Hz, 2H, *H7*). **^{13}C NMR** (101 MHz, DMSO-d_6) δ 161.8 (*C9*), 156.4 (*C13*), 147.9 (*C2*) 147.7 (*C15*), 146.3 (*C19*), 115.6 *(C17*), 112.7 (*C18*), 94.8 (*C8*), 40.5 (*C4*), 21.4 (*C5*), 20.7 (*C6*); **LR-ESI-MS** $C_{13}H_{13}N_3O_2$ $[M+H]^+$ *m/z* found 260.4 cald 260.1; **HR-ESI-MS** $C_{13}H_{13}N_3O_2$ $[M+H]^+$ *m/z* found 260.0310 cald 260.1035.

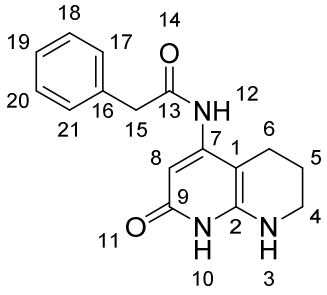
N-(2-oxo-1,2,5,6,7,8-hexahydro-1,8-naphthyridin-4-yl)benzamide, 49



Synthesised as per the general procedure to give **49** as an off-white solid (1.9 mg, 14%)

Mpt (°C) 262–265; **V_{max} (cm^{-1})** 3311, 2971, 2875, 1977, 1845, 1377, 1090, 1049, 881; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 10.39 (br s, 1H, *H10*), 9.48 (s, 1H, *H12*), 7.94–7.87 (m, 2H, *H16, 20*), 7.63–7.55 (m, 1H, *H18*), 7.55–7.47 (m, 2H, *H17, 19*), 5.97 (s, 1H, *H3*), 5.79 (s, 1H, *H8*), 3.20 (dt, *J* = 2.9, 6.3 Hz, 2H, *H4*), 2.40 (t, *J* = 6.3 Hz, 2H, *H6*), 1.71 (p, *J* = 6.3 Hz, 2H, *H5*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 165.8 (*C13*), 161.8 (*C9*), 134.9 (*C15*), 132.2 (*C18*), 128.9 (*C17, 19*), 128.2 (*C16, 20*), 40.5 (*C4*), 21.4 (*C5*), 21.0 (*C6*); **LR-ESI-MS** $C_{15}H_{15}N_3O_2$ $[M+H]^+$ *m/z* found 270.4 cald 270.1; **HR-ESI-MS** $C_{15}H_{15}N_3O_2$ $[M+H]^+$ *m/z* found 270.0490 cald 270.1243.

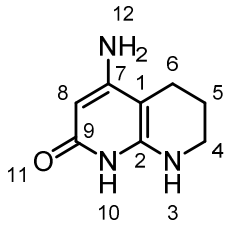
N-(2-oxo-1,2,5,6,7,8-hexahydro-1,8-naphthyridin-4-yl)-2-phenylacetamide, 50



Synthesised as per the general procedure to give **50** as an off-white solid (2.6 mg, 13%).

Mpt (°C) 268–270; **V_{max} (cm^{-1})** 3307, 2974, 2881, 1661, 1087, 1045, 879; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 10.16 (br s, 1H, *H10*), 8.99 (s, 1H, *H12*), 7.36–7.15 (m, 5H, *aryl*), 5.99 (s, 1H, *H8*), 5.87 (br s, 1H, *H3*), 3.69 (s, 2H, *H15*), 3.17 (dt, *J* = 2.9, 6.3 Hz, 2H, *H4*), 2.32 (t, *J* = 6.3 Hz, 2H, *H6*), 1.71 (p, *J* = 6.3 Hz, 2H, *H5*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 170.0 (*C13*), 161.9 (*C9*), 149.4 (*C2*), 148.4 (*C7*), 136.4 (*C16*), 129.6 (*C17, 21*), 128.8 (*C18, 20*), 127.0 (*C19*), 94.0 (*C8*), 88.9 (*C1*), 43.5 (*C15*), 40.3 (*C4*) 21.4 (*C5*), 20.6 (*C6*); **LR-ESI-MS** $C_{16}H_{17}N_3O_2$ $[M-H]^-$ *m/z* found 282.2 cald 282.1; **HR-ESI-MS** $C_{16}H_{17}N_3O_2$ $[M+H]^+$ *m/z* found 284.0607 cald 284.1399.

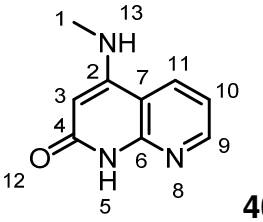
4-amino-5,6,7,8-tetrahydro-1,8-naphthyridin-2(1H)-one, 51



17b (35 mg, 0.123 mmol) was dissolved in TFA (2 mL) and stirred at 70 °C for 3 h. The TFA was evaporated under reduced pressure and the final product was purified by preparative HPLC to give **7** as green waxy solid (14.6 mg, 72%).

V_{max} (cm^{-1}) 3266, 2973, 2885, 1670, 1183, 1044; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 6.67 (s, 2H, *H12*), 5.36 (s, 1H, *H8*), 3.22 (t, *J* = 6.2 Hz, 2H, *H4*), 2.26 (t, *J* = 6.2 Hz, 2H, *H6*), 1.77 (m, 2H, *H5*). **^{13}C NMR** (101 MHz, DMSO-*d*6) δ 158.5 (*C9*), 156.3 (*C7*), 148.0 (*C2*), 87.9 (*C8*), 82.1 (*C1*), 40.1 (*C4*), 20.5 (*C5*), 19.5 (*C6*); **LR-ESI-MS** $C_8H_{11}N_3O$ $[M+H]^+$ *m/z* found 166.3 cald 166.1; **HR-ESI-MS** $C_8H_{11}N_3O$ $[M+H]^+$ *m/z* found 166.0506 cald 166.0980.

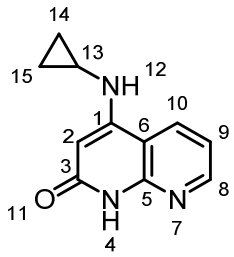
4-(methylamino)-1,8-naphthyridin-2(1H)-one, 67



40 (572 mg, 3.17 mmol) was added methylamine in EtOH (33%, 7 mL) and DMSO (8 mL) and stirred at 60 °C for 18 h. The reaction was cooled and the precipitate filtered under pressure to give **67** as a white solid (545 mg, 98%).

Mtp (°C) 338–349; **V_{max} (cm^{-1})** 3292, 2945, 2830, 1423, 1023, 629; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 11.04 (s, 1H, *H5*), 8.45 (dd, *J* = 4.7, 1.7 Hz, 1H, *H11*), 8.29 (dd, *J* = 8.0, 1.7 Hz, 1H, *H9*), 7.18 (d, *J* = 4.6 Hz, 1H, *H13*), 7.16 (dd, *J* = 8.0, 4.7 Hz, 1H, *H10*), 5.23 (s, 1H, *H3*), 2.80 (d, *J* = 4.6 Hz, 3H, *H1*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 164.1 (*C4*), 151.9 (*C2*), 150.7 (*C11*), 150.5 (*C6*), 131.2 (*C9*), 117.3 (*C10*), 109.5 (*C7*), 90.9 (*C3*), 29.9 (*C1*); **LR-ESI-MS** $C_9H_{13}N_3O$ $[M+H]^+$ *m/z* found 176.5 cald 176.1; **HR-ESI-MS** $C_9H_{13}N_3O$ $[M+H]^+$ *m/z* found 176.0812 cald 176.0824.

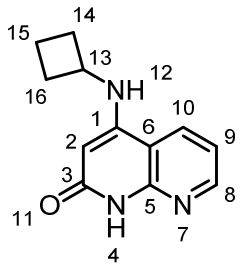
4-(cyclopropylamino)-1,8-napthyridin-2-ol, 68



40 (63 mg, 0.35 mmol) was added cyclopropylamine (720 µL, 10 mmol) and DMSO (1.3 mL) and stirred at 46 °C for 16 h. The reaction mixture was evaporated to dryness and the crude purified by flash column chromatography (0-5% MeOH in DCM) to give **68** as a white solid (9.6 mg, 14%).

Mtp (°C) 234–237; **V_{max}(cm^{-1})** 3278, 2942, 2830, 1024, 421; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 11.12 (s, 1H, *H4*), 8.44 (dd, *J* = 4.7, 1.7 Hz, 1H, *H10*), 8.33 (dd, *J* = 8.2, 1.6 Hz, 1H, *H8*), 7.34 (s, 1H, *H12*), 7.15 (dd, *J* = 8.0, 4.7 Hz, 1H, *H9*), 5.64 (s, 1H, *H2*), 1.24 (s, 1H, *H13*), 0.88-0.80 (m, 2H, *H14'*, *15"*), 0.61-0.52 (m, 2H, *H14"*, *15'*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 164.0 (*C3*), 152.1 (*C1*), 150.7 (*C10*), 150.4 (*C5*), 131.4 (*C8*), 117.3 (*C9*), 93.1 (*C2*), 24.8 (*C13*), 7.05 (*C14*, *15*); **LR-ESI-MS** $C_{11}H_{11}N_3O$ $[M+H]^+$ *m/z* found 202.3 cald 202.1; **HR-ESI-MS** $C_{11}H_{11}N_3O$ $[M+H]^+$ *m/z* found 202.0971 cald 202.0980.

4-(cyclobutylamino)-1,8-naphthyridin-2(1H)-one, 69

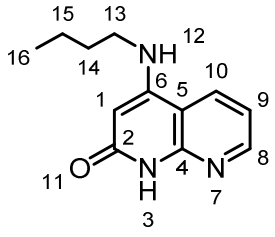


40 (50 mg, 0.27 mmol) was added cyclobutylamine (940 µL, 11 mmol) and DMSO (1.2 mL) and stirred at 80 °C for 16 h. The reaction was cooled, partitioned between DCM and water, and the organic evaporated to dryness. The final product

was purified by flash column chromatography (0-5% MeOH in DCM) to give **69** as a white solid (39 mg, 66%).

Mtp (°C) 265–276; **V_{max}(cm^{-1})** 3261, 2938, 2828, 1025, 605; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 11.07 (s, 1H, *H4*), 8.46 (q, *J* = 1.7 Hz, 1H, *H8*), 8.44 (s, 1H, *H10*), 7.21–7.12 (m, 1H, *H9*), 7.10 (d, *J* = 5.7 Hz, 1H, *H12*), 5.17 (d, *J* = 1.5 Hz, 1H, *H2*), 3.95 (h, *J* = 7.7 Hz, 1H, *H13*), 2.45–2.32 (m, 1H, *H16*), 2.13–1.99 (m, 2H, *H14'*, *15"*), 1.87–1.71 (m, 2H, *H14"*, *15'*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 164.0 (*C3*), 150.7 (*C1*), 150.5 (*C10*), 149.9 (*C5*), 131.7 (*C8*), 117.2 (*C9*), 109.5 (*C6*), 92.2 (*C2*), 48.2 (*C13*), 29.8 (*C16*), 15.6 (*C14*, *15*); **LR-ESI-MS** $C_{12}H_{13}N_3O$ $[M+H]^+$ *m/z* found 216.6 cald 216.1; **HR-ESI-MS** $C_{12}H_{13}N_3O$ $[M+H]^+$ *m/z* found 216.1129 cald 216.1137.

4-(butylamino) -1,8-napthyridin-2(*1H*)-one, 70

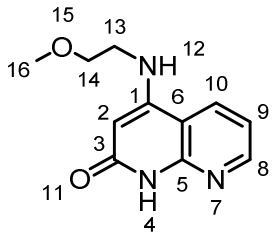


40 (64 mg, 0.34 mmol) was added *n*-butylamine (1.0 mL, 11 mmol) and DMSO (1.6 mL) and stirred at 45 °C for 4 h. The reaction was cooled to 0 °C and the resulting precipitate collected by vacuum filtration to give **70** as a white solid (41 mg, 53%)

Mtp (°C) 220–230; **V_{max}(cm^{-1})** 3300, 2943, 2831, 1023, 628; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 11.03 (s, 1H, *H3*), 8.44 (dd, J = 4.7, 1.6 Hz, 2H, *H10*), 8.40 (d, J = 7.9 Hz, 1H, *H8*), 7.17 (dd, *J* = 8.0, 4.7 Hz, 1H, *H9*), 6.99 (t, *J* = 5.3 Hz, 1H, *H12*), 5.28 (d, *J* = 6.8 Hz, 1H, *H1*), 3.16 (q, *J* = 6.8 Hz, 2H, *H13*), 1.62 (p, *J* = 7.5 Hz, 2H, *H14*), 1.40 (dq, *J* = 14.6, 7.3 Hz, 2H, *H15*), 0.94 (t, *J* = 7.3 Hz, 3H, *H16*); **^{13}C NMR** (101 MHz, DMSO-

d$_6$) δ 164.1 (*C2*), 151.0 (*C6*), 150.7 (*C10*), 150.5 (*C4*), 131.5 (*C8*), 117.2 (*C9*), 109.6 (*C5*), 91.0 (*C1*), 42.8 (*C13*), 30.1 (*C14*), 20.3 (*C15*), 14.2 (*C16*); **LR-ESI-MS** $C_{12}H_{15}N_3O$ $[M+H]^+$ *m/z* found 218.3 cald 218.1; **HR-ESI-MS** $C_{12}H_{15}N_3O$ $[M+H]^+$ *m/z* found 218.1288 cald 218.1293.

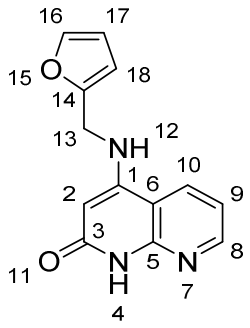
4-((2-methoxyethyl)amino)-1,8-naphthyridin-2(1H)-one, 71



40 (62 mg, 0.34 mmol) was added 2-methoxyethylamine (1.2 mL, 14 mmol) and DMSO (1.6 mL) and stirred at 90 ºC for 4 h. The reaction was cooled and partitioned between DCM and water and the organic filtered through a phase separator and the solvent evaporated. The crude was purified by flash column chromatography (0-5% MeOH in DCM) to give **71** as a white solid (53 mg, 71%).

Mtp (˚C) 178–204; **V_{max}(cm^{-1})** 3278, 3940, 2830, 1023, 631; **^{1}H NMR** (400 MHz, DMSO-*d*6) δ 11.07 (s, 1H, *H4*), 8.45 (dd, *J* = 4.7, 1.6 Hz, 1H, *H10*), 8.40 (dd, *J* = 8.2, 1.7 Hz, 1H, *H8*), 7.18 (dd, *J* = 8.1, 4.6 Hz, 1H, *H9*), 7.05 (t, *J* = 5.4 Hz, 1H, *H12*), 5.36–5.32 (m, 1H, *H2*), 3.57 (t, *J* = 5.6 Hz, 2H, *H14*), 3.37 (t, *J* = 5.5 Hz, 2H, *H13*), 3.30 (s, 3H, *H16*); **^{13}C NMR** (101 MHz, DMSO-*d*6) δ 164.1 (*C3*), 150.9 (*C1*), 150.7 (*C10*), 150.5 (*C5*), 131.5 (*C8*), 117.3 (*C9*), 109.5 (*C6*), 91.3 (*C2*), 69.8 (*C14*), 58.6 (*C16*), 42.8 (*C13*); **LR-ESI-MS** $C_{11}H_{13}N_3O_2$ $[M+H]^+$ *m/z* found 220.9 cald 220.1; **HR-ESI-MS** $C_{11}H_{13}N_3O_2$ $[M+H]^+$ *m/z* found 220.1079 cald 220.1086.

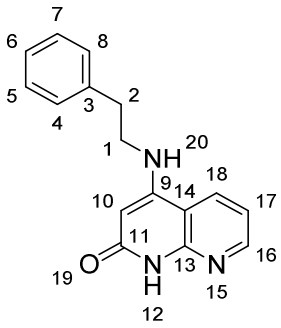
4-((furan-2-ylmethyl)amino)-1,8,napthyridin-2-ol , 72



To a solution of naphthyridine **40** (41 mg, 0.23 mmol) in DMSO (1.0 mL) was added furfurylamine (0.80 mL, 9.04 mmol) and the reaction stirred at 45 °C for 4 h. The reaction was cooled to 0 °C and the resulting precipitate collected by vacuum filtration to give **22** as a pale yellow solid (38 mg, 70%).

Mtp (ºC) 210–215; **V_{max}(cm^{-1})** 3292, 2943, 2830, 1023, 640; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 11.15 (s, 1H, *H4*), 8.46 (dd, *J* = 4.7, 1.6 Hz, 1H, *H10*), 8.39 (dd, *J* = 8.1, 1.7 Hz, 1H, *H8*), 7.64–7.59 (m, 2H, *H12*, *H16*), 7.19 (dd, *J* = 8.0, 4.7 Hz, 1H, *H9*), 6.42 (dd, *J* = 3.3, 1.8 Hz, 1H, *H17*), 6.41–6.38 (m, 1H, *H18*), 5.43 (d, *J* = 1.7 Hz, 1H, *H2*), 4.42 (d, *J* = 5.6 Hz, 2H, *H13*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 164.0 (*C3*), 151.8 (*C1*), 150.8 (*C10*), 150.6 (*C14*), 150.4 (*C5*), 142.9 (*C16*), 131.4 (*C8*), 117.4 (*C9*), 110.9 (*C17*), 109.5 (*C6*), 108.2 (*C18*), 92.3 (*C2*), 39.7 (*C13*); **LR-ESI-MS** $C_{13}H_{11}N_3O_2$ $[M+H]^+$ *m/z* found 242.3 cald 242.1; **HR-ESI-MS** $C_{13}H_{11}N_3O_2$ $[M+H]^+$ *m/z* found 242.0923 cald 242.0930.

4-(phenethylamino)-1,8-naphthyridin-2-ol, 73



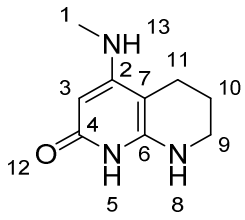
40 (62 mg, 0.34 mmol) was added phenethylamine (1.7 mL, 14 mmol) and DMSO (1.6 mL) and stirred at 78 ºC for 16 h. The reaction was cooled, evaporated to dryness and re-suspended in MeOH. The resulting precipitate was filtered under vacuum to give **73** as a white solid (43 mg, 47%).

Mtp (ºC) 150–163; **V_{max}(cm^{-1})** 3302, 2942, 2830, 1448, 1023, 632, 431; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 11.08 (s, 1H, *H12*), 8.45 (dd, *J* = 4.7, 1.6 Hz, 1H, *H18*), 8.35 (dd, *J* = 8.2, 1.7 Hz, 1H, *H16*), 7.31 (d, *J* = 5.6 Hz, 1H, *H17*), 7.35–7.15 (m, 5H, *H4*, *5*, *6*, *7*, *8*), 7.13 (t, *J* = 5.4 Hz, 1H, *H20*), 5.37 (s, 1H, *H10*), 3.41 (q, *J* = 6.7 Hz, 2H, *H1*), 2.95 (t, J = 6.5 Hz, 2H, *H2*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 164.1 (*C11*), 150.8 (*C9*), 150.7 (*C18*), 150.5 (*C13*), 139.9 (*C3*), 131.4 (*C16*), 129.2 (*C5*, *7*), 128.9 (*C4*, *8*), 126.7 (*C6*), 117.3 (*C17*), 109.6 (*C14*), 91.4 (*C10*), 44.5 (*C1*), 34.0 (*C2*); **LR-ESI-MS** $C_{16}H_{15}N_3O$ $[M+H]^+$ *m/z* found 266.4 cald 266.1; **HR-ESI-MS** $C_{16}H_{15}N_3O$ $[M+H]^+$ *m/z* found 266.1291 cald 266.1293.

General procedure for compounds 52–58

Napthyridones **67**–**73** were dissolved in MeOH (0.18 M) and formic acid (0.07 M) and subjected to flow hydrogenation with a Pd cartridge until complete conversion by LCMS analysis (40 °C, 20 bar, 0.5 mL/min). The resulting solutions were evaporated to dryness and purified by preparative HPLC.

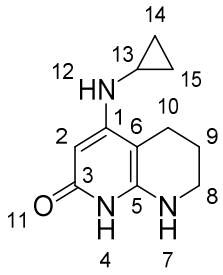
4-(methylamino)-5,6,7,8-tetrahydro-1,8-napthyridin-2(*1H*)-one, 52



Synthesised as per the general procedure from **67** to give the **52** as a red solid (116 mg, 69%).

Mtp (°C) 198–202; **V_{max}(cm^{-1})** 3297, 2940, 2830, 1447, 1024, 632; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 5.58 (q, *J* = 4.9 Hz, 1H. *H13*), 5.33 (s, 1H, *H8*), 4.52 (s, 2H, *H3*), 3.10 (q, *J* = 2.2 Hz, 2H, *H9*), 2.60 (d, *J* = 4.7 Hz, 3H, *H1*), 2.15 (t, *J* = 6.4 Hz, 2H, *H11*), 1.73 (p, *J* = 6.2 Hz, 2H, *H10*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 161.8 (*C4*), 157.2 (*C2*), 145.5 (*C6*), 81.5 (*C7*), 80.4 (*C3*), 40.0 (*C9*), 29.2 (*C1*), 21.3 (*C10*), 19.0 (*C11*); **LR-ESI-MS** $C_9H_{13}N_3O$ $[M+H]^+$ *m/z* found 180.3 cald 180.1; **HR-ESI-MS** $C_9H_{13}N_3O$ $[M+H]^+$ *m/z* found 180.1121 cald 180.1137.

4-(cyclopropylamino)-5,6,7,8-tetrahydro-1,8-napthyridin-2(*1H*)-one, 53

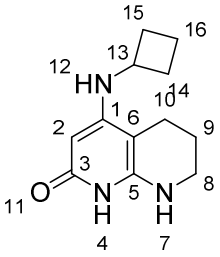


Synthesised as per the general procedure from **68** to give the **53** as an orange wax (19 mg, 41%).

V_{max}(cm^{-1}) 3309, 2940, 2830, 1409, 1113, 1052, 1024, 657; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 9.38 (s, 1H, *H4*), 5.74 (s, 1H, *H12*), 5.32 (s, 1H, *H7*), 4.95 (d, J = 1.7 Hz, 1H, *H2*), 3.11 (d, J = 6.2 Hz, 2H, *H8*), 2.25 (s, 1H, *H13*), 2.13 (t, J = 6.3 Hz, 2H, *H10*), 1.79–1.65 (m, 2H, *H9*), 0.69–0.60 (m, 2H, *H14', 15"*), 0.42 (dt, J = 4.1, 2.1 Hz, 2H, *H14", 15'*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 161.6 (*C3*), 157.3 (*C1*), 145.2 (*C5*), 82.4

(*C6*), 81.5 (*C2*), 39.6 (*C8*), 24.3 (*C13*), 21.2 (*C9*), 19.0 (*C10*), 6.81 (*C14*, *15*); **LR-ESI-MS** $C_{11}H_{15}N_3O$ $[M+H]^+$ *m/z* found 206.3 cald 206.1; **HR-ESI-MS** $C_{11}H_{15}N_3O$ $[M+H]^+$ *m/z* found 206.1288 cald 206.1293.

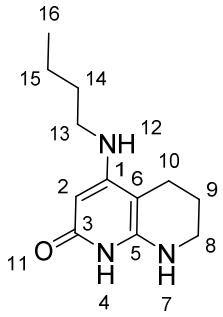
(cyclobutylamino)-5,6,7,8-tetrahydro-1,8-napthyridin-2(*1H*)-one, 54



Synthesised as per the general procedure from **69** to give the **54** as a pink solid (10 mg, 19%).

Mtp (°C) 287–299; **V_{max}(cm^{-1})** 3277, 2940, 2830, 1398, 1023, 645; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 9.38 (s, 1H, *H4*), 5.49–5.38 (m, 1H, *H12*), 5.33 (s, 1H, *H7*), 4.48 (s, 1H, *H2*), 3.74 (h, J = 7.6 Hz, 1H, *H13*), 3.12 (dq, J = 6.7, 4.1 Hz, 2H, *H8*), 2.30–2.10 (m, 4H, *H14'*, *14"*, *15'*, *15"*), 2,19 (t, J = 6.2 Hz, 2H, *H10*), 2.00–1.85 (m, 2H, *H16*), 1.77–1.58 (m, 2H, *H9*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 161.7 (*C3*), 155.3 (*C1*), 145.8 (*C5*), 81.5 (*C6*), 81.4 (*C2*), 47.6 (*C13*), 39.5 (*C8*), 30.0 (*C14*, *15*), 21.2 (*C9*), 19.1 (*C10*), 14.9 (*C16*); **LR-ESI-MS** $C_{12}H_{17}N_3O$ $[M+H]^+$ *m/z* found 220.4 cald 220.1; **HR-ESI-MS** $C_{12}H_{17}N_3O$ $[M+H]^+$ *m/z* found 220.1441 cald 220.1450.

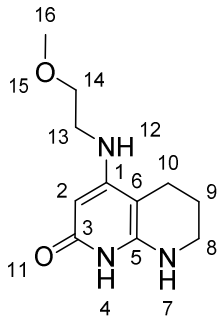
4-(butylamino)-5,6,7,8-tetrahydro-1,8-napthyridin-2(*1H*)-one, 55



Synthesised as per the general procedure from **70** to give the **55** as an off-white solid (30 mg, 55%).

Mtp (°C) 97–129; **V_{max}(cm^{-1})** 3309, 2939, 2830, 1445, 1022, 627; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 9.29 (s, 1H, *H4*), 5.37 (s, 1H, *H12*), 5.29 (s, 1H, *H7*), 4.56 (s, 1H, *H2*), 3.11 (s, 2H, *H8*), 2.96 (q, J = 6.6 Hz, 2H, *H13*), 2.16 (t, J = 6.4 Hz, 2H, *H10*), 1.73 (dd, J = 6.0, 5.3 Hz, 2H, *H9*), 1.49 (p, J = 7.3 Hz, 2H, *H14*), 1.30 (dt, J = 14.6, 7.3 Hz, 2H, *H15*), 0.90 (q, J = 7.6 Hz, 3H, *H16*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 156.9 (*C1*), 146.2 (*C5*), 82.5 (*C6*), 80.8 (*C2*), 42.3 (*C13*), 40.9 (*C8*), 31.0 (*C14*), 21.7 (*C9*), 20.2 (*C15*), 19.5 (*C10*), 14.3 (*C16*); **LR-ESI-MS** $C_{12}H_{19}N_3O$ $[M+H]^+$ *m/z* found 222.4 cald 222.2; **HR-ESI-MS** $C_{12}H_{19}N_3O$ $[M+H]^+$ *m/z* found 222.1598 cald 222.1606.

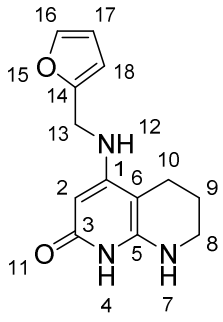
4-((2-methoxyethyl)amino)-5,6,7,8-tetrahydro-1,8-napthyridin-2(*1H*)-one, 56



Synthesised as per the general procedure from **71** to give the **56** as a purple wax (16 mg, 49%).

V_{max}(cm^{-1}) 3303, 2941, 2830, 1447, 1116, 1023, 631; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 9.34 (s, 1H, *H4*), 5.35 (s, 1H, *H12*), 5.33 (t, J = 5.5 Hz, 1H, *H7*), 4.59 (s, 1H, *H2*), 3.43 (t, J = 5.9 Hz, 2H, *H14*), 3.26 (s, 3H, *H16*), 3.16 (t, J = 5.9 Hz, 2H, *H13*), 3.15–3.06 (m, 2H, *H8*), 2.16 (t, J = 6.4 Hz, 2H, *H10*), 1.74 (dt, J = 6.4, 4.2 Hz, 2H, *H9*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 161.7 (*C3*), 156.1 (*C1*), 145.8 (*C5*), 81.4 (*C6*), 80.5 (*C2*), 70.0 (*C14*), 58.0 (*C16*), 41.7 (*C13*), 39.7 (*C8*), 21.2 (*C9*), 19.0 (*C10*); **LR-ESI-MS** $C_{12}H_{19}N_3O$ $[M+H]^+$ *m/z* found 224.4 cald 224.1; **HR-ESI-MS** $C_{12}H_{19}N_3O$ $[M+H]^+$ *m/z* found 224.1393 cald 224.1399.

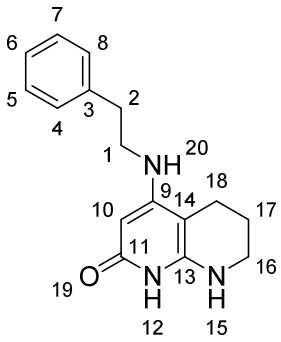
4-((furan-2-ylmethyl)amino)-5,6,7,8-tetrahydro-1,8-napthyridin-2(*1H*)-one, 57



Synthesised as per the general procedure from **72** to give the **57** as a purple wax (7 mg, 23%).

^{1}H NMR (400 MHz, DMSO-d_6) δ 9.52 (s, 1H, *H4*), 7.56 (d, J = 1.9 Hz, 1H, *H16*), 6.38 (dd, J = 3.2, 1.9 Hz, 1H, *H17*), 6.24 (d, J = 3.3 Hz, 1H, *H18*), 6.03 (t, J = 5.9 Hz, 1H, *H7*), 5.44 (t, J = 5.1 Hz, 1H, *H12*), 4.67 (s, 1H, *H2*), 4.20 (d, J = 5.8 Hz, 1H, *H13*), 3.12 (s, 2H, *H8*), 2.21 (t, J = 6.4 Hz, 2H, *H10*), 1.76–1.72 (m, 2H, *H9*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 162.1 (*C3*), 156.5 (*C1*), 153.3 (*C14*), 146.3 (*C5*), 142.3 (*C16*), 110.8 (*C17*), 107.3 (*C18*), 81.8 (*C2*), 81.7 (*C6*), 40.3 (*C8*), 39.7 (*C13*), 21.7 (*C9*), 19.5 (*C10*); **LR-ESI-MS** $C_{12}H_{19}N_3O$ $[M+H]^+$ *m/z* found 246.4 cald 246.1; **HR-ESI-MS** $C_{12}H_{19}N_3O$ $[M+H]^+$ *m/z* found 246.1238 cald 246.1243.

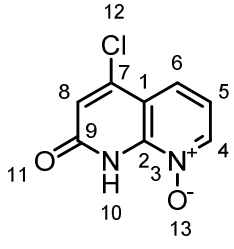
4-(phenethylamino)-5,6,7,8-tetrahydro-1,8-napthyridin-2(*1H*)-one, 58



Synthesised as per the general procedure from **73** to give the **58** as a red wax (9 mg, 96%).

V_{max}(cm^{-1}) 3302, 2944, 2830, 1472, 1024, 642; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 7.35–7.16 (m, 5H, *H4*, *H5*, *H6*, *H7*, *H8*), 5.44 (t, J = 5.6 Hz, 1H, *H20*), 5.37 (t, J = 3.4 Hz, 1H, *H15*), 4.66 (s, 1H, *H10*), 3.20 (dt, J = 8.1, 6.1 Hz, 2H, *H1*), 3.11 (t, J = 7.3 Hz, 2H, *H16*), 2.81 (t, J = 6.4 Hz, 2H, *H2*), 2.14 (t, J = 6.4 Hz, 2H, *H18*), 1.73 (dt, J = 12.2, 5.7 Hz, 2H, *H17*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 161.8 (*C11*), 156.0 (*C9*), 145.8 (*C13*), 139.6 (*C3*), 128.6 (*C4*, *C8*), 128.4 (*C5*, *C7*), 126.1 (*C6*), 82.0 (*C14*), 80.6 (*C10*), 43.8 (*C1*), 39.7 (*C16*), 34.5 (*C2*), 21.5 (*C17*), 18.9 (*C18*); **LR-ESI-MS** $C_{16}H_{19}N_3O$ $[M+H]^+$ *m/z* found 270.4 cald 270.2; **HR-ESI-MS** $C_{16}H_{19}N_3O$ $[M+H]^+$ *m/z* found 270.1601 cald 270.1606.

5-chloro-7-oxo-7,8-dihydro-1,8-naphthyridine 1-oxide, 87

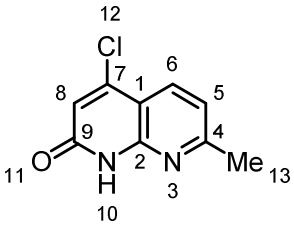


mCPBA (75%, 956 mg, 4.15 mmol) in a solution of DCM (0.4 mL) was added dropwise to a solution of **40** (500 mg, 2.77 mol) in DCM (0.4 mL) at 0 °C. The reaction was stirred at r.t. for 15 h, added sodium thiosulfate, and the reaction filtered

through a phase separator and evaporated to dryness. The product was purified by flash column chromatography (0-10% MeOH in DCM) to give the product as a white solid (223 mg, 41%).

V_{max}(cm^{-1}) 3360, 3114, 3023, 1632, 1398, 610; **^{1}H NMR** (400 MHz, DMSO-d_6) δ 12.10 (s, 1H, *NH*), 8.60 (d, *J* = 6.3 Hz, 1H, *H4*), 7.82 (d, *J* = 8.2 Hz, 1H, *H6*), 7.33 (t, *J* = 7.3 Hz, 1H, *H5*), 7.06 (s, 1H, *H8*); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 159.8 (*C9*), 143.1 (*C7*), 142.5 (*C2*), 139.3 (*C4*), 124.6 (*C8*), 122.7 (*C6*), 118.7 (*C5*), 116.1 (*C1*); **LR-ESI-MS** $C_8H_5ClN_2O_2$ $[M+H]^+$ *m/z* found 196.8 cald 196.0; **HR-ESI-MS** $C_8H_5ClN_2O_2$ $[M+H]^+$ *m/z* found 224.1393 cald 224.1399.

4-chloro-7-methyl-1,8-naphthyridin-2(1H)-one, 88



87 (17.0 mg, 0.09 mmol), copper (I) chloride (0.9 mg, 0.01 mmol) and magnesium chloride (12.3 mg, 0.130 mmol) was dissolved in diethyl ether (900 µL) and MeMgBr (3M in diethyl ether, 58 µL**,** 0.173 mmol) was added dropwise. The reaction was added ammonium chloride, water and extracted with DCM (3x). The organics were washed with brine (2x), filtered through a phase separator and evaporated to dryness. The product was purified by flash column chromatography (0–10% MeOH in DCM) to give an off white waxy solid (~1.0 mg, 6%).

Based on the availability of material, only NMR and LCMS analysis was performed. **^{1}H NMR** (400 MHz, DMSO-d_6) δ 12.32 (s, 1H, *NH*), 8.13 (d, *J* = 8.1 Hz, 1H, *H6*), 7.24 (d, *J* = 8.1 Hz, 1H, *H5*), 6.80 (s, 1H, *H8*), 2.56 (s, 3H, *H13*); **^{13}C NMR** (101 MHz,

DMSO-d_6) δ 170.8 (*C9*), 162.3 (*C4*), 161.7 (*C2*), 149.3 (*C7*), 134.4 (*C6*), 121.5 (*C8*), 119.3 (*C5*), 111.1 (*C1*), 24.6 (*C13)*; **LR-ESI-MS** $C_9H_7ClN_2O$ $[M+H]^+$ *m/z* found 194.3 cald 194.0.

7.2 Computational docking and visualisation

Computational docking and visualisation was performed in the Maestro Suite from Schrodinger.[125] Proteins were prepared with the Protein Preparation Wizard using standard settings and ligand docking performed with Glide Docking.[311] Covalent docking was performed with CovDock.[312] Large ligand libraries were enumerated by reaction using reactant files downloaded from Fluorochem and Enamine.

7.3 BODIPY-GTP nucleotide exchange assay

7.3.1 General protocol

Synthesised compounds were tested at 500 µM or 250 µM final concentration using a nucleotide exchange assay developed by Carmen Jimenez Antunez. Dose-response curves were obtained using a 10-point 2-fold serial dilution with a high concentration of 500 µM. 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 (including control wells). The assay buffer (20 mM Tris pH 7.5, 50 mM NaCl, 1mM $MgCl_2$, 100 µg/mL BSA and 1 mM DTT) was dispensed into sample and negative control wells (10 µL), and positive control wells added GppNHp (400 µM in assay buffer, 10 µL, Jena BioScience) to give a final concentration of 200 µM. The enzymes were mixed to a concentration of 4 µM GTPase and 0.1 µM GEF in assay buffer and dispensed into all wells (5 µL) to make a final concentration of 1 µM and 0.025 µM respectively. The plate was centrifuged at 2,000 rpm for 5 mins. After 2 h incubation at room temperature, BODIPY^{TM} FL GTP (2 µM in assay buffer, 5 µL, ThermoFisher Scientific) was dispensed to all wells to give a final concentration of 0.5 µM. The plate was shaken for 10 seconds and the FI signal measured using the Pherastar FSX (BMG Labtech, FI 485/520, Gain 300) every 0.5–2.0 min. The initial rate was calculated and plotted using GraphPad Prism 9, the values being normalised to the positive and negative controls. All values were plotted as normalised initial rate against each compound or concentration in the case of dose responses (FI/s). Kalirin-GEF(1) (amino acids 1273–1581) and full length Rac1 were produced and supplied by Dr. Janine Gray.

7.3.2 K_{inact}/Ki

k_{inact}/Ki was calculated using the equation reported by Krippendorf[235] and the FI nucleotide exchange assay. IC_{50} values of both **RS-009** and **MC-278** were recorded at time periods of 10–120 min and plotted. The k_{inact} and Ki values were calculated in GraphPad Prism 9 in min^{-1} μM^{-1} and converted to s^{-1} M^{-1} prior to reporting.

7.3.3 Limited selectivity panel

The limited Rho selectivity panel described in Section 2.4.1 was performed with the described BODIPY-GTP nucleotide exchange assay. Full length Rac1 was produced and supplied by Dr. Janine Gray, RhoA by Dr. Vittorio Katis, and Cdc42 by Dr. Andrew Thompson, all of the CMD. Kalirin-GEF(1) (amino acids 1273–1581) was supplied by Dr. Janine Gray, and Kalirin-GEF(2) construct **B1** was produced as described in Section 7.5. The GEF domains of Tiam1, Dbs, and Vav1 were purchased from Universal Biologicals, and c9orf72 from Origene. Rab39B was supplied by the Huber Group of the CMD. All GTPases were used at a concentration of 1 µM, and GEFs at 25–250 nM. **MC-278** was tested against each GEF/GTPase pair at 500 µM, and if activity was observed, a dose-response curve was obtained. The data was processed as described in Section 7.2.

7.4 Activity-based protein profiling

7.4.1 Competitive-ABPP with MC-278

Activity-based protein profiling experiments were performed using a protocol optimised by Dr. Elisabeth Mira Rothweiler of the CMD. A frozen cell pellet (K-562 cell line) supplied by Dr. Elisabeth Mira Rothweiler was added PBS (10 mL, 1 µL Benzonase) and lysed by sonication (2x, 10 second, 20% amplitude) and filtered through a syringe (10x). The solution was left on ice for 30 mins, and centrifuged. The supernatant was collected and the lysate diluted to 1 mg/mL. For each heavy and light sample required, 2 mL of lysate was used. To one sample, **MC-278** (1 µL of 100 mM solution in DMSO, to give a final concentration of 100 µM) was added, and the control sample added DMSO (1 µL). The samples were incubated for 1–2 h while rotating. Both samples were added IA-alkyne (40 µL of a 5 mM DMSO stock, MedChemExpress) and incubated for 1 h. To the **MC-278** experiment, a master mix containing the heavy isoDTB tag (20 µL, 5 mM DMSO stock), TBTA (60 µL, 0.9 mg/mL stock in DMSO/t-BuOH), $CuSO_4$ (20 µL, 12.5 mg/mL in water) and TCEP (20 µL, 13 mg/mL in water) was added and incubated for 1 h at 750 rpm (shaking plate). The equivalent was performed for the control sample, albeit with the light isoDTB tag. The heavy/light samples were combined into cold acetone (8 mL), vortexed and stored at -20 °C overnight.

The resulting solution was centrifuged (3500 G, 5 min) and the supernatant removed. The precipitate was added MeOH (1 mL) and resuspended by sonication (10 s, 20% amplitude). The solution was centrifuged (3500 G, 5 min) and the supernatant removed. The precipitate was added 8 M urea in 0.1 M TEAB (300 µL) and sonicated (20 s, 25% amplitude), and added 900 µL TEAB.

To prepare the streptavidin high capacity agarose beads, 50 µL (per each sample) was washed with NP40 (0.2% in PBS, 3x, aspirate and centrifuge at 0.8 G, 3 min) and finally added NP40 (0.2% in PBS,100 µL per 50 µL bead bed). The samples prepared previously were added to the beads (1.2 mL) and incubated at room temperature while rotating. The samples were centrifuged (1000 G, 15 s) and aspirated. The beads were transferred to Biospin columns with NP40 (0.1% in PBS, 600 µL) and centrifuged (0.8 G, 3 s) and aspirated. This method of cleaning the beads was performed with NP40 (0.1% in PBS, 600 µL, 3x), PBS (600 µL, 3x) and water (600 µL, 3x). Finally the beads were transferred to a LoBind Eppendorf with 8M urea (in 0.1 M TEAB, 600 µL), centrifuged (0.8 G, 3 min) and aspirated. 8 M urea (in 0.1 M TEAB, 300 µL) was added, followed by the addition of DTT (15 µL, 31 mg/mL stock in water) and the samples incubated for 45 min while rotating. Iodoacetamide (15 µL, 74 mg/mL stock in water) was added and the samples incubated for 30 min while rotating. Finally, 0.1 M TEAB (900 µL) was added, the samples centrifuged (0.8 G, 3 min) and aspirated. The beads were added 1 M urea (in 0.1 M TEAB, 900 µL) and chymotrypsin (0.5 mg/mL stock in 1 mM HCl, 5 µL) and incubated overnight at r.t. at 1000 rpm (shaking plate).

The chymotryptic digests were added NP40 (0.1% in PBS, 400 µL) and the solutions transferred to Biospin columns. The beads were washed as described previously with NP40 (0.1% in PBS, 600 µL, 3x), PBS (600 µL, 3x) and water (600 µL, 3x). The column was transferred to a fresh LoBind Eppendorf, and the column added 0.1% TFA in 1:1 MeCN/water (200 µL) and left to slowly drip through. Two more elutions were performed (each 100 µL), and the beads eluted completely by centrifuge (3000 G, 3 min). The resulting solution was evaporated in a speed vac (30 °C, V-AQ).

7.4.2 MS sample preparation

The obtained peptides were reconstituted in a solution of 5% DMSO and 5% formic acid (20 µL) by shaking plate (1000 rpm) and sonication (bath, 5 min). An Ultrafree-MC Centrifugal Filter (Merck Millipore) was washed with the resuspension buffer (300 µL) and centrifuged (17,000 G, 2 min). The peptide sample was added to the filter and centrifuged (17,000 G, 2 min) and the filtrate transferred to a MS sample vial ready for analysis.

7.4.3 Sample analysis by LC-MS/MS

MS data was collected at the Discovery Proteomics Facility at the University of Oxford. The suspended peptides were trapped on an Acclaim™ PepMap™ 100 C18 HPLC Columns (5µm x 0.1mm x 20mm, Thermo Fisher Scientific) using solvent A (0.1 % formic acid in water) at a pressure of 60 bar and separated on an Ultimate 3000 UHPLC system (Thermo Fischer Scientific) coupled to a QExactive mass spectrometer (Thermo Fischer Scientific). The peptides were separated on an Easy Spray PepMap RSLC column (75 µm i.d. x 2 µm x 50 mm, 100 Å, Thermo Fisher) and electro sprayed directly into an QExactive mass spectrometer (Thermo Fisher Scientific) through an EASY-Spray nano-electrospray ion source (Thermo Fisher Scientific) using a gradient of solvent A (0.1 % formic acid in water) and solvent B solvent B (0.1 % formic acid in acetonitrile and 5 % dimethyl sulfoxide), flow rate: 250 nL/min. Gradient: 5 % solvent B for 7 min followed by a gradient from 5 % to 40 % solvent B over 105 min, an increase to 60 % B in 10 min and another increase to 90% B in 10 min. 90 % B was held for 10 min, then decreased to 5 % in 0.1 min and held at 5 % for 9.9 min. The raw data was acquired on the mass spectrometer in a data-dependent mode (DDA). Full scan MS spectra were acquired in the Orbitrap (scan

range 300-1500 m/z, resolution 70000, AGC target $3e^6$, maximum injection time 80 ms). After the MS scans, the 15 most intense peaks were selected for HCD fragmentation at 27 % of normalised collision energy. HCD spectra were also acquired in the Orbitrap (resolution 17500, AGC target 1e5, maximum injection time 100 ms) with first fixed mass at 100 m/z. Charge exclusion was selected for unassigned and 1+ ions.

7.4.4 Data evaluation by MaxQuant

Data analysis of ABPP experiments was performed as described previously by Zanon, involving the use of a 'U proteome' in order to distinguish peptides with two or more cysteines.[313] MS raw data was analysed using MaxQuant v 2.4.13.0.[240] Standard settings were used except for the following changes: a modified FASTA database with individual substitutions of cysteines with the placeholder "U" was used and obtained from the ProteomeXchange Consurtium via PRIDE[314] (identifiers PXD038301 and PXD038375).[315] Labels were set on the placeholder amino acid "U" for the light isoDTB tag as light label (C(28) H(46) N(10) O(6) S(1) Se(-1)) and the heavy isoDTB tag as heavy label (C(24) Cx(4) H(46) N(8) Nx(2) O(6) S(1) Se(-1)) where 'x' indicates an isotopically heavy atom. A multiplicity of 2 was set and a maximum number of labelled amino acids of 1. The digestion enzyme was set to Chymotrypsin+ with a maximum number of missed cleavages of 2. No variable modifications were included. "Re-quantify" was enabled. Carbamidomethyl (C_2H_3NO) was used as a variable modification on cysteine. Contaminants were included. 'Second peptides' was unselected. After the run was complete, identified peptides were observed using the 'peptides.txt' file and filtering out any peptides without the 'U' placeholder. The results were not processed further due to the limited number of peptides identified.

7.5 GppNHp competition assay

RapidFire experiments were performed in duplicate on an Agilent 6530 Accurate Mass QTOF and the data processed by the Agilent MassHunter Workstation software. Full length Rac1 (20 µL), Kalirin-GEF(1) (amino acids 1273–1581) (20 µL), MC-278 (20 µL) and GppNHp (20 µL) were incubated for 24 h at 4 °C in assay buffer (20 mM Tris PH 7.5, 50 mM NaCl, 1mM $MgCl_2$) to give final concentrations of 1.25 µM, 1.25 µM, 200 µM and 0 – 200 µM respectively. The abundances of WT-Rac1 and MC-278-Rac1 by mass were calculated and used to determine the percentage of Rac1 labelling. For dose response curves, a 10-point 2-fold serial dilution was performed for MC-278 with a high concentration of 200 µM. The measurement was repeated in the presence of GppNHp at a final concentration of 10 µM. Control wells absent of Kalirin-GEF(1) or GppNHp were displaced with buffer (20 µL).

7.6 Production of Kalirin-GEF(2) constructs

7.6.1 Expression

Kalirin-GEF(2) genes were purchased from Twist Biosciences and constructs were cloned in p-NIC28-Bsa4 expression vectors (N-terminal His6, TEV-cleavable (23 aa) MHHHHHHSSGVDLGTENLYFQ*SM, Kanamycin resistant) acquired from Biotech SRF and test expressions performed by Leela Shrestha. Of those that were successful, cloned DNA constructs were transformed into BL21 DE3 Rosetta cells in LB with kanamycin (50 ug/mL) and chloramphenicol (25 ug/mL) at 37 °C overnight. Expression was performed using the auto induction method, inoculating the transformed cells (1 mL) into TB (100 mL) containing kanamycin, antifoam (0.01 vol%) and autoinduction mix (0.8% glycerol, 2 g lactose, 0.5 g glucose per litre). The mixture was shaken for 6 h at 37 °C, and a further 66 h at 18 °C. The mixture was subsequently centrifuged and the cells harvested. The acquired pellets were frozen at -80 °C.

7.6.2 Purification

The frozen pellets were suspended in base buffer (10 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 0.5 mM TCEP) and added lysozyme (0.5 mg/mL), Benzonase (0.1 mg/mL) and 1 in 1000 protease inhibitor cocktail (Merck). The mixtures were vortexed to dissolve the pellet, followed by sonication for 5 min with a 5 sec on/off cycle to lyse the cells. Triton X-100 (10%) was added and the mixture stirred. Imidazole was added to give a final concentration of 20 mM post-lysis, the cells centrifuged and the supernatant collected. The soluble fraction was then poured onto a His GraviTrap column (GE healthcare) fitted with LabMate extender, and washed with 20 mM imidazole in base buffer twice. The protein was subsequently eluted using 500 mM imidazole in base buffer onto a PD10 column fitted with LabMate extender (pre-

equilibrated with 20 mM imidazole). The PD10 column was eluted with 20 mM imidazole. Tag cleavage was achieved with TEV protease overnight at 4 ^{o}C. Finally, the solution was added over a His GraviTrap column equilibrated in 20 mM imidazole, washed with 20 mM imidazole and the resultant protein concentrated. The proteins were further purified using size exclusion chromatography, eluting with base buffer. Intact mass spectrometry confirmed expected masses and purity by SDS-PAGE.

7.7 Crystallisation and structure determination

7.7.1 General experimental

Proteins were crystallised using the sitting-drop vapour diffusion method in 96 well, 3 lens SWISSCI plates (SWISSCI AG). Proteins were dispensed in ratios of 2:1, 1:1, and 1:2 to precipitant in 150 nL drops with the TTP Labtech's Mosquito Crystal liquid handling robot. Crystallisation screens (JCSG7 and Morpheus) were obtained from Molecular Dimensions. Crystallisation plates were incubated at 20 or 4 °C and imaged using an automatic inspection system (Rigaku). The use of microcrystal seeds was sometimes used to improve crystal growth and quality. Obtained crystals were cryoprotected with 25–35% ethylene glycol and flash-cooled in liquid nitrogen. X-ray diffraction data was collected at Diamond Light Source on the I03-1 beamline and auto-processed (using xia2[316], DIALS[317], XDS[318], POINTLESS[319], CCP4[320]). The structures were phased and refined using PHENIX[321] and manual model correction was performed in Coot.[322] REFMAC[323] was used to model covalent linkages.

7.7.2 Co-crystallisation of MC-278 with Kalirin-GEF(1)/Rac1

Kalirin-GEF(1)/Rac1 crystals were obtained by sitting-drop vapour diffusion at 20 °C with 100 nL of protein (1:1 Kalirin-GEF(1)/Rac1, 15.0 mg/mL) and 50 nL reservoir solution consisting of 0.1 M citrate pH 4.2, 0.2 M NaCl and 18 % PEG 8K. After 6 days, the droplet was added 13 µL reservoir solution, 6 µL ethylene glycol, and 1 µL MC-278 (100 mM in DMSO) to give a final concentration of 5 mM and left to incubate at ambient temperature for 2 h. The resulting crystals were flash cooled in liquid nitrogen. X-ray diffraction data was collected at Diamond Light Source on the I03-1 beamline. The structure was phased and refined using PHENIX[321], the covalent bond modelled with REFMAC[323], and manual model correction performed in Coot.[322]

7.7.3 Crystallisation of the Kalirin DH2 domain

Kalirin-GEF(2) crystals were obtained by sitting-drop vapour diffusion at 20 °C with 50 nL of protein (1:1 Kalirin-GEF(2)/RhoA, 18 mg/mL total) and 150 nL reservoir solution consisting of 20% w/v PEG3350 and 0.1 M succinic acid. After 4 days, the droplet was cryoprotected with 25% ethylene glycol and flash cooled in liquid nitrogen. X-ray diffraction data was collected at Diamond Light Source on the I03-1 beamline. The structure was phased and refined using PHENIX[321] and manual model correction performed in Coot.[322]

7.8 Rho Dbl-GEF Alignment

7.8.1 General protocol

Rho Dbl GEFs were selected based on three criteria: (i) the availability of either a crystal structure or AlphaFold prediction of the Rho Dbl GEF DH domain, and if using the AlphaFold prediction, (ii) a per-residue confidence score (pLDDT) of the secondary structure greater than 70 and (iii) the presence of a complete DH domain.[82] This resulted in 70 GEF structures that were truncated as per the DH domain boundaries dictated in UniProt.[239] The structures were aligned utilising the Protein Structure Alignment tool in the Maestro suite (Schrödinger).[125] The Kalirin-GEF(1)/Rac1 structure was used as the alignment reference and model system (PDB: 5O33, Chain B).[125] Residue side chains on the helices overlapped well (RMSD 0.711–3.940 Å for alpha and beta carbons). A spatially conserved residue was flagged when >80% of GEFs comprised the same type of residue at the same position: acidic, basic, polar, non-polar, and aromatic. All residues were referenced to the corresponding position on Kalirin-GEF(1) and denoted by that position. All percentage calculations and data visualisation was performed in Microsoft Excel.

8 References

1. R. S. Doody, R. Raman, M. Farlow, T. Iwatsubo, B. Vellas, S. Joffe, K. Kieburtz, F. He, X. Sun, R. G. Thomas, P. S. Aisen, E. Siemers, G. Sethuraman and R. Mohs, *New Engl. J. Med.*, 2013, **369**, 341-350.
2. E. Karran and B. De Strooper, *Nat. Rev. Drug Discov.*, 2022, **21**, 306-318.
3. J.-Y. Hur, *Exp. Mol. Med.*, 2022, **54**, 433-446.
4. S. Tagami, K. Yanagida, T. S. Kodama, M. Takami, N. Mizuta, H. Oyama, K. Nishitomi, Y. W. Chiu, T. Okamoto, T. Ikeuchi, G. Sakaguchi, T. Kudo, Y. Matsuura, A. Fukumori, M. Takeda, Y. Ihara and M. Okochi, *Cell Rep.*, 2017, **21**, 259-273.
5. S. M. Paul, D. S. Mytelka, C. T. Dunwiddie, C. C. Persinger, B. H. Munos, S. R. Lindborg and A. L. Schacht, *Nat. Rev. Drug Discov.*, 2010, **9**, 203-214.
6. C. Smith, *Nature*, 2003, **422**, 342-345.
7. C. Finan, A. Gaulton, F. A. Kruger, R. T. Lumbers, T. Shah, J. Engmann, L. Galver, R. Kelley, A. Karlsson, R. Santos, J. P. Overington, A. D. Hingorani and J. P. Casas, *Sci. Transl. Med.*, 2017, **9**.
8. P. Workman and I. Collins, *Chem. Biol.*, 2010, **17**, 561-577.
9. C. H. Arrowsmith, J. E. Audia, C. Austin, J. Baell, J. Bennett, J. Blagg, C. Bountra, P. E. Brennan, P. J. Brown, M. E. Bunnage, C. Buser-Doepner, R. M. Campbell, A. J. Carter, P. Cohen, R. A. Copeland, B. Cravatt, J. L. Dahlin, D. Dhanak, A. M. Edwards, M. Frederiksen, S. V. Frye, N. Gray, C. E. Grimshaw, D. Hepworth, T. Howe, K. V. M. Huber, J. Jin, S. Knapp, J. D. Kotz, R. G. Kruger, D. Lowe, M. M. Mader, B. Marsden, A. Mueller-Fahrnow, S. Müller, R. C. O'Hagan, J. P. Overington, D. R. Owen, S. H. Rosenberg, R. Ross, B. Roth, M. Schapira, S. L. Schreiber, B. Shoichet, M. Sundström, G. Superti-Furga, J. Taunton, L. Toledo-Sherman, C. Walpole, M. A. Walters, T. M. Willson, P. Workman, R. N. Young and W. J. Zuercher, *Nat. Chem. Biol.*, 2015, **11**, 536-541.
10. P. Filippakopoulos, J. Qi, S. Picaud, Y. Shen, W. B. Smith, O. Fedorov, E. M. Morse, T. Keates, T. T. Hickman, I. Felletar, M. Philpott, S. Munro, M. R. McKeown, Y. Wang, A. L. Christie, N. West, M. J. Cameron, B. Schwartz, T. D. Heightman, N. La Thangue, C. A. French, O. Wiest, A. L. Kung, S. Knapp and J. E. Bradner, *Nature*, 2010, **468**, 1067-1073.
11. E. Nicodeme, K. L. Jeffrey, U. Schaefer, S. Beinke, S. Dewell, C.-w. Chung, R. Chandwani, I. Marazzi, P. Wilson, H. Coste, J. White, J. Kirilovsky, C. M. Rice, J. M. Lora, R. K. Prinjha, K. Lee and A. Tarakhovsky, *Nature*, 2010, **468**, 1119-1123.
12. P. V. Fish, P. Filippakopoulos, G. Bish, P. E. Brennan, M. E. Bunnage, A. S. Cook, O. Federov, B. S. Gerstenberger, H. Jones, S. Knapp, B. Marsden, K. Nocka, D. R. Owen, M. Philpott, S. Picaud, M. J. Primiano, M. J. Ralph, N. Sciammetta and J. D. Trzupek, *J. Med. Chem.*, 2012, **55**, 9831-9837.
13. J. Blagg and P. Workman, *Cancer Cell*, 2017, **32**, 9-25.
14. D. Greiner, T. Bonaldi, R. Eskeland, E. Roemer and A. Imhof, *Nat. Chem. Biol.*, 2005, **1**, 143-145.
15. F. L. Cherblanc, K. L. Chapman, R. Brown and M. J. Fuchter, *Nat. Chem. Biol.*, 2013, **9**, 136-137.
16. A. A. Antolin, D. Sanfelice, A. Crisp, E. Villasclaras Fernandez, I. L. Mica, Y. Chen, I. Collins, A. Edwards, S. Müller, B. Al-Lazikani and P. Workman, *Nucleic Acids Res.*, 2022, **51**, D1492-D1502.
17. M. E. Bunnage, E. L. P. Chekler and L. H. Jones, *Nat. Chem. Biol.*, 2013, **9**, 195-199.

18. J. Singh, R. C. Petter, T. A. Baillie and A. Whitty, *Nat. Rev. Drug Discov.*, 2011, **10**, 307-317.
19. L. Boike, N. J. Henning and D. K. Nomura, *Nat. Rev. Drug Discov.*, 2022, **21**, 881-898.
20. F. Sutanto, M. Konstantinidou and A. Dömling, *RSC Med. Chem.*, 2020, **11**, 876-884.
21. F. Huang, X. Han, X. Xiao and J. Zhou, *Molecules (Basel, Switzerland)*, 2022, **27**.
22. J. R. Vane, *Nat. New. Biol.*, 1971, **231**, 232-235.
23. M. L. Wang, S. Rule, P. Martin, A. Goy, R. Auer, B. S. Kahl, W. Jurczak, R. H. Advani, J. E. Romaguera, M. E. Williams, J. C. Barrientos, E. Chmielowska, J. Radford, S. Stilgenbauer, M. Dreyling, W. W. Jedrzejczak, P. Johnson, S. E. Spurgeon, L. Li, L. Zhang, K. Newberry, Z. Ou, N. Cheng, B. Fang, J. McGreivy, F. Clow, J. J. Buggy, B. Y. Chang, D. M. Beaupre, L. A. Kunkel and K. A. Blum, *N. Engl. J. Med.*, 2013, **369**, 507-516.
24. R. Lonsdale and R. A. Ward, *Chem. Soc. Rev.*, 2018, **47**, 3816-3830.
25. R. Lagoutte, R. Patouret and N. Winssinger, *Curr. Opin. Chem. Biol.*, 2017, **39**, 54-63.
26. W. Lu, M. Kostic, T. Zhang, J. Che, M. P. Patricelli, L. H. Jones, E. T. Chouchani and N. S. Gray, *RSC Chem. Biol.*, 2021, **2**, 354-367.
27. J. N. Spradlin, E. Zhang and D. K. Nomura, *Acc. Chem. Res.*, 2021, **54**, 1801-1813.
28. G. Drewes and S. Knapp, *Trends Biotechnol.*, 2018, **36**, 1275-1286.
29. R. P. Joyce, V. W. Hu and J. Wang, *Med. Chem. Res.*, 2022, **31**, 1637-1646.
30. J. M. McKenna and R. S. J. Proctor, *CHIMIA*, 2023, **77**, 537.
31. J. Liu, X. Pan, S. Zhang, M. Li, K. Ma, C. Fan, Y. Lv, X. Guan, Y. Yang, X. Ye, X. Deng, Y. Wang, L. Qin, Z. Xia, Z. Ge, Q. Zhou, X. Zhang, Y. Ling, T. Qi, Z. Wen, S. Huang, L. Zhang, T. Wang, Y. Liu, Y. Huang, W. Li, H. Du, Y. Chen, Y. Xu, Q. Zhao, R. Zhao, D. Annane, J. Qu and D. Chen, *Lancet Reg.*, 2023, **33**.
32. Faridoon, R. Ng, G. Zhang and J. J. Li, *Med. Chem. Res.*, 2023, **32**, 1039-1062.
33. R. A. M. Serafim, L. Haarer, J. G. B. Pedreira and M. Gehringer, *Curr. Res. Chem. Biol.*, 2023, **3**, 100040.
34. N. London, R. M. Miller, S. Krishnan, K. Uchida, J. J. Irwin, O. Eidam, L. Gibold, P. Cimermančič, R. Bonnet, B. K. Shoichet and J. Taunton, *Nat. Chem. Biol.*, 2014, **10**, 1066-1072.
35. J. P. Hughes, S. Rees, S. B. Kalindjian and K. L. Philpott, *Br. J. Pharmacol.*, 2011, **162**, 1239-1249.
36. R. Macarron, M. N. Banks, D. Bojanic, D. J. Burns, D. A. Cirovic, T. Garyantes, D. V. S. Green, R. P. Hertzberg, W. P. Janzen, J. W. Paslay, U. Schopfer and G. S. Sittampalam, *Nat. Rev. Drug Discov.*, 2011, **10**, 188-195.
37. A. Barker, J. G. Kettle, T. Nowak and J. E. Pease, *Drug Discov. Today*, 2013, **18**, 298-304.
38. D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke and H. Jhoti, *Nat. Rev. Drug Discov.*, 2016, **15**, 605-619.
39. J. Boström, D. G. Brown, R. J. Young and G. M. Keserü, *Nat. Rev. Drug Discov.*, 2018, **17**, 709-727.
40. P. J. Hajduk and J. Greer, *Nat. Rev. Drug Discov.*, 2007, **6**, 211-219.
41. A. H. Lipkus, Q. Yuan, K. A. Lucas, S. A. Funk, W. F. Bartelt, III, R. J. Schenck and A. J. Trippe, *J. Org. Chem.*, 2008, **73**, 4443-4451.
42. M. Dow, M. Fisher, T. James, F. Marchetti and A. Nelson, *Org. Biomol. Chem.*, 2012, **10**, 17-28.
43. J. B. Baell and J. W. M. Nissink, *ACS Chem. Biol.*, 2018, **13**, 36-44.
44. J. Baell and M. A. Walters, *Nature*, 2014, **513**, 481-483.

45. J. L. Dahlin, J. W. M. Nissink, J. M. Strasser, S. Francis, L. Higgins, H. Zhou, Z. Zhang and M. A. Walters, *J. Med. Chem.*, 2015, **58**, 2091-2113.
46. J. B. Baell and G. A. Holloway, *J. Med. Chem.*, 2010, **53**, 2719-2740.
47. K. Schorpp, I. Rothenaigner, E. Salmina, J. Reinshagen, T. Low, J. K. Brenke, J. Gopalakrishnan, I. V. Tetko, S. Gul and K. Hadian, *J. Biomol. Screen.*, 2014, **19**, 715-726.
48. P. Kirsch, A. M. Hartman, A. K. H. Hirsch and M. Empting, *Molecules (Basel, Switzerland)*, 2019, **24**.
49. M. J. Harner, A. O. Frank and S. W. Fesik, *J. Biomol. NMR*, 2013, **56**, 65-75.
50. D. Patel, J. D. Bauman and E. Arnold, *Prog. Biophys. Mol. Biol.*, 2014, **116**, 92-100.
51. Q. Li, *Front. Mol. Biosci.*, 2020, **7**, 180.
52. M. Bon, A. Bilsland, J. Bower and K. McAulay, *Mol. Oncol.*, 2022, **16**, 3761-3777.
53. I. J. P. de Esch, D. A. Erlanson, W. Jahnke, C. N. Johnson and L. Walsh, *J. Med. Chem.*, 2022, **65**, 84-99.
54. R. Guha, *Methods Mol. Biol.*, 2013, **993**, 81-94.
55. A. Bancet, C. Raingeval, T. Lomberget, M. Le Borgne, J.-F. Guichou and I. Krimm, *J. Med. Chem.*, 2020, **63**, 11420-11435.
56. O. Ichihara, J. Barker, R. J. Law and M. Whittaker, *Mol. Inform.*, 2011, **30**, 298-306.
57. P. O. Nikiforov, S. Surade, M. Blaszczyk, V. Delorme, P. Brodin, A. R. Baulard, T. L. Blundell and C. Abell, *Org. Biomol. Chem.*, 2016, **14**, 2318-2326.
58. B. Benner, L. Good, D. Quiroga, T. E. Schultz, M. Kassem, W. E. Carson, M. A. Cherian, S. Sardesai and R. Wesolowski, *Drug. Des. Devel. Ther.*, 2020, **14**, 1693-1704.
59. A. Kim and M. S. Cohen, *Expert Opin. Drug Discov.*, 2016, **11**, 907-916.
60. C. W. Murray, D. R. Newell and P. Angibaud, *MedChemComm*, 2019, **10**, 1509-1511.
61. W. J. Fairbrother, J. D. Leverson, D. Sampath and A. J. Souers, in *Successful Drug Discovery*, 2019, DOI: https://doi.org/10.1002/9783527814695.ch9, pp. 225-245.
62. B. A. Lanman, J. R. Allen, J. G. Allen, A. K. Amegadzie, K. S. Ashton, S. K. Booker, J. J. Chen, N. Chen, M. J. Frohn, G. Goodman, D. J. Kopecky, L. Liu, P. Lopez, J. D. Low, V. Ma, A. E. Minatti, T. T. Nguyen, N. Nishimura, A. J. Pickrell, A. B. Reed, Y. Shin, A. C. Siegmund, N. A. Tamayo, C. M. Tegley, M. C. Walton, H.-L. Wang, R. P. Wurz, M. Xue, K. C. Yang, P. Achanta, M. D. Bartberger, J. Canon, L. S. Hollis, J. D. McCarter, C. Mohr, K. Rex, A. Y. Saiki, T. San Miguel, L. P. Volak, K. H. Wang, D. A. Whittington, S. G. Zech, J. R. Lipford and V. J. Cee, *J. Med. Chem.*, 2020, **63**, 52-65.
63. J. Schoepfer, W. Jahnke, G. Berellini, S. Buonamici, S. Cotesta, S. W. Cowan-Jacob, S. Dodd, P. Drueckes, D. Fabbro, T. Gabriel, J.-M. Groell, R. M. Grotzfeld, A. Q. Hassan, C. Henry, V. Iyer, D. Jones, F. Lombardo, A. Loo, P. W. Manley, X. Pellé, G. Rummel, B. Salem, M. Warmuth, A. A. Wylie, T. Zoller, A. L. Marzinzik and P. Furet, *J. Med. Chem.*, 2018, **61**, 8120-8135.
64. L. G. Mureddu and G. W. Vuister, *Front. Mol. Biosci.*, 2022, **9**, 834453.
65. A. M. Petros, J. Dinges, D. J. Augeri, S. A. Baumeister, D. A. Betebenner, M. G. Bures, S. W. Elmore, P. J. Hajduk, M. K. Joseph, S. K. Landis, D. G. Nettesheim, S. H. Rosenberg, W. Shen, S. Thomas, X. Wang, I. Zanze, H. Zhang and S. W. Fesik, *J. Med. Chem.*, 2006, **49**, 656-663.
66. S. B. Shuker, P. J. Hajduk, R. P. Meadows and S. W. Fesik, *Science*, 1996, **274**, 1531-1534.
67. A. Divakaran, S. E. Kirberger and W. C. K. Pomerantz, *Acc. Chem. Res.*, 2019, **52**, 3407-3418.

68. G. E. de Kloe, D. Bailey, R. Leurs and I. J. P. de Esch, *Drug Discov. Today*, 2009, **14**, 630-646.
69. C. W. Murray and D. C. Rees, *Nature Chemistry*, 2009, **1**, 187-192.
70. L. Maveyraud and L. Mourey, *Molecules (Basel, Switzerland)*, 2020, **25**.
71. M. S. Smyth and J. H. Martin, *Mol. Pathol.*, 2000, **53**, 8-14.
72. E. F. Garman, *Science*, 2014, **343**, 1102-1108.
73. H. J. Bernstein, A. Förster, A. Bhowmick, A. S. Brewster, S. Brockhauser, L. Gelisio, D. R. Hall, F. Leonarski, V. Mariani, G. Santoni, C. Vonrhein and G. Winter, *IUCrJ*, 2020, **7**, 784-792.
74. U. B. Gawas, V. K. Mandrekar and M. S. Majik, in *Advances in Biological Science Research*, eds. S. N. Meena and M. M. Naik, Academic Press, 2019, DOI: https://doi.org/10.1016/B978-0-12-817497-5.00005-7, pp. 69-84.
75. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235-242.
76. V. V. Krishnan and B. Rupp, in *eLS*, 2012, DOI: https://doi.org/10.1002/9780470015902.a0002716.pub2.
77. J.-P. Renaud, A. Chari, C. Ciferri, W.-t. Liu, H.-W. Rémigy, H. Stark and C. Wiesmann, *Nat. Rev. Drug Discov.*, 2018, **17**, 471-492.
78. T. B. Acton, R. Xiao, S. Anderson, J. Aramini, W. A. Buchwald, C. Ciccosanti, K. Conover, J. Everett, K. Hamilton, Y. J. Huang, H. Janjua, G. Kornhaber, J. Lau, D. Y. Lee, G. Liu, M. Maglaqui, L. Ma, L. Mao, D. Patel, P. Rossi, S. Sahdev, R. Shastry, G. V. T. Swapna, Y. Tang, S. Tong, D. Wang, H. Wang, L. Zhao and G. T. Montelione, in *Methods in Enzymology*, ed. L. C. Kuo, Academic Press, 2011, vol. 493, pp. 21-60.
79. Y. Liu, D. T. Huynh and T. O. Yeates, *Nat. Commun.*, 2019, **10**, 1864.
80. P. Savitsky, J. Bray, C. D. O. Cooper, B. D. Marsden, P. Mahajan, N. A. Burgess-Brown and O. Gileadi, *J. Struct. Bio.*, 2010, **172**, 3-13.
81. A. G. Murachelli, G. Damaskos and A. Perrakis, *Acta Crystallogr. D Struct. Biol.*, 2021, **77**, 992-1000.
82. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis, *Nature*, 2021, **596**, 583-589.
83. E. A. McKenzie and W. M. Abbott, *Methods*, 2018, **147**, 40-49.
84. M. N. Baeshen, A. M. Al-Hejin, R. S. Bora, M. M. Ahmed, H. A. Ramadan, K. S. Saini, N. A. Baeshen and E. M. Redwan, *J. Microbiol. Biotechnol.*, 2015, **25**, 953-962.
85. R. Baghban, S. Farajnia, M. Rajabibazl, Y. Ghasemi, A. Mafi, R. Hoseinpoor, L. Rahbarnia and M. Aria, *Mol. Biotechnol.*, 2019, **61**, 365-384.
86. P. T. Wingfield, *Curr. Protoc. Protein Sci.*, 2015, **80**, 6.1.1-6.1.35.
87. R. R. Burgess, *Protein Expression Purif.*, 2018, **150**, 81-85.
88. J. Y. Nehete, R. S. Bhambar, M. R. Narkhede and S. R. Gawali, *Pharmacogn. Rev.*, 2013, **7**, 107-116.
89. H. Hou, M. Shi, S.-Y. Hu, F. Ahmad, B. Zhang, Z.-H. Chen and D.-C. Yin, *J. Cryst. Growth*, 2019, **521**, 1-8.
90. M. Benvenuti and S. Mangani, *Nat. Protoc.*, 2007, **2**, 1633-1651.

91. S. Karthika, T. K. Radhakrishnan and P. Kalaichelvi, *Cryst. Growth Des.*, 2016, **16**, 6663-6681.
92. F. Gorrec, *Drug Discov. Today*, 2016, **21**, 819-825.
93. P. H. M. Torres, A. C. R. Sodero, P. Jofily and F. P. Silva-Jr, *Int. J. Mol. Sci.*, 2019, **20**.
94. M. Senda, T. Hayashi, M. Hatakeyama, K. Takeuchi, A. T. Sasaki and T. Senda, *Cryst. Growth Des.*, 2016, **16**, 1565-1571.
95. J. W. Pflugrath, *Acta Crystallogr. F Struct. Biol. Commun.*, 2015, **71**, 622-642.
96. J. Miyahara, K. Takahashi, Y. Amemiya, N. Kamiya and Y. Satow, *Nucl. Instrum. Methods Phys. Res., Sect. A*, 1986, **246**, 572-578.
97. S. M. Gruner, *Curr. Opin. Struct. Biol.*, 1994, **4**, 765-769.
98. M. Rowicka, A. Kudlicki and Z. Otwinowski, *Acta Crystallogr. A*, 2003, **59**, 172-182.
99. Z. Dauter and A. Wlodawer, *Protein Pept. Lett.*, 2016, **23**, 201-210.
100. G. L. Taylor, *Acta Crystallogr. D Biol. Crystallogr.*, 2010, **66**, 325-338.
101. B. Qian, S. Raman, R. Das, P. Bradley, A. J. McCoy, R. J. Read and D. Baker, *Nature*, 2007, **450**, 259-264.
102. R. Schwarzenbacher, A. Godzik, S. K. Grzechnik and L. Jaroszewski, *Acta Crystallogr. D Biol. Crystallogr.*, 2004, **60**, 1229-1236.
103. A. Giorgetti, D. Raimondo, A. E. Miele and A. Tramontano, *Bioinformatics*, 2005, **21**, 72-76.
104. P. Evans and A. McCoy, *Acta Crystallogr. D Biol. Crystallogr.*, 2008, **64**, 1-10.
105. A. D. Crawshaw, A. Baslé and P. S. Salgado, *Acta Crystallogr. D Struct. Biol.*, 2020, **76**, 261-271.
106. I. G. Shabalin, P. J. Porebski and W. Minor, *Crystallogr. Rev.*, 2018, **24**, 236-262.
107. J. Wang, *Protein Sci.*, 2015, **24**, 661-669.
108. A. Wlodawer, W. Minor, Z. Dauter and M. Jaskolski, *FEBS J.*, 2008, **275**, 1-21.
109. E. N. Brown and S. Ramaswamy, *Acta Crystallogr. D Biol. Crystallogr.*, 2007, **63**, 941-950.
110. J. M. Holton, S. Classen, K. A. Frankel and J. A. Tainer, *FEBS J.*, 2014, **281**, 4046-4060.
111. J. T. C. Wennmacher, C. Zaubitzer, T. Li, Y. K. Bahk, J. Wang, J. A. van Bokhoven and T. Gruene, *Nat. Commun.*, 2019, **10**, 3316.
112. G. N. Ramachandran, C. Ramakrishnan and V. Sasisekharan, *J. Mol. Biol.*, 1963, **7**, 95-99.
113. Z. Sun, Q. Liu, G. Qu, Y. Feng and M. T. Reetz, *Chemical Reviews*, 2019, **119**, 1626-1665.
114. H. Berman, K. Henrick and H. Nakamura, *Nat. Struct. Biol.*, 2003, **10**, 980.
115. J. Revillo Imbernon, C. Jacquemard, G. Bret, G. Marcou and E. Kellenberger, *RSC Med. Chem.*, 2022, **13**, 300-310.
116. A. Carbery, R. Skyner, F. von Delft and C. M. Deane, *J. Med. Chem.*, 2022, **65**, 11404-11413.
117. B. Wienen-Schmidt, M. Oebbeke, K. Ngo, A. Heine and G. Klebe, *ChemMedChem*, 2021, **16**, 292-300.
118. I. Müller, *Acta Crystallogr. D Struct. Biol.*, 2017, **73**, 79-92.
119. P. M. Collins, J. T. Ng, R. Talon, K. Nekrosiute, T. Krojer, A. Douangamath, J. Brandao-Neto, N. Wright, N. M. Pearce and F. von Delft, *Acta Crystallogr. D Struct. Biol.*, 2017, **73**, 246-255.
120. A. Douangamath, A. Powell, D. Fearon, P. M. Collins, R. Talon, T. Krojer, R. Skyner, J. Brandao-Neto, L. Dunnett, A. Dias, A. Aimon, N. M. Pearce, C. Wild, T. Gorrie-Stone and F. von Delft, *J. Vis. Exp.*, 2021, DOI: 10.3791/62414.

121. N. D. Wright, P. Collins, L. Koekemoer, T. Krojer, R. Talon, E. Nelson, M. Ye, R. Nowak, J. Newman, J. T. Ng, N. Mitrovich, H. Wiggers and F. von Delft, *Acta Crystallogr. D Struct. Biol.*, 2021, **77**, 62-74.
122. T. Krojer, R. Talon, N. Pearce, P. Collins, A. Douangamath, J. Brandao-Neto, A. Dias, B. Marsden and F. von Delft, *Acta Crystallogr. D Struct. Biol.*, 2017, **73**, 267-278.
123. N. M. Pearce, T. Krojer, A. R. Bradley, P. Collins, R. P. Nowak, R. Talon, B. D. Marsden, S. Kelm, J. Shi, C. M. Deane and F. von Delft, *Nat. Commun.*, 2017, **8**, 15123.
124. M. L. Boby, D. Fearon, M. Ferla, M. Filep, L. Koekemoer, M. C. Robinson, C. M. C. The, J. D. Chodera, A. A. Lee, N. London, A. von Delft, F. von Delft, H. Achdout, A. Aimon, D. S. Alonzi, R. Arbon, J. C. Aschenbrenner, B. H. Balcomb, E. Bar-David, H. Barr, A. Ben-Shmuel, J. Bennett, V. A. Bilenko, B. Borden, P. Boulet, G. R. Bowman, L. Brewitz, J. Brun, S. Bvnbs, M. Calmiano, A. Carbery, D. W. Carney, E. Cattermole, E. Chang, E. Chernyshenko, A. Clyde, J. E. Coffland, G. Cohen, J. C. Cole, A. Contini, L. Cox, T. I. Croll, M. Cvitkovic, S. De Jonghe, A. Dias, K. Donckers, D. L. Dotson, A. Douangamath, S. Duberstein, T. Dudgeon, L. E. Dunnett, P. Eastman, N. Erez, C. J. Eyermann, M. Fairhead, G. Fate, O. Fedorov, R. S. Fernandes, L. Ferrins, R. Foster, H. Foster, L. Fraisse, R. Gabizon, A. García-Sastre, V. O. Gawriljuk, P. Gehrtz, C. Gileadi, C. Giroud, W. G. Glass, R. C. Glen, I. Glinert, A. S. Godoy, M. Gorichko, T. Gorrie-Stone, E. J. Griffen, A. Haneef, S. Hassell Hart, J. Heer, M. Henry, M. Hill, S. Horrell, Q. Y. J. Huang, V. D. Huliak, M. F. D. Hurley, T. Israely, A. Jajack, J. Jansen, E. Jnoff, D. Jochmans, T. John, B. Kaminow, L. Kang, A. L. Kantsadi, P. W. Kenny, J. L. Kiappes, S. O. Kinakh, B. Kovar, T. Krojer, V. N. T. La, S. Laghnimi-Hahn, B. A. Lefker, H. Levy, R. M. Lithgo, I. G. Logvinenko, P. Lukacik, H. B. Macdonald, E. M. MacLean, L. L. Makower, T. R. Malla, P. G. Marples, T. Matviiuk, W. McCorkindale, B. L. McGovern, S. Melamed, K. P. Melnykov, O. Michurin, P. Miesen, H. Mikolajek, B. F. Milne, D. Minh, A. Morris, G. M. Morris, M. J. Morwitzer, D. Moustakas, C. E. Mowbray, A. M. Nakamura, J. B. Neto, J. Neyts, L. Nguyen, G. D. Noske, V. Oleinikovas, G. Oliva, G. J. Overheul, C. D. Owen, R. Pai, J. Pan, N. Paran, A. M. Payne, B. Perry, M. Pingle, J. Pinjari, B. Politi, A. Powell, V. Pšenák, I. Pulido, R. Puni, V. L. Rangel, R. N. Reddi, P. Rees, S. P. Reid, L. Reid, E. Resnick, E. G. Ripka, R. P. Robinson, J. Rodriguez-Guerra, R. Rosales, D. A. Rufa, K. Saar, K. S. Saikatendu, E. Salah, D. Schaller, J. Scheen, C. A. Schiffer, C. J. Schofield, M. Shafeev, A. Shaikh, A. M. Shaqra, J. Shi, K. Shurrush, S. Singh, A. Sittner, P. Sjö, R. Skyner, A. Smalley, B. Smeets, M. D. Smilova, L. J. Solmesky, J. Spencer, C. Strain-Damerell, V. Swamy, H. Tamir, J. C. Taylor, R. E. Tennant, W. Thompson, A. Thompson, S. Tomásio, C. W. E. Tomlinson, I. S. Tsurupa, A. Tumber, I. Vakonakis, R. P. van Rij, L. Vangeel, F. S. Varghese, M. Vaschetto, E. B. Vitner, V. Voelz, A. Volkamer, M. A. Walsh, W. Ward, C. Weatherall, S. Weiss, K. M. White, C. F. Wild, K. D. Witt, M. Wittmann, N. Wright, Y. Yahalom-Ronen, N. K. Yilmaz, D. Zaidmann, I. Zhang, H. Zidane, N. Zitzmann and S. N. Zvornicanin, *Science*, **382**, eabo7201.
125. Schrödinger, Schrödinger Release 2022-3: Maestro, Schrödinger, LLC, New York, NY, 2021).
126. R. Abagyan, M. Totrov and D. Kuznetsov, *J. Comput. Chem.*, 1994, **15**, 488-506.
127. L. G. Ferreira, R. N. Dos Santos, G. Oliva and A. D. Andricopulo, *Molecules (Basel, Switzerland)*, 2015, **20**, 13384-13421.
128. X. Y. Meng, H. X. Zhang, M. Mezei and M. Cui, *Curr. Comput. Aided Drug. Des.*, 2011, **7**, 146-157.

129. B. J. Bender, S. Gahbauer, A. Luttens, J. Lyu, C. M. Webb, R. M. Stein, E. A. Fink, T. E. Balius, J. Carlsson, J. J. Irwin and B. K. Shoichet, *Nat. Protoc.*, 2021, **16**, 4799-4832.
130. J. Li, A. Fu and L. Zhang, *Interdiscip. Sci. Comput. Life Sci.*, 2019, **11**, 320-328.
131. I. A. Vakser, *Curr. Opin. Struct. Biol.*, 2020, **64**, 160-165.
132. K. W. Lexa and H. A. Carlson, *Q. Rev. Biophys.*, 2012, **45**, 301-343.
133. D. Gioia, M. Bertazzo, M. Recanatini, M. Masetti and A. Cavalli, *Molecules (Basel, Switzerland)*, 2017, **22**.
134. V. Limongelli, L. Marinelli, S. Cosconati, C. La Motta, S. Sartini, L. Mugnaini, F. Da Settimo, E. Novellino and M. Parrinello, *Proc. Natl. Acad. Sci.*, 2012, **109**, 1467-1472.
135. H. R. Bourne, D. A. Sanders and F. McCormick, *Nature*, 1991, **349**, 117-127.
136. K. Wennerberg, K. L. Rossman and C. J. Der, *J. Cell. Sci.*, 2005, **118**, 843-846.
137. A. M. Rojas, G. Fuentes, A. Rausell and A. Valencia, *J. Cell Biol.*, 2012, **196**, 189-201.
138. J. Colicelli, *Sci. STKE*, 2004, **2004**, Re13.
139. D. K. Simanshu, D. V. Nissley and F. McCormick, *Cell*, 2017, **170**, 17-33.
140. G. Garcia-Rostan, H. Zhao, R. L. Camp, M. Pollan, A. Herrero, J. Pardo, R. Wu, M. L. Carcangiu, J. Costa and G. Tallini, *J. Clin. Oncol.*, 2003, **21**, 3226-3235.
141. Y. Homma, S. Hiragi and M. Fukuda, *FEBS J.*, 2021, **288**, 36-55.
142. Z. Boudhraa, E. Carmona, D. Provencher and A. M. Mes-Masson, *Front. Cell Dev. Biol.*, 2020, **8**, 345.
143. K. B. Matchett, S. McFarlane, S. E. Hamilton, Y. S. Eltuhamy, M. A. Davidson, J. T. Murray, A. M. Faheem and M. El-Tanani, *Adv. Exp. Med. Biol.*, 2014, **773**, 323-351.
144. E. Sztul, P. W. Chen, J. E. Casanova, J. Cherfils, J. B. Dacks, D. G. Lambright, F. S. Lee, P. A. Randazzo, L. C. Santy, A. Schürmann, I. Wilhelmi, M. E. Yohe and R. A. Kahn, *Mol. Biol. Cell*, 2019, **30**, 1249-1271.
145. C. D'Souza-Schorey and P. Chavrier, *Nat. Rev. Mol. Cell Biol.*, 2006, **7**, 347-358.
146. T. R. Stankiewicz and D. A. Linseman, *Front. Cell Neurosci.*, 2014, **8**, 1-14.
147. A. L. Bishop and A. Hall, *Biochem. J.*, 2000, **348 Pt 2**, 241-255.
148. A. B. Jaffe and A. Hall, *Annu. Rev. Cell Dev. Biol.*, 2005, **21**, 247-269.
149. H. R. Bourne, D. A. Sanders and F. McCormick, *Nature*, 1990, **348**, 125-132.
150. J. L. Bos, H. Rehmann and A. Wittinghofer, *Cell*, 2007, **129**, 865-877.
151. J. Cherfils and M. Zeghouf, *Physiol. Rev.*, 2013, **93**, 269-309.
152. A. D. Cox, S. W. Fesik, A. C. Kimmelman, J. Luo and C. J. Der, *Nat. Rev. Drug Discov.*, 2014, **13**, 828-851.
153. P. A. Konstantinopoulos, M. V. Karamouzis and A. G. Papavassiliou, *Nat. Rev. Drug Discov.*, 2007, **6**, 541-555.
154. A. L. Ho, G. J. Hanna, C. R. Scholz, A. Gualberto and S. H. Park, *J. Clin. Oncol.*, 2020, **38**, 6504-6504.
155. A. L. Ho, I. Brana, R. Haddad, J. Bauman, K. Bible, S. Oosting, D. J. Wong, M.-J. Ahn, V. Boni, C. Even, J. Fayette, M. J. Flor, K. Harrington, D. S. Hong, S.-B. Kim, L. Licitra, I. Nixon, N. F. Saba, S. Hackenberg, P. Specenier, F. Worden, B. Balsara, M. Leoni, B. Martell, C. Scholz and A. Gualberto, *J. Clin. Oncol.*, 2021, **39**, 1856-1864.
156. L. Santarpia, S. M. Lippman and A. K. El-Naggar, *Expert Opin. Ther. Targets*, 2012, **16**, 103-119.
157. E. M. Terrell and D. K. Morrison, *Cold Spring Harb. Perspect. Med.*, 2019, **9**.
158. M. Beeram, A. Patnaik and E. K. Rowinsky, *J. Clin. Oncol.*, 2005, **23**, 6771-6790.

159. E. C. Nakajima, N. Drezner, X. Li, P. S. Mishra-Kalyani, Y. Liu, H. Zhao, Y. Bi, J. Liu, A. Rahman, E. Wearne, I. Ojofeitimi, L. T. Hotaki, D. Spillman, R. Pazdur, J. A. Beaver and H. Singh, *Clin. Cancer Res.*, 2022, **28**, 1482-1486.
160. F. Skoulidis, B. T. Li, G. K. Dy, T. J. Price, G. S. Falchook, J. Wolf, A. Italiano, M. Schuler, H. Borghaei, F. Barlesi, T. Kato, A. Curioni-Fontecedro, A. Sacher, A. Spira, S. S. Ramalingam, T. Takahashi, B. Besse, A. Anderson, A. Ang, Q. Tran, O. Mather, H. Henary, G. Ngarmchamnanrith, G. Friberg, V. Velcheti and R. Govindan, *New Engl. J. Med.*, 2021, **384**, 2371-2381.
161. J. Liu, R. Kang and D. Tang, *Cancer Gene Ther.*, 2022, **29**, 875-878.
162. M. Z. Guo, K. A. Marrone, A. Spira and S. Rosner, *Future Oncol.*, 2023, DOI: 10.2217/fon-2022-1106.
163. Y. Oya and T. Mitsudomi, *Transl. Lung Cancer Res.*, 2023, **12**, 940-943.
164. J. L. Gray, F. von Delft and P. E. Brennan, *Angew. Chem. Int. Ed.*, 2020, **59**, 6342-6366.
165. K. L. Rossman, C. J. Der and J. Sondek, *Nat. Rev. Mol. Cell Biol.*, 2005, **6**, 167-180.
166. J. Cherfils and M. Zeghouf, *Physiol. Rev.*, 2013, **93**, 269-309.
167. J. L. Gray, C. Jimenez Antunez, T. Krojer, M. Fairhead, N. Burgess-Brown, F. von Delft and P. E. Brennan, Human Kalirin/RAC1 GEF/GTPase Complex, DOI: https://doi.org/10.5281/zenodo.3266933).
168. Y. Gao, J. B. Dickerson, F. Guo, J. Zheng and Y. Zheng, *Proc. Natl. Acad. Sci.*, 2004, **101**, 7618-7623.
169. T. Humphries-Bickley, L. Castillo-Pichardo, E. Hernandez-O'Farrill, L. D. Borrero-Garcia, I. Forestier-Roman, Y. Gerena, M. Blanco, M. J. Rivera-Robles, J. R. Rodriguez-Medina, L. A. Cubano, C. P. Vlaar and S. Dharmawardhane, *Mol. Cancer Ther.*, 2017, **16**, 805-818.
170. N. Bouquier, E. Vignal, S. Charrasse, M. Weill, S. Schmidt, J.-P. Léonetti, A. Blangy and P. Fort, *Chemistry & Biology*, 2009, **16**, 657-666.
171. F. Lemtiri-Chlieh, L. Zhao, D. D. Kiraly, B. A. Eipper, R. E. Mains and E. S. Levine, *BMC Neurosci.*, 2011, **12**, 126.
172. E. E. Bosco, J. F. Mulloy and Y. Zheng, *Cell Mol. Life Sci.*, **66**, 370-374.
173. H. Marei and A. Malliri, *Small GTPases*, 2017, **8**, 90-99.
174. J. Didsbury, R. F. Weber, G. M. Bokoch, T. Evans and R. Snyderman, *J. Bio. Chem.*, 1989, **264**, 16378-16382.
175. J. K. Westwick, T. Lambert Q, G. J. Clark, M. Symons, L. Van Aelst, R. G. Pestell and C. J. Der, *Mol. Cell Biol.*, 1997, **17**, 1324-1335.
176. D. J. Marston, K. L. Anderson, M. F. Swift, M. Rougie, C. Page, K. M. Hahn, N. Volkmann and D. Hanein, *Proc. Natl. Acad. Sci.*, 2019, **116**, 1267-1272.
177. R. Kozma, S. Ahmed, A. Best and L. Lim, *Mol. Cell. Biol.*, 1995, **15**, 1942-1952.
178. J. M. Bristow, M. H. Sellers, D. Majumdar, B. Anderson, L. Hu and D. J. Webb, *J. Cell Sci.*, 2009, **122**, 4535-4546.
179. N. Ma, E. Xu, Q. Luo and G. A.-O. Song, *Molecules (Basel, Switzerland)*, 2023, **27**, 2976.
180. J. Liang, L. Oyang, S. Rao, Y. Han, X. Luo, P. Yi, J. Lin, L. Xia, J. Hu, S. Tan, L. Tang, Q. Pan, Y. Tang, Y. Zhou and Q. Liao, *Front. Oncol.*, 2021, **17**.
181. N. Ferri, C. A., S. K. Bernini and A. Corsini, *J. Cardiovasc. Pharmacol.*, 2013, **5**, 425-435.
182. M. Satoh, H. Ogita, K. Takeshita, Y. Mukai, D. Kwiatkowski and J. K. Liao, *Proc. Natl. Acad. Sci. U S A*, 2006, **103**, 7432-7437.
183. M. Kikuchi, M. Sekiya, N. Hara, A. Miyashita, R. Kuwano, T. Ikeuchi, K. M. Iijima and A. Nakaya, *Hum. Mol. Genet.*, 2020, **29**, 817-833.

184. B. J. Aguilar, Y. Zhu and Q. Lu, *Alzheimer's Res. Ther.*, 2017, **9**, 97.
185. M. Borin, C. Saraceno, M. Catania, E. Lorenzetto, V. Pontelli, A. Paterlini, S. Fostinelli, A. Avesani, G. Di Fede, G. Zanusso, L. Benussi, G. Binetti, S. Zorzan, R. Ghidoni, M. Buffelli and S. Bolognin, *Acta Neuropathol. Commun.*, 2018, **6**, 61.
186. A. Tousley, M. Iuliano, E. Weisman, E. Sapp, N. Zhang, P. Vodicka, J. Alexander, H. Aviolat, L. Gatune, P. Reeves, X. Li, A. Khvorova, L. M. Ellerby, N. Aronin, M. DiFiglia and K. B. Kegel-Gleason, *J. Huntingtons Dis.*, 2019, **8**, 53-69.
187. A. Hayashi-Takagi, M. Takaki, N. Graziane, S. Seshadri, H. Murdoch, A. J. Dunlop, Y. Makino, A. J. Seshadri, K. Ishizuka, D. P. Srivastava, Z. Xie, J. M. Baraban, M. D. Houslay, T. Tomoda, N. J. Brandon, A. Kamiya, Z. Yan, P. Penzes and A. Sawa, *Nat. Neurosci.*, 2010, **13**, 327-332.
188. P. Madaule and R. Axel, *Cell*, 1985, **41**, 31-40.
189. M. P. Girouard, M. Pool, R. Alchini, I. Rambaldi and A. Fournier, *PLoS One*, 2016, **11**, e0168641.
190. R. C. Nayak, K. Chang, N. Vaitinadin and J. A. Cancelas, *Immunol. Rev.*, 2013, **256**, 255-268.
191. Anne J. Ridley, *Cell*, 2011, **145**, 1012-1022.
192. A. J. Ridley and A. Hall, *Cell*, 1992, **70**, 389-399.
193. A. Hall, *Science*, 1998, **279**, 509-514.
194. J. L. MacKay and S. Kumar, *Integr. Biol.*, 2014, **6**, 885-894.
195. M. J. Marinissen, M. Chiariello and J. S. Gutkind, *Genes. Dev.*, 2001, **1**, 535-553.
196. H. Warner, B. J. Wilson and P. T. Caswell, *Curr. Opin. Cell Biol.*, 2019, **56**, 64-70.
197. R. Marjoram, E. C. Lessey and K. Burridge, *Curr. Mol. Med.*, 2014, **14**, 199-208.
198. S. I. Schmidt, M. Blaabjerg, K. Freude and M. Meyer, *Cells*, 2022, **1**, 1520.
199. Z. Zhou, J. Kim, R. Insolera, X. Peng, D. Fink and M. Mata, *Mol. Cell. Nueurosci.*, 2011, **48**, 29-37.
200. Y. Zhou, Y. Su, Li B, F. Liu, J. Ryder, X. Wu, P. Gonzalez-DeWhitt, V. Gelfanova, J. Hale, P. May, S. Paul and B. Ni, *Science*, 2003, **302**, 1215-1217.
201. P. J. Chacon, R. Garcia-Mejias and A. Rodriguez-Tebar, *Mol. Neurodegener.*, 2011, **6**, 14.
202. M. Takata, H. Tanaka, M. Kimura, Y. Nagahara, K. F. Tanaka, K. Kawasaki, M. Seto, K. Tsuruma, M. Shimazawa and H. Hara, *Br. J. Pharmacol.*, 2013, **170**, 341-351.
203. M. R. Alam, R. C. Johnson, D. N. Darlington, R. E. Mains and B. A. Eipper, *J. Biol. Chem.*, 1996, **271**, 28636-28640.
204. E. Parnell, L. P. Shapiro, R. A. Voorn, M. P. Forrest, H. A. Jalloul, D. D. Loizzo and P. Penzes, *Gene*, 2020, DOI: https://doi.org/10.1016/j.gene.2020.145306, 145306.
205. X.-M. Ma, D. D. Kiraly, E. D. Gaier, Y. Wang, E.-J. Kim, E. S. Levine, B. A. Eipper and R. E. Mains, *J. Neurosci.*, 2008, **28**, 12368-12382.
206. A. J. Deo, M. E. Cahill, S. Li, I. Goldszer, R. Henteleff, J.-E. VanLeeuwen, I. Rafalovich, R. Gao, E. K. Stachowski, A. R. Sampson, D. A. Lewis, P. Penzes and R. A. Sweet, *Neurobiol.*, 2012, **45**, 796-803.
207. Z. Xie, D. Srivastava, H. Photowala, L. Kai, M. E. Cahill, K. M. Woolfrey, C. Y. Shum, D. J. Surmeier and P. Penzes, *Nueron.*, 2007, **56**, 640-656.
208. A. W. Mould, N. Al-Juffali, A. von Delft, P. E. Brennan and E. M. Tunbridge, *CNS Drugs*, 2022, **36**, 1-16.
209. J. D. Paskus, B. E. Herring and K. W. Roche, *Trends. Neurosci.*, 2020, **43**, 505-518.
210. C. Remmers, R. A. Sweet and P. Penzes, *Brain Res. Bull.*, 2014, **103**, 29-38.

211. P. Mandela and X. M. Ma, *Nueral. Plast.*, **2012**.
212. Y. Li, Y. Pustovalova, T. I. Doukov, J. C. Hoch, R. E. Mains, B. A. Eipper and B. Hao, *Nat. Commun.*, 2023, **14**, 96.
213. X.-M. Ma, M. B. Miller, K. S. Vishwanatha, M. J. Gross, Y. Wang, T. Abbott, T. T. Lam, R. E. Mains and B. A. Eipper, *Molecular Biology of the Cell*, 2014, **25**, 1458-1471.
214. K. S. Vishwanatha, Y. P. Wang, H. T. Keutmann, R. E. Mains and B. A. Eipper, *Biochem.*, 2012, **51**, 5663-5673.
215. M. R. Schiller, F. Ferraro, Y. Wang, X.-m. Ma, C. E. McPherson, J. A. Sobota, N. I. Schiller, R. E. Mains and B. A. Eipper, *Exp. Cell Res.*, 2008, **314**, 2674-2691.
216. N. Kurochkina and U. Guha, *Biophys Rev*, 2013, **5**, 29-39.
217. T. Kawai, H. Sanjo and S. Akira, *Gene*, 1999, **227**, 249-255.
218. P. Penzes, R. Johnson, M. R. Alam, V. Kambampati, R. E. Mains and B. A. Eipper, *J. Biol. Chem.*, 2000, **275**, 6395-6403.
219. R. C. Johnson, P. Penzes, B. A. Eipper and R. E. Mains, *J. Biol. Chem.*, 2000, **275**, 19324-19333.
220. G. Powis, E. J. Meuillet, M. Indarte, G. Booher and L. Kirkpatrick, *Biomed. Pharmacother.*, 2023, **165**, 115024.
221. Z. Chen, F. Medina, M. Y. Liu, C. Thomas, S. R. Sprang and P. C. Sternweis, *J. Biol. Chem.*, 2010, **285**, 21070-21081.
222. M. K. Chhatriwala, L. Betts, D. K. Worthylake and J. Sondek, *J. Mol. Biol.*, 2007, **368**, 1307-1320.
223. M. A. Lemmon, *Biochem. Soc. Symp.*, 2007, **74**, 81-93.
224. A. P. Thompson, C. Bitsina, J. L. Gray, F. von Delft and P. E. Brennan, *J. Biol. Chem.*, 2021, **296**, 100521.
225. M. Laurin and J. F. Côté, *Genes Dev.*, 2014, **28**, 533-547.
226. D. E. Hansel, M. E. Quiñones, G. V. Ronnett and B. A. Eipper, *Journal of Histochemistry & Cytochemistry*, 2001, **49**, 833-844.
227. C. M. Mazzone, P. Taylor, D. Kiraly, B. A. Eipper and R. E. Mains, *Mol. Pharmacol.*, 2012, **82**, 1241-1249.
228. J. J. Hill, T. Hashimoto and D. A. Lewis, *Mol. Psychiatry*, 2006, **11**, 557-566.
229. M. D. Rubio, J. H. Haroutunian V Fau - Meador-Woodruff and J. H. Meador-Woodruff.
230. I. R. Vetter and A. Wittinghofer, *Science*, 2001, **294**, 1299-1304.
231. A. Wittinghofer, *Ras Superfamily Small G Proteins: Biology and Mechanisms 1*, Springer Vienna, 2014.
232. K. L. Rossman, D. K. Worthylake, J. T. Snyder, D. P. Siderovski, S. L. Campbell and J. Sondek, *EMBO J*, 2002, **21**, 1315-1326.
233. Y. Wei, Y. Zhang, U. Derewenda, X. Liu, W. Minor, R. K. Nakamoto, A. V. Somlyo, A. P. Somlyo and Z. S. Derewenda, *Nat. Struct. Biol.*, 1997, **4**, 699-703.
234. E. Mons, S. Roet, R. Q. Kim and M. P. C. Mulder, *Curr. Protoc.*, 2022, **2**, e419.
235. B. F. Krippendorff, R. Neuhaus, P. Lienau, A. Reichel and W. Huisinga, *SLAS Discov.*, 2009, **14**, 913-923.
236. H. Melissa, G. Tarikere, K. Taisei, P. D. James, J. H. Ronald and M. Ann, *J. Pharmacol. Exp. Ther.*, 2020, **372**, 331.
237. H. Lin, M. Wang, Y. W. Zhang, S. Tong, R. A. Leal, R. Shetty, K. Vaddi and J. I. Luengo, *ACS Med. Chem. Lett.*, 2019, **10**, 1033-1038.
238. A. E. Speers and B. F. Cravatt, *Curr. Protoc. Chem. Biol.*, 2009, **1**, 29-41.
239. T. U. Consortium, *Nucleic Acids Res.*, 2022, **51**, D523-D531.

240. J. Cox and M. Mann, *Nat. Biotechnol.*, 2008, **26**, 1367-1372.
241. A. T. Kong, F. V. Leprevost, D. M. Avtonomov, D. Mellacheruvu and A. I. Nesvizhskii, *Nature Methods*, 2017, **14**, 513-520.
242. E. Weerapana, C. Wang, G. M. Simon, F. Richter, S. Khare, M. B. D. Dillon, D. A. Bachovchin, K. Mowen, D. Baker and B. F. Cravatt, *Nature*, 2010, **468**, 790-795.
243. D. Leung, C. Hardouin, D. L. Boger and B. F. Cravatt, *Nat. Biotechnol.*, 2003, **21**, 687-691.
244. M. M. Walker, B. Koronkiewicz, S. Chen, K. N. Houk, J. M. Mayer and J. A. Ellman, *J. Am. Chem. Soc.*, 2020, **142**, 8194-8202.
245. J. W. Beatty and C. R. J. Stephenson, *Acc. Chem. Res.*, 2015, **48**, 1474-1484.
246. K. Yamamoto, M. Kuriyama and O. Onomura, *Chem. Rec.*, 2021, **21**, 2239-2253.
247. I. Habib, K. Singha and M. Hossain, *ChemistrySelect*, 2023, **8**, e202204099.
248. F. Zhang and X.-F. Duan, *Org. Lett.*, 2011, **13**, 6102-6105.
249. H. Andersson, R. Olsson and F. Almqvist, *Org. Biomol. Chem.*, 2011, **9**, 337-346.
250. T. Shono, H. Hamaguchi and Y. Matsumura, *J. Am. Chem. Soc.*, 1975, **97**, 4264-4268.
251. S. Wang, Y. Tian, M. Wang, M. Wang, G.-b. Sun and X.-b. Sun, *Front. Pharmacol.*, 2018, **9**, 353.
252. M. Békés, D. R. Langley and C. M. Crews, *Nat. Rev. Drug Discov.*, 2022, **21**, 181-200.
253. X. Sun, H. Gao, Y. Yang, M. He, Y. Wu, Y. Song, Y. Tong and Y. Rao, *Signal Transduct. Ther.*, 2019, **4**, 64.
254. K. M. Sakamoto, K. B. Kim, A. Kumagai, F. Mercurio, C. M. Crews and R. J. Deshaies, *Proc. Natl. Acad. Sci.*, 2001, **98**, 8554-8559.
255. X. He, Y. C. Kuo, T. J. Rosche and X. Zhang, *Structure*, 2013, **21**, 355-364.
256. U. Derewenda, A. Oleksy, A. S. Stevenson, J. Korczynska, Z. Dauter, A. P. Somlyo, J. Otlewski, A. V. Somlyo and Z. S. Derewenda, *Structure*, 2004, **12**, 1955-1965.
257. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass and M. J. E. Sternberg, *Nat. Prot.*, 2015, **10**, 845-858.
258. X. Liu, H. Wang, M. Eberstadt, A. Schnuchel, E. T. Olejniczak, R. P. Meadows, J. M. Schkeryantz, D. A. Janowick, J. E. Harlan, E. A. Harris, D. E. Staunton and S. W. Fesik, *Cell*, 1998, **95**, 269-277.
259. G. M. C. Gasmi-Seabrook, C. B. Marshall, M. Cheung, B. Kim, F. Wang, Y. J. Jang, T. W. Mak, V. Stambolic and M. Ikura, *J. Biol. Chem.*, 2010, **285**, 5137-5145.
260. D. K. Worthylake, K. L. Rossman and J. Sondek, *Nature*, 2000, **408**, 682-688.
261. A. Oleksy, Ł. Opaliński, U. Derewenda, Z. S. Derewenda and J. Otlewski, *J. Biol. Chem.*, 2006, **281**, 32891-32897.
262. J. T. Snyder, D. K. Worthylake, K. L. Rossman, L. Betts, W. M. Pruitt, D. P. Siderovski, C. J. Der and J. Sondek, *Nat. Struct. Biol.*, 2002, **9**, 468-475.
263. L. Cheng, K. L. Rossman, G. M. Mahon, D. K. Worthylake, M. Korus, J. Sondek and I. P. Whitehead, *Mol. Cell. Biol.*, 2002, **22**, 6895-6905.
264. J. E. Chrencik, A. Brooun, H. Zhang, I. I. Mathews, G. L. Hura, S. A. Foster, J. J. P. Perry, M. Streiff, P. Ramage, H. Widmer, G. M. Bokoch, J. A. Tainer, G. Weckbecker and P. Kuhn, *J. Mol. Biol.*, 2008, **380**, 828-843.
265. Y. Gao, J. Xing, M. Streuli, T. L. Leto and Y. Zheng, *J. Biol. Chem.*, 2001, **276**, 47530-47541.
266. A. E. Karnoub, D. K. Worthylake, K. L. Rossman, W. M. Pruitt, S. L. Campbell, J. Sondek and C. J. Der, *Nat. Struct. Mol. Biol.*, 2001, **8**, 1037-1041.
267. R. Kristelly, G. Gao and J. J. Tesmer, *J. Biol. Chem.*, 2004, **279**, 47352-47362.

268. W. T. Arthur, S. M. Ellerbroek, C. J. Der, K. Burridge and K. Wennerberg, *J. Biol. Chem.*, 2002, **277**, 42964-42972.
269. J. Senderek, *Eur. J. Neurol.*, 2021, **28**, 1106-1107.
270. Y. Miao, M. Yu, L. Meng, W. Zhang, H. Lv, Z. Wang and Y. Yuan, *Clin. Neuropathol.*, 2021, **40**, 328-332.
271. T. Hirano, Y. Shinsato, K. Tanabe, N. Higa, M. Kamil, K. Kawahara, M. Yamamoto, K. Minami, M. Shimokawa, T. Arigami, S. Yanagita, D. Matushita, Y. Uenosono, S. Ishigami, Y. Kijima, K. Maemura, I. Kitazono, A. Tanimoto, T. Furukawa and S. Natsugoe, *Oncogenesis*, 2020, **9**, 13.
272. Z. H. Chen and L. H. Wang, *Am. J. Dermatopathol.*, 2019, **41**, 908-913.
273. M. Cooke, G. Kreider-Letterman, M. J. Baker, S. Zhang, N. T. Sullivan, E. Eruslanov, M. C. Abba, S. M. Goicoechea, R. García-Mata and M. G. Kazanietz, *Cell Rep.*, 2021, **37**, 109905.
274. Y. Li, Z. Ye, S. Chen, Z. Pan, Q. Zhou, Y. Z. Li, W. D. Shuai, C. M. Kuang, Q. H. Peng, W. Shi, X. Mao, R. Y. Liu and W. Huang, *Int. J. Cancer.*, 2018, **142**, 1379-1391.
275. L. Niu, Y. Zhou, W. Zhang, Y. Yan and Y. Ren, *Physiol. Int.*, 2021, DOI: 10.1556/2060.2021.00187.
276. Z. Wang, G. Kan, C. Sheng, C. Yao, Y. Mao and S. Chen, *Biochem. Biophys. Res. Commun.*, 2020, **533**, 792-799.
277. K. L. Rossman, L. Cheng, G. M. Mahon, R. J. Rojas, J. T. Snyder, I. P. Whitehead and J. Sondek, *J. Biol. Chem.*, 2003, **278**, 18393-18400.
278. A. M. Crompton, L. H. Foley, A. Wood, W. Roscoe, D. Stokoe, F. McCormick, M. Symons and G. Bollag, *J. Biol. Chem.*, 2000, **275**, 25751-25759.
279. T. H. Chuang, X. Xu, V. Kaartinen, N. Heisterkamp, J. Groffen and G. M. Bokoch, *Proc. Natl. Acad. Sci. U S A*, 1995, **92**, 10282-10286.
280. M. Jaiswal, R. Dvorsky and M. R. Ahmadian, *J. Biol. Chem.*, 2013, **288**, 4486-4500.
281. M. J. Hart, S. Sharma, N. elMasry, R.-G. Qiu, P. McCabe, P. Polakis and G. Bollag, *J. Biol. Chem.*, 1996, **271**, 25452-25458.
282. Z. Wang, Y. Kumamoto, P. Wang, X. Gan, D. Lehmann, A. V. Smrcka, L. Cohn, A. Iwasaki, L. Li and D. Wu, *J. Biol. Chem.*, 2009, **284**, 28599-28606.
283. M. E. Yohe, K. L. Rossman, O. S. Gardner, A. E. Karnoub, J. T. Snyder, S. Gershburg, L. M. Graves, C. J. Der and J. Sondek, *J. Biol. Chem.*, 2007, **282**, 13813-13823.
284. E. Manser, T.-H. Loo, C.-G. Koh, Z.-S. Zhao, X.-Q. Chen, L. Tan, I. Tan, T. Leung and L. Lim, *Mol. Cell*, 1998, **1**, 183-192.
285. A. S. Alberts and R. Treisman, *EMBO J*, 1998, **17**, 4075-4085.
286. M. Mohl, S. Winkler, T. Wieland and S. Lutz, *Naunyn-Schmiedeberg's Archives of Pharmacology*, 2006, **373**, 333-341.
287. S. Winkler, M. Mohl, T. Wieland and S. Lutz, *Biochem. Biophys. Res. Commun.*, 2005, **335**, 1280-1286.
288. D. Diviani, J. Soderling and J. D. Scott, *J. Biol. Chem.*, 2001, **276**, 44247-44257.
289. A. W. Oliver, X. He, K. Borthwick, A. J. Donne, L. Hampson and I. N. Hampson, *Br. J. Cancer*, 2011, **104**, 324-331.
290. N. Hiramoto-Yamaki, S. Takeuchi, S. Ueda, K. Harada, S. Fujimoto, M. Negishi and H. Katoh, *J. Cell Biol.*, 2010, **190**, 461-477.
291. U. Rümenapp, A. Freichel-Blomquist, B. Wittinghofer, K. H. Jakobs and T. Wieland, *Biochem. J.*, 2002, **366**, 721-728.
292. K.-I. Nagata and M. Inagaki, *Oncogene*, 2005, **24**, 65-76.

293. M. J. Hart, A. Eva, D. Zangrilli, S. A. Aaronson, T. Evans, R. A. Cerione and Y. Zheng, *J. Biol. Chem.*, 1994, **269**, 62-65.
294. A. Debant, C. Serra-Pagès, K. Seipel, S. O'Brien, M. Tang, S. H. Park and M. Streuli, *Proc. Natl. Acad. Sci. U S A*, 1996, **93**, 5466-5471.
295. M. Souchet, E. Portales-Casamar, D. Mazurais, S. Schmidt, I. Léger, J.-L. Javré, P. Robert, I. Berrebi-Bertrand, A. Bril, B. Gout, A. Debant and T. P. G. Calmels, *J. Cell Sci.*, 2002, **115**, 629-640.
296. S. M. Ellerbroek, K. Wennerberg, W. T. Arthur, J. M. Dunty, D. R. Bowman, K. A. DeMali, C. Der and K. Burridge, *Mol. Biol. Cell*, 2004, **15**, 3309-3319.
297. S. M. Shamah, M. Z. Lin, J. L. Goldberg, S. Estrach, M. Sahin, L. Hu, M. Bazalakova, R. L. Neve, G. Corfas, A. Debant and M. E. Greenberg, *Cell*, 2001, **105**, 233-244.
298. F. P. van Horck, M. R. Ahmadian, L. C. Haeusler, W. H. Moolenaar and O. Kranenburg, *J. Biol. Chem.*, 2001, **276**, 4948-4956.
299. M. J. Hamann, C. M. Lubking, D. N. Luchini and D. D. Billadeau, *Mol. Cell Biol.*, 2007, **27**, 1380-1393.
300. P. A. Solski, R. S. Wilder, K. L. Rossman, J. Sondek, A. D. Cox, S. L. Campbell and C. J. Der, *J. Biol. Chem.*, 2004, **279**, 25226-25233.
301. M. Umikawa, H. Obaishi, H. Nakanishi, K. Satoh-Horikawa, K. Takahashi, I. Hotta, Y. Matsuura and Y. Takai, *J. Biol. Chem.*, 1999, **274**, 25197-25200.
302. S. Park, Y. Guo, J. Negre, J. Preto, C. C. Smithers, A. K. Azad, M. Overduin, A. G. Murray and G. Eitzen, *Small GTPases*, 2021, **12**, 147-160.
303. D. Wu, M. Asiedu and Q. Wei, *Oncogene*, 2009, **28**, 2219-2230.
304. R. E. Joseph and F. A. Norris, *J. Biol. Chem.*, 2005, **280**, 27508-27512.
305. M. Kiyono, T. Satoh and Y. Kaziro, *Proc. Natl. Acad. Sci. U S A*, 1999, **96**, 4826-4831.
306. W.-T. Fan, C. A. Koch, C. L. de Hoog, N. P. Fam and M. F. Moran, *Curr. Biol.*, 1998, **8**, 935-939.
307. M. Hoshino, M. Sone, M. Fukata, S. Kuroda, K. Kaibuchi, Y.-i. Nabeshima and C. Hama, *J. Biol. Chem.*, 1999, **274**, 17837-17844.
308. J. Heo, R. Thapar and S. L. Campbell, *Biochem.*, 2005, **44**, 6573-6585.
309. K. Abe, K. L. Rossman, B. Liu, K. D. Ritola, D. Chiang, S. L. Campbell, K. Burridge and C. J. Der, *J. Biol. Chem.*, 2000, **275**, 10141-10149.
310. N. Movilla and X. R. Bustelo, *Mol. Cell Biol.*, 1999, **19**, 7870-7885.
311. R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *J. Med. Chem.*, 2004, **47**, 1739-1749.
312. K. Zhu, K. W. Borrelli, J. R. Greenwood, T. Day, R. Abel, R. S. Farid and E. Harder, *J. Chem. Inf. Model.*, 2014, **54**, 1932-1940.
313. P. R. A. Zanon, L. Lewald and S. M. Hacker, *Angew. Chem. Int. Ed.*, 2020, **59**, 2829-2836.
314. Y. Perez-Riverol, J. Bai, C. Bandla, D. García-Seisdedos, S. Hewapathirana, S. Kamatchinathan, D. J. Kundu, A. Prakash, A. Frericks-Zipper, M. Eisenacher, M. Walzer, S. Wang, A. Brazma and J. A. Vizcaíno, *Nucleic Acids Res.*, 2022, **50**, D543-d552.
315. R. N. Reddi, A. Rogel, R. Gabizon, D. G. Rawale, B. Harish, S. Marom, B. Tivon, Y. S. Arbel, N. Gurwicz, R. Oren, K. David, J. Liu, S. Duberstein, M. Itkin, S. Malitsky, H. Barr, B.-Z. Katz, Y. Herishanu, I. Shachar, Z. Shulman and N. London, *J. Am. Chem. Soc.*, 2023, **145**, 3346-3360.
316. G. Winter, *J. Appl. Crystallogr.*, 2010, **43**, 186-190.

317. D. G. Waterman, G. Winter, R. J. Gildea, J. M. Parkhurst, A. S. Brewster, N. K. Sauter and G. Evans, *Acta Crystallogr D*, 2016, **72**, 558-575.
318. W. Kabsch, *Acta Crystallogr. D Biol. Crystallogr.*, 2010, **66**, 125-132.
319. P. Evans, *Acta Crystallogr. D Biol. Crystallogr.*, 2006, **62**, 72-82.
320. M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson, *Acta Crystallogr. D Biol. Crystallogr.*, 2011, **67**, 235-242.
321. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart, *Acta Crystallogr. D Biol. Crystallogr.*, 2010, **66**, 213-221.
322. P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta Crystallogr. D Biol. Crystallogr.*, 2010, **66**, 486-501.
323. A. A. Vagin, R. A. Steiner, A. A. Lebedev, L. Potterton, S. McNicholas, F. Long and G. N. Murshudov, *Acta Crystallogr. D Biol. Crystallogr.*, 2004, **60**, 2184-2195.