BERGISCHE
UNIVERSITÄT
WUPPERTAL

QSAR-GUIDED DESIGN AND SYNTHESIS OF NOVEL INHIBITORS TARGETING KRAS PROTEIN-PROTEIN INTERACTIONS

DISSERTATION

Submitted for the Degree of Dr. rer. nat.

Benjamin Thäbas Nava Höer

Wuppertal June 2025

First reviewer:	Prof. Dr. Jürgen Scherkenbeck
	Organic Chemistry
	University of Wuppertal
Second reviewer:	Prof. Dr. Fabian Mohr
	Inorganic Chemistry
	University of Wuppertal

Acknowledgements

Thank yous

A raft through space and time. These lines
Are laden with precious meaning, meticulously collected
Through shared experiences—condensed into letters.
They won't survive—neither the lines nor their cargo,
Indefinitely. Sadly. Thankfully.

A raft through space and time. These lines
Will speak to those who want to listen, because
They were there, or would like to be then
For a moment. A shy glance at the
Echo of connection left by the author. I, grateful.

A raft through space and time. These lines
Hide the hardship and lose it along the way.
Heavy load aground. Weighing down itself only
Buoyant lines reach distant shores—as intended.
"We were here, together, gratefully."

A raft through space and time. These lines
Share the context we shared, selflessly.
Feel the cohesion we felt, profoundly.
Embody what we cling to, relentlessly.
Accompany us on our journeys, constantly.
Thankfully.

I owe so unspeakably much to my mother Heidi, my father Lalo, my love Merle and her family, as well as my friends Jonas, Seppi, Manuel, Cenk, Daniel, Sebi, Luis, Rici, Lutz, Sascha, Manuela, Oliver, Gerrit, Daniel, and those friends who have not been listed here. Finally, I'd like to thank the staff at the BUW.

Thank you all.

Abstract

The recent identification of the first non-covalent $KRAS^{G12D}$ inhibitor exhibiting nanomolar potency constitutes a significant paradigm shift in the therapeutic targeting of KRAS, a protein historically regarded as *undruggable*. This breakthrough was exemplified by MRTX-1133, a compound operating within the chemical space associated with protein-protein interaction inhibitors (PPIIs). The present thesis is centred on the systematic exploration of this PPII chemical space with the objective of discovering novel scaffolds capable of inhibiting KRAS.

The first project within this thesis was directed towards the expansion of a project compound library with KRAS-targeting molecules. Sixteen analogues based on two scaffold classes, i.e. biazoles and zafirlukast, were synthesised and evaluated for their ability to inhibit KRAS activity. Several derivatives demonstrated inhibition of SOS-mediated nucleotide exchange on $KRAS^{G12D}$, with half-maximal inhibitory concentrations (IC_{50}) in the low micromolar range. These compounds also elicited a reduction in cell viability in KRAS-mutant cancer cell lines. Structure-activity relationship (SAR) analysis revealed a positive correlation between the presence of carboxylic acid bioisosteres and the inhibition of nucleotide exchange.

Molecular docking studies provided insights into the potential binding site and mechanism of action of the biazole derivatives. In contrast, docking and biochemical assay data for zafirlukast analogues did not yield similarly conclusive results. Despite these limitations, this project successfully enriched the project library with valuable SAR data pertaining to KRAS-targeted PPIIs.

The second project leveraged both the SAR data from the project library and curated data from the ChEMBL database to construct a quantitative structure-activity relationship (QSAR) model of high predictive accuracy. Advanced machine learning methodologies, including nested cross-validation and mutual information-based feature selection, were employed to optimise model performance.

This model was subsequently applied to predict IC_{50} values for over seven million compounds sourced from ten structurally diverse or PPII-focused virtual libraries. Additionally, two comprehensive *in silico* libraries comprising (click) cyclic tetrapeptides

(cyctetpep) built from the twenty canonical amino acids were generated. Eleven scaffolds emerged as top candidates from the QSAR screening, with cyctetpep distinguished by their privileged three-dimensional shape, modular synthesis, and promising predicted inhibitory potency.

A robust synthetic route to access click-cyclised tetrapeptides was developed, culminating in the successful synthesis of three candidate derivatives. However, biochemical assays revealed an absence of meaningful KRAS inhibition, likely attributable to the limited solubility of the synthesised click cyctetpep.

Despite this outcome, the advantageous shape and modular synthesis of click cyctetpep suggest they remain attractive candidates for further optimisation. The iterative refinement of this KRAS PPII screening protocol through successive prediction-synthesis-evaluation cycles offers a promising strategy for the efficient discovery of potent KRAS inhibitors.

Table of Contents

1 Biazole and Zafirlukast SAR

1.1 Introduction

The small guanine nucleotide-binding protein KRAS was initially identified in the Kirsten rat sarcoma virus. It is now recognised as a central component of the signal transduction machinery in human cells. Approximately 14% of all cancers carry KRAS mutations, corresponding to an estimated 2.6 million new cases of KRAS mutant cancers globally each year. The development of the first clinically approved KRAS inhibitor, sotorasib (**1**), by the United States Food and Drug Administration (FDA) in 2018 marked the culmination of over five decades of research since the protein's discovery in 1967 (**Figure 1**).[1]

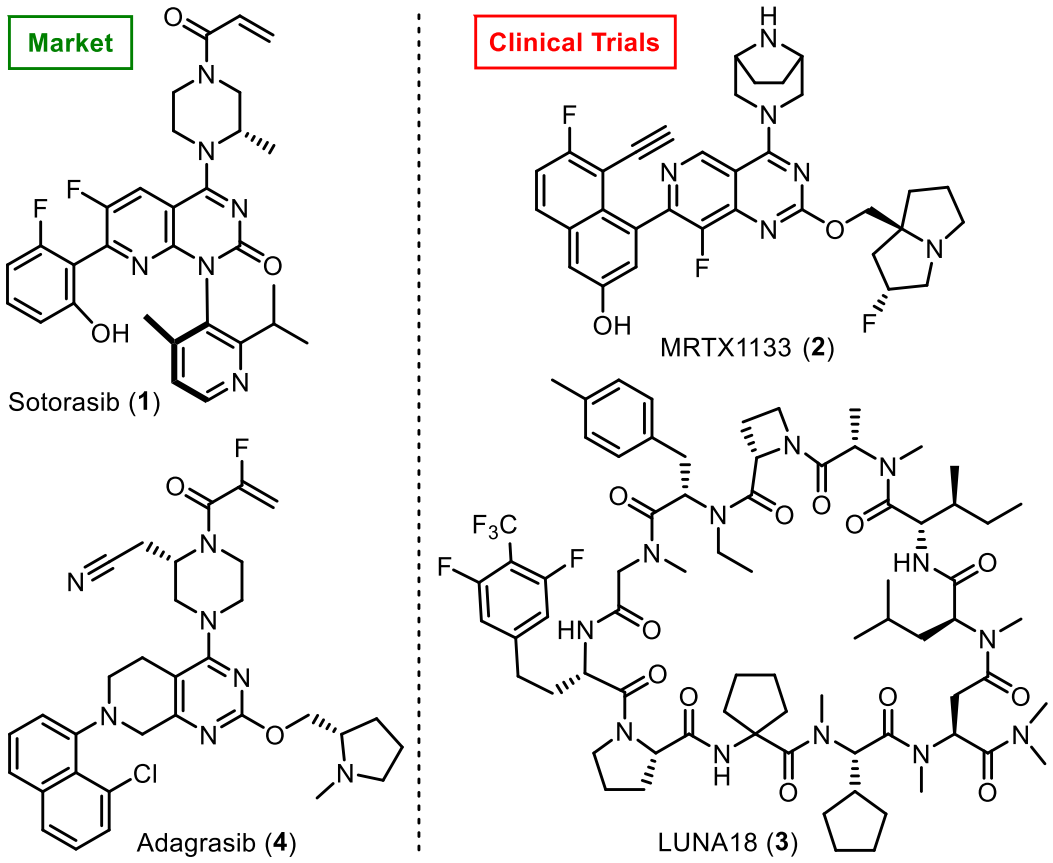


Figure 1: Four KRAS protein-protein interaction inhibitors. Only sotorasib and adagrasib have so far been approved by the FDA. MRTX-1133 and LUNA18 are currently in clinical trials.

Since the approval of sotorasib, numerous KRAS inhibitors have progressed into clinical trials, including MRTX-1133 (**2**) and LUNA18 (**3**). However, only one additional inhibitor, adagrasib (**4**), has received market authorisation to date. The demand for novel KRAS-targeting scaffolds therefore remains substantial.[2]

This introduction is structured to address the broad and complex subject of KRAS by dividing it into four key sections. First, KRAS will be contextualised within the framework of protein-protein interactions (PPIs). This will be followed by a description of its molecular structure and a mechanistic account of its role in cellular signalling pathways. Lastly, the structural characteristics and modes of action of KRAS-directed protein-protein interaction inhibitors (PPIIs) will be outlined. From this foundation, the necessity of developing further potent KRAS PPIIs and more effective screening pipelines for such scaffolds will emerge as a logical conclusion.

1.1.1 KRAS – A PPI Challenge

In recent decades, considerable efforts have been dedicated to sequencing the human genome. Approximately 19,370 protein-coding genes have been identified to date. Of these, around 18,000 (~90%) of these are predicted to encode functional proteins (**Figure 2**, blue/green).[3]

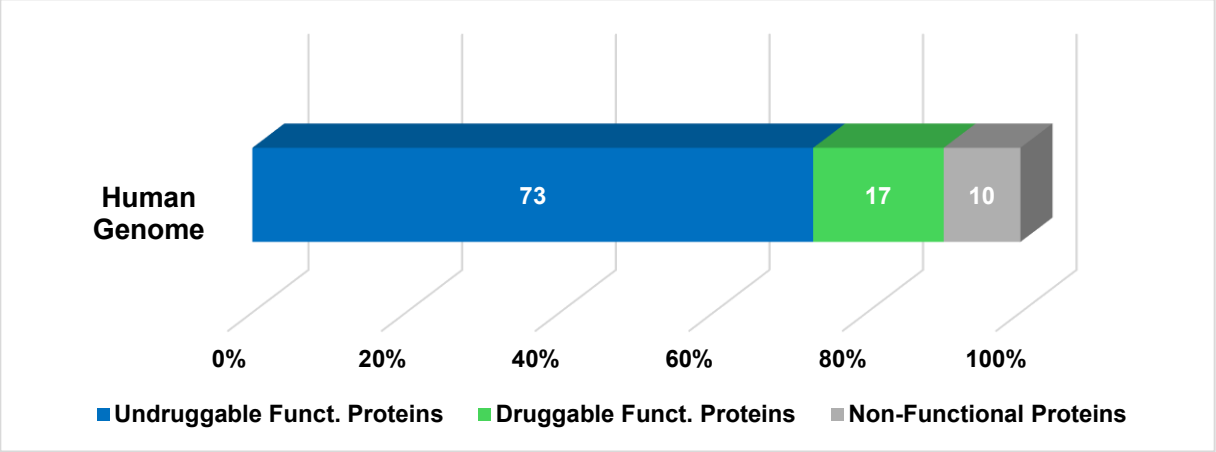


Figure 2: Percentages of the human genome currently termed (un)druggable and (non-)functional.

At present, only approximately 3,000 (~17%) of these proteins are estimated to be druggable, typically including classes such as enzymes. (Figure 2, green).[3] Proteins for which orthosteric inhibition has proven unsuccessful, such as KRAS, are frequently referred to as *undruggable* (blue). Nevertheless, recent advances have enabled the allosteric inhibition of KRAS, reclassifying it as a *druggable* and functional protein (blue to green). The following concepts are essential for understanding the challenges associated with this achievement. While the discussion is centred on KRAS, the principles are broadly applicable to the discovery PPIIs.

The first essential step is the elucidation of the target's interactome. The full spectrum of possible PPIs involving KRAS is still being mapped, although a substantial number of interaction partners have already been identified.[4–7] In parallel, hot spot amino acids at PPI interfaces, as well as potential allosteric sites, must be characterised. One approach entails the systematic mutation of individual residues followed by measurement of changes in PPI binding free energy. Although the experimental burden is considerable, the insights gained are of high value. The resulting spatial map of interaction interfaces and allosteric sites constitutes a critical resource for PPII discovery. Only recently has such an allosteric atlas been reported for KRAS (**Figure 16**).[8,9]

Second, proteins exhibit inherent conformational flexibility, contrary to the static representations implied by crystal structures. For $KRAS^{GTP}$ alone, two distinct conformations have been identified.[10] When complexed with various PPI partners, a broad spectrum of conformational states can arise (**Figure 3**, green/blue).

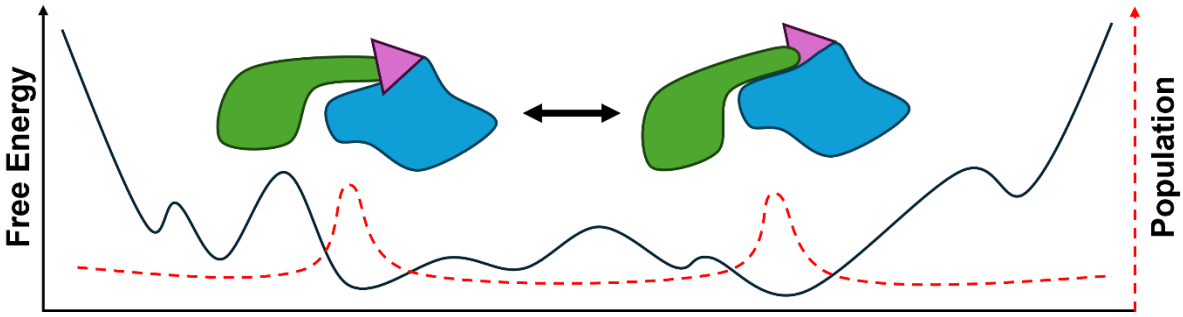


Figure 3: Schematic free energy and population distributions of a protein-protein complex ensemble. Proteins (blue and green) with PPII (magenta). Modified from KESKIN *et al*.[9]

Protein-protein interaction inhibitors (magenta) are employed to modulate the conformational distribution of such protein-protein complexes. Unlike orthosteric enzyme inhibitors, which typically target a single active site, PPIIs must contend with numerous potential allosteric binding sites, significantly increasing the complexity of the task.[9] Upon binding, PPIIs induce a shift in the conformational ensemble of the target complex. Notably, the conformation stabilised by a given PPII may be only marginally populated, or even virtually absent, in the unbound state.[9] In this manner, ligand binding can effectively give rise to otherwise absent binding pockets.[11]

Accounting for this dynamic behaviour has proven particularly useful in the context of KRAS-targeted PPII discovery. FENG *et al*. employed molecular dynamics simulations (MDS) to evaluate the persistence of pocket three in $KRAS^{G12D}$ in solution over a 200 ns timescale (**Figure 16**). The insights obtained enabled the rational design of a PPII with micromolar affinity for $KRAS^{G12D}$.[12] Despite such progress, efficiently modulating the conformational equilibrium of specific pathogenic mutants remains a major obstacle in PPII discovery.

Third, PPIIs generally exhibit structural characteristics that distinguish them from orthosteric inhibitors. While the latter typically bind within well-defined, deep active site pockets, PPIIs often interact with broad and shallow surface regions on target proteins.[13] As a result, PPIIs tend to display increased molecular weight (MW), greater topological polar surface area, and a more pronounced T-shaped geometry.[14] In analogy to Lipinski's Rule of Five (RO5), a corresponding set of guidelines has been proposed for effective PPIIs, referred to as the Rule of Four (RO4) (**Table 1**).[15] In order to facilitate the exploration of this distinct chemical space, dedicated screening libraries have been developed.[16–20]

	MW [Da]	**Log*P***	**HBA**	**HBD**	**Ring count**
RO5	≤ 500	≤ 5	≤ 10	≤ 5	-
RO4	> 400	> 4	> 4	-	> 4

Table 1: Overview of the RO5 and RO4.[15,21]

Macrocyclic scaffolds, i.e. cyclic peptides, have emerged as particularly promising scaffolds for the development of PPIIs.[22] Among these, the $KRAS^{G12C/D/V}$ PPII LUNA18 with nanomolar activity has entered clinical evaluation (**Figure 1**, compound **3**).[23] Notably, LUNA18 demonstrates unusually high oral bioavailability for a macrocyclic compound. Typically, macrocyclic inhibitors exhibit limited bioavailability and cellular permeability due to their large size (MW > 500 Da).[22] Current research efforts seek to overcome these limitations through the design and synthesis of smaller, strained macrocycles such as cyclic tetrapeptides.[24] In the second project of this thesis, a modular synthesis route towards a series of click-cyclised tetrapeptides was successfully established.

1.1.2 Primary Structure of RAS

Binary molecular switches are an integral part of the biochemical self-regulation of cells. The RAS superfamily of GTPases is a prime example of such binary molecular switches. In their 'on'-state they are bound to guanosine-5'-triphosphate (GTP) (RAS^{GTP}) and interact with proteins involved in cellular self-regulation. Ras proteins switch 'off' by hydrolysing GTP to GDP (RAS^{GDP}).[25] The regions responsible for the RAS GTPase activity are highly conserved among the RAS superfamily (**Figure 4**).

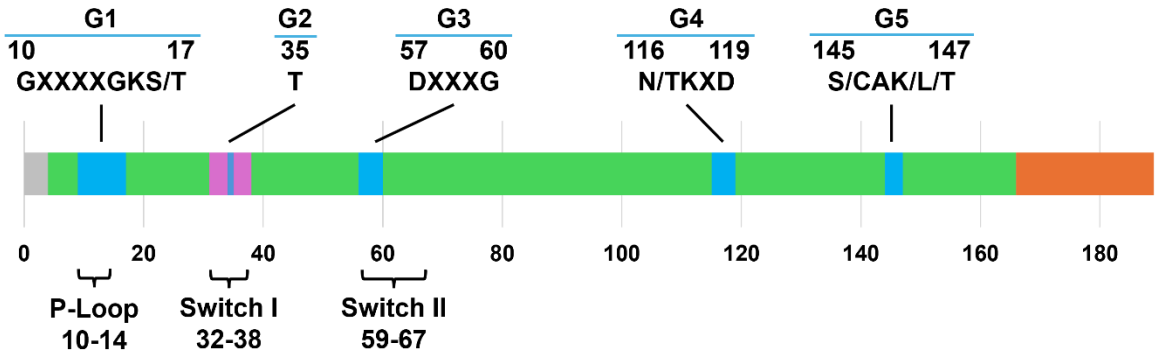


Figure 4: Domains of the primary structure of RAS GTPases. They consist of a G domain (5-166, green) and a hyper variable region (HVR, 167-188/189, orange). The former contains G motifs (blue) and a core effector domain (32-40, magenta). Modified from WENNERBERG, ROSSMAN and DER.[26]

Figure 4 depicts a schematic representation of the average primary structure of the RAS superfamily, comprising the RAS, RHO, RAB, RAN and ARF subfamilies. The G domain (**Figure 4**, green) contains highly conserved G motifs (blue), while less conserved AAs are marked as 'X'. The G motifs are responsible for binding of GDP/GTP and Mg^{2+} as well as for catalysing GTP hydrolysis. The switch regions and the P-loop lack a fixed secondary structure and undergo substantial conformational changes depending on the bound nucleotide and PPIs. Only RAS^{GTP} interacts with effector proteins *via* the switch regions, thereby initiating downstream signalling. The core effector domain (magenta) contains some of the crucial AAs for this interaction. Binding to the plasma membrane is also critical for the RAS life cycle. The less conserved hyper variable region (HVR, 167-188/189, orange) serves as a membrane anchor following posttranslational farnesylation and/or palmitoylation.[26–28]

The RAS subfamily consists of three isoforms primarily found in mammals, i.e. KRAS, NRAS and HRAS. Two different splice variants exist of KRAS, i.e. KRAS4A and KRAS4B. These splice variants differ in the final 15 residues of the G domain and the HVR, but both variants have been implicated in various cancers.

KRAS4B, in particular, displays higher expression levels in humans and has thus become the primary focus of KRAS-related research in recent decades.[29] This thesis focuses on the development of PPIIs for KRAS4B. Henceforth, KRAS will be used synonymously with KRAS4B.

1.1.3 Secondary and Tertiary Structure of RAS

The secondary structure of KRAS comprises five α helices and six β sheets. The positions of the secondary structural elements along the AA sequence of KRAS are depicted in **Figure 5**.

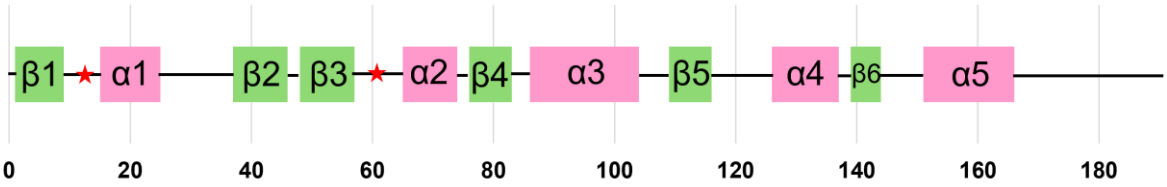


Figure 5: Secondary structure elements of KRAS along its primary structure. The red stars mark Gly12, Gly13 and Gln61. Modified from NUSSINOV, TSAI and JANG.[30]

The α helices are composed of the following AAs: 16–25, 66–74, 87–104, 127–137, 155–166. The β sheets are formed by the AA sequences: 2–9, 38–47, 49–57, 77–83, 110–116, 140–144. Disordered regions, represented by black lines, include the switch regions and the P-loop, as shown in **Figure 4**.The red stars in **Figure 5** highlight Gly12, Gly13 (P-loop) and Gln61 (switch I), which are frequently mutated in RAS-driven cancers.[30] The corresponding tertiary structure of KRAS is depicted in **Figure 6**.

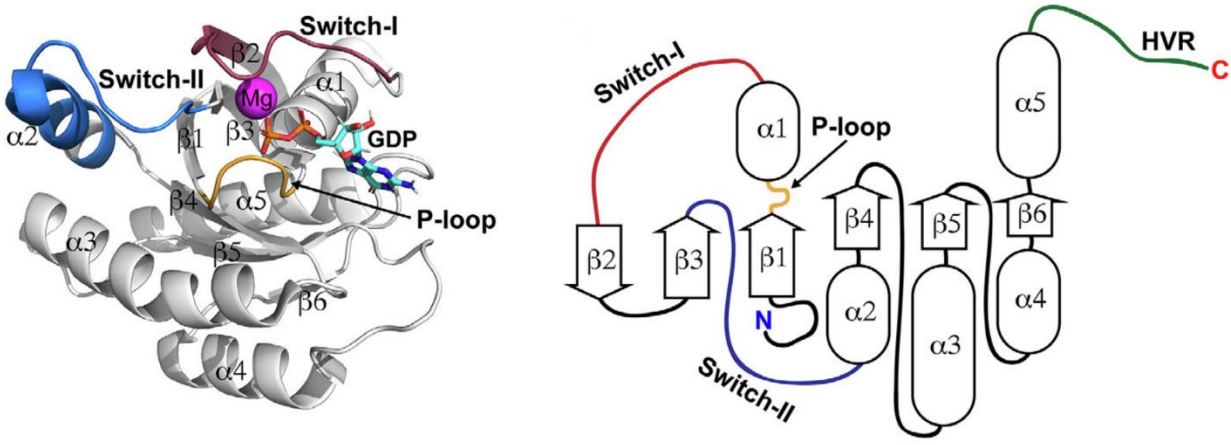


Figure 6: Crystal structure of KRAS4B$^{WT, GDP}$ (PDB ID: 4OBE, left) with schematic representation of its tertiary structure (right). Reprinted from PANTSAR.[28]

The cofactor Mg^{2+} is essential for the GTPase activity of RAS. It plays a dual role: firstly, it is crucial for nucleotide binding, where a network of electrostatic interactions exists between Mg^{2+}, GDP/GTP, and RAS.[31] HALL and SELF observed a 10-fold increase in exchange rate for GDP in $NRAS^{GDP}$ upon the addition of ethylenediaminetetraacetate (EDTA), which complexes Mg^{2+} and destabilises the RAS-nucleotide complex.[32] Secondly, Mg^{2+} is involved in catalysing GTP hydrolysis by stabilising the negative partial charges generated during nucleophilic attack by water on the γ-phosphate of GTP. GTP hydrolysis is particularly influenced by switch II and the P-loop of RAS.[33] Mutations at Gly12, Gly13, and Gln61 impair the ability of RAS to hydrolyse GTP to GDP, leading to the persistent activation of RAS^{GTP}. In this active state, effector proteins can bind to switch I, resulting in uncontrolled cell proliferation, differentiation, and migration, which underlie the pathogenesis of various cancers.[34]

1.1.4 KRAS in Cancer

The three RAS isoforms HRAS, NRAS and KRAS dominate the literature about RAS-driven cancers.[34–36] According to estimates from the year 2020, approximately 19% of cancer patients harbour a mutation in one of these three RAS isoforms with KRAS being most frequently mutated (75%), followed by NRAS (17%) and HRAS (7%). The distribution of these mutations is shown in **Figure 7**.[37]

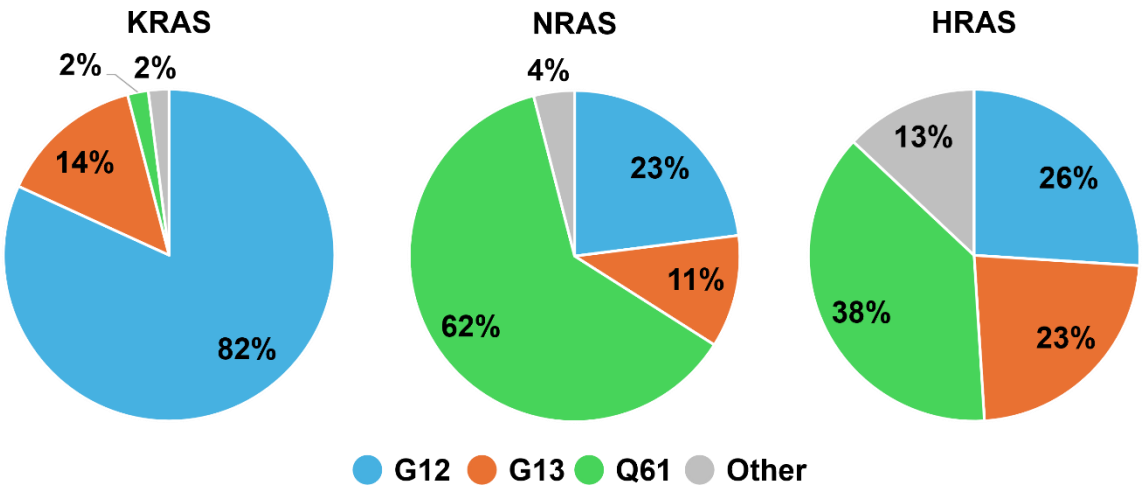


Figure 7: Portions of RAS mutants in human cancers in COSMIC database. Modified from PRIOR, HOOD and HARTLEY.[37]

About 70% of RAS-mutated cancers feature one of five mutations, i.e. G12D, G12V, G12C, G13D and Q61R. Each mutation, depending on the isoform, results in slight variations in GTPase activity and PPI affinity.[37,38] This thesis focusses on $KRAS^{G12D/V}$.

In contrast to $KRAS^{G12C}$, $KRAS^{G12D/V}$ mutants are more evenly distributed across colon, lung, bone marrow and pancreas tissues.[39] $KRAS^{G12D/V}$ mutations lead to reduced intrinsic GTPase activity and decreased GAP affinity. A detailed comparison of the differences in effector interaction between $KRAS^{G12D/V}$ and $KRAS^{WT}$ extends beyond the scope of this introduction.[40]

1.1.5 RAS Cascade

The interactome of KRAS is an extremely complex and intricate network of interacting biomolecules, which remains far from being elucidated entirely.[5,7] Nevertheless, researchers in the drug development field require a foundational understanding of the RAS cascade. One of the commonly encountered schematic depictions of the RAS cascade is shown in **Figure 8**.

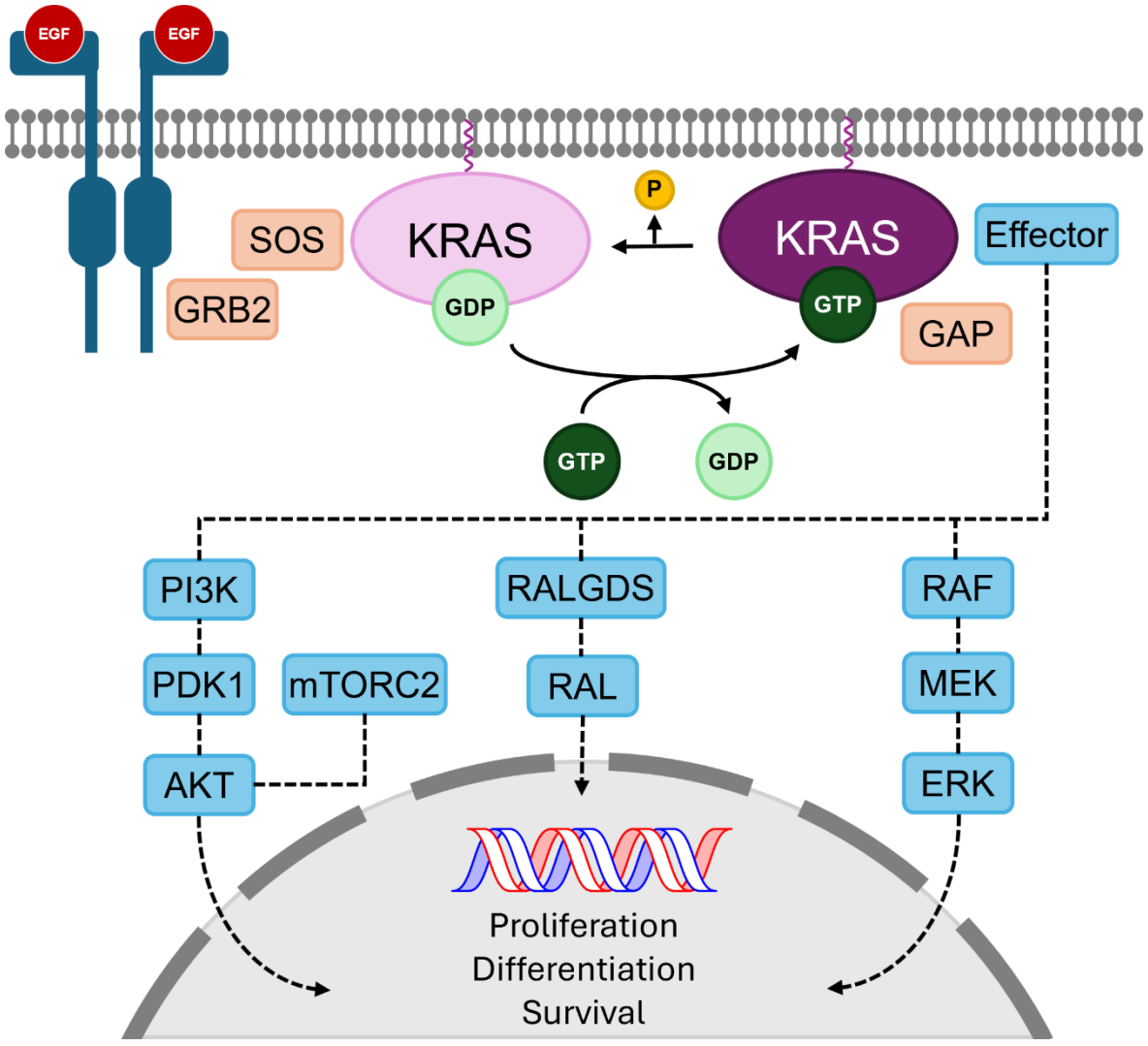


Figure 8: Schematic overview of the oncogenic RAS signalling cascade. Magenta: KRAS, green: nucleotide, yellow: phosphate, orange: GEFs and GAPs, blue: effector proteins, grey: cell membrane and nucleus. Modified from NUSSINOV and JANG, as well as HUANG *et al.*[34,41]

Numerous comprehensive reviews have been published on the RAS cascade.[34,41–43] In this context, particular attention is given to the three best-characterised downstream pathways: phosphoinositide 3-kinase (PI3K), mitogen-activated protein kinase (MAPK), and RAS-like GTPase (RAL). Additionally, the RAS-SOS interaction is of particular relevance to this thesis, as it constitutes the central focus of the drug development efforts described herein.

1.1.6 SOS Activates RAS

The RAS signalling cascade commences at the membrane, where RAS is localised through its HVR (**Figure 4** and **Figure 6**). The HVR consists of ~20 AAs and at least one farnesyl residue covalently linked through a thioester bond to the Cys closest to the C-terminus, serving as a membrane anchor. Depending on the specific RAS isoform, additional lipid modifications may occur, such as palmitoylation, farnesylation, or geranylation.[30] This membrane anchoring results in a high local concentration of RAS, thereby promoting dimerization and oligomerisation of RAS proteins (**Figure 9**).[44]

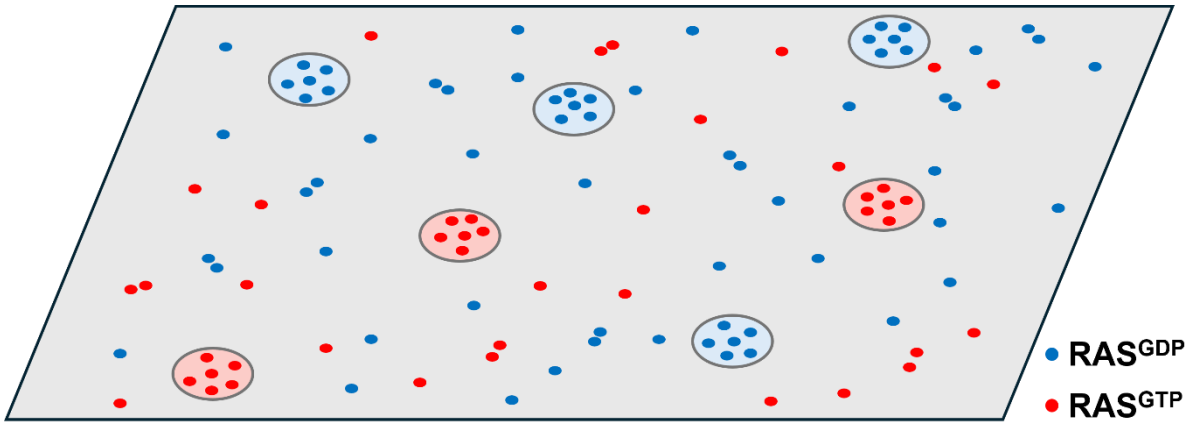


Figure 9: Spatial organisation of RAS on the plasma membrane. About 56% of RAS proteins exist as monomers. The diameter of RAS nanoclusters is about 18 nm. Modified from ZHOU *et al.*[44]

Approximately 56% of membrane-bound RAS proteins exist as monomers, while the remaining ~44% are organised into GDP/GTP-loaded oligomeric assemblies (blue/red), typically forming either dimers or nanoclusters comprising five to six RAS molecules. The RAS^{GTP} oligomers serve as key intermediates in signal transduction, rendering the abundance of such clusters on the plasma membrane a critical determinant of signalling intensity.[44]

For RAS to participate in downstream signalling, it must adopt the active, GTP-bound conformation. This activation is mediated by guanine nucleotide exchange factors

(GEFs), such as Son of Sevenless (SOS). Two orthologues, SOS1 and SOS2, have been identified, with SOS1 exhibiting greater physiological and pathological relevance and henceforth being referred to as SOS.[45] SOS is recruited to the plasma membrane *via* interaction with the epidermal growth factor receptor (EGFR), facilitated by adaptor proteins such as Grb2, which serve as molecular linkers between EGFR and SOS (**Figure 8**).[46] This recruitment increases the local concentration of both RAS and SOS, leading to an estimated 1,000-fold enhancement in their association rate relative to that observed in the cytosol.[47] Upon formation of the RAS:SOS complex, SOS induces opening of the nucleotide-binding cleft of RAS. This is achieved through insertion of a helical hairpin (αH) of SOS, which disrupts the electrostatic interaction network stabilising GDP and Mg^{2+} within the RAS active site **(Figure 10**).[48]

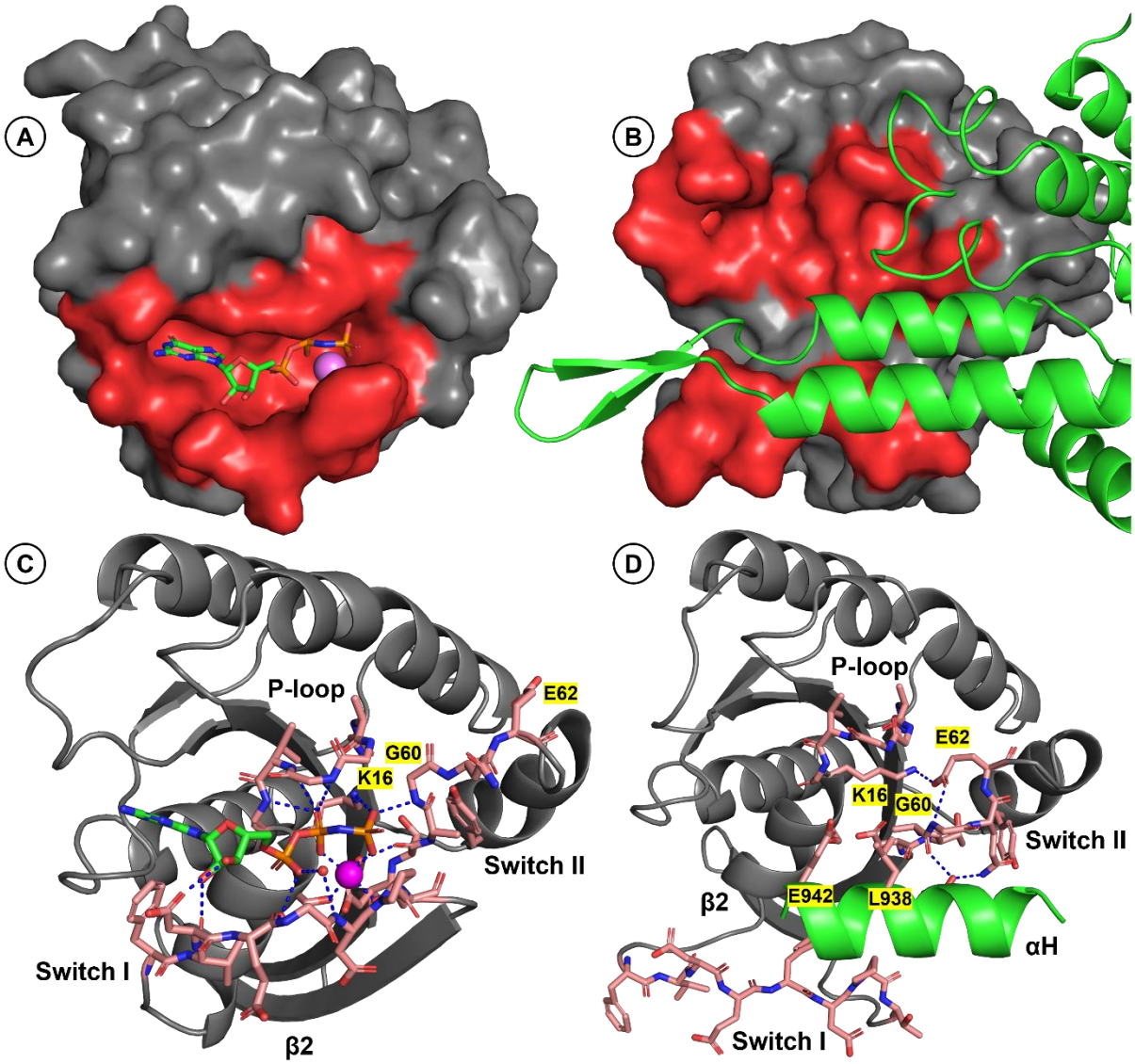


Figure 10: A) $HRAS^{GTP}$ with Mg^{2+} as magenta sphere and nucleotide binding pocket coloured in red. B) $HRAS^{cat}$:SOS complex with former binding site coloured in red. C) Interaction network in $HRAS^{GTP}$. D) Interaction network in $HRAS^{cat}$:SOS complex. Modified from BORIACK-SJODIN *et al*.[48]

Figure 10 shows a comparison between $HRAS^{GTP}$ (A) and $HRAS^{cat}$ complexed at the catalytic site of SOS (B). GTP and Mg^{2+} are shown as green sticks and a magenta sphere, respectively. The nucleotide binding site is highlighted in red to allow visual comparison.[48] Not shown in B and D is the second HRAS protein bound to the allosteric site of SOS. The allosteric RAS^{allo} is essential for the catalytic function of SOS. Recent findings from our research group suggest that the GEF activity of SOS can be inhibited by binding of betulinic acid monophthalates at the $KRAS^{allo}$:SOS interface.[49] The lower half of **Figure 10** illustrates the changes in the HRAS interaction network upon SOS binding. Switch I is displaced by αH, which inserts into the nucleotide-binding cleft. The two AAs Glu942 and Leu938 of αH occupy the positions usually held by the nucleotide α-phosphate and Mg^{2+}, respectively. The backbone of switch II also undergoes significant conformational rearrangement. As a result, Glu62 coordinates with Gly60 and Lys16 on HRAS, two AAs that normally form key interactions with the nucleotide in $HRAS^{GTP}$. In the $HRAS^{cat}$:SOS complex, the interaction network involving GTP/Mg^{2+} and the switch regions is disrupted.[48] This accelerates the GDP release by a factor of 10^5.[50] The drastic rate enhancement results from the immense affinity of RAS for GDP/GTP in the absence of GEFs. For $HRAS^{WT}$ at 4°C in the presence of Mg^{2+}, the dissociation constant (K_D) of and GDP/GTP was reported as ~10^{-11} M.[51] For comparison, one of the strongest, non-covalent interactions between streptavidin and biotin exhibits a K_D of $4x10^{-14}$ M at pH 7 and 25°C.[52] Only three orders of magnitude separate these affinities. Following release of GDP and Mg^{2+}, a GTP:Mg^{2+} complex binds to $HRAS^{cat}$:SOS.[53] The bias towards GTP-arises from the cytosolic GTP/GDP ratio of approximately 10.[54] Upon complete exchange, active RAS^{GTP} is released from SOS. The pivotal role of SOS in RAS activation has made the RAS:SOS PPI a prime target in cancer research.[55–60]

1.1.7 GAPs Deactivate RAS

The counterpart to GEFs are GTPase activating proteins (GAPs), such as GAP-334. While GEFs promote the formation of RAS^{GTP}, GAPs catalyse the hydrolysis of RAS^{GTP} to RAS^{GDP}.[61] Notably, RAS GTPases are capable of slowly hydrolysing GTP in the absence of a GAP. JOHN *et al.* reported the intrinsic $HRAS^{GTP}$ hydrolysis rate at 37°C as 0.028 min^{-1}.[62] In the presence of a GAP at 25°C, this rate increases up to 19 s^{-1} by a factor of 10^5.[63] This considerable rate enhancement results from the introduction of an Arg residue into the nucleotide binding cleft (**Figure 11**).[61]

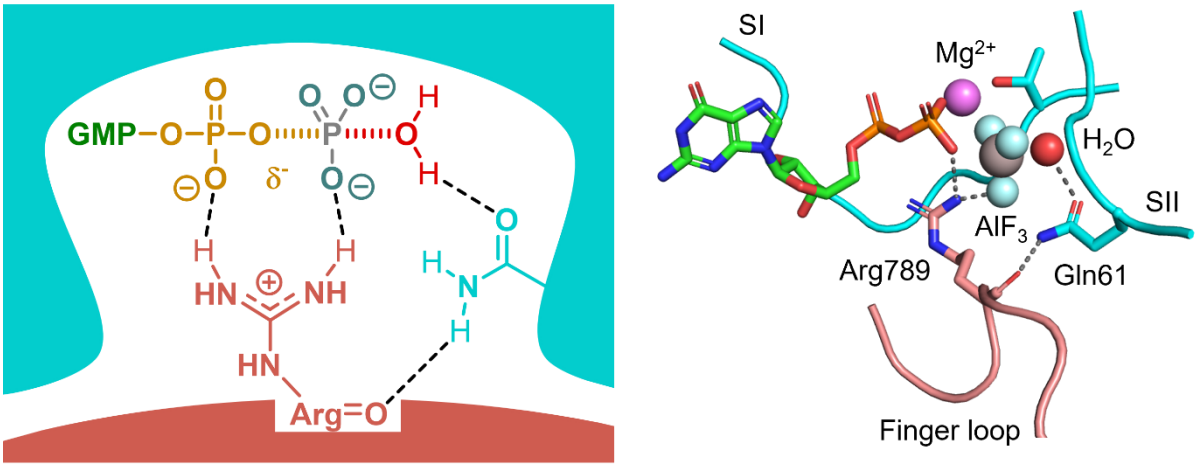


Figure 11: Structural basis for the GTPase rate enhancement in RAS:GAP complexes. Left: Schematic diagram of GAP action. Modified from BOS, REHMANN and WITTINGHOFER.[61] Right: Crystal structure of the active site in $HRAS^{GDP,\ AlF3}$:GAP-334 complex.[33] Colours: HRAS (cyan cartoon), GAP-334 (salmon cartoon), H_2O (red sphere), AlF3/γ-phosphate (grey and cyan spheres), Mg^{2+} (magenta sphere), guanosine (green sticks), phosphates (red and orange sticks).

SCHEFFZEK *et al.* crystallised $HRAS^{GDP}$ with GAP-334 and AlF_3 (**Figure 11**, right).[33] The aluminium fluoride is positioned roughly where the γ-phosphate would be in $HRAS^{GTP}$ (left). AlF_3 adopts a trigonal bipyramidal geometry with the terminal GDP oxygen and a H_2O oxygen, which occupies the same position as observed in the crystal structure of $HRAS^{GPPNP}$.[64] Therein, 5'-guanylyl imidodiphosphate (GPPNP) is a non-hydrolysable GTP analogue. Viewed as one molecule, GDP and AlF_3 serve as a transition state mimic of the GAP-induced GTP hydrolysis (left). The nucleophilic H_2O is held in place and activated for substitution by Gln61. This interaction is believed to be the structural basis for the oncogenicity of RAS^{Q61} mutations. Gln61 in RAS is fixed by an H-bond to Arg789 in GAP-334, which neutralises the partial negative charges that develop during nucleophilic attack of H_2O on the γ-phosphate.[33] Molecular dynamics simulations confirmed the aforementioned role of Gln61 and further demonstrated that common mutations at Gly12 and Gly13 in RAS impair the correct positioning of both Gln61 and Arg789.[65] Almost all RAS-driven cancers harbour mutations at positions 12, 13, or 61. The resulting reduction in both intrinsic and GAP-induced GTPase activity renders these RAS mutants constitutively active.[37]

1.1.8 Active RAS^{GTP} Binds to Downstream Effectors

As long as RAS^{GTP} is not hydrolysed to RAS^{GDP}, it engages specific signalling proteins, so-called effectors. Such effector proteins are involved in a multitude of biochemical processes, including cell proliferation, differentiation and survival. Over 50 potential RAS effectors have been identified. Among the best understood protein families are the rapidly accelerated fibrosarcoma kinase (RAF), PI3K and RAL. Generally, effectors bind to RAS through RAS binding domains (RBDs) (**Figure 12**).[66] In 2023, JUNK and KIEL have analysed the RBD:RAS binding interface in 54 crystal structures. They found that the RBD:RAS binding mode is highly conserved. The interfaces consist of the two switch regions (**Figure 12**, top right oval) and the β-sheets β2 on RAS and β2 on the RBD (left oval). Crucial AAs for binding include on RAS Ile36, Asp38 and Tyr40 (bottom right oval).[67] Nonetheless, the RAS-effector interaction is influenced not solely by RAS:RBD binding. Other interfaces, membrane association and clustering play an important part as well.[47,66]

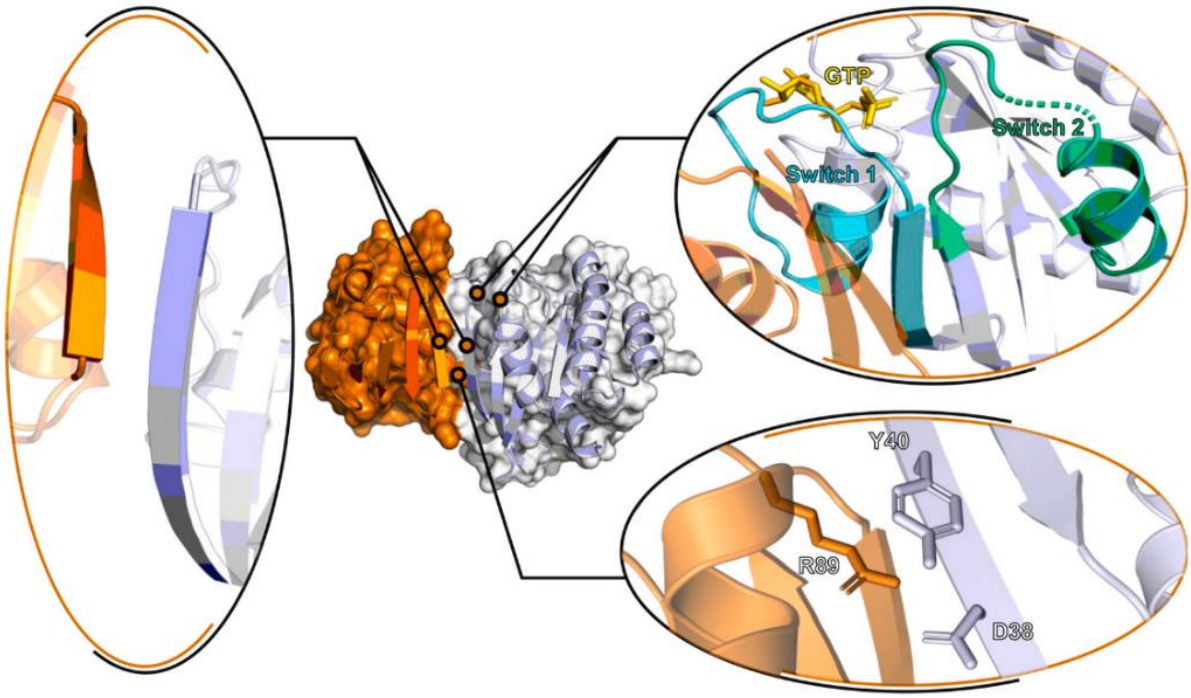


Figure 12: Highly conserved features of RBD:RAS interface (orange and grey, respectively) in 54 crystal structures. Reprinted from JUNK and KIEL.[67]

1.1.9 MAPK Pathway

Dimers of active RAS^{GTP} are the starting point of the MAPK signalling pathway. Dimerization and nanoclustering occur at the membrane (**Figure 9**). RAS^{GTP} dimers bind two RAF kinase proteins through their RAS binding domain (RBD) and cysteine-rich domain (CRD).[68] ARAF, BRAF and CRAF make up the RAF protein family.[42] **Figure 13** shows BRAF in complex with a KRAS^{GTP} dimer.

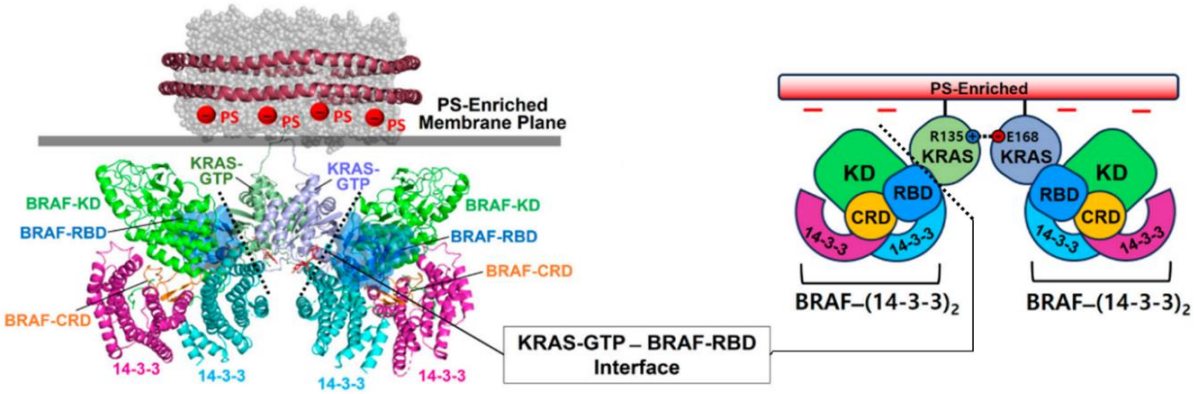


Figure 13: Left: NMR-derived KRAS^{GTP} dimer on PS enriched membrane superimposed with KRAS:BRAF-14-3-3 complex and RBD and CRD of BRAF. Right: Schematic view of the complex. Modified from LEE.[69]

LEE has recently studied the structural basis of the KRAS^{GTP}:BRAF complexation mechanism. In NMR studies he identified KRAS^{GTP} dimers on anionic membranes containing phosphatidyl serine lipids. KRAS^{GTP} was observed to dimerize through its α interface, consisting of the α4 and α5 helices (**Figure 6**). This allows binding of two BRAF proteins through their RBD and CRD, leading to their dimerization.[69] Only in such a dimerized state RAF can phosphorylate its downstream signalling partners, the mitogen activated protein kinase kinases MEK1 and MEK2, which phosphorylate the extracellular signal-regulated kinases ERK1 and ERK2.[68,70] In the context of KRAS PPII development, ERK1/2 phosphorylation levels serve as a readout for MAPK pathway inhibition. A potent RAS:RAF PPII reduces phosphorylated ERK (pERK) levels without affecting total ERK (tERK) levels, indicating specificity.[71–73] Phosphorylated ERK mediates the phosphorylation of more than 250 known target proteins, including nuclear transcription factors. As such, RAS-driven hyperactivation of MAPK signalling enhances the expression of cell proliferation regulators.[74] A detailed discussion of ERK-related PPIs falls outside the scope of this introduction and has been covered elsewhere.[42,75]

1.1.10 PI3K Pathway

The lipid kinase family PI3K is divided into three classes, of which Class I is the most extensively characterised and most directly associated with human cancers. Class I PI3Ks are multi-subunit enzymes composed of a catalytic subunit p110 (**Figure 14**, red and yellow), a membrane-binding domain C2 (blue) and a RBD (pink). In addition, they associate with a regulatory subunit, p85, which is not shown here.[76,77]

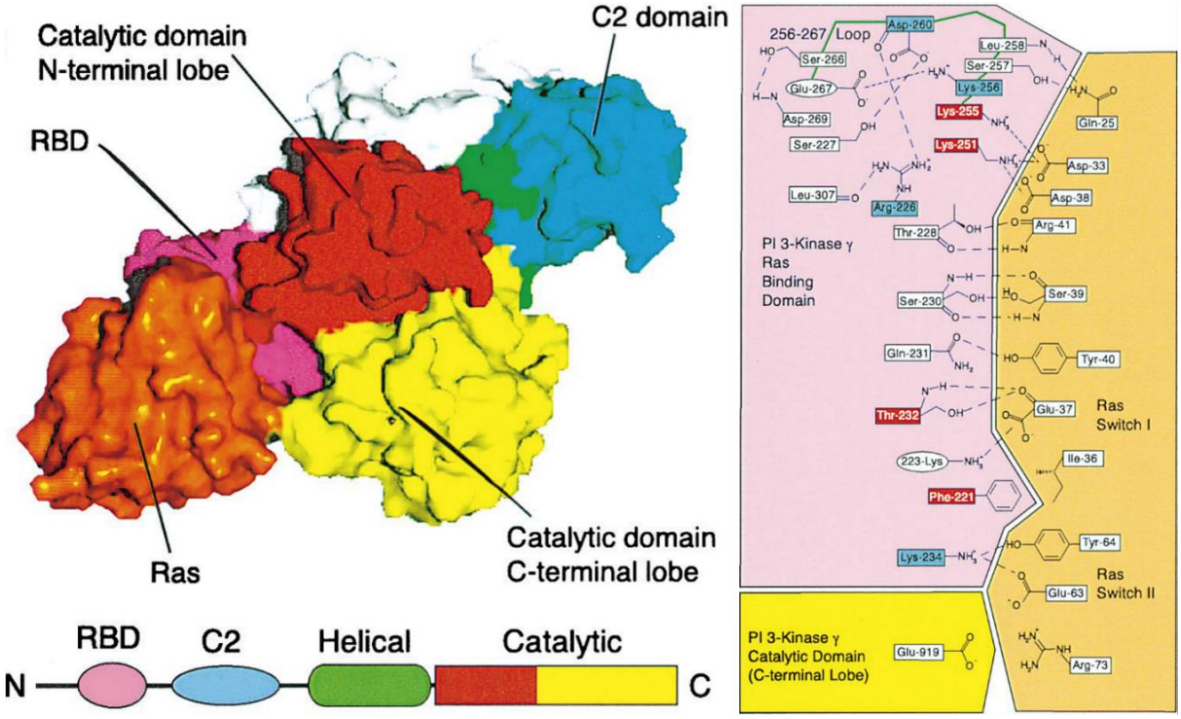


Figure 14: Crystal structure and interacting AAs in $HRAS^{GTP\gamma S}$:PI3Kγ complex. Reprinted from PACOLD *et al.*[76]

The presence of a dedicated membrane binding domain (blue) in PI3K underscores the importance of membrane association in Ras:PI3K complexation. At the membrane, PI3K binds to RAS^{GTP}. The right half of Figure 14 illustrates the key AA interactions. Switch I is the primary complex interface, while switch II of RAS and the catalytic domain of PI3K are involved as well. Only in active RAS^{GTP} are the switch regions appropriately arranged to enable RBD complexation. In the inactive state, the regulatory subunit p85 suppresses the kinase activity of the catalytic subunit p110. Upon RAS^{GTP} binding, p110 is activated and phosphorylates its substrate, phosphatidylinositol-4,5-bisphosphate (PIP_2), to generate phosphatidylinositol-3,4,5-trisphosphate (PIP_3). PIP_3 subsequently recruits 3-phosphoinositide-dependent protein kinase-1 (PDK1) to the plasma membrane, enabling phosphorylation of AKT by PDK1 (**Figure 8**). AKT is phosphorylated a second time by mammalian target of

rapamycin complex 2 (mTORC2).[77–79] Analogous to ERK, phosphorylated AKT (pAKT) regulates over 200 downstream targets, many of which are involved in promoting cell survival. Aberrant PI3K signalling driven by mutant RAS contributes to a wide spectrum of tumorigenic processes.[74] In this context, pAKT levels serve as a key readout for PI3K pathway activity, analogous to pERK in the MAPK pathway.[71,80] Total AKT (tAKT) and pAKT levels are commonly used to assess the efficacy of PI3K pathway inhibitors. A detailed discussion of AKT and its PPIs can be found in dedicated reviews.[77–79]

1.1.11 RAL Pathway

The RAL cascade commences with binding of RAS^{GTP} to the RBD of a RAS-like guanine nucleotide dissociation stimulator protein (RALGDS).[81] **Figure 15** shows the RBD of RALGDS (orange) in complex with $RAS^{E31K,\ GPPNP}$ (green). The AAs involved in the interaction are highlighted in grey, with their specific interactions detailed in the right half of **Figure 15**. The RAS^{E31K} mutant was chosen for crystallisation over RAS^{WT} due to its increased affinity for RALGDS-RBD; Lys31 forms one ionic interaction and two hydrogen bonds, strengthening the complex. A comparison between **Figure 12** and **Figure 15** underscores once more the conserved structural features of RAS:RBD binding across different effectors. The β2 sheets of RAS and RBD, along with switch I of RAS make up most of the interface. However, this structural similarity does not imply conservation of the specific AA composition at the interface. A comparison between **Figure 14** and **Figure 15** highlights these sequence-level differences. Thus, while the overall binding architecture is conserved, the unique residue compositions confer high specificity to each RAS:RBD interaction, supporting the rationale for designing effector-selective RAS PPI inhibitors (PPIIs).[67,81,82]

In the RAL signalling cascade, the primary role of RAS is to recruit RALGDS to the plasma membrane. This membrane translocation is essential for RALGDS activation and has been demonstrated using RALGDS mutants engineered with membrane anchors analogous to those of RAS. These membrane-anchored RALGDS variants can activate RAL independently of RAS. Once at the membrane, RALGDS acts as a guanine nucleotide exchange factor (GEF) for the RAL GTPases, RALA and RALB; two isoforms that, like RAS, are membrane-associated small GTPases. The catalytic domain of RALGDS promotes the exchange of GDP for GTP, generating the active RAL^{GTP} form.[81,83]

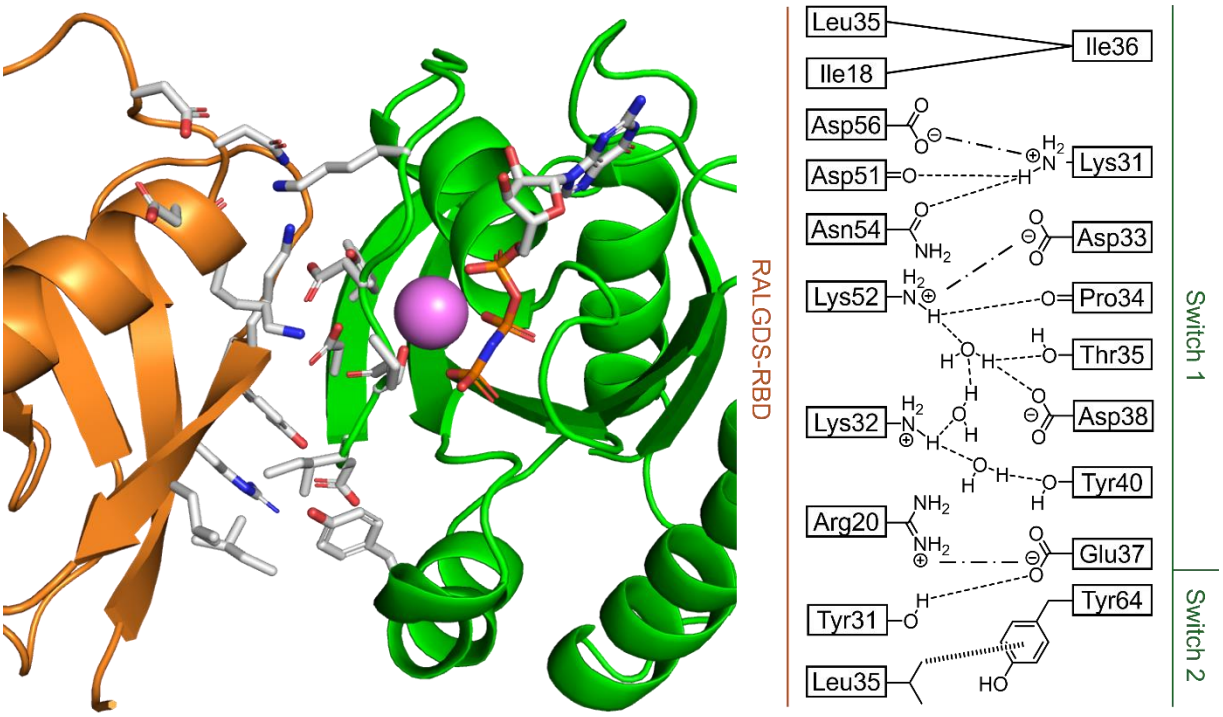


Figure 15: RALGDS-RBD (orange) in complex with RAS$^{E31K,\ GPPNP}$ (green) with interface AAs (grey). The interface AAs and their interactions are highlighted on the right. Hydrophobic interactions (solid line), ionic (long dash), H-bond (short dash) and π-interaction (broad dash). Modified from HUANG *et al.*[81]

The effectiveness of PPIIs targeting the RAL pathway can be assessed by quantifying RAL^{GTP} levels in treated *vs.* untreated cells.[84] Active RAL^{GTP} binds to a multitude of effectors that regulate endo-/exocytosis, gene expression and actin organisation. Additionally, there is evidence of crosstalk between the RAL, MAPK, and PI3K signalling pathways, suggesting that perturbation of one pathway may influence the other.[83,85,86]

1.1.12 Orthosteric RAS PPIIs

To date, no orthosteric protein–protein interaction inhibitors (PPIIs) targeting RAS have reached clinical approval. Their limited success can be attributed primarily to two factors. First, the intracellular concentration of GTP in mammalian cells is in the micromolar range, making it extremely difficult for non-covalent GTP analogues to compete effectively.[87] The inhibitor concentrations required to outcompete endogenous GTP are impractically high for therapeutic use. Second, RAS exhibits exceptionally high affinity and selectivity for GDP and GTP. Attempts to design covalent GDP/GTP analogues necessitate structural modification of the guanosine scaffold.

One such example is the covalent inhibitor SML-8-73-1 (**Figure 17**, compound **5**), which features a reactive, covalent warhead attached to the β-phosphate of GDP and is designed to bind covalently to Cys12 in $KRAS^{G12C}$. In a crystal structure of the $KRAS^{G12C}$:SML-8-73-1 conjugate, $KRAS^{G12C}$ was found to be in its inactive conformation and is unable to associate productively with its downstream effectors.[88,89] The conjugate forms through initial reversible binding of SML-8-73-1 at the nucleotide binding site of $KRAS^{G12C}$ followed by thioether linkage between Cys12 and the Michael acceptor moiety in SML-8-73-1. Unfortunately, the reversible binding affinity between $KRAS^{G12C}$ and GTP mimics similar to compound **5** was reported to be approximately 10^4 times lower than that of GDP/GTP. As a result, multiple association and disassociation events occur in competition with GDP/GTP before alkylation proceeds, resulting in ineffective inhibition.[90] These findings are representative for the limited overall efficacy of orthosteric covalent inhibition strategies targeting KRAS.

In contrast, allosteric KRAS PPIIs avoid direct competition with GDP/GTP, bypassing the affinity barrier. As a result, allosteric inhibition has emerged as the leading strategy for targeting KRAS therapeutically.

1.1.13 Allosteric RAS PPIIs

In 2023, WENG *et al*. have published a landmark study that systematically mapped the allosteric landscape of KRAS inhibition.[8] In their comprehensive approach, they engineered over 26,000 KRAS mutants and assessed their energetic differences regarding protein folding and binding to RAF1, PIK3CG, RALGDS, SOS1 and the designed ankyrin repeat proteins (DARPins) K27 and K55. They identified four distinct pockets with maximum allosteric inhibitory effect on all six binding partners. Their findings are in line with the sites previously proposed by GRANT *et al.* (**Figure 16**).[91]

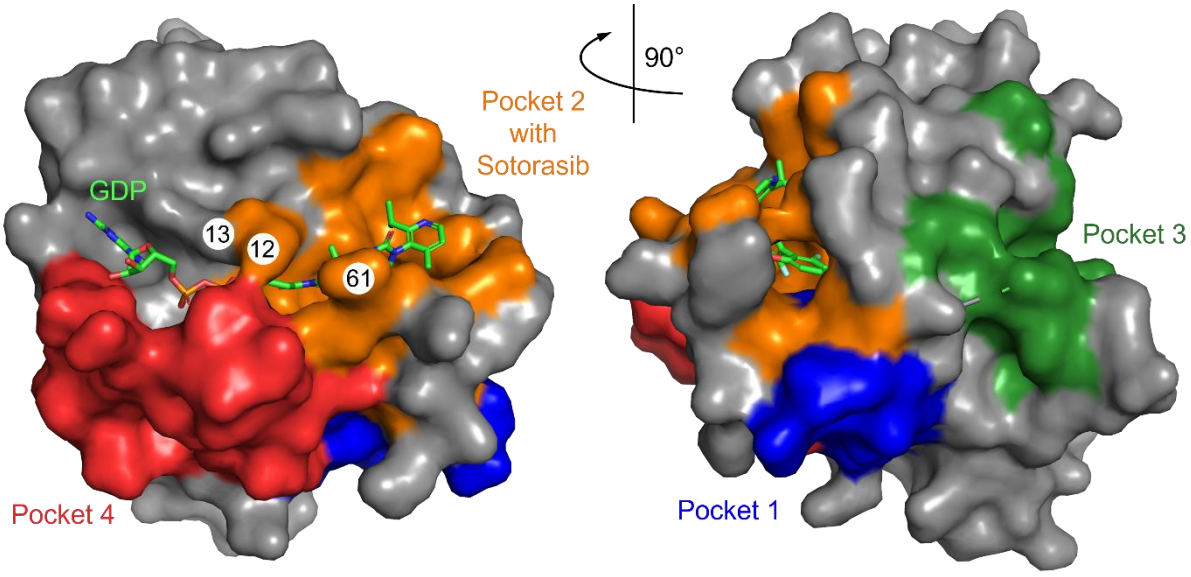


Figure 16: Allosteric binding pockets for PPIIs on KRAS. Created by superimposing 6OIM ($KRAS^{G12C,\ GDP}$:sotorasib) and 6VJJ ($KRAS^{WT,\ GPPNP}$:RAF1-RBD). RAF1-RBD (grey) is shown as reference. GDP and sotorasib are shown as green sticks. G12, G13 and Q61 are marked with numbers. Modified from WENG *et al*.[8]

Among the identified allosteric sites, Pockets 1 and 2 (blue and orange) have garnered the most attention in past drug development efforts. These sites are commonly referred to as the switch I/II pocket (P1) and the switch II pocket (P2). Notably, the first clinically approved covalent $KRAS^{G12C}$ inhibitor sotorasib binds at P2 (**Figure 17**, **1**).[92]

Cell-based assays demonstrated that sotorasib effectively inhibits ERK phosphorylation with IC_{50} = 68 nM. Its exceptional selectivity for $KRAS^{G12C}$ arises from its covalent mode of inhibition. Building on this success, Mirati Therapeutics developed MRTX-1133 (**2**), a non-covalent $KRAS^{G12D}$ inhibitor.[72] MRTX-1133 also binds at P2, forming a key H-bond between the NH of the bridged piperazine moiety and Asp12. This interaction confers 500-fold selectivity for $KRAS^{G12D}$ over $KRAS^{WT}$. In cellular assays, MRTX-1133 inhibits ERK phosphorylation with IC_{50} = 2 nM and also disrupts

the KRAS^{G12D}:SOS1 interaction. As of now, MRTX-1133 is undergoing clinical evaluation.[93] The KRAS^{G12X} mutant selectivity of sotorasib and MRTX-1133 is primarily due to the accessibility of AA12 from P2. Likewise, AA13 and AA61 are also exposed at P2, enabling selective inhibition of KRAS$^{G13X/Q61X}$ mutants. This accessibility and versatility make P2 a privileged site for drug targeting.[94]

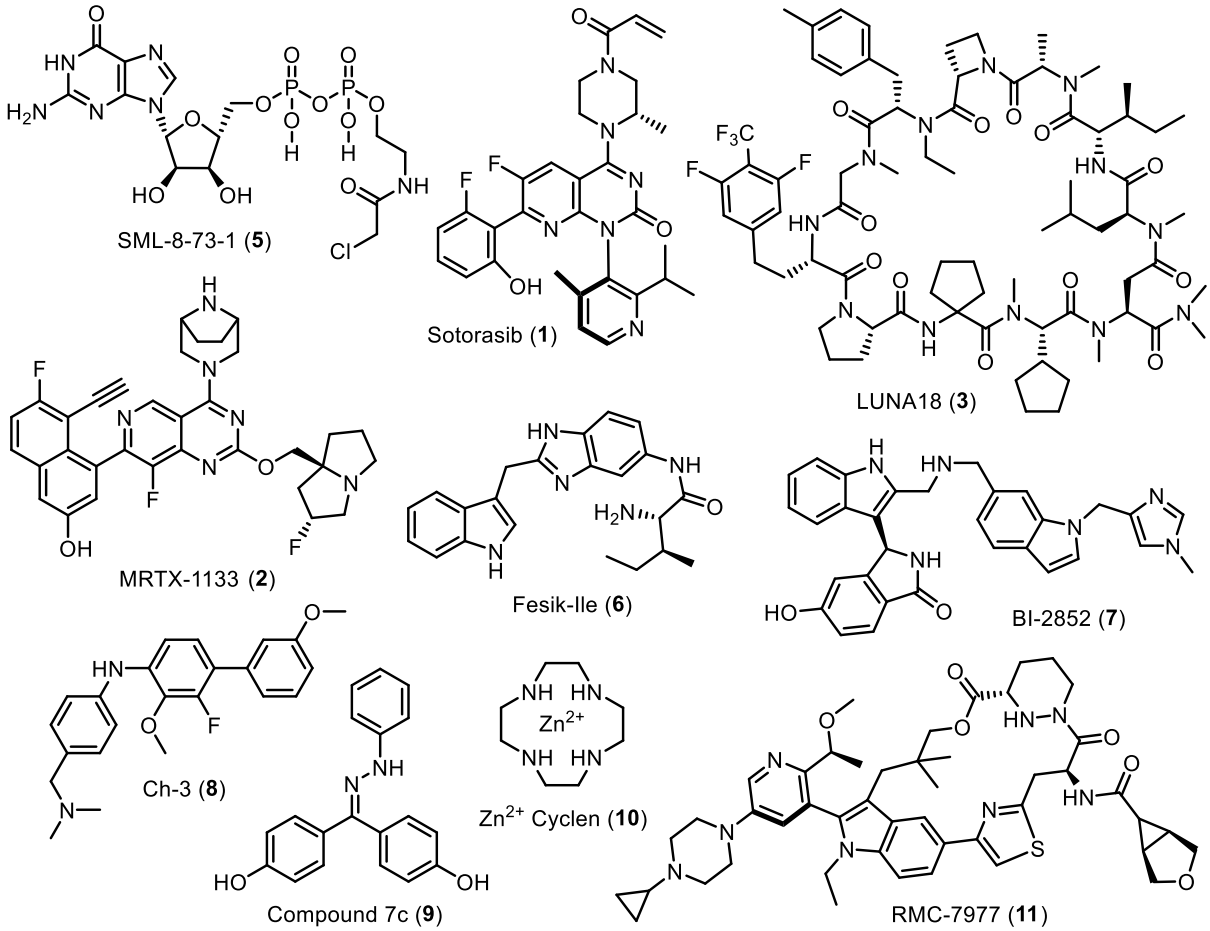


Figure 17: Structures of the orthosteric RAS PPII **5** and the allosteric RAS PPIIs **1**–**3** and **6**–**11**.

Direct interaction with AA12/13/61 is a key determinant for mutant-selective KRAS inhibition. However, potent inhibition can also be achieved in their absence. The cyclic peptide LUNA18 (**3**) binds at P2 without directly engaging AA12/13/61. LUNA18 potently inhibits the KRAS^{G12D}:SOS PPI with $IC_{50} < 2$ nM. Comparable inhibitory concentrations were observed in cellular assays for several KRAS$^{G12D/V/C}$ mutants. Due to its broad activity across KRAS mutants, LUNA18 is considered a promising candidate for pan-KRAS^{Mutant} inhibition However, its selectivity over KRAS^{WT} has not yet been reported. Notably, LUNA18 exhibits excellent oral bioavailability for a cyclic peptide, ranging from 21–47%, making it a strong candidate for clinical application.[23]

In contrast to P2, P1 is smaller in volume and highly conserved across all three RAS isoforms, irrespective of the bound nucleotide or interacting ligand. P1 exists constitutively in $KRAS^{GDP/GTP}$ conformations and is not induced by ligand binding. Its structural features, a shallow, lipophilic core surrounded by a hydrophilic rim, make it a privileged site for indole-containing compounds.[95] Using NMR-based fragment screening, the Fesik group identified Fesik-Ile (**6**) with moderate binding affinity for $KRAS^{G12D}$ with K_D = 190 µM and 78±8% inhibition of SOS-mediated nucleotide exchange at a concentration of 1 mM.[55]

Building on this hit, KESSLER *et al*. developed BI-2852 (**7**), a non-covalent PPII with the highest known affinity for P1 on $KRAS^{G12D, GCP}$, i.e. K_D = 750 nM (GCP is a non-hydrolysable GTP analogue).[73] Notably, BI-2852 exhibits ~10-fold selectivity for $KRAS^{G12D, GCP}$ over $KRAS^{WT, GCP}$ and binds to NRAS, HRAS and KRAS. Despite the high affinity of BI-2852, only moderate inhibition of ERK phosphorylation was observed in cell assays with EC_{50} = 6.7 µM.[95] Ultimately, the compound's limited cellular efficacy was considered insufficient to warrant further clinical development.

A P1 ligand with notable inhibitory potency in cell-based assays is Ch-3 (**8**), developed by CRUZ-MIGONI *et al*.[71] In $KRAS^{G13D}$ mutant cells, treatment with 20 µM of Ch-3 led to near-complete inhibition of both ERK and AKT phosphorylation. Additionally, Ch-3 disrupted the interaction between RALGDS and all three RAS isoforms. However, no data were reported regarding its mutant/WT selectivity.

Our own research group, in collaboration with the Stoll laboratory at the Ruhr University Bochum, has identified dihydroxybenzophenone phenylhydrazone (**9**) as a ligand for P1.[96,97] Compound **9** inhibits the SOS-mediated nucleotide exchange on $KRAS^{G12D}$ with IC_{50} = 413 µM and demonstrates ~2.5-fold selectivity for $KRAS^{G12D}$ over $KRAS^{WT}$. Despite this moderate mutant/WT selectivity, cell-based viability assays with $KRAS^{G12D/V/C/S}$ and $KRAS^{G13D}$-mutant cell lines all showed IC_{50} values in the range 17.6–33.7 µM, indicating a lack of mutant selectivity.

Overall, the relatively modest inhibitory potency of P1-targeting PPIIs compared to P2-targeting compounds like sotorasib and MRTX-1133 underscores why drug development efforts continue to focus predominantly on P2. Moreover, the potential for PPIIs binding at P2 to interact with AA12/12/61 enables exceptional mutant/WT selectivity, a critical criterion for clinical application.

Similar to P1, both P3 and P4 are positioned unfavourably with respect to the mutation hot spots in KRAS, reducing their potential for direct mutant-specific targeting. P3 is the most distal pocket relative to the nucleotide- and effector-binding sites, located near the C-terminus of RAS.[41,98] To date, limited drug discovery efforts have focused on this region. One of the few ligands reported to engage P3 is Zn^{2+} cyclen (**10**).[99] NMR studies revealed that the Zn^{2+} ion not only binds at P3 but also coordinates to the γ-phosphate of RAS^{GTP}, thereby enabling simultaneous interaction with AA12/13/61 of RAS. While Zn^{2+} cyclen interferes with effector binding, its relatively low binding affinity to RAS precludes it from therapeutic use in its current form, though it may serve as a lead for more potent derivatives.

Analogously, the P4 pocket has seen limited exploration in ligand development. A notable exception is the recently reported macrocycle RMC-7977 (**11**), which binds at the P4 region.[100] Macrocycle **11** first binds to cyclophilin A with K_D(CYPA) = 195 nM. The resulting binary complex then associates with RAS, forming a ternary assembly with $K_D(KRAS^{G12V})$ = 85 nM. Located at the RAS:CYPA interface, macrocycle **11** acts as a so-called *molecular glue*, stabilising the tri-complex and thereby inhibiting KRAS-effector PPIs. As a result, phosphorylation of ERK was suppressed with EC_{50} = 0.421 nM, exhibiting ~10-fold selectivity for $KRAS^{G12X}$ mutants over the WT. Remarkably, macrocycle **11** also enhances the intrinsic GTPase activity of KRAS, with pronounced selectivity for the $KRAS^{G12D}$. This dual mechanism, PPI inhibition and GTPase activation, renders RMC-7977 a highly promising candidate currently undergoing clinical evaluation.[101]

While the list above is not exhaustive, it includes several of the most advanced and promising KRAS protein-protein interaction inhibitors identified to date, notably compounds **1**, **2**, **3**, and **11**.

1.2 Aim

Over the course of the past decade, the Scherkenbeck group and the company Lead Discovery Center in Dortmund (LDC) have collected a project library of ~1,200 KRAS PPIIs. The syntheses of selected derivatives and their *in vitro* efficacy has been discussed in several publications.[49,96,97,102–104] The Scherkenbeck group has largely focused on the syntheses, while LDC has provided their assay expertise. Recently, a joint high throughput screening (HTS) of 250 000 compounds targeting the SOS-mediated nucleotide exchange on KRAS identified several hits, which inhibited the. Two hit structures were selected for synthesis and subsequent derivatisation in the context of this project, i.e. LDC151135 (**12**) and the known leukotriene receptor antagonist zafirlukast (**13**) (**Figure 18**).

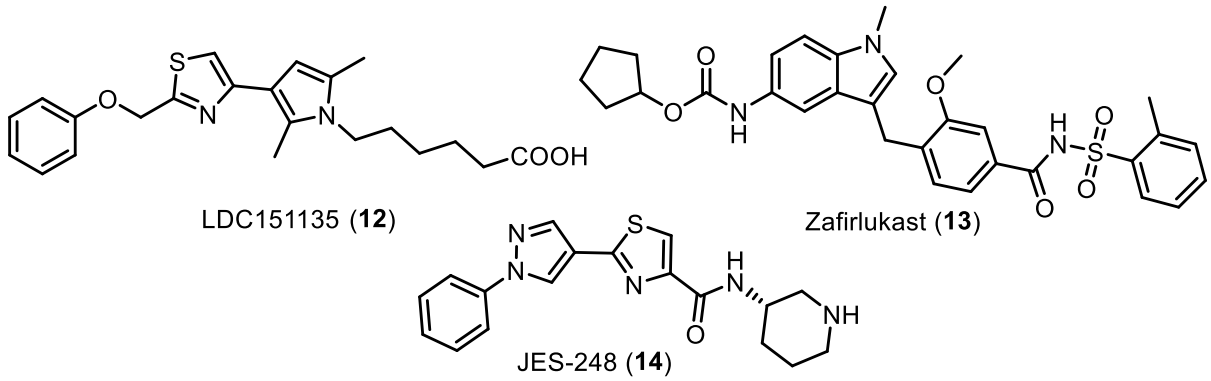


Figure 18: HTS hit **12** and **13** and *in silico* HTS hit **14** were selected as targets for this project.

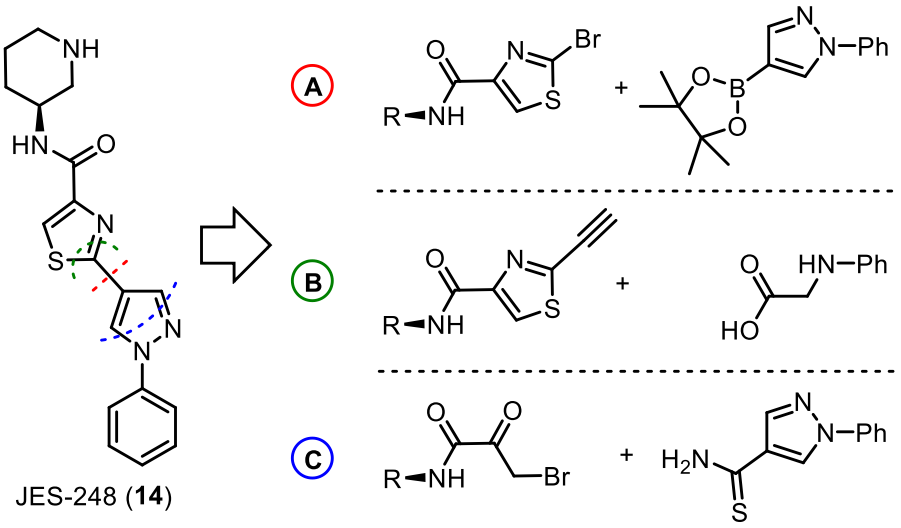
Furthermore, the hit compound JES-248 (**14**) was identified through an *in silico* HTS conducted by our former colleague Dr. Jeuken and was selected for synthesis in this project.[105] In this context, the objectives of this project are as follows:

1. Establish syntheses towards the biazoles **12** and **14**.
2. Synthesise derivatives of the zafirlukast scaffold to gain insights into their SAR
3. Study the mode of action of all derivatives through *in silico* modelling.
4. Draw comparisons between the assay and modelling results.
5. Enrich the project compound library with SAR data of the biazole and zafirlukast scaffolds. The resulting data serve as a foundation for the second project of this thesis.

1.3 Results and Discussion

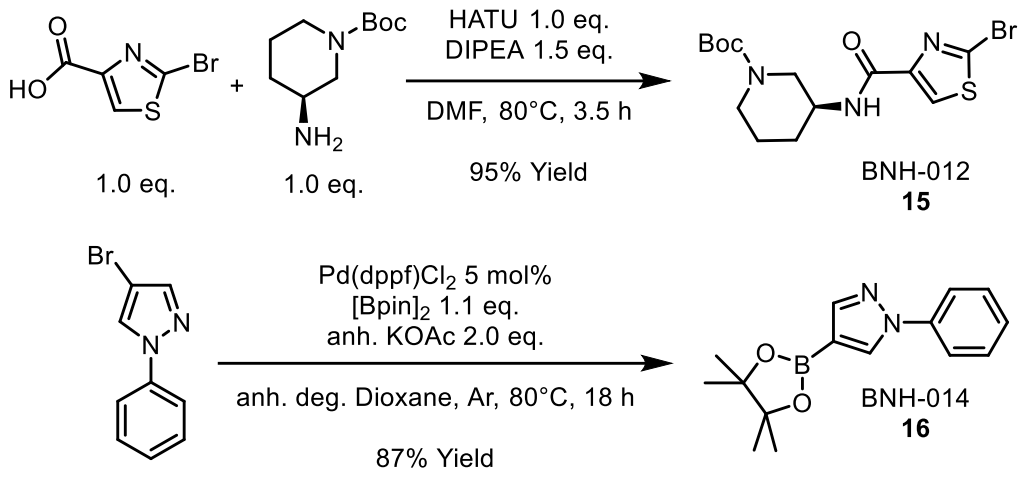
1.3.1 Synthesis of JES-248 (14)

Biazoles such as JES-248 and LDC151135 have garnered considerable attention in medicinal chemistry due to their structural resemblance to marine natural products that exhibit notable antimicrobial, antifungal, and antitumor activities.[106] Milligram-scale quantities of both hit compounds were initially procured and fully consumed during primary screening efforts. To enable further biological evaluation of these hits and related analogues, the development of suitable synthetic routes became necessary. Established strategies for biazole synthesis typically involve either the coupling of two heterocyclic precursors (**Scheme 1**, A) or the construction of the heterocyclic cores via cyclization reactions (B and C).



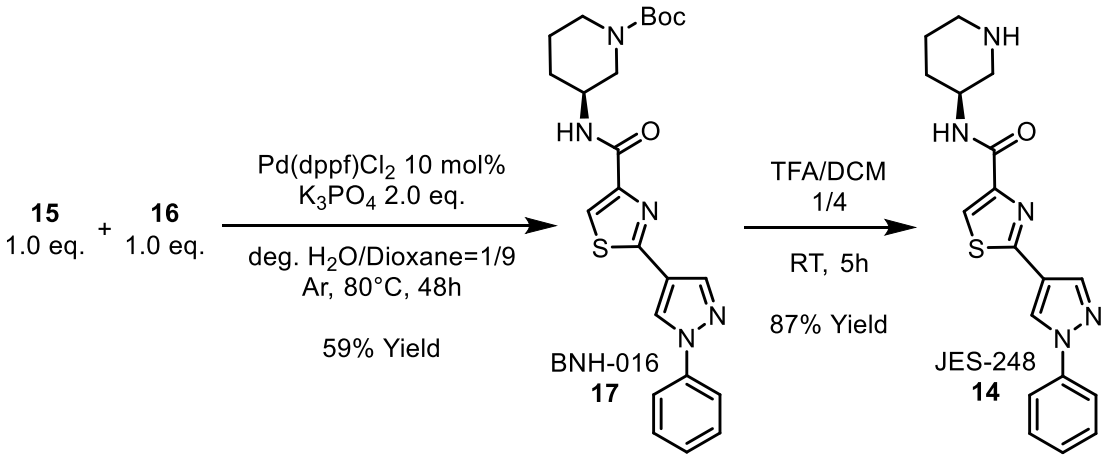
Scheme 1: Three initial retrosynthetic disconnections were considered for JES-248: A) coupling of two heterocyclic precursors,[107] B) pyrazole synthesis using thiazole acetylene and *N*-phenylglycine,[108] C) Hantzsch thiazole synthesis.[106]

For JES-248, neither retrosynthetic pathway B nor C were pursued, as both involve elaborate precursors and/or exotic reaction conditions.[106,108] Instead, pathway A was selected as the most viable route, owing to the ready availability of both coupling partners (**Scheme 2**). Bromide **15** was synthesised in excellent yields from 2-bromothiazole-4-carboxylic acid and (*S*)-*tert*-butyl 3-aminopiperidine-1-carboxylate under standard amide coupling conditions.[109] The pyrazole **16** was employed as the boronic ester coupling partner, as Suzuki-Miyaura couplings generally provide higher yields when thiazoles are used as halide-containing substrates rather than as boronic acid derivatives.[107]



Scheme 2: Synthesis of the precursors required for pathway A, i.e. bromide **15** and boronic ester **16**.

The pyrazole **16** was prepared *via* Miyaura borylation of 4-bromo-1-phenyl-1*H*-pyrazole. An initial borylation attempt with $Pd_2(dba)_3$ (10 mol%), $[Bpin]_2$ (1.2 eq.) and anhydrous KOAc (1.5 eq.) in dry, degassed dioxane failed to afford the desired product **16**. However, under identical conditions using only 5 mol% catalyst, $Pd(dppf)Cl_2$ successfully catalysed the formation of boronic ester **16**. The resulting coupling partners, **15** and **16**, were then subjected to Suzuki-Miyaura cross-coupling reactions (**Scheme 3**).



Scheme 3: Suzuki-Miyaura coupling of bromide **15** with boronic ester **16** to afford Boc-JES-248 (**17**), followed by Boc deprotection to yield the target compound **14**.

The reaction conditions tested for the Suzuki-Miyaura coupling reactions shown in **Scheme 3** are summarised in **Table 2**. The use of K_3PO_4 as base, in combination with 10% H_2O in the solvent mixture, resulted in a threefold increase in the yield of Boc-protected JES-248 (**17**). K_3PO_4 is commonly used in Suzuki-Miyaura couplings, as it

generates the requisite hydroxide anions in the presence of water, which are essential for the catalytic cycle.[110,111]

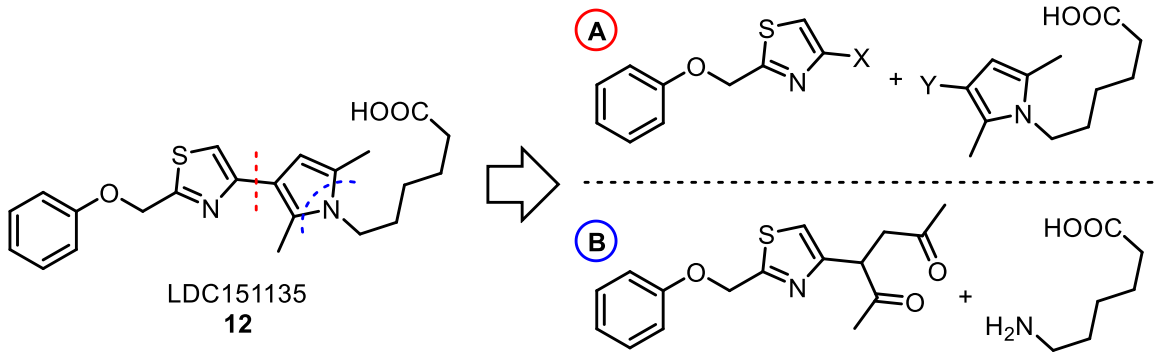
Catalyst	Base	Solvent	Yield
$Pd(dppf)Cl_2$ (10 mol%)	Cs_2CO_3 (3 eq.)	anh. deg. dioxane	19%
$Pd(OAc)_2$ (20 mol%) XPhos (20 mol%)	Cs_2CO_3 (3 eq.)	anh. deg. dioxane	20%
$Pd(dppf)Cl_2$ (5 mol%)	Cs_2CO_3 (2 eq.)	dist. deg. H_2O/Dioxane = 1/9	42%
$Pd(dppf)Cl_2$ (5 mol%)	K_3PO_4 (2 eq.)	dist. deg. H_2O/Dioxane = 1/9	59%

Table 2: Tested reaction parameters for the Suzuki-Miyaura coupling of bromide **15** with boronic ester **16** to synthesise Boc-JES-248 (**17**).

In the final step, the Boc protecting group of compound **17** was removed using 25% trifluoroacetic acid in dichloromethane, affording the free amine **14**. The overall yield of **14** was 48% over three steps starting from 2-bromothiazole-4-carboxylic acid.

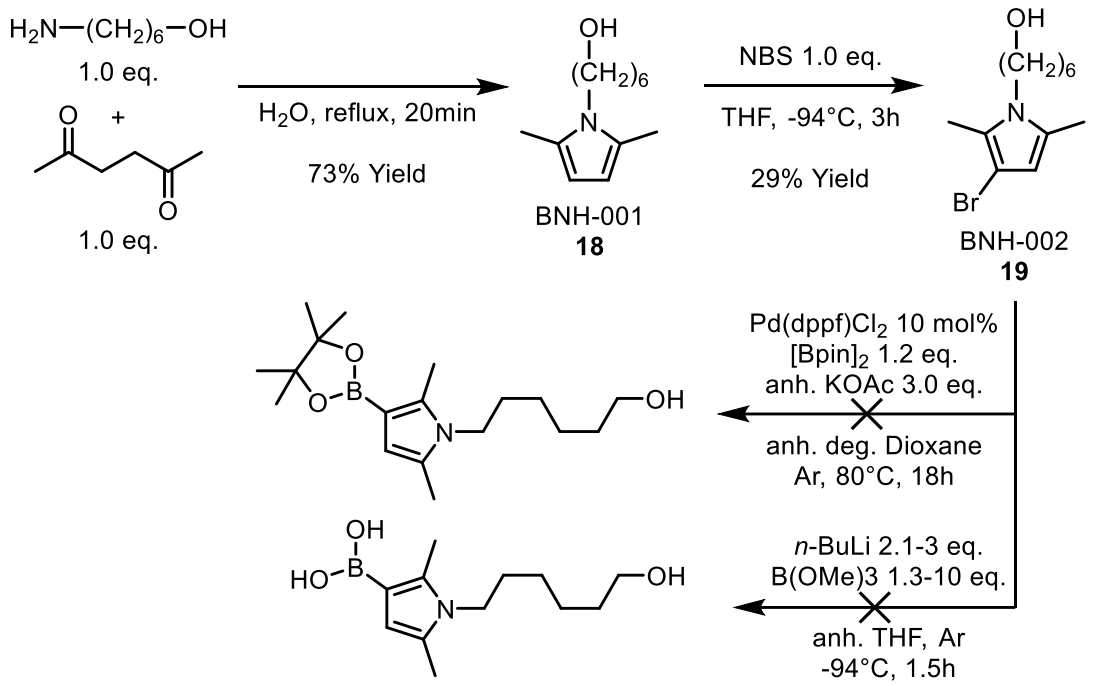
1.3.2 Attempted Syntheses Towards LDC151135 (12)

The synthesis of the target biazole LDC151135 proved substantially more challenging than that of JES-248. Two principal synthetic approaches were investigated: coupling of two heterocyclic fragments and the Paal-Knorr pyrrole synthesis (**Scheme 4**, A and B).



Scheme 4: Two retrosynthetic disconnections explored for LDC151135. A) coupling of two heterocycles and B) Paal-Knorr pyrrole synthesis.

For pathway A, the Bromide precursor **19** was obtained starting with a Paal-Knorr synthesis of the dimethyl pyrrole **18** from 2,5-hexanedione and 6-amino-1-hexanol, as shown in **Scheme 5**.

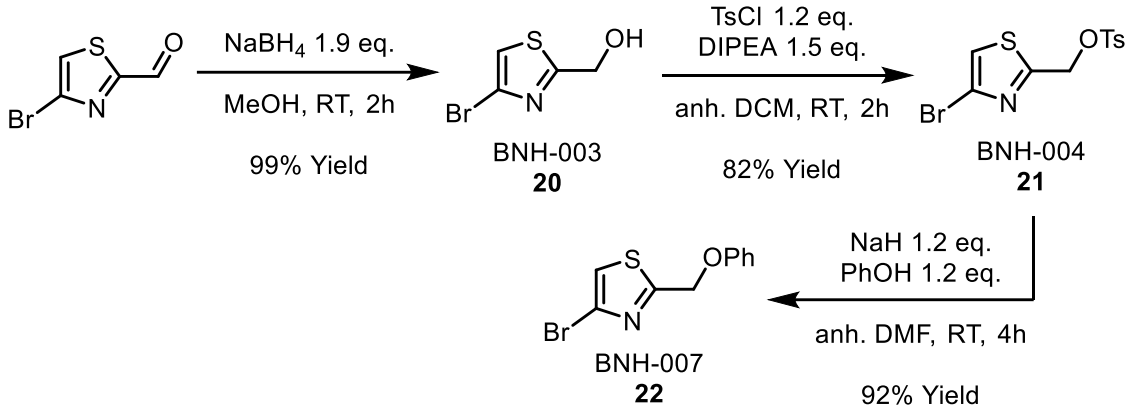


Scheme 5: Synthesis of pyrrole bromide **19** and subsequent borylation attempts.

Condensation of the starting materials in refluxing water proceeded smoothly, affording the product in good yield. While 6-amino-1-hexanoic acid would have been the ideal starting material, the corresponding alcohol was readily available in-house and thus employed. The resulting dimethyl pyrrole was initially a colourless oil but rapidly darkened upon exposure to air and silica. Alkylpyrroles are highly susceptible to oxidation under a broad range of conditions. Upon oxidation, pyrroles tend to undergo polymerisation, forming a tar-like material commonly referred to as *pyrrole black*, a phenomenon attributed to the high π-electron density of the pyrrole ring.[112]

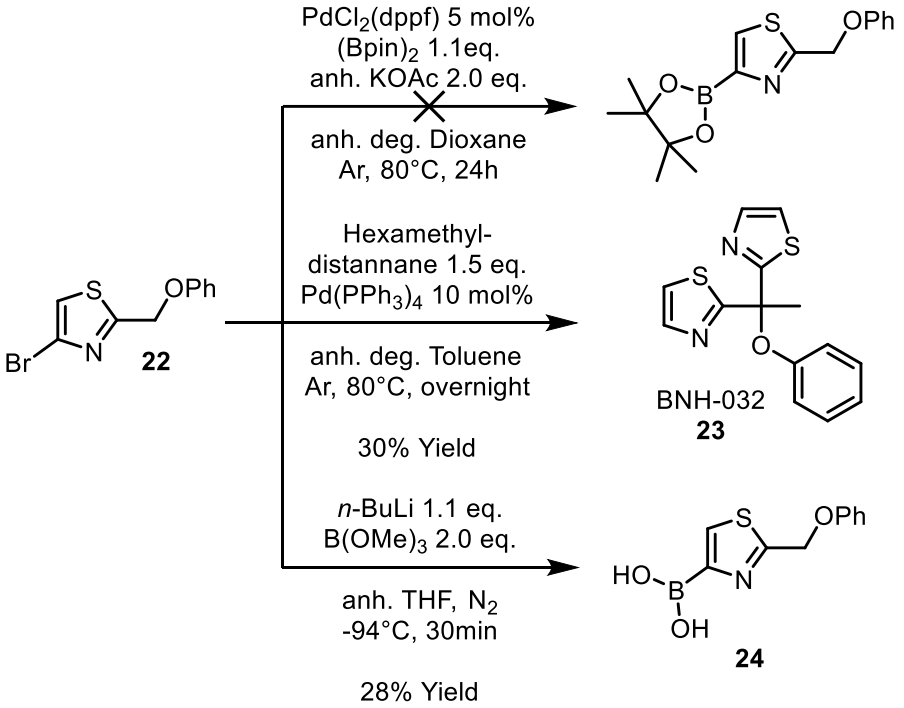
In the present study, the formation of high-molecular-weight black by-products and reduced yields were consistently observed in reactions involving alkylpyrroles. Additional side reactions stemmed from the electron-rich nature of the pyrrole core, which readily undergoes electrophilic substitution. *N*-Bromosuccinimide (NBS), a mild oxidant and source of electrophilic bromine,[113] was used to brominate compound **1**. However, reaction with one equivalent of NBS led to the formation of *pyrrole black*, along with mono- and 3,4-dibrominated derivatives of pyrrole **1**. As a result, the target monobromide **19** was obtained in low yield. These observations are consistent with previous reports employing similar pyrrole substrates.[114]

Attempts to borylate **19** under Miyaura conditions or *via* a lithiation-borylation sequence failed to yield the desired boron-containing intermediate. Consequently, the synthesis was redirected towards preparation of a thiazole boronic acid. This began with the reduction of 4-bromo-2-formylthiazole using $NaBH_4$, affording the corresponding alcohol in excellent yield, in accordance with literature precedent (**Scheme 6**).[115]



Scheme 6: Synthesis of thiazole bromide **22** from 4-bromo-2-formylthiazole.

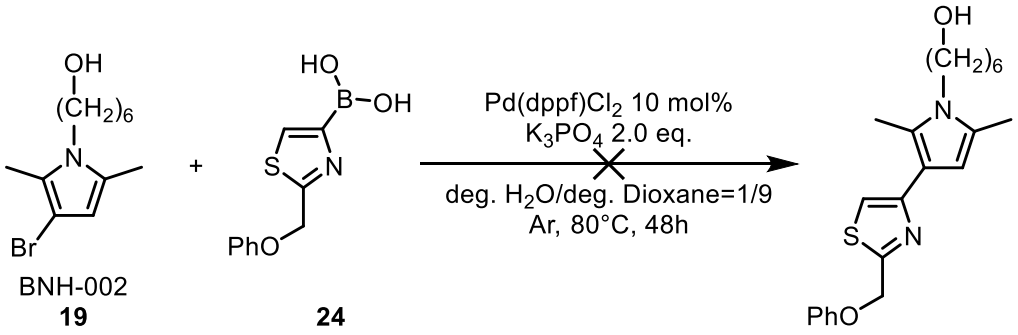
Tosylation of the alcohol **20** to afford tosylate **21**, followed by a Williamson ether synthesis, provided the thiazole ether **22** in good yield. However, subsequent Miyaura borylation of ether **22** under the same conditions previously employed for the synthesis of boronic ester **16** failed to produce the desired product (**Scheme 7**).



Scheme 7: Attempted syntheses of thiazole-based coupling partners for reaction with pyrrole bromide **19**.

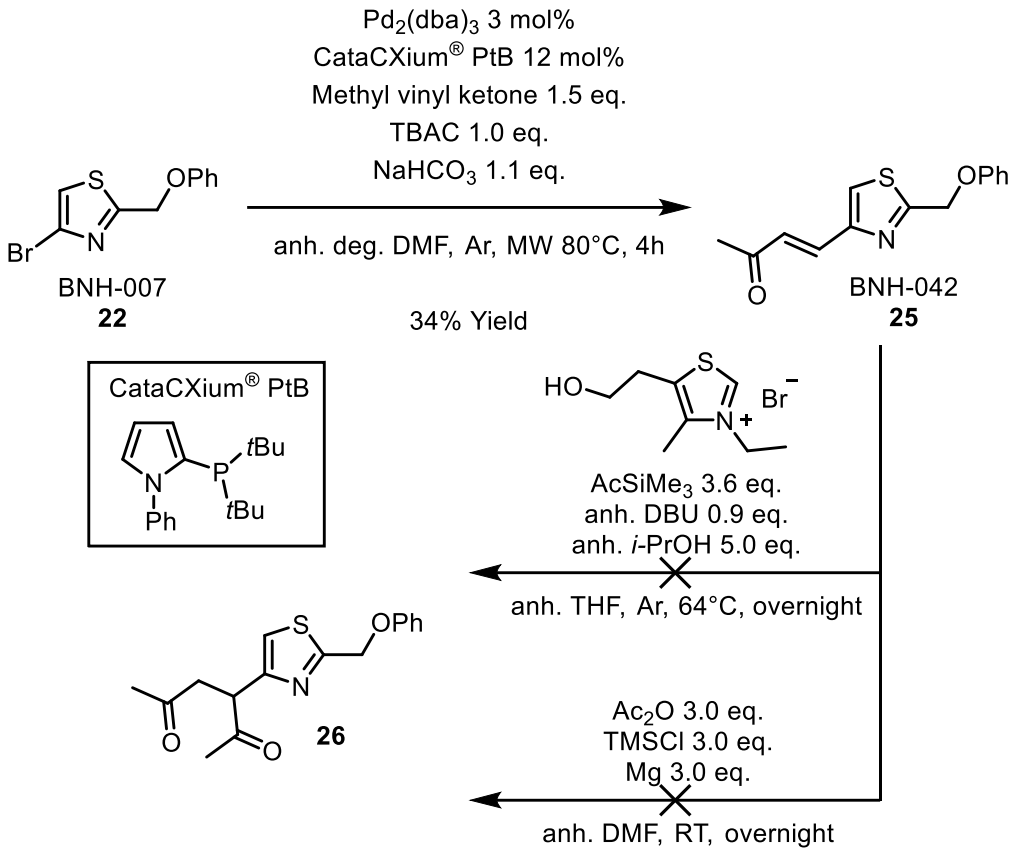
Surprisingly, Pd-catalysed stannylation of bromide **22** with hexamethyldistannane afforded BNH-032 (**23**) as the major product albeit in low yield. Organodistannanes are known sources of $R_3Sn•$ radicals, with the Sn-Sn bond susceptible to homolytic cleavage upon exposure to heat or light.[116] Compound **23** was likely formed *via* a multi-step radical process originating from bromide **22**. This assumption is supported by the structure of compound **23**, as both the bromide and the benzylic position in precursor **22** are particularly prone to radical transformation. In product **23**, the bromide functionality is absent, and both a thiazole ring and a methyl group have been introduced at the benzylic position of the debrominated thiazole core.

Fortunately, a complementary coupling partner for pyrrole bromide **19** was successfully synthesised from bromide **22**. Deprotonation with *n*-BuLi followed by reaction with $B(OMe)_3$ afforded thiazole boronic acid **24** in low yield. This compound was used directly in a Suzuki-Miyaura coupling with bromide **19**, without further purification or characterisation. However, LC/MS analysis indicated that the reaction did not furnish the desired biazole product (**Scheme 8**).



Scheme 8: Attempted Suzuki-Miyaura coupling with bromide **19**.

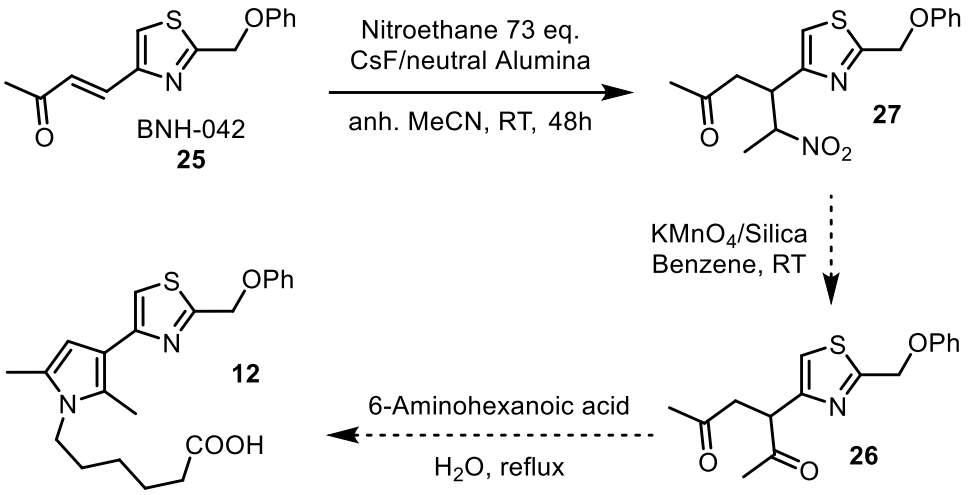
Given the sensitivity of the pyrrole intermediates and the low-yielding borylation of thiazole bromide **22**, pathway A in **Scheme 4** was not pursued further. Instead, attention was redirected to pathway B, as the sensitive pyrrole ring could be synthesised in the last step through a Paal-Knorr synthesis. Literature reports on related scaffolds suggest that the requisite 1,4-diketone can be synthesised from α,β-unsaturated ketones, i.e. compound **25** in **Scheme 9**. The α,β-unsaturated ketone **25** was obtained in moderate yield through Heck coupling between bromide **22** and methyl vinyl ketone. The modest yield is likely attributable to the volatility of methyl vinyl ketone, which has a boiling point of 81.4 °C.[117] The elevated temperatures required for the Heck reaction are not ideally compatible with this volatile reagent, thereby limiting overall conversion.



Scheme 9: Attempted syntheses towards the 1,4-diketone **26** *via* the α,β-unsaturated ketone **25**.

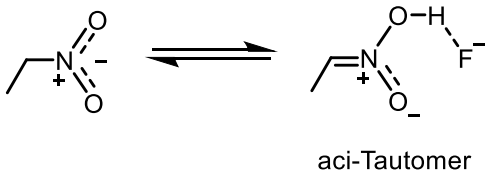
Subsequent conversion of the α,β-unsaturated ketone **25** into the corresponding 1,4-diketone **26** was attempted through a catalytic Sila-Stetter reaction.[118] However, the thiazolium-catalysed addition of acetyltrimethylsilane to α,β-unsaturated ketone **25** afforded the desired 1,4-diketone **26** only in trace amounts, as determined by LC/MS analysis. An alternative method involving Mg-promoted addition of acetic anhydride to α,β-unsaturated ketone **25** likewise failed to provide an improved yield.[119]

A more promising two-step strategy for converting α,β-unsaturated ketones into 1,4-diketones was reported by CLARK, MILLER and SO.[120] The first step involves a Michael-addition of nitroethane to the internal alkene, followed by conversion of the resulting nitro intermediate into the corresponding ketone *via* a Nef reaction. Stirring α,β-unsaturated ketone **25** with excess nitroethane and CsF on alumina led to quantitative formation of the nitro intermediate **27**, while minor amounts of the 1,4-diketone **26** were detected as well (**Scheme 10**).



Scheme 10: Attempted synthesis of nitro intermediate **27** from α,β-unsaturated ketone **25** through Michael addition of nitroethane. The subsequent Nef reaction to access the1,4-diketone **26** as well as the Paal-Knorr synthesis towards target compound **12** were not attempted.

The Michael addition of nitroalkanes to α,β-unsaturated ketones is facilitated by fluoride anions, which shift the nitro-*aci*-nitro tautomerism in favour of the *aci*-nitro form (**Scheme 11**).



Scheme 11: Stabilisation of the *aci*-nitro tautomer of 2-nitroethane by F^-.[121]

Hydrogen bonding between the fluoride anion and the *aci*-nitro alcohol has been reported as the key stabilising interaction.[121] The predominance of the *aci*-nitro tautomer increases the reactivity of the nitroalkane towards Michael acceptors, i.e. ketone **25**. Subsequent oxidation of the nitro intermediate **27** *via* a Nef reaction using $KMnO_4$ on silica and final Paal-Knorr pyrrole synthesis is a promising strategy towards, followed by a final Paal-Knorr pyrrole synthesis, represents a promising strategy for the preparation of LDC151135 (**12**).[120]

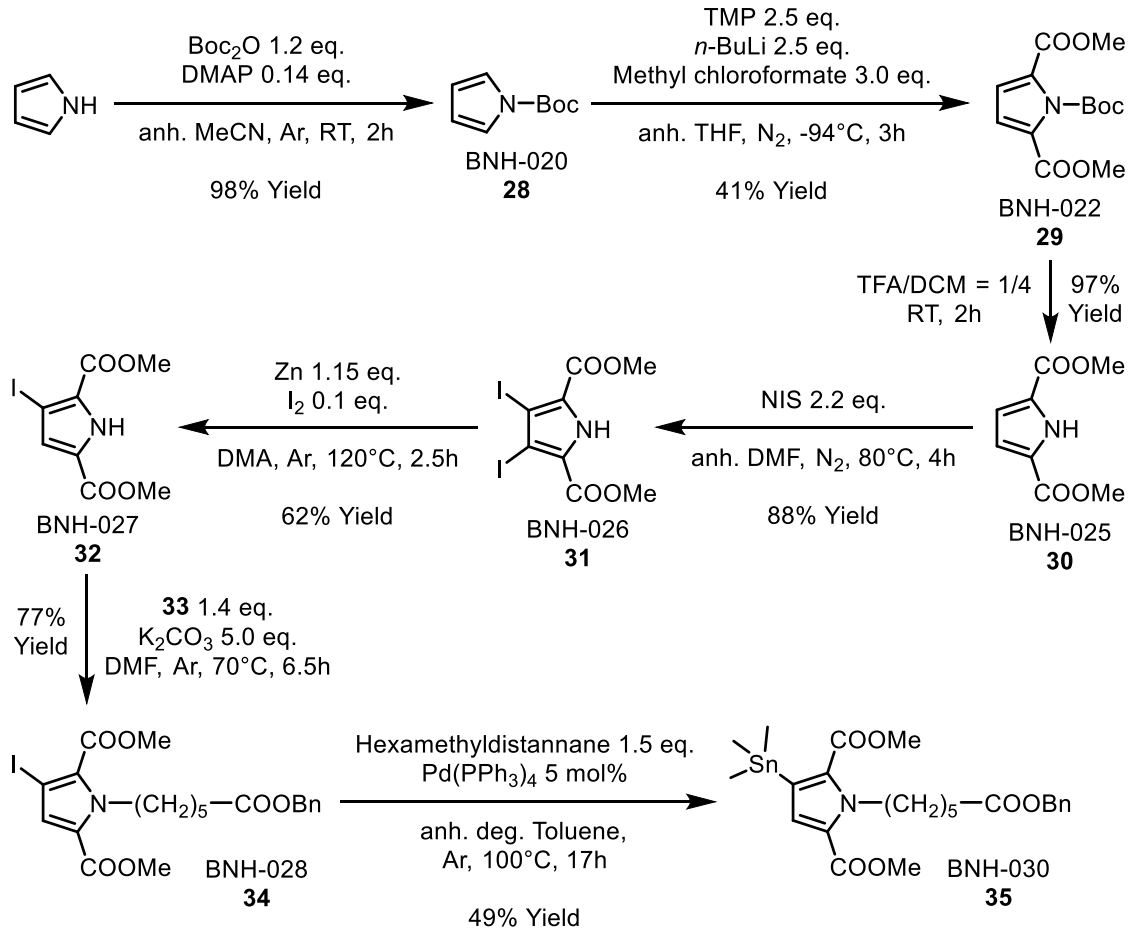
However, at the time of the synthetic efforts towards compound **12**, our collaboration partner LDC observed significant decomposition of the screening sample. Repeated purity assessments of the DMSO stock solution over several months indicated that LDC151135 is not stable in solution. This observation is consistent with the findings reported here and is likely attributable to oxidative degradation and/or polymerisation

of the electron-rich pyrrole ring. As a result, further synthesis of compound **12** was not pursued.

To address the stability issue, a derivative of compound **12** bearing electron-withdrawing substituents in place of the methyl groups on the pyrrole ring was targeted, with the aim of improving solution stability. Synthetic efforts were subsequently redirected toward this more robust analogue.

1.3.3 Synthesis of BNH-039 (37)

The lack of success in synthesising LDC151135 through pathway A in **Scheme 4** was attributed to the intrinsic instability of the pyrrole-based intermediates involved. Consequently, the same coupling strategy was revisited using an electron-deficient pyrrole diester, with the expectation that its reduced electron density would confer enhanced chemical stability and facilitate the successful synthesis of a derivative of LDC151135. The synthesis commenced with the introduction of a Boc protecting group into freshly distilled pyrrole, following a published protocol (**Scheme 12**).[122]

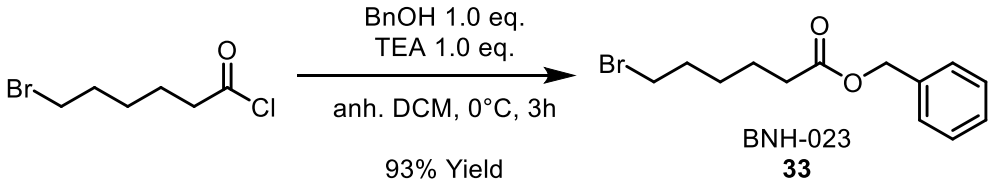


Scheme 12: Synthesis of stannane **35** from pyrrole.

The *N*-Boc-protected pyrrole **28** was subsequently 2,5-dicarboxylated according to a published procedure.[123] 2,2,6,6-tetramethylpiperidine (TMP) was deprotonated using *n*-BuLi to generate a sterically hindered base, which was then employed to selectively deprotonate the C2 and C5 positions of pyrrole **28**. Carboxylation with methyl chloroformate furnished the diester **29** in moderate yield. Removal of the Boc group with 20% TFA in DCM afforded the free amine **30**.

Subsequent diiodination of pyrrole **30** using NIS at 80°C yielded the diiodide **31**, which was selectively monodeiodinated with Zn powder in *N*,*N*-dimethylacetamide (DMA) at 120°C to give the monoiodide **32** in moderate yield and good purity. This two-step iodination sequence was employed in preference to direct monoiodination, which has been reported to afford a mixture of diester **30** and diiodide **31**. [124,125]

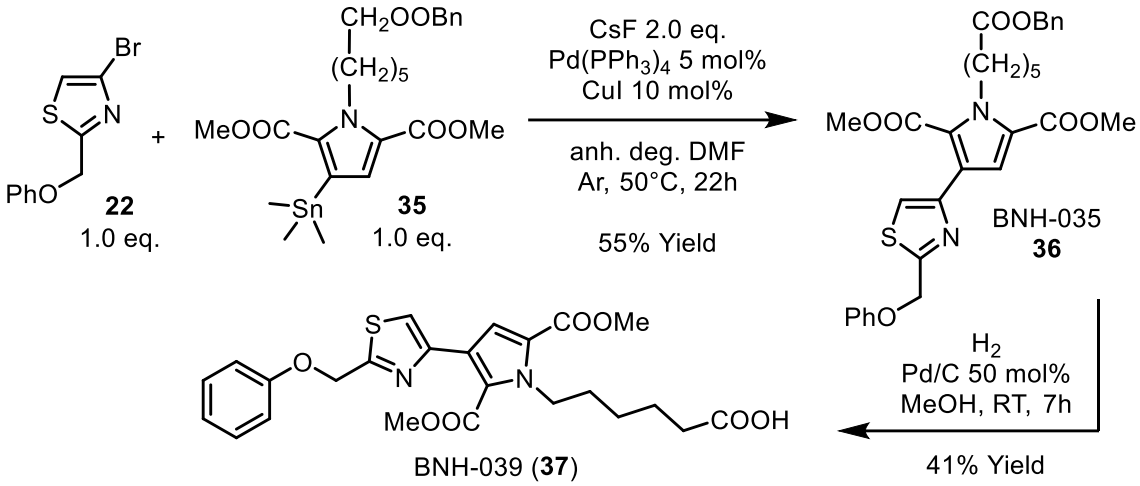
Notably, no decomposition of the pyrrole diester intermediates was observed under elevated temperatures, exposure to air, or on silica, indicating significantly greater stability than previously synthesised alkyl-substituted pyrroles. Prior to *N*-alkylation of the pyrrole **32**, the requisite benzyl ester **33** was synthesised from the corresponding acid chloride and benzyl alcohol according to a literature protocol (**Scheme 13**).[126]



Scheme 13: Synthesis of benzyl ester **33** from 6-bromohexanoyl chloride

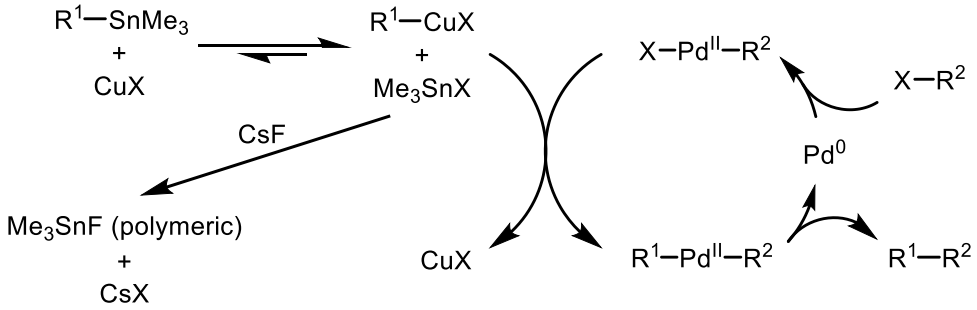
Ester **33** was employed to *N*-alkylate pyrrole **32**, affording the triester **34** in good yield. ZHANG *et al*. have reported the successful application of Stille cross-coupling reactions to pyrrole diesters structurally analogous to iodide, supporting the viability of this approach in the present synthesis.[124] With thiazole bromide **22** in hand, the required stannane **35** was synthesised through Pd-catalysed stannylation of iodide **34** using hexamethyldistannane, yielding the desired product in moderate yield. Deiodination of **34** was identified as the primary competing side reaction, as confirmed by LC/MS and NMR analysis of the crude mixture.

In the next step, the stannane **35** was coupled with bromide **22** under Stille cross-coupling conditions (**Scheme 14**).



Scheme 14: Synthesis of the target diester **37** through Stille coupling of bromide **22** and stannane **35**, followed by catalytic hydrogenolysis of the benzyl ester.

The desired benzyl ester **36** was obtained in moderate yield, accompanied by significant formation of destannylated diester **35** as major side-product. MEE *et al*. demonstrated that the addition of CuI and CsF exerts a positive, synergistic effect on the efficiency of Pd-catalysed cross-couplings between organostannanes and aryl halides.[127] Their mechanistic hypothesis involves initial transmetallation between Cu^{I} and the organostannane (R^1-$SnMe_3$), generating a more reactive organocopper intermediate (**Scheme 15**).



Scheme 15: Proposed mechanism for the positive, synergistic effect of Cu^{I} and CsF on the yields of Stille coupling reactions. Modified from MEE *et al*.[127]

The transmetallation step concurrently produces Me_3SnI, which reacts with CsF to form R_3SnF, an insoluble by-product that precipitates from the toluene reaction mixture. This precipitation shifts the equilibrium, driving the transmetallation forward by removing Me_3SnI from solution. Promotion of the transmetallation step, in turn, facilitates the Pd-catalysed Stille coupling cycle. The observed moderate yield of compound **36** was

unexpected, as destannylation has not been reported as a major side reaction under comparable conditions.[127] A plausible explanation is that the reaction parameters were primarily optimised for benzene derivatives, whereas the electronic properties and reactivity of pyrrole **35** differ significantly.

The final step in the synthesis towards acid **37** involved catalytic hydrogenation of the benzyl ester **36** (**Scheme 14**). Appropriate hydrogenation conditions had to be established to achieve selective cleavage of the benzyl ester while avoiding reduction of the phenol-methylthiazole ether moiety. A total of nine hydrogenation conditions were evaluated, as summarised in **Table 3**.

Conditions	**LC/MS Results**
10% Pd/C (54 mol%), 1,4-Cyclohexadiene (10 eq.), Ethanol[128]	No conversion of ester **36**
$NiCl_2x6H_2O$ (6 eq.), $NaBH_4$ (18 eq.), MeOH[129]	Acid **37** not detected
10% Pd/C(en) (200 mol%.), H_2, MeOH[130]	Complex mixture
10% Pd/C (25 mol%), H_2, MeOH	50% conversion of ester **36** after 18h
10% Pd/C (50 mol%), H_2, MeOH	50% conversion of ester **36** after 7h
10% Pd/C (50 mol%), H_2, AcOH, MeOH	50% conversion of ester **36** after 3h
10% Pd/C (75 mol%), H_2, MeOH	Acid **37** with side products
10% Pd/C (100 mol%), H_2, MeOH	Acid **37** with side products
10% Pd/C (200 mol%), H_2, MeOH	Red. of Bn-ester and PhO-ether in **36**

Table 3: Tested reaction parameters for the catalytic hydrogenation of benzyl ester **36**.

The most effective conditions employed 50 mol% of Pd/C (10%) in MeOH for 3 hours. However, complete conversion was not observed, even after extended hydrogenation overnight. This incomplete conversion suggests potential catalyst poisoning, likely caused by minor side-products generated during the hydrogenation.

Complete conversion was ultimately achieved only after repeated isolation of the starting material and subsequent hydrogenation, affording acid **37** in a moderate yield of 41%. The overall yield of compound **37** over the nine-step synthetic sequence, starting from pyrrole, was 1%. Despite the modest yield, the final product was obtained in sufficient quantity for all planned biological evaluations and exhibited excellent solution stability over prolonged storage.

1.3.4 Derivatisation of Zafirlukast (13)

A comprehensive understanding of the structure-activity relationship (SAR) of the zafirlukast scaffold is essential for the successful evaluation of its KRAS PPI inhibitory activity. Fortunately, the synthesis of various zafirlukast derivatives has been reported in the literature.[131–137] In the present study, previously published synthetic protocols were optimised, and new synthetic routes were developed to access novel derivatives. The zafirlukast analogues prepared in this work are summarised in **Table 4**.

BNH-	R^1	R^2	R^3	R^4	R^5
057 (**38**)	NO_2	Me	CH_2	Me	TSA
Ox. sen. (**39**)	NH_2	Me	CH_2	Me	TSA
059 (**40**)	MC	Me	CH_2	Me	TSA
031 (**41**)	CPC	Me	CH_2	H	TSA
054 (**42**)	CPC	H	CH_2	Me	TSA
053 (**43**)	CPC	Me	CH_2	Me	OH
051 (**44**)	CPC	Me	CH_2	Me	OMe
055 (**45**)	CPC	Me	CH_2	Me	MSA
044 (**46**)	NO_2	Me	CH_2	Me	OMe
049 (**47**)	NO_2	Me	CH_2	Me	OH
047 (**48**)	NH_2	Me	CH_2	Me	OMe
050 (**49**)	NH_2	Me	CH_2	Me	OH
081 (**50**)	NO_2	Me	C=O	Me	OMe

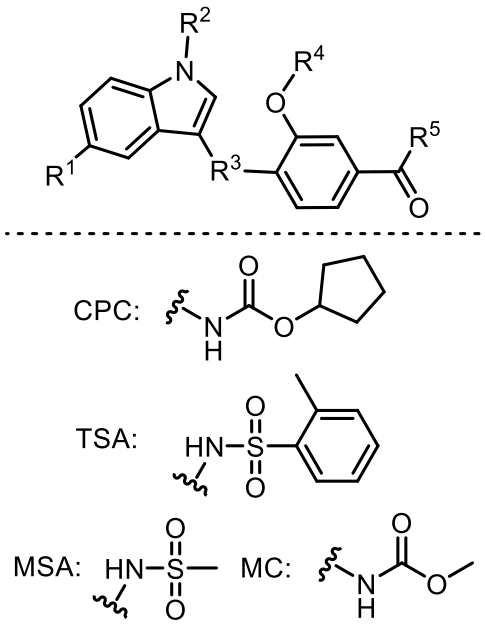
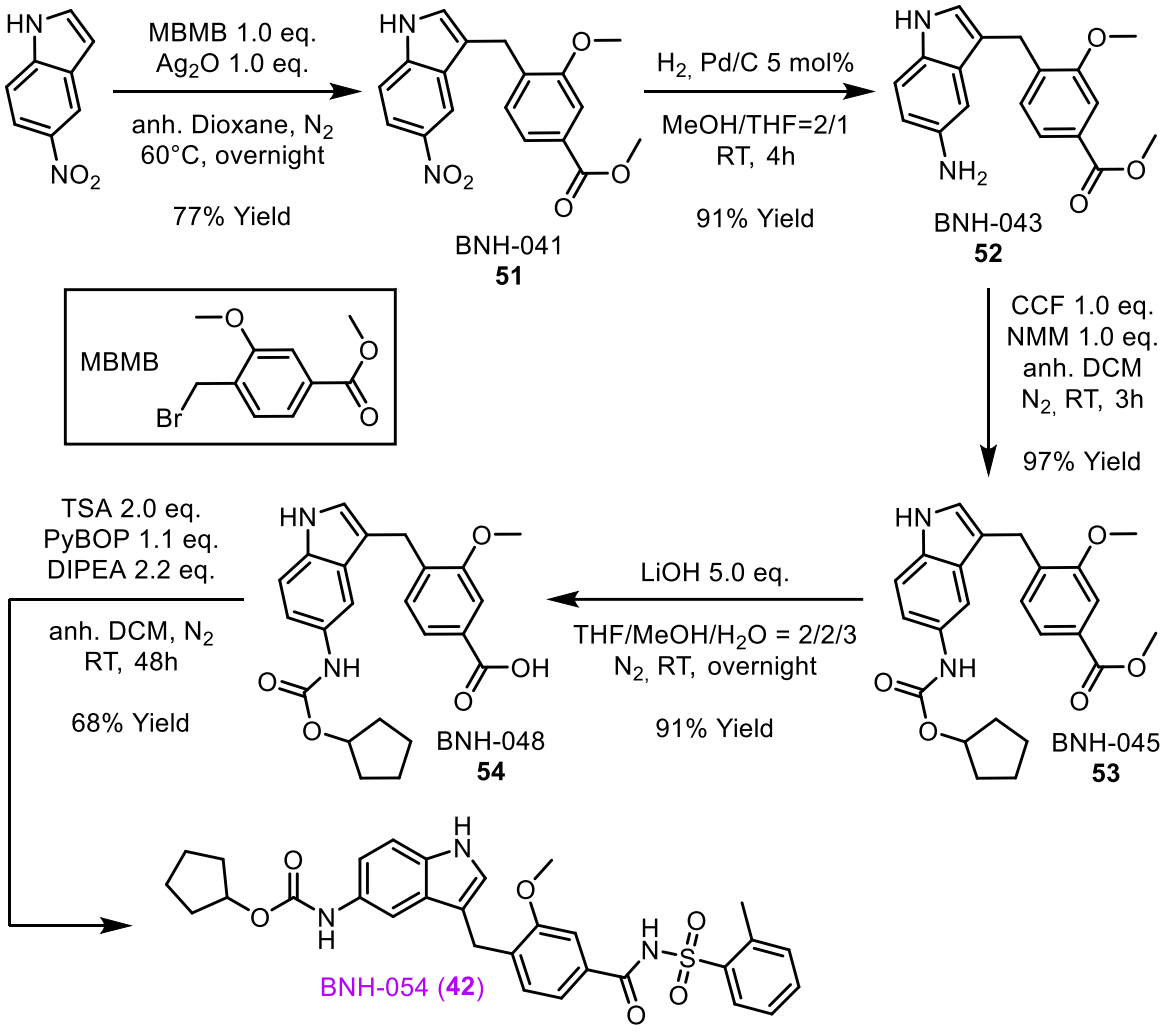


Table 4: Target derivatives of the zafirlukast SAR grouped according to variations in substituent residues. Abbreviations: methyl/cyclopentyl carbamate (MC/CPC) and methane/*o*-toluenesulfonamide (MSA/TSA).

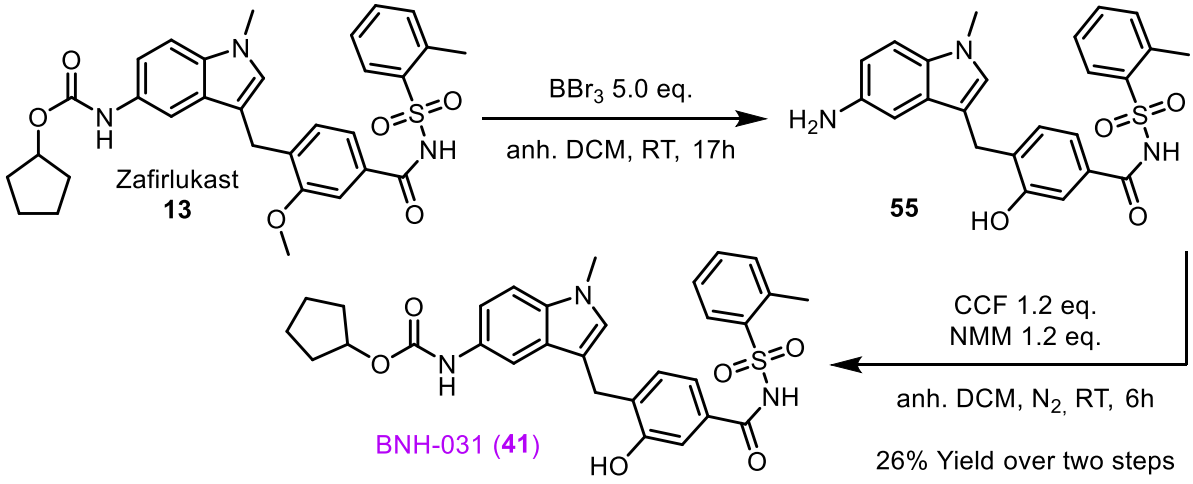
The green series comprises derivatives with modifications at R^1, i.e. nitro indole **38**, oxidation sensitive amine **39**, and the methyl carbamate (MC) **40**. The magenta series features variations at R^2 and R^4, i.e. *N*/*O*-desmethyl derivatives **41** and **42**. The blue series includes modifications at R^5, i.e. carboxylic acid **43**, methyl ester **44** and methanesulfonamide (MSA) **45**. Notably, acyl sulfonamides are considered bioisosteres of carboxylic acids due to their low pka values in the range 4–5.[138] The orange series comprises all four combinations of nitro/amino indole with benzoic acid/methyl benzoate, i.e. **46**-**49**. Additionally, a ketone at R^3 is featured in derivative **50**. Derivatives **41**, **42**, **45**, and **49** are novel, while the remaining compounds have been reported previously.[131–137] The *N*-desmethyl analogue **42** can be conveniently synthesised from 5-nitroindole (**Scheme 16**).



Scheme 16: Synthesis of target derivative **42** of the magenta series, commencing from 5-nitroindole.

Alkylation of 5-nitroindole at C3 using methyl 4-(bromomethyl)-3-methoxybenzoate (MBMB) in the presence of Ag_2O gave the zafirlukast core scaffold **51**. Unreacted starting material was recovered and re-subjected to alkylation, resulting in a combined overall yield of 77%. Catalytic hydrogenation of the nitro group then provided the primary amine **52** in good yield. All synthesised zafirlukast derivatives were found to be prone to oxidation in air and on silica, particularly the primary amines. As a result, purification was only possible by HPLC, followed by lyophilisation of the appropriate fractions. Carbamate formation proceeded smoothly upon treatment of amine **52** with cyclopentyl chloroformate (CCF) and N-methylmorpholine (NMM), furnishing methyl ester **53**. Notably, the carbamate moiety appeared to impart oxidative stability to the scaffold, likely through electron withdrawal from the indole ring system. Subsequent hydrolysis of **53** afforded the carboxylic acid **54**. Final coupling with o-toluenesulfonamide (TSA) under standard peptide coupling conditions yielded the

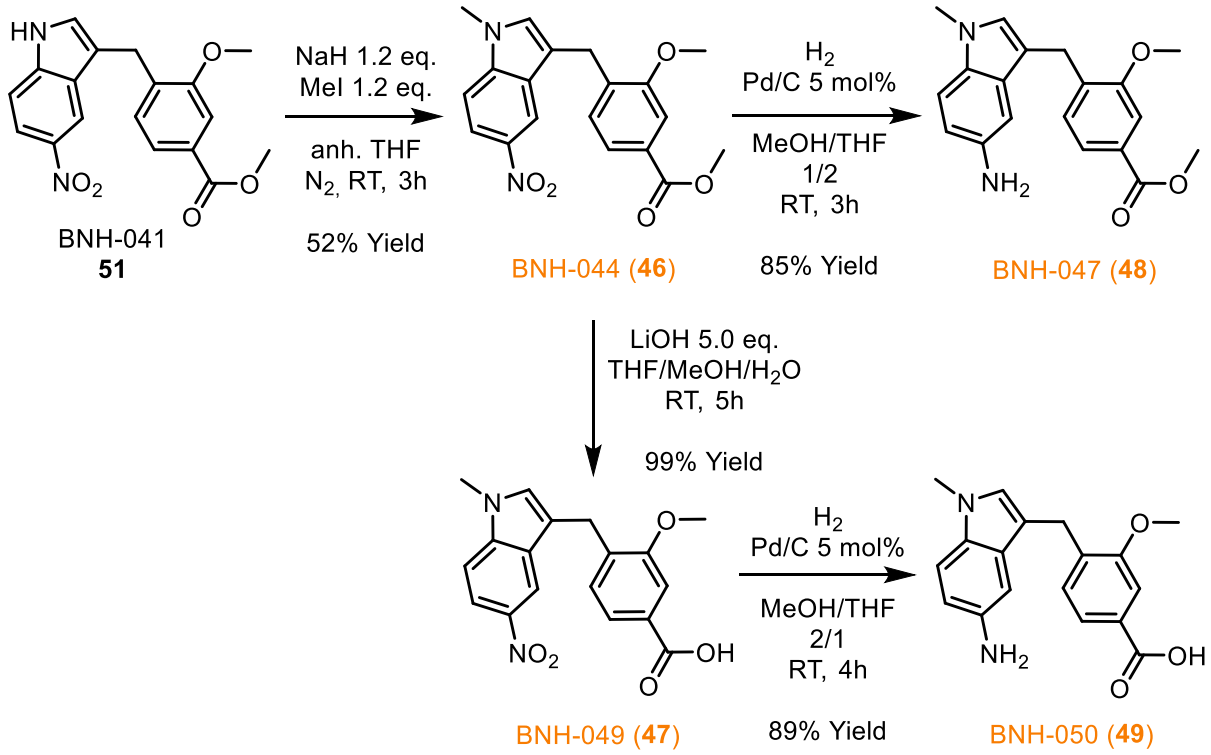
target acyl sulfonamide BNH-054 (**42**). An initial attempt using HATU as the coupling reagent led to undesired homocoupling between the indole amine and the carboxylic acid of intermediate **54**, as evidence by LC/MS analysis. This side reaction likely occurred during preactivation of compound **54** with HATU, prior to addition of TSA, in line with the known reactivity of HATU with both amines and acids.[139] In contrast, PyBOP does not react with amines enabling all reagents to be combined from the outset. This allowed the excess TSA to promote the desired heterocoupling. In comparison, the *O*-desmethyl derivative **41** was prepared from zafirlukast through a novel, two-step synthesis (**Scheme 17**).



Scheme 17: Two-step synthesis of target derivative **41** of the magenta series starting from zafirlukast.

A commonly employed method for the *O*-demethylation of phenol ethers involves treatment of the starting material with the strong Lewis acid BBr_3.[140] When this protocol was applied to zafirlukast, cleavage of both the methyl ether and the carbamate functionality was observed. The resulting intermediate **55** displayed poor solubility and pronounced sensitivity towards oxidation, rendering it unsuitable for further use in the zafirlukast SAR. Consequently, intermediate **55** was carried forward directly used into the next step without purification. The primary amine was then successfully transformed into the desired cyclopentyl carbamate **41** using cyclopentyl chloroformate (CCF) and *N*-methyl-morpholine (NMM).

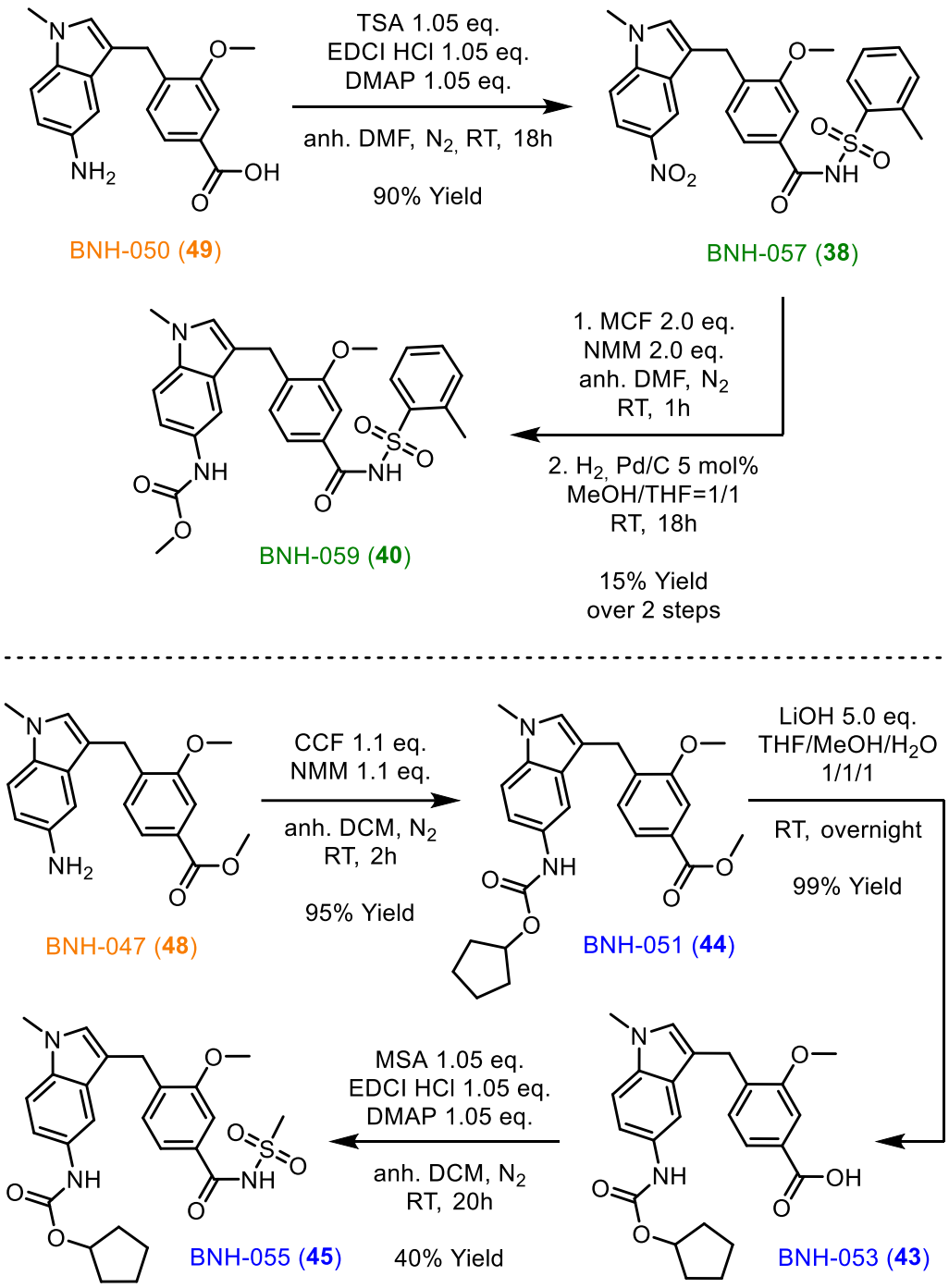
The subsequent set of derivatives, i.e. the orange series, was synthesised starting from intermediate **51**, employing modified versions of the protocols shown in **Scheme 16** and **Scheme 17** (**Scheme 18**).



Scheme 18: Synthesis of the orange target derivative series, i.e. compounds **46**-**49**, starting from indole **51**.

The nitroindole **51** was *N*-methylated to afford the target derivative **46** in moderate yield. Subsequent catalytic reduction of the nitro group furnished the amine **48**. In parallel, methyl ester **46** was hydrolysed under basic conditions to yield the carboxylic acid **47** in excellent yield. Catalytic hydrogenation of the nitro derivative **47** gave the amino acid **49** in good yield. This amino acid **49** an the corresponding methyl ester **48** served as starting materials for the preparation of the green and blue target derivative series (**Scheme 19**).

Amide coupling of derivative **49** with TSA furnished the acyl sulfonamide **38** in good yield. Catalytic hydrogenation of the nitro group compound **38** gave the corresponding primary amine **39**, which is listed in **Table 4**, but not shown in **Scheme 19**. Similar to intermediate **55**, amine **39** exhibited low solubility and high susceptibility to oxidative degradation, preventing its isolation in sufficient purity for biological evaluation. Therefore, amine **39** was used directly in the subsequent step without further purification.

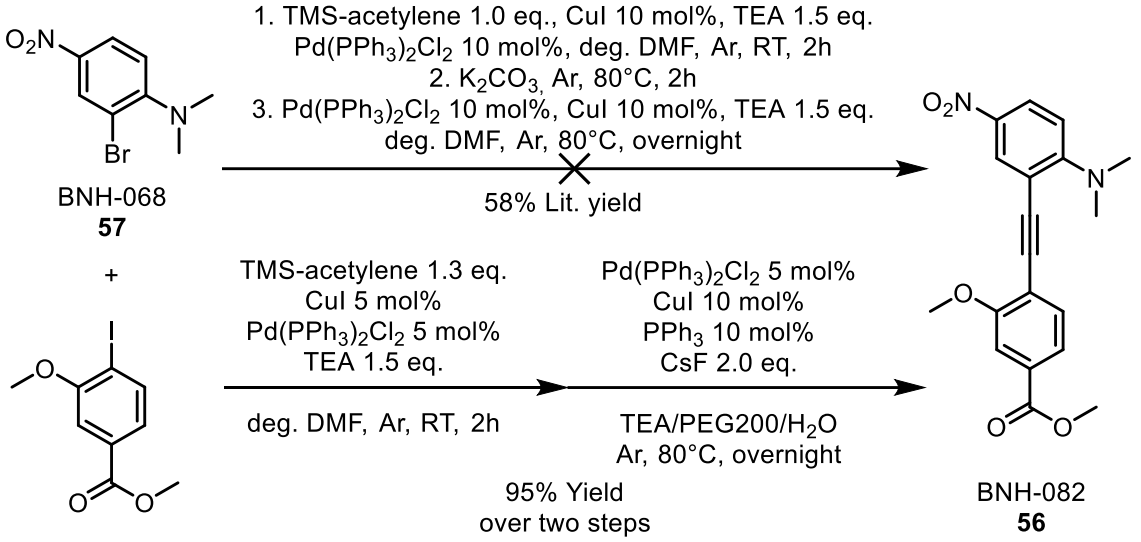


Scheme 19: Top: synthesis of the green target derivative series, i.e. compounds **38** and **40**, starting from derivative **49**. Bottom: synthesis of the blue target derivative series, i.e. compounds **43**-**45**, starting from derivative **48**.

Conversion of the amine **39** into the corresponding methyl carbamate **40** using methyl chloroformate (MCF) proceeded with 15% overall yield across the two steps. In contrast, transformation of the amine group in derivative **48** into the corresponding cyclopentyl carbamate **44** was achieved in excellent yield. Subsequent hydrolysis of the methyl ester afforded the corresponding carboxylic acid **43**, which was coupled with methanesulfonamide (MSA) in standard amide coupling condition to yield the

target derivative **45**. The moderate yield in this step is probably a result of the relatively low nucleophilicity of sulfonamides compared to amines.

The final target derivative **50** features a ketone in place of the methylene bridge in present in zafirlukast. This substitution introduces a hydrogen bond acceptor (HBA) between the two aromatic rings, making ketone **50** the only compound in this zafirlukast SAR series to incorporate such a feature. It is therefore a key intermediate for this zafirlukast SAR. Fortunately, the synthesis of ketone **50** has been reported recently and involves the alkyne intermediate **56** (**Scheme 20**, top arrow).[136]



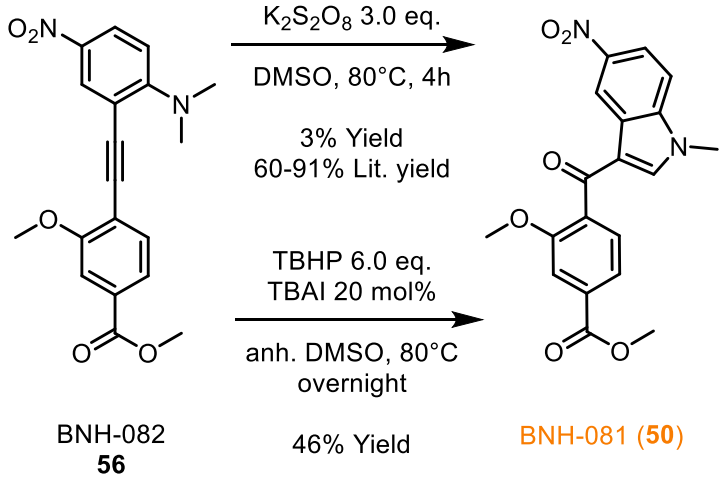
Scheme 20: Top arrow: attempted synthesis of alkyne **56** through a published one-pot domino Sonogashira coupling protocol over three steps. Bottom arrows: successful synthesis of alkyne **56** through an adapted sequential Sonogashira coupling protocol.

The synthesis of alkyne **56** was originally published as a one-pot, three-step, also referred to as domino Sonogashira coupling.[136,141] The required 2-bromo-*N*,*N*-dimethyl-4-nitroaniline **57** was synthesised through monobromination of *N*,*N*-dimethyl-4-nitroaniline (see Experimental Section). In the published protocol, bromide **57** was first converted into the corresponding trimethyl silyl (TMS) alkyne. In the second step, the TMS group was removed using K_2CO_3, and the resulting the terminal alkyne was immediately subjected to Sonogashira coupling with methyl 4-iodo-3-methoxy-benzoate. A final yield of 58% was reported for alkyne **56**.

However, this procedure could not be reproduced successfully in our laboratory. As an alternative, we employed a modified synthetic route, beginning with the Sonogashira coupling of the more reactive methyl 4-iodo-3-methoxybenzoate with TMS-acetylene. The resulting TMS-alkyne intermediate was purified *via* flash chromatography and then

used in a sequential Sonogashira coupling. In this adapted method, the TMS group was cleaved *in situ* using CsF, and the resulting terminal alkyne was immediately coupled in a one-pot reaction.[141,142] Using this approach, the internal alkyne **56** was obtained in excellent yield over two steps. This improved yield can be attributed to the low concentration of terminal alkyne in the reaction mixture, which suppresses the competing Glaser homocoupling.[142]

With the alkyne **56** in hand, the subsequent oxidation to ketone **50** was attempted following a published procedure (**Scheme 21**, top arrow).[136]



Scheme 21: Comparison of two published oxidation protocols from alkyne **56** towards ketone **50**.

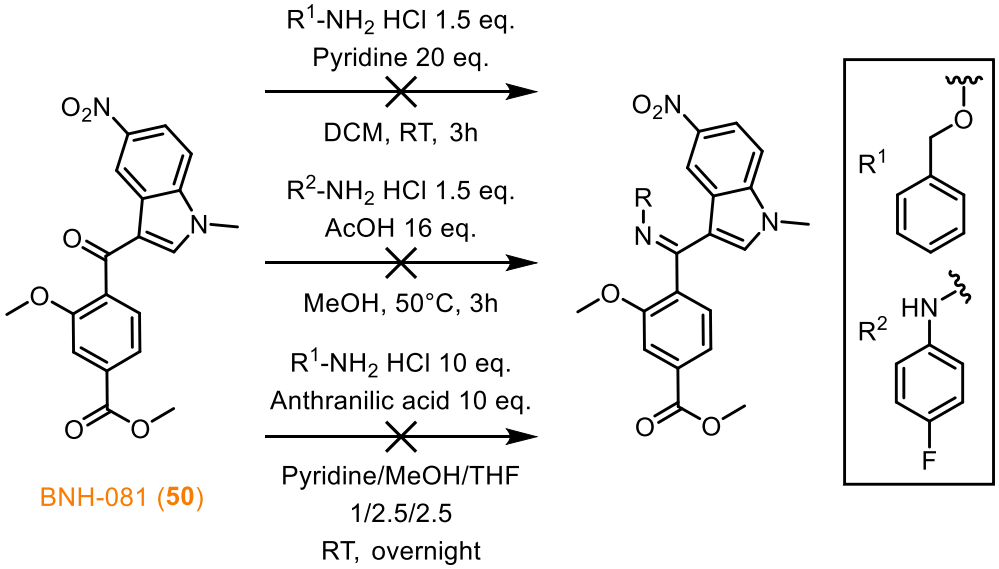
In the published procedure, the alkyne **50** was oxidised using $K_2S_2O_8$ in DMSO at 80°C for 4–5 hours with a reported yield of 60–91%. However, under identical conditions, our attempts resulted in a complex product mixture and a significantly lower yield of only 3%. This discrepancy suggests that $K_2S_2O_8$ may be too potent an oxidant in this context, leading to overoxidation and side reactions that compromise selectivity.

Fortunately, an alternative protocol was identified, which reported successful oxidation of an alkyne structurally similar to compound **56**, using *tert*-butyl hydroperoxide (TBHP) in combination with tetrabutylammonium iodide (TBAI).[143] When applied to alkyne **56**, this milder procedure led to a much cleaner conversion affording the target ketone **50** 47% yield.

Notably, both derivatives **50** and **56** represent valuable scaffolds for further derivatisation, particularly due to the presence of central functional groups, i.e. an alkyne and a carbonyl moiety. As discussed in the introduction, effective PPIIs often

adopt L- or T-shaped geometries.[14] The central alkyne or carbonyl moieties in derivatives **50** and **56** offer ideal vectors for installing a third substituent, thereby enabling the construction of such geometries in a rational design approach to PPIIs.

Toward this goal, attempts were made to convert ketone **50** into the corresponding oxime or hydrazone derivatives by stirring with *O*-benzylhydroxylamine or *p*-fluorophenylhydrazine, in the presence of either base or acid. Unfortunately, these transformations did not yield the desired products (**Scheme 22**).

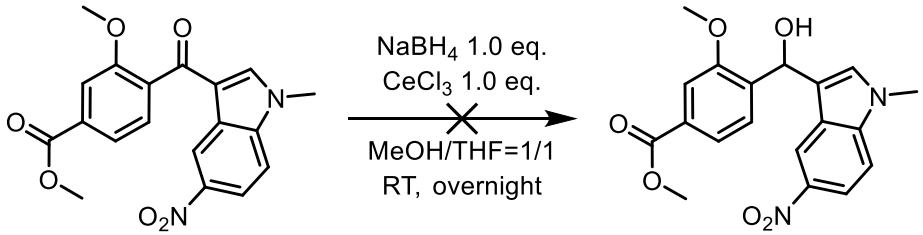


Scheme 22: Attempted conversions of ketone **50** into the corresponding benzyl oxime and *p*-fluorophenyl hydrazone.

Notably, no conversion of ketone **50** was observed by LC/MS analysis under the tested reaction conditions for oxime or hydrazone formation. CRISALLI *et al.* reported that oxime formation from aldehydes can be significantly accelerated in the presence of anthranilic acid as a nucleophilic catalyst.[144] Motivated by this finding, the same catalytic conditions were applied to ketone **50**; however, once again, no reaction was observed by LC/MS analysis.

A likely explanation for the lack of reactivity lies in the electronic nature of the carbonyl group in ketone **50**. The ketone is flanked by two aromatic rings, and conjugation with these electron-rich systems considerably reduces the electrophilicity of the carbonyl carbon. As a result, its reactivity towards nucleophiles such as hydroxylamines or hydrazines is diminished, even under catalysed conditions.

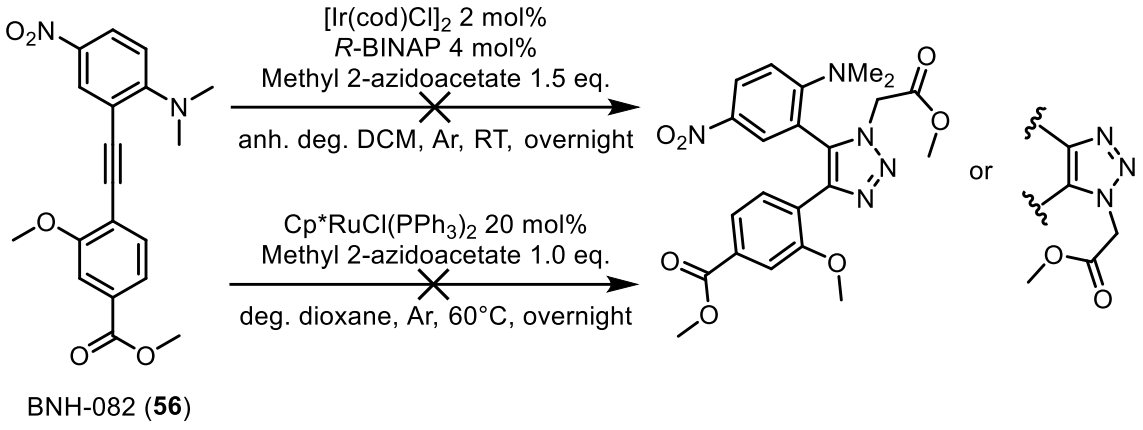
To circumvent this issue, an alternative strategy was pursued: the reduction of ketone **50** to the corresponding secondary alcohol, which would serve as a precursor for further functionalisation *via* alkylation. For this purpose, a Luche reduction was attempted under standard conditions (**Scheme 23**).



Scheme 23: Attempted Luche reduction of ketone **50** towards the corresponding alcohol.

The Lewis acid $CeCl_3$ has been reported to activate carbonyl groups toward nucleophilic attack by hard nucleophiles, i.e. H^-.[145] However, no conversion of ketone **50** was observed by LC/MS analysis following overnight stirring at room temperature. This lack of reactivity highlights the reduced electrophilicity of the carbonyl, likely due to resonance stabilisation by the flanking aromatic systems. Future work should therefore focus on applying stronger, more reactive hydride donors, i.e. DIBAL-H and $LiAlH_4$, which may overcome the electronic deactivation of the ketone.

The synthetic work of this project concluded with two attempted conversion of alkyne **56** into the corresponding 1,4,5- or 3,4,5-trisubstituted 1,2,3-triazoles (**Scheme 24**).



Scheme 24: Attempted conversions of alkyne **56** into the corresponding 1,4,5- or 3,4,5-trisubstituted 1,2,3-triazoles.

Standard Cu-catalysed azide-alkyne cycloaddition (CuAAC) is a widely employed and reliable method for synthesising 1,4-disubstituted 1,2,3-triazoles from terminal alkynes. However, internal alkynes such as compound **56** are generally unreactive under CuAAC conditions and do not yield the corresponding trisubstituted triazoles. In contrast, alternative cycloaddition methods using Ir and Ru catalysts have been reported to successfully convert internal alkynes into trisubstituted 1,2,3-triazoles, often with good yields.[146,147] The regioselectivity of these transformations depends significantly on the substituents present on the alkyne. In both published protocols, ethyl 2-azidoacetate was employed as the azide partner. As this compound was not readily available in our laboratory, the corresponding methyl ester was used instead. Unfortunately, under both Ir- and Ru-catalysed conditions, no conversion of alkyne **56** into the desired trisubstituted 1,2,3-triazoles was observed.

1.3.5 Biochemical Data and Docking Results

1.3.5.1 Assays and Reference Molecules

Our project library currently comprises ~1,200 molecules with 1 to 30+ assay datapoints each. As the primary screening method, we employ the KRAS-SOS nucleotide exchange assay to evaluate initial compound activity. Compounds exhibiting IC_{50} values in the low micromolar range are subsequently advanced to cellular assays. Compounds that show little to no inhibitory activity are typically not subjected to further evaluation, with the exception of reference compounds or molecules of particular interest. The relevant reference compounds used to benchmark assay outcomes have been outlined previously in the introductory section 1.1.13 (**Figure 19**).

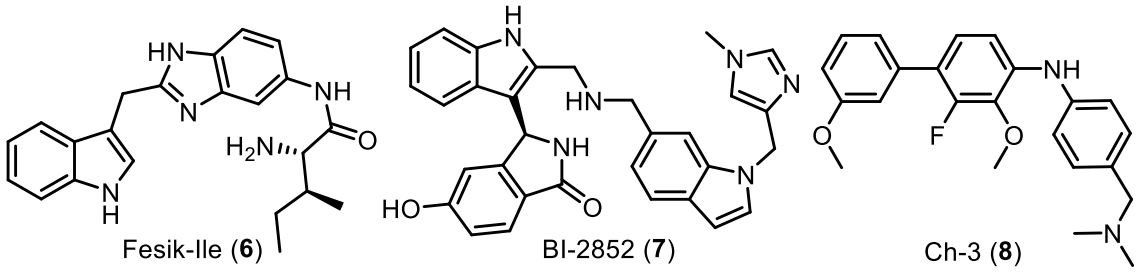


Figure 19: Relevant reference molecules for the discussion of the biochemical data and docking results of the biazole and zafirlukast derivatives.

The biazole and zafirlukast scaffolds constitute rather linear small molecules sharing notable structural similarities with the reference compounds Fesik-Ile (**6**), BI-2852 (**7**) and Ch-3 (**8**), which are included in the following SAR discussion. The analysis is based on two assays: $KRAS^{G12D/V/WT}$-SOS nucleotide exchange ($NE^{G12D/V/WT}$) and CellTiter Glo (CTG) viability of SNU-1($KRAS^{G12D}$) and RKo ($KRAS^{WT}$) cell lines. The $NE^{G12D/G12V/WT}$ assays quantify the SOS1-catalysed GDP/GTP exchange in the presence of a PPII. The CTG SNU-1/RKo assays measure the amount of living cells in a cell culture 72h after PPII treatment. The assay protocols are described in detail in the Experimental Section.

A summary of the NE^{G12D} assay results for the biazole and zafirlukast series, along with the outcome of rigid receptor docking of these ligands at P1 on $KRAS^{G12D, GDP}$ is presented in **Table 5**.

ID	NE^{G12D} IC_{50} [µM]	DS [kcal/mol]	Glide emodel
12	3.84 ± 2.14	-7.77	-74.2
37	100 ± 32.1	-6.91	-80.2
14	>300	-5.15	-56.5
23	>300	-4.67	-42.9
58	>300*	-5.72	-56.0
59	>300*	-4.39	-45.4
60	>300*	-5.17	-54.5
61	>300*	-4.58	-44.9
62	>300*	-4.97	-50.0
63	19.8*	-4.30	-36.3
13	12.7 ± 6.22	-3.89	-43.2
38	16.5 ± 1.78	-4.24	-54.4
40	66.7 ± 30.9	-3.71	-48.8
41	11.9 ± 5.99	-2.13	-43.4
42	10.4 ± 0.909	-1.93	-41.3
43	82.7 ± 40.7	-2.24	-38.1
44	>300	-4.14	-41.3
45	55.4 ± 2.89	-3.22	-42.1
46	>300	-3.70	-38.7
47	48.7 ± 9.96	-4.01	-36.6
48	>300	-4.98	-44.7
49	80.1 ± 65.1	-4.93	-41.1
50	>300*	-4.32	-42.0
56	>300*	-3.75	-36.7
6	>300	-5.77	-63.6
7	17.0 ± 9.31	-6.13	-79.3
8	>300	-4.86	-45.3

Table 5: Left: nucleotide exchange assay results of the biazole (orange) and zafirlukast (green) series. The mean is reported with the standard deviation. Single measurements are marked with an asterisk. Right: docking scores (DS) and Glide emodel values of those ligand conformations with the largest negative emodel values in **Figure 22**. The references are shaded white.

Molecules with an ID starting with “LDC”, i.e. biazoles **58**–**63**, have been provided by our collaboration partner. Their syntheses are not covered in this thesis. In the following discussion all compounds with $IC_{50}(NE^{G12D})$ > 300 µM are classified as inactive, while those with lower IC_{50} values are considered active. The structures of all derivatives listed in **Table 5** are shown in **Figure 20**.

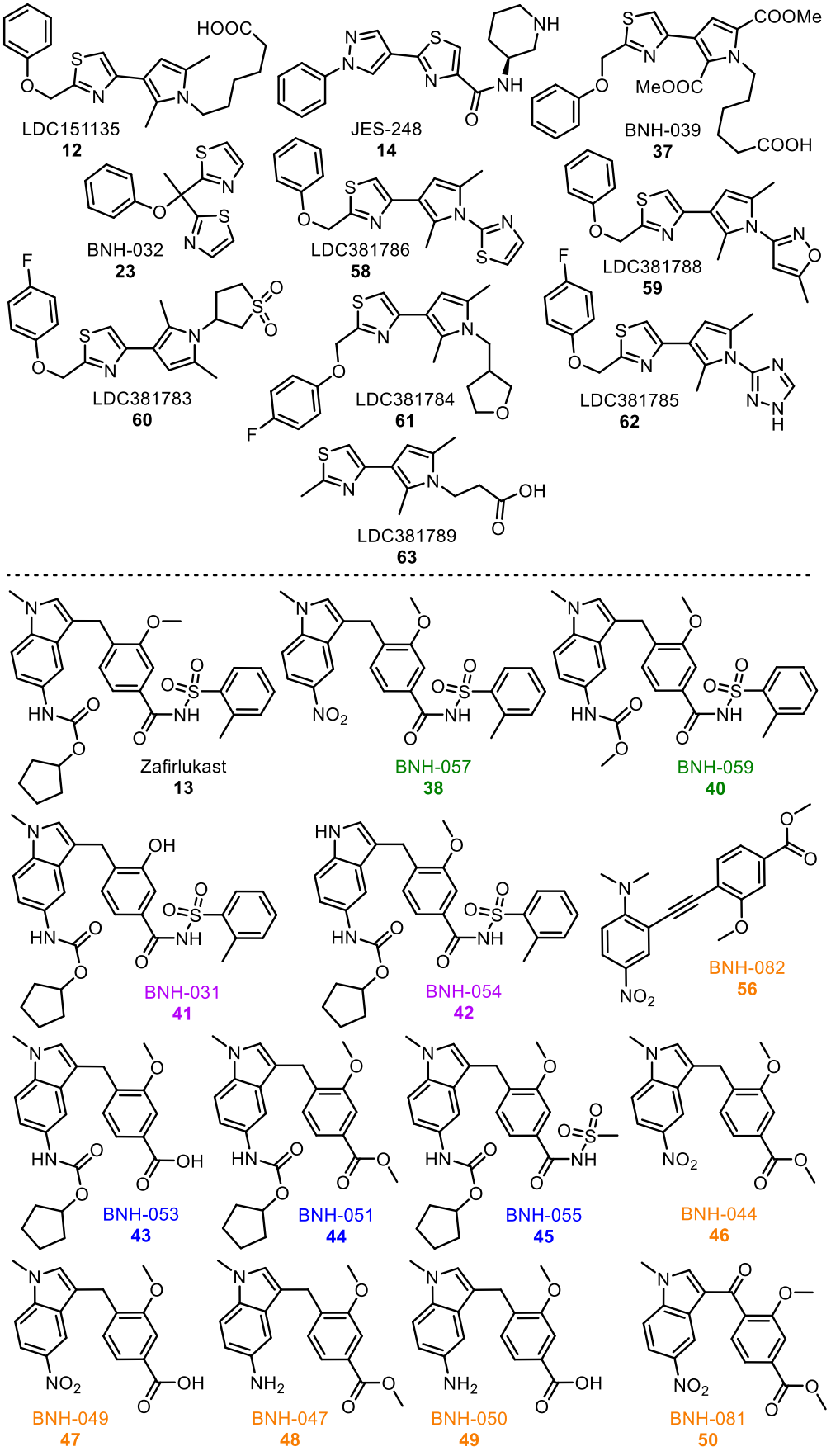


Figure 20: Biazole and zafirlukast derivatives synthesised in this project. The zafirlukast derivatives are coloured according to their structural differences, analogous to **Table 4**.

1.3.5.2 Rigid Receptor Docking

Visualising the protein:ligand complexes of the derivatives listed in **Table 5** greatly facilitates interpretation of the corresponding assay data. To this end, our research group routinely employs the Schrödinger Maestro Suite 2018 for the *in silico* simulation of protein:ligand interactions. In this section, the generated complexes serve as visual and interpretive tools for analysing $KRAS^{G12D}$:ligand binding.

Rigid receptor docking was performed using the crystal structure of $KRAS^{G12D,\ C118S}$ in complex with BI-2852 (**7**) (PDB ID: 6ZL5).[95] The C118S mutation was introduced to enhance protein stability, as previously reported.[55] Rigid receptor docking involves generating and ranking potential binding poses of a ligand (any listed in **Table 5**) within a static binding site on a protein (P1 in 6ZL5).[148] The resulting poses were evaluated based on their Glide emodel values, a dimensionless metric that reflects the likelihood of a ligand adopting a particular binding pose. Ideally, a single binding pose with a highly negative emodel value is observed, indicating a strongly favoured conformation.

Additionally, the docking score (DS), which estimates the binding free energy (ΔG_{bind}) in kcal/mol, was used to compare binding affinities across chemically diverse ligands. This score accounts for various factors including hydrogen bonding, desolvation, and metal-ligand interactions. The DS thus provides a comparative measure of complex stability and potential biological relevance.[148,149]

Detailed protocols for the performed calculations are provided in the Experimental Section. The results of the rigid receptor docking are summarised in **Table 5**.

Before discussing the data presented in **Table 5**, two critical aspects of the docking process must be highlighted. First, P1 on $KRAS^{G12D,\ GDP}$ was chosen as binding site for the rigid receptor docking based on the structural similarity of the biazole and zafirlukast derivatives with established P1 ligands, specifically references **6**–**8**. Among these, BI-2852 demonstrates the highest binding affinity for P1, making the corresponding crystal structure 6ZL5 the most appropriate binding site model for this rigid receptor docking study.

Second, following the identification of P1 as a probable binding site for the synthesised ligands, the docking parameters were systematically optimised to accurately reproduce the binding pose of the co-crystallised reference ligand BI-2852. The finalised docking parameters are provided in the Experimental Section.

The left panel of **Figure 21** illustrates the overlay of the co-crystallised conformation of BI-2852 (red) with its docked pose (green) within P1 on KRAS$^{G12D,\ GDP}$. The right panel of **Figure 21** displays the allosteric binding pockets of KRAS, coloured as in **Figure 16**, to provide spatial orientation and facilitate interpretation of ligand placement within the protein surface.[8]

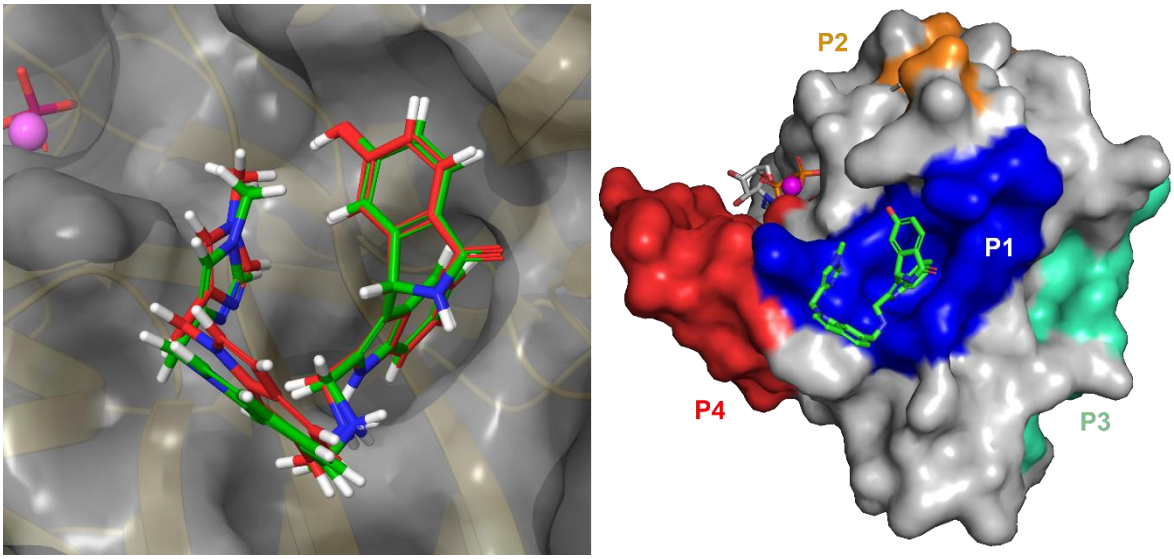


Figure 21: Left: superimposed conformations of BI-2852 (**7**) co-crystallised (red) and docked (green) to P1 at KRAS$^{G12D,\ GDP}$ (PDB ID: 6ZL5). GDP is depicted as red sticks in the background. Mg^{2+} is a magenta sphere. Right: binding pockets of 6ZL5 coloured according to WENG *et al.*[8]

The close spatial alignment of both conformations of Bi-2852 is an indication for the suitability of the established docking model for predicting accurate docking poses at P1 for ligands structurally related to BI-2852.

1.3.5.3 Discussion

In order to highlight the correlations between our *in vitro* and *in silico* findings, the DS and emodel values of up to ten poses per ligand were plotted alongside box plots summarising the NE^{G12D} assay results of the ligands listed in **Table 5** (**Figure 22**).

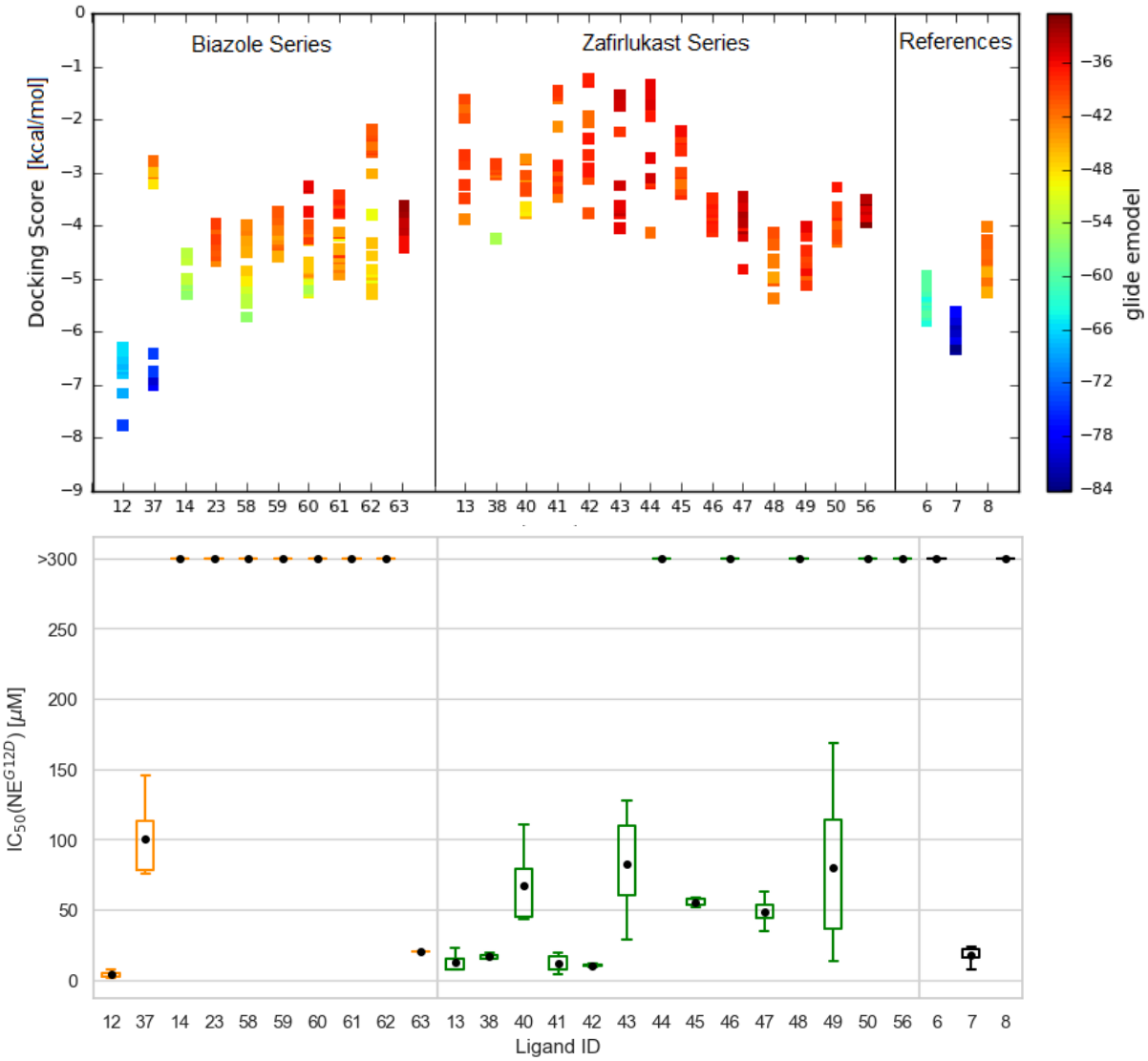


Figure 22: Top: summary of the DS and Glide emodel values of up to ten conformations per ligand, resulting from docking the biazole and zafirlukast series with published reference ligands to P1 in 6ZL5. Bottom: box plots of the assay results in **Table 5**, divided by scaffolds analogous to the upper plot. The mean is indicated by a black dot, the central 50% of the data (interquartile range) is indicated by a box, and the whiskers extend up to the 1.5-fold interquartile range.

The poses of biazoles **12** and **37**, as well as the reference **7**, are distinguished by their consistently large, negative docking scores (DS) and emodel values within narrowly distributed ranges. This suggests energetically favourable binding of these ligands at the P1 pocket. These computational findings align with their *in vitro* NE^{G12D} assay

results, which yielded IC_{50} values of 3.84±2.14, 100±32.1, and 17.0±9.31 µM, respectively. In contrast, the remaining biazole derivatives performed comparatively poorly both in silico and in vitro. A notable exception is derivative **63**, which demonstrated moderate NE^{G12D} inhibition with IC_{50} = 19.8 µM, despite DS (~4 kcal/mol) and emodel values (~36) that typically correspond to weak binders.

Nevertheless, the docking model appears sufficiently robust to differentiate between active and inactive biazole ligands at P1. A key discriminatory interaction seems to be the ionic coordination between the carboxylic acid moieties of active biazoles **12** and **37** and the Mg^{2+} cation situated opposite the β-phosphate of GDP (**Figure 23**).

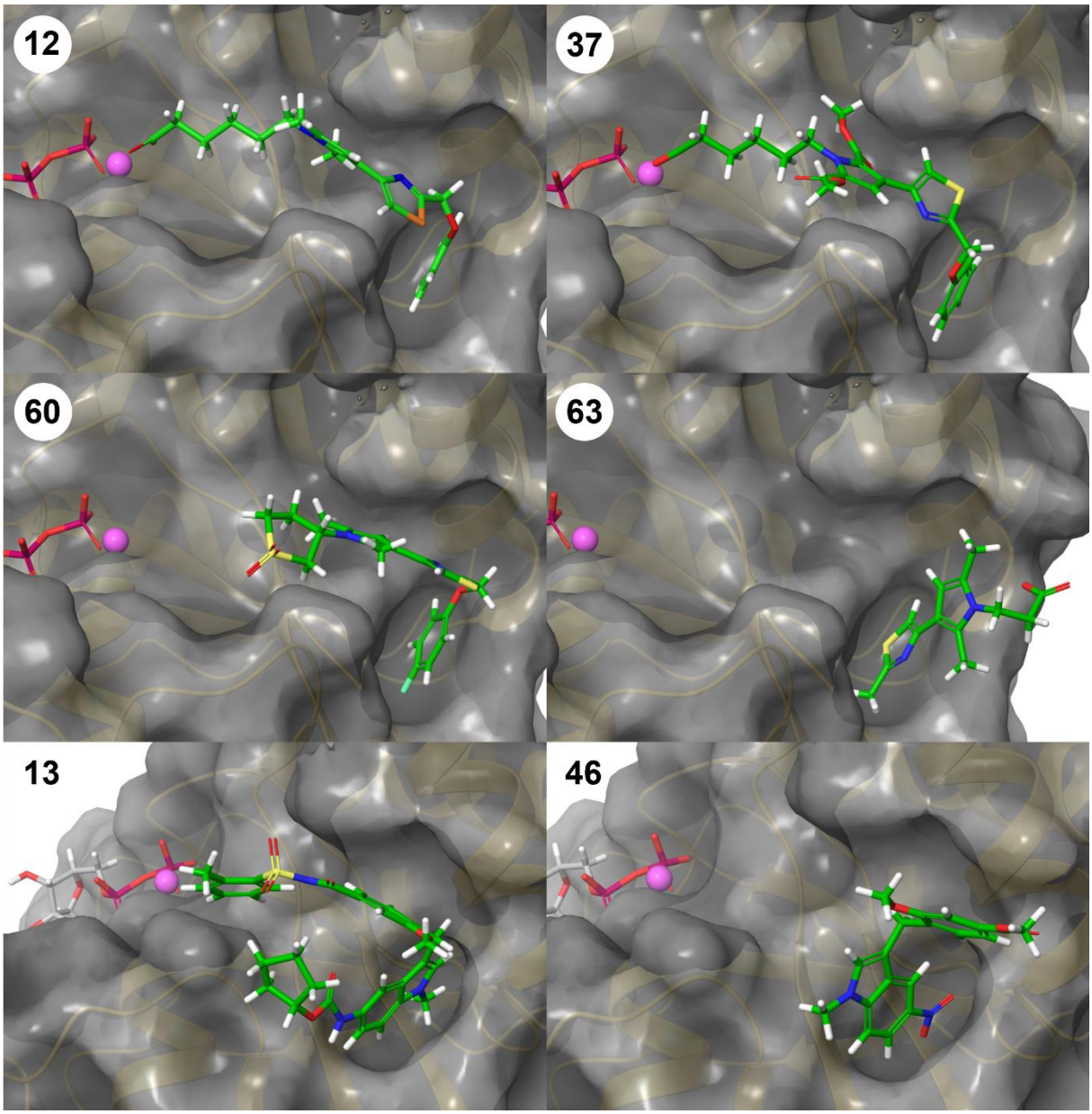


Figure 23: Predicted binding poses of biazoles **12**, **37**, **60**, and **63** as well as of zafirlukast (**13**) and its derivative **46** at P1 in 6ZL5.

The flexible hexanoic acid moiety in biazoles **12** and **37** enables simultaneous coordination of the Mg^{2+} cation and optimal positioning of the aromatic ring within the small, apolar P1 pocket. This dual interaction potentially contributes to the observed activity of these ligands in the NE^{G12D} assay. The preference of P1 for small aromatic scaffolds has been described previously by KESSLER *et al*.[95] Biazoles lacking the carboxylic acid moiety, i.e. derivatives **14**, **23**, and **58**–**62**, are inactive in the NE^{G12D} assay, presumably due to their inability to engage in Mg^{2+} coordination. Correspondingly, these compounds exhibit comparatively poor DS and emodel values. The docking pose of biazole **60** in **Figure 23** serves as a representative example. The docking poses of the remaining biazoles are similarly unremarkable and are therefore provided in the Appendix for completeness.

Notably, a similar NE^{G12D} inhibition mechanism, based on ionic interaction with the Mg^{2+} cation, was previously proposed by our research group for betulinic acid derivatives acting on the KRAS:SOS complex.[49] The apparent inconsistency between the docking and assay results for biazole **63** may stem from constraints imposed on the docking protocol to prioritise binding poses at P1. While biazole **63** may indeed operate *via* Mg^{2+} coordination, its structural length likely precludes simultaneous interaction with both P1 and the metal centre. Consequently, the Glide scoring function may disfavour such conformations energetically.

The successful synthesis of pyrrole diester **37**, a stable analogue of screening hit **12**, combined with the elucidation of a plausible mode of action for active biazoles, marks a significant advancement in the SAR understanding of this ligand class.

In contrast to the biazole series, no meaningful correlation was observed between the *in vitro* and *in silico* results for the zafirlukast scaffold, as illustrated in **Figure 22**. Specifically, none of the docking poses indicated complexation with the Mg^{2+} cation or other strong interactions, as exemplified by the representative poses of zafirlukast and its derivative **46** in **Figure 23**. The docking conformations of the remaining zafirlukast derivatives were similarly unremarkable and are presented in the Appendix.

Despite the lack of supportive docking data, a distinct pattern is apparent in the NE^{G12D} assay results: active zafirlukast derivatives consistently contain either a carboxylic acid moiety, i.e. derivatives **43**, **47**, and **49**, or an acyl sulfonamide group, i.e. derivatives **13**, **38**, **40**, **41**, **42**, and **45**. Acyl sulfonamides are regarded as bioisosteres of

carboxylic acids, typically exhibiting pKa values in the range 4–6.[138] Conversely, inactive derivatives lack such acidic functionalities. This suggests that NE^{G12D} inhibition in the zafirlukast series may depend on the presence of an acidic group capable of interacting with Mg^{2+}.

However, the absence of Mg^{2+} complexation in the docking poses implies that P1 is unlikely to be the relevant binding site for the zafirlukast derivatives. Instead, a different binding pocket, potentially one still proximal to the Mg^{2+} ion, may be responsible for the observed biological activity. Consequently, while the docking model is effective for predicting biazole interactions at P1, it appears unsuitable for zafirlukast-based ligands and should not be relied upon to infer binding conformations for this scaffold at the P1 site of $KRAS^{G12D}$ in 6ZL5.

Furthermore, the most potent NE^{G12D} inhibitors in the zafirlukast series (IC_{50}< 20 µM), i.e. derivatives **13**, **38**, **41**, and **42**, share a common *o*-toluenesulfonamide moiety. In contrast, derivatives **43**, **45**, **47**, and **49**, which lack this structural feature, showed reduced activity. This trend suggests that the *o*-toluenesulfonamide group plays a critical role in enhancing KRAS:ligand interactions. Conversely, structural features such as the cyclopentyl carbamate moiety and the *N*/*O*-methyl substituents present in zafirlukast appear non-essential for NE^{G12D} inhibition.

Importantly, although a reliable docking model for zafirlukast derivatives at P1 remains elusive, the biological data are nonetheless promising. Specifically, compounds **41** and **42** not only demonstrated superior NE^{G12D} inhibition compared to the P1 reference PPII, BI-2852, but even slightly exceeded the activity of their parent compound zafirlukast. This represents a significant advance in the SAR study of this scaffold.

Notably, an additional pattern emerges from the extended assay data of the biazole and zafirlukast derivatives, as summarised in **Table 6**. Specifically, derivatives that **reduce cell viability** in the **SNU-1** and/or **RKO** cell lines, i.e. derivatives **14**, **46**, and **48**, were found to be inactive in the NE^{G12D} assay. Conversely, derivatives that were inactive in the CTG assays, i.e. derivatives **12**, **13**, **37**, **41**, **47**, and **49**, demonstrated inhibitory activity in the NE^{G12D} assay.

ID	NE^{G12V} [µM]	NE^{WT} [µM]	CTG SNU-1 [µM]	CTG RKO [µM]
12	0.842*	9.97 ± 9.76	>30	>30
14	>300*	>300*	16.2 ± 1.15	n.d.
37	n.d.	283*	>30	>30
13	14.9 ± 6.11	14.03*	>30	>30
38	17.4*	25.3*	n.d.	n.d.
40	38.0*	61.7*	n.d.	n.d.
41	n.d.	15.0*	>30	>30
42	12.0*	13.9*	n.d.	n.d.
43	n.d.	46.6*	n.d.	n.d.
45	n.d.	80.4*	n.d.	n.d.
46	n.d.	>300*	1.51 ± 0.100	2.93 ± 2.22
47	n.d.	91.9*	>30	>30
48	n.d.	>300	3.72 ± 0.355	2.30 ± 0.185
49	n.d.	203*	>30	>30
50	n.d.	>300*	n.d.	n.d.
6	n.d.	>300	9.52 ± 1.52	>30
7	10.9 ± 0.655	33.1 ± 9.71	>30	>30
8	n.d.	>300	7.19 ± 0.836	7.28 ± 0.150

Table 6: Assay results of the biazole (orange) and zafirlukast (green) series and the reference molecules (white). The mean is reported with the standard deviation. Single measurements are marked with an asterisk.

One possible explanation for this inverse relationship is the potentially limited cell permeability of the active NE^{G12D} inhibitors. These compounds contain acidic functional groups, which are deprotonated at physiological pH, resulting in negatively charged conjugate bases that may hinder cellular uptake. This physicochemical property could explain their lack of cytotoxicity in cell-based assays, despite their efficacy in inhibiting nucleotide exchange *in vitro*.

The methyl esters **46** and **48** are expected to be uncharged under physiological conditions, which likely enhances their cell permeability compared to zafirlukast (**13**) and its derivative **41**. Once inside the cell, the esters may undergo hydrolysis, liberating the corresponding carboxylate anions, which could then engage in Mg^{2+} complexation, thereby inhibiting the nucleotide exchange.

In contrast, biazole **14** does not have a carboxylic acid or bioisosteric acid group, making Mg^{2+} complexation an unlikely mode of action. Given its structural resemblance to reference **6**, it is reasonable to assume that they share a common mode of action.

Finally, comparison of the NE IC_{50} values presented in **Table 5** and **Table 6** reveals that all tested derivatives, with the exception of compound **43**, exhibit either no or weak

selectivity (≤ two-fold) for $KRAS^{G12D}$ over the WT. This selectivity is comparable to the behaviour of reference **7**. In contrast, compound **43** demonstrated a two-fold preference for $KRAS^{WT}$. In the CTG assays, no selectivity was observed for any derivative, indicating comparable effects across cell lines, regardless of KRAS mutation status.

1.4 Summary and Outlook

In the first project of this thesis, a total of 16 potential KRAS PPIIs were designed and synthesised. Among these were six biazoles provided by LDC (compounds **58**-**63**) and three biazole derivatives (**14**, **37**, and **23**) synthesised in our laboratory *via* convergent Pd-catalysed coupling of five-membered heterocycles. Notably, the challenging nine-step synthesis of the pyrrole diester **37** represents a significant achievement in SAR exploration of the biazole series. The electron-withdrawing methyl esters on the pyrrole ring in compound **37** notably improved its solution stability compared to the oxidation sensitive HTS hit **12**.

Unfortunately, the NE^{G12D} inhibition observed for biazoles **12** and **37** did not translate into significant selectivity between KRAS mutants or into cellular activity through inhibition of downstream effector interaction. However, rigid receptor docking of all biazole derivatives at the P1 site on $KRAS^{G12D}$ (PDB ID: 6GJ8) proved to be a suitable *in silico* model for discriminating between biazoles with/without NE^{G12D} inhibition. The proposed key interactions contributing to NE^{G12D} inhibition include complexation of the Mg^{2+} cation by a carboxylate anion, alongside hydrophobic interaction of a benzene ring with the P1 pocket.

Among the synthesised molecules were also 13 zafirlukast derivatives, i.e. compounds **38**, **40**-**50**, and **56**. Notably, a significant yield improvement in the synthesis of the key intermediate **56**, up to 95% over two steps, was accomplished through sequential Sonogashira coupling. The NE^{G12D} inhibition SAR of the zafirlukast derivatives suggests a similar complexation of the Mg^{2+} cation as observed in the biazole series, with IC_{50} values in the low micromolar range. This pattern, however, was not reflected in the rigid receptor docking results of the zafirlukast derivatives at the P1 site in 6GJ8. Particularly noteworthy are derivatives **41** and **42**, which not only demonstrated superior NE^{G12D} inhibition compared to their parent compound zafirlukast, but also outperformed BI-2852, the currently most potent KRAS PPII binding at the P1 site. The

lack of mutant/WT selectivity of the zafirlukast remains an area for future optimisation of this promising scaffold.

Overall, the goal of studying the SAR of the biazole and zafirlukast scaffolds through synthesis of derivatives and their analysis *in vitro* and *in silico* was achieved successfully. In the process, the project library was enriched with both active and inactive NE^{G12D} inhibitors, providing valuable datapoints for the following QSAR studies.

2 Design of a KRAS PPII Prediction-Synthesis Pipeline

2.1 Introduction

Despite recent advances in KRAS PPII development, the demand for evermore streamlined screening pipelines persists. Modern medicinal chemistry has progressed well beyond brute-force screening approaches towards more efficient strategies that leverage the synergy between *in silico* modelling and laboratory synthesis. This project focussed on employing state-of-the-art machine learning techniques to derive quantitative structure-activity relationships (QSARs) from our project library. These QSARs enable the rapid evaluation of the inhibitory potential of millions of candidate KRAS PPIIs. Notably, libraries based on modular scaffolds, such as cyclic peptides, offer the opportunity to synthesise a broad spectrum of structurally diverse PPIIs using a limited set of robust synthetic protocols. The first part of this introduction aims to clarify key cheminformatics terminology, as some core concepts used in this work are less familiar to the average medicinal chemist. The second part outlines the advantages and limitations of the cyclic tetrapeptide (cyctetpep) scaffold as a platform for designing PPIIs.

2.1.1 Medicinal Chemistry in the Age of Data

Developing a new small-molecule drug in the United States costs, on average, approximately US$2 billion and takes around 15 years. While the profit margins of successful drugs are substantial, the pressure to continuously optimise the drug development process remains immense. Despite their decades-long history, computer-aided drug discovery tools have only in recent years gained widespread recognition as key driving forces for drug discovery in both academia and industry. This shift is largely attributed to recent technological breakthroughs in computing power and artificial intelligence.[150]

A particularly promising approach involves conducting *in silico* and *in vitro* screening campaigns in parallel. Virtual libraries allow the exploration of vast chemical space at low cost, while continuous feedback from targeted *in vitro* assays helps refine compound selection and keep drug development efforts aligned.[151] Although the methodologies used by medicinal chemists to conduct screenings have advanced significantly, the diversity of molecules selected for these campaigns has not. A comprehensive analysis of the ZINC database, which contains over 800 million unique

molecules, revealed a striking pattern: the majority of screening efforts still draw from a narrow range of scaffolds and structural motifs, particularly anilides and amides (**Figure 24**).[152]

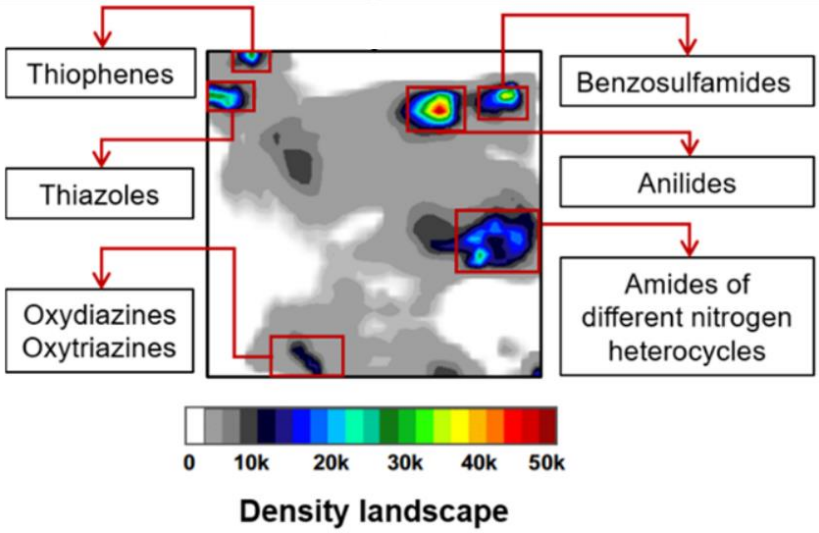


Figure 24: Density map of lead-like molecules in ZINC database with most frequent structural motifs. Reprinted from ZABOLOTNA *et al.*[152]

The widespread use of certain scaffolds in medicinal chemistry is largely due to their accessibility *via* reliable and well-established synthetic protocols, which significantly reduces the resources required for derivatisation. Moreover, the hybridisation of known bioactive scaffolds remains a common and effective strategy for lead optimisation.[153] However, this conventional approach has only scratched the surface of chemical space. Of the estimated 10^{33} possible drug-like molecules, only a minute fraction has been synthesised and characterised.[152]

Machine learning now offers the capability to identify promising, unexplored scaffolds beyond this narrow domain with greater reliability. This allows medicinal chemists to allocate resources more strategically, particularly toward overcoming potential synthetic challenges associated with novel structures. The rapidly evolving field of cheminformatics is at the forefront of these efforts, dedicated to developing tools that enhance scaffold discovery.[154]

The following three sections introduce key concepts in cheminformatics that underpin this project's approach.

2.1.2 Descriptors and Similarity

Molecular descriptors are numerical representations of structural features of molecules.[155] A familiar example is the logarithm of the partition coefficient (log*P*), a floating-point number classified as a zero-dimensional (0D) descriptor. This project primarily utilises 1D descriptors, specifically molecular fingerprints. Descriptors of higher dimensionality also exist, often represented as two-dimensional or higher-order matrices. A comprehensive overview of commonly used molecular descriptors is provided by TODESCHINI and CONSONNI.[156]

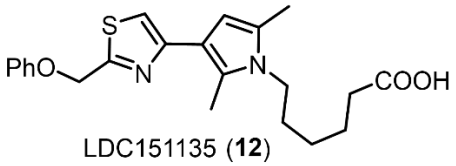
Fingerprints are a key class of 1D descriptors. They are Boolean arrays of fixed length, consisting of binary digits (*on* bits as 1s and *off* bits as 0s), where each position corresponds to a specific structural feature. The presence of a feature in a molecule is indicated by an *on* bit at the relevant position.

Figure 25 illustrates how structural similarity between molecules can be efficiently computed using fingerprints, i.e. RDKit fingerprints. A widely used similarity metric is the Tanimoto coefficient (T_c). It is calculated as the ratio of the number of shared *on* bits between two fingerprints (bothAB) to the total number of unique *on* bits in both molecules (onlyA + onlyB + bothAB) (**Equation 1**).[157,158]

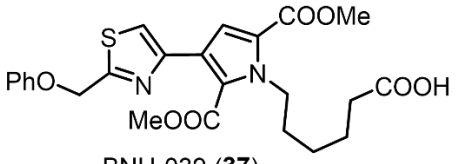
$$Tanimoto\ coefficient = \frac{bothAB}{onlyA + onlyB + bothAB} = \frac{473}{0 + 28 + 473} = 0.94$$

Equation 1: Definition of the T_c and its calculation for the RDKit fingerprints in **Figure 25**.[157]

The example in **Equation 1** shows T_c = 0.94 for biazoles **12** and **37** in **Figure 25**, reflecting their high structural similarity. The T_c ranges from 0 (no similarity) to 1 (identical structures).



LDC151135 (**12**)



BNH-039 (**37**)

Figure 25: RDKit fingerprints of length 512 of biazoles **12** and **37**. The fingerprints differ in 28 positions (red) and have 473 *on* bits in common.

It is important to note that different fingerprints can yield varying T_c values for the same pair of molecules. Consequently, selecting an appropriate fingerprinting algorithm is essential for the accuracy and relevance of the analysis. In this project, both RDKit and the extended-connectivity fingerprints (ECFPs) were utilised. The ECFP is among the most widely used and best-performing fingerprints in ligand-based virtual screening and target prediction.[159,160] In contrast, the RDKit fingerprint, when combined with the Tanimoto coefficient, has demonstrated superior performance in representing the similarity and diversity within chemical datasets.[161]

The ECFP and RDKit FP differ significantly in their generation algorithms. The ECFP is a circular fingerprinting approach based on a modified version of the Morgan algorithm.[162] It encodes structural information by iteratively updating a numerical identifier assigned to each atom, incorporating details of its atomic environment within a defined radius. At each iteration, the radius, representing the topological distance in terms of bonds, expands, progressively including more neighbouring atoms.[163] The top half of **Figure 26** schematically illustrated this process for the quaternary carbon atom of the benzene ring in benzoic acid amide.

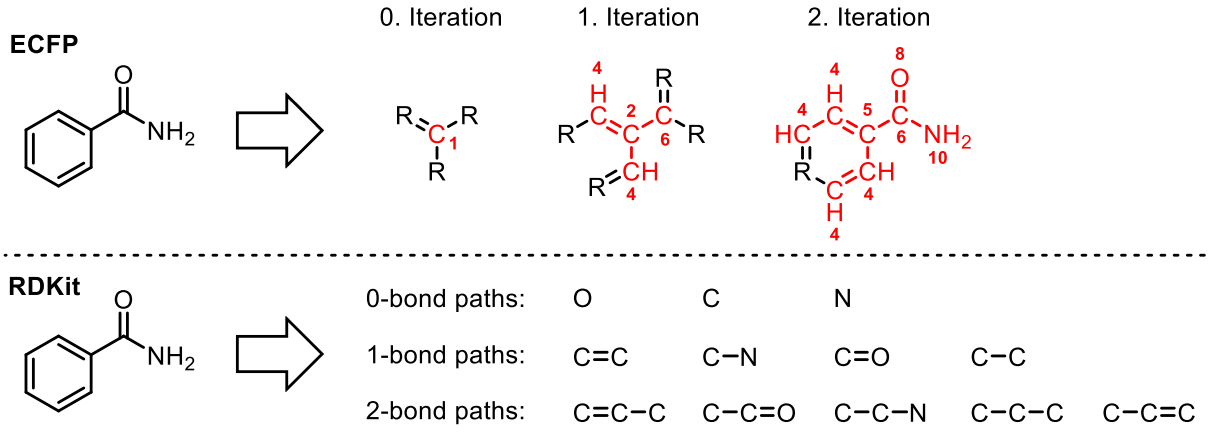


Figure 26: Schematic representation of the generation of ECFP and RDKit FP (top and bottom). Modified from ROGERS and HAHN as well as Daylight Chemical Information Systems website.[163,164]

It is important to note that the position indices depicted in **Figure 26** do not correspond to the actual numerical identifiers. In each iteration, only the atoms highlighted in red contribute to the updated identifiers. After all atoms in a molecule (e.g. benzoic acid amide) are processed, their final identifiers are hashed and aggregated to construct the ECFP. Commonly, a radius of two bonds and a fingerprint length of 512 bits are used, although these parameters can be adjusted depending on the application.

In contrast, the RDKit fingerprint is a path-based method that represents a modified implementation of the Daylight fingerprinting algorithm.[164] These subgraphs are then converted into short bit arrays through a process known as *hashing*. In the final step, the resulting bit arrays are processed and merged to yield the RDKit fingerprint of fixed length. Both the ECFP and RDKit FP methods, however, have a notable limitation: they do not inherently account for chirality.[165] Although recent advancements have introduced chirality-aware fingerprints such as MAP4C, effective utilisation of such methods requires a sufficiently large and diverse dataset containing chiral molecules.[166] Ideally, this includes both potent chiral inhibitors and their corresponding inactive stereoisomers to facilitate meaningful SAR modelling.

At present, the project library lacks an adequate number of chiral compounds to support the development of robust chirality-sensitive models. Moreover, as discussed in the introduction to the first project, potent KRAS inhibitors remain scarce overall, irrespective of their stereochemistry. Consequently, this project deliberately omits stereochemical information and focuses exclusively on atomic connectivity. Nonetheless, chirality-aware fingerprints such as MAP4C represent a promising future direction in KRAS PPII development, contingent upon the expansion of available chiral ligand datasets

2.1.3 Quantitative Structure Activity Relationship (QSAR)

The calculated descriptors define a multidimensional space, referred to as *descriptor space*, in which each dimension corresponds to one molecular descriptor. Since each descriptor captures a distinct structural aspect of a molecule, any compound can be represented as a unique point within this space (**Figure 27**). In a labelled dataset, each molecule is also associated with an activity measure, which is illustrated as colour coding in **Figure 27**. When the activity measure is categorical, such as *active* (red), *moderately active* (green), or *inactive* (blue), a classification algorithm is typically employed. This algorithm assigns each point in descriptor space to a discrete activity category. Conversely, if the activity measure is continuous, a regression algorithm is applied, assigning each point to a specific value within a predefined activity range.

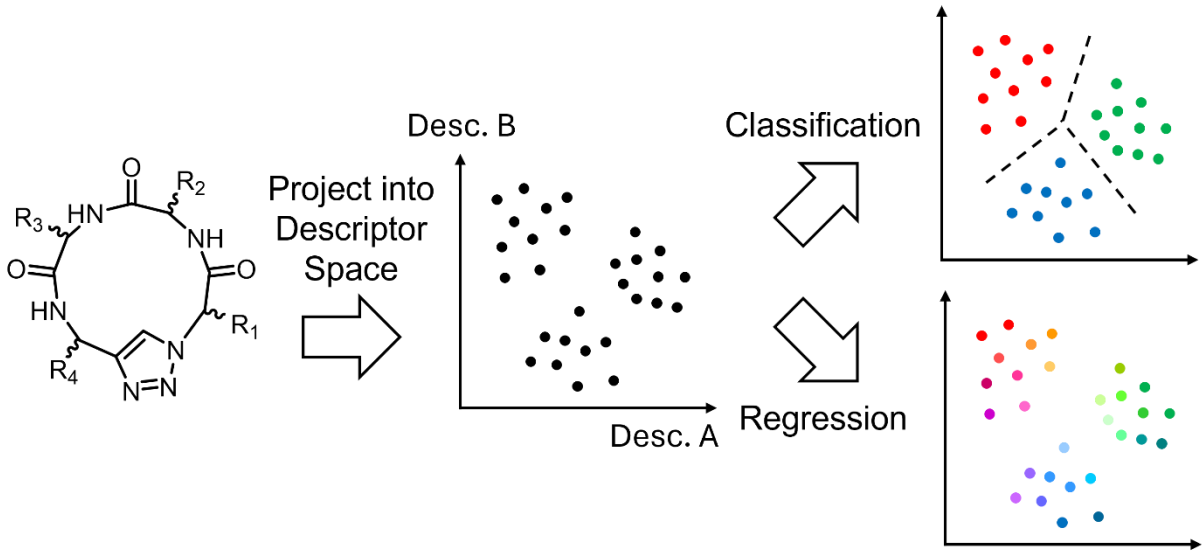


Figure 27: A library of click cyctetpep is represented as points in descriptor space. Applying a classification or regression algorithm to the obtained data points yields individually labelled data points/molecules.

The correlation between molecular structure of biological activity, as represented within descriptor space, is known as the quantitative structure-activity relationship (QSAR). For any given set of molecules, this relationship can be formalised as a mathematical function, where molecular descriptor values serve as inputs and the corresponding activity values as outputs (**Equation 2**).[167]

$$f(Descriptors) = Activity$$

Equation 2: QSAR models can be understood as a function of the employed descriptors. Structure-dependent variables can be predicted by such models, i.e. *in vitro* assay activities.

The function that most accurately estimates activity values for a labelled dataset is selected through a process known as *fitting*. At this stage, the computer has effectively "learned" the QSAR for that dataset. Subsequently, any unlabelled molecule, i.e. one without an associated activity value, can be projected into descriptor space, and an activity value can be assigned based on the learned model.

A widely used algorithm for both classification and regression tasks is the *Random Forest* algorithm. This approach involves constructing an ensemble of decision trees, often numbering 100 or more. **Figure 28** illustrates a hypothetical regression decision tree that employs some of the molecular descriptors used in this project; the specific definitions of these descriptors are not relevant at this point.

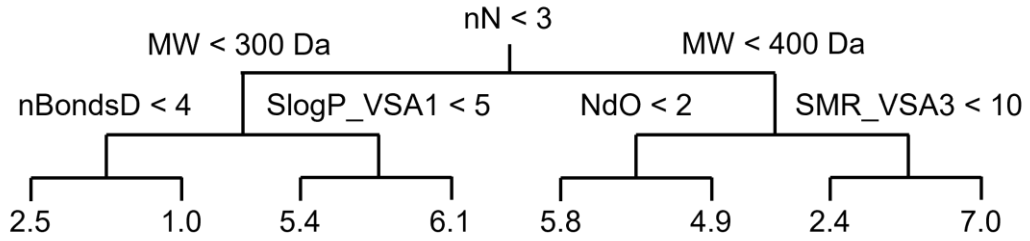


Figure 28: A regression decision tree utilising some of descriptors listed in **Table 8**.

The training dataset, such as the project library, is recursively divided into two subsets through a process known as *recursive binary splitting*. At each node, the algorithm selects a descriptor and a corresponding threshold that minimises the variation in activity values within the resulting subsets, relative to their mean activity. This splitting continues until a predefined stopping criterion is met. The mean activity values of the final subsets, referred to as *leaves*, are shown at the bottom of the tree.

In the case of a classification decision tree, each leaf would represent the predominant class among the data points it contains. When generating a Random Forest of, for example, 100 regression trees from a single dataset, many of the trees might otherwise be very similar. To introduce diversity, each decision tree is trained on a random one percent subset of the labelled dataset. Furthermore, during each recursive binary split, only a random subset of the available descriptors is considered. This dual randomisation, of data and descriptors, decorrelates the individual trees, significantly improving the robustness and generalisability of the model.

For any unlabelled molecule, the activity prediction is obtained by averaging the predictions from all trees in the forest.[168] However, QSAR can only identify promising PPIIs if the screened libraries contain compounds structurally suitable for KRAS inhibition. Without relevant chemical diversity in the input space, even the most sophisticated model cannot predict effective ligands.

2.1.4 Characterisation of Molecular Libraries

Understanding the composition of the project library is essential for selecting external libraries with complementary characteristics. In the context of molecular collections, the term *chemical space* broadly refers to the entirety of all conceivable molecules and their properties.[169] A standard approach to analysing the chemical space of a given library involves projecting a limited number of molecular descriptors into two or three dimensions (**Figure 29**).[170]

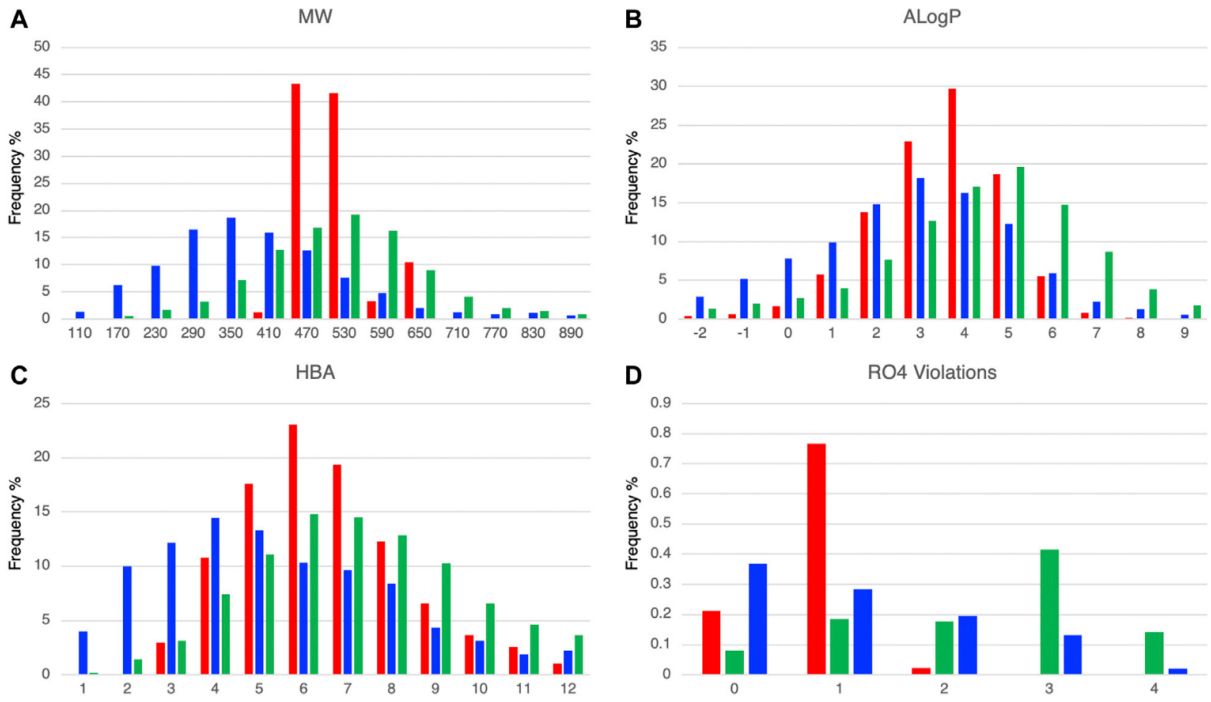


Figure 29: Comparison of three molecular libraries: Approved small molecule drugs (blue), DLiP-PPI (red) and known PPI modulators (green). Histograms of four molecular descriptors were compared: A) MW, B) AlogP (calculated log*P* value)[171], C) numbers of HBA and D) RO4 violations. Reprinted from IKEDA *et al.*[16]

In a recent study, IKEDA *et al*. illustrated the chemical space occupied by their DLip-PPI library (red), comparing it to libraries of approved small-molecule drugs (blue) and established PPI modulators (green).[16] The histograms in **Figure 29** highlight the partial overlap between the chemical spaces of small molecules and PPIIs. While small molecules typically conform to the rule of five (RO5), PPIIs tend to follow the rule of four (RO4), though a substantial number of PPIIs in the DLiP-PPI library deviate from the RO4, as seen in panel D. The DLiP-PPI library was intentionally designed to bridge the gap between RO5- and RO4-compliant compounds. In general, large molecules such as PPIIs suffer from low aqueous solubility and poor cell permeability, whereas small molecules frequently display inadequate binding at the wide, shallow surfaces characteristic of PPI interfaces. The authors aimed to identify compounds that combine the most favourable features of both chemical spaces and consolidated them into the innovative DLiP-PPI library. However, macrocycles were intentionally excluded from this collection of small PPIIs to maintain synthetic accessibility. While this decision is justifiable from a practical standpoint, it overlooks the significant potential of small macrocyclic structures as KRAS PPIIs; a potential that arguably outweighs the synthetic challenges they present.

2.1.5 Cyclic Peptides – A Privileged PPII Scaffold

The potential of macrocyclic scaffolds as PPIIs was briefly outlined in Section 1.1.1 The wide and shallow topology of typical PPII binding sites has been linked to the distinct structural characteristics of PPIIs compared to traditional small-molecule inhibitors. This has prompted the formulation of the RO4, which define a lower threshold of physicochemical properties associated with high affinity PPIIs (**Table 1**).[172]

Many cyclic peptides fulfil the RO4, making them well-suited for targeting protein-protein interactions. In addition, two key structural features distinguish them as privileged scaffolds for PPII development. First, they are composed of AA building blocks, which enables modular design. A diverse array of natural and non-natural AAs is commercially available, and further specialty AA analogues can be synthesised efficiently in the laboratory (**Figure 30**, left).[173]

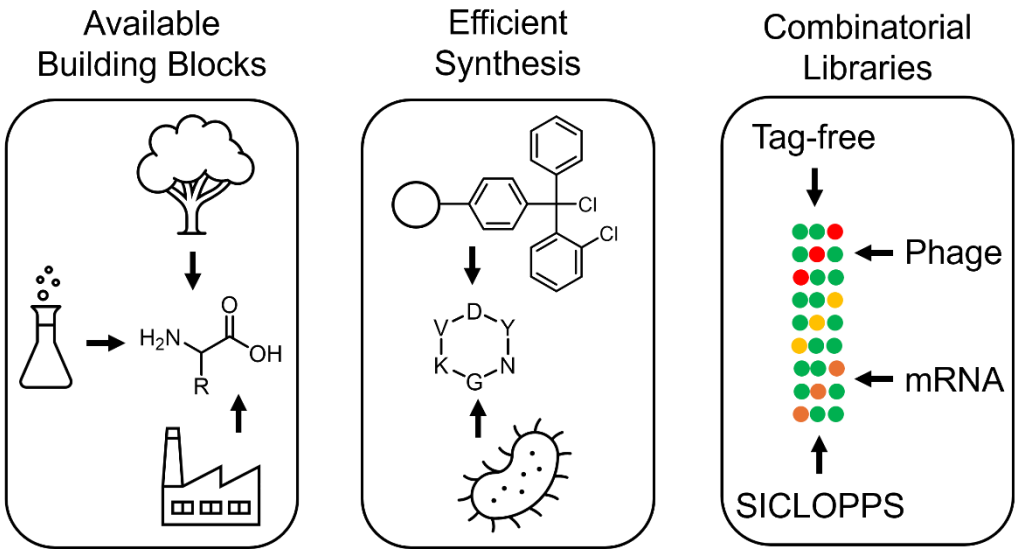


Figure 30: Advantages of AA-based macrocycles: Readily available AA building blocks (left), efficient (bio)chemical synthesis (middle) and diverse strategies for combinatorial library generation (right).

The AA building blocks can be coupled efficiently using well-established amide coupling chemistry (middle). When employing solid-phase peptide synthesis (SPPS), workup steps can be omitted, and on-resin cyclisation can prevent polymerisation of the linear, bifunctional precursors.[24,174] Both linear and some cyclic peptides are also accessible through biosynthetic expression.[175] This synthetic accessibility has enabled strategies for the combinatorial generation of cyclic peptide libraries. One such approach, the split-and-pool method, produces mixture containing up 10^6 unique cyclic peptide sequences.[176] High affinity binders are selected through target binding and

identified *via* LC/MS-MS analysis. Peptide identification can also be achieved using mRNA or phage display tags.[177,178] A particularly promising approach is the split-intein circular ligation of peptides and proteins (SICLOPPS). Therein, plasmids are employed to synthesise the cyclic peptides in the cells which contain the desired assay, which allows for rapid screening of up 10^9 sequences.[179]

The second key structural advantage of cyclic peptides lies in their conformational rigidity. Compared to their linear counterparts, cyclic peptides are more conformationally constrained, due to the amide-iminol tautomerism of peptide bonds and transannular interactions. As illustrated in **Figure 3**, only specific protein:ligand conformations are energetically favourable. Adopting these conformations results in an entropic cost for both ligand and protein, but the entropy loss is significantly lower for cyclic peptides. MILLWARD *et al*. reported a 15-fold enhancement in binding affinity upon cyclisation.[180] Additionally, the highly ordered conformations of cyclic peptides enable them to mimic secondary structural motifs of proteins, i.e. turns, helices and β-strands (**Figure 31**, left).[181] This structural mimicry facilitates high-affinity binding to wide, shallow protein interfaces, in some cases rivalling the binding affinities of antibodies.[180]

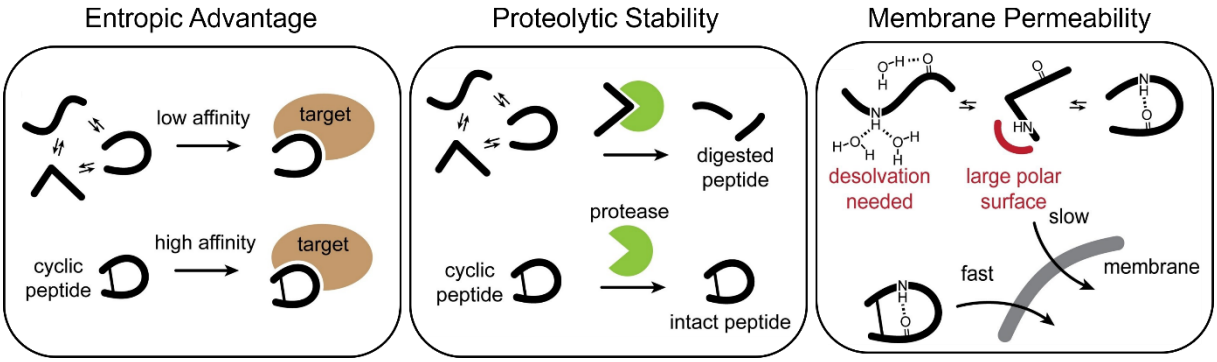


Figure 31: Advantageous effects of cyclisation: Higher binding affinity (left), proteolytic stability (middle) and increased membrane permeability (right). Modified from JI, NIELSEN and HEINIS.[182]

Cyclisation of peptides mitigates one of their key limitations, their limited serum stability. Linear peptides composed of L-AAs are rapidly degraded by proteases under physiological conditions, typically exhibiting serum half-lives of only 5–30 minutes. In contrast, cyclic peptides are far more resistant to enzymatic hydrolysis, resulting in significantly prolonged serum stability.[183] For instance, QIAN *et al*. reported a 40-fold increase in serum half-life, from 15 minutes to 10 hours, upon cyclisation.[184]

Beyond proteolytic stability, cyclisation facilitates the spatial proximity of hydrogen bond donors (HBDs) and acceptors (HBAs), promoting the formation of stable intramolecular hydrogen-bonding networks. This conformational constraint reduces the polar surface area and diminishes the hydration shell, both of which are known barriers to passive membrane diffusion. PRICE *et al*. demonstrated that linearisation of the cyclic peptide cyclosporin A led to a ten-fold reduction in cell permeability.[185] However, findings on this topic remain mixed. A systematic study by KWON *et al*. on the cell permeability of several cyclic peptides and their linear analogues found no consistent correlation between cyclisation and increased permeability.[186]

The mechanisms underlying oral bioavailability and membrane permeability of cyclic peptides remain an active area of research.[187–189] These properties represent the two most significant challenges impeding the clinical application of cyclic peptides.[190]

2.1.6 Cyclic Peptide KRAS PPIIs

TANADA *et al*. have reported the identification of a narrow region within the chemical space of cyclic peptide KRAS PPIIs that exhibit desirable pharmacokinetic properties. The most prominent example from their study is LUNA18 (**3**).[23] **Figure 32** presents a curated selection of cyclic peptide KRAS PPIIs discovered over the past two decades.

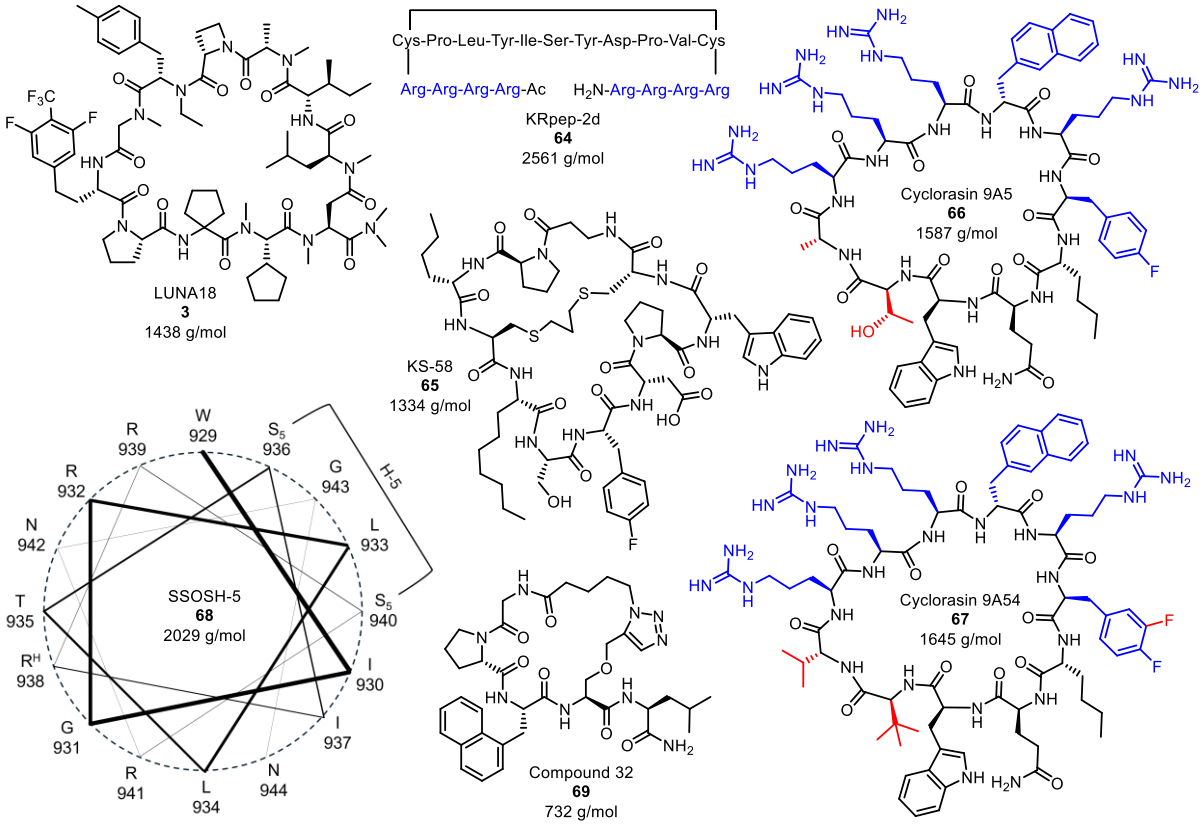


Figure 32: Selection of cyclic peptide KRAS PPIIs.

The 11-mer cyclic peptide **3** stands out with an oral bioavailability of 21–47% and a cell permeability of 2.3 x 10^{-6} cm/s. It inhibits the $KRAS^{G12D}$:SOS PPI with an IC_{50} < 2 nM and exhibits comparable potency in AsPc-1 cells ($KRAS^{G12D}$). However, mutant/WT selectivity was not reported. Crystallographic analysis of LUNA18 bound to $KRAS^{G12D}$ confirmed binding at the P2 site (**Figure 16**), and clinical trials are currently underway.[23] The 19-mer peptide KRpep-2d (**64**) shows similar inhibition of the $KRAS^{G12D}$:SOS PPI with an IC_{50} = 1.6 nM, while also exhibiting a 26-fold selectivity for $KRAS^{G12D}$ over the WT.[191] Although KRpep-2d reduces proliferation of A427 cells ($KRAS^{G12D}$) at 30 µM, no effect was observed in A549 cells ($KRAS^{G12C}$).[192] The inclusion of a cell-penetrating peptide (CPP) motif (blue) in peptide **64** significantly enhances its cell permeability; however, CPPs have been linked to adverse effects

such as mast cell degranulation.[193] Furthermore, the disulfide bridge in KRpep-2d contributes to plasma instability prompting the development of KS-58 (**65**), which binds at P2 similarly to KRpep-2d.[194,195]

The 11-mer **65** inhibits ERK phosphorylation in A427 cells ($KRAS^{G12D}$) to 26.0 ± 6.0 % at 30 μM, and to 57.6 ± 7.6 % in PANC-1 cells ($KRAS^{G12D}$) under the same conditions. KS-58 exhibits 3.6-fold selectivity for $KRAS^{G12D}$ over the WT, although its low water solubility limits *in vivo* application. To address this, injectable nano-formulations have been evluated.[196] A structurally similar CPP motif (blue) is also present in the 11-mer peptides cyclorasin 9A5 and 9A54 (**66** and **67**). Small structural modifications (red) between peptides **66** and **67** drastically improve inhibition of the $KRAS^{G12V}$:RAF PPI from an IC_{50} = 0.12 μM to 18 nM, although this enhancement comes at the cost of a five-fold reduction in cell permeability. Later findings suggest that compound **66** induces nonspecific protein unfolding, rather than specific binding to KRAS, a property likely shared by peptide **67**.[197–199]

Similarly, a stapled helix analogue of SSOSH (**68**) has also been shown to bind non-specifically.[199] Compound **68**, a pan-RAS inhibitor, mimics the αH helix of SOS1 (**Figure 10**, D, green) and reduces viability to <10% in a panel of RAS-mutant cell lines ($RAS^{WT/G12C/G12S}$, $NRAS^{Q61K/G13D}$, $HRAS^{G12V}$).[200]

At the opposite end of the molecular size spectrum, FUMAGALI *et al*. reported one of the smallest cyclic peptide KRAS PPIIs in 2021: the stapled 6-mer peptide **69**. It inhibits the KRAS:SOS or KRAS:RAF PPI with an IC_{50} = 2.4 μM and its cell permeability was demonstrated *via* antiproliferative activity in NCI-H358 ($KRAS^{G12C}$) and PC9 ($KRAS^{WT}$) cells at 50 μM.[201]

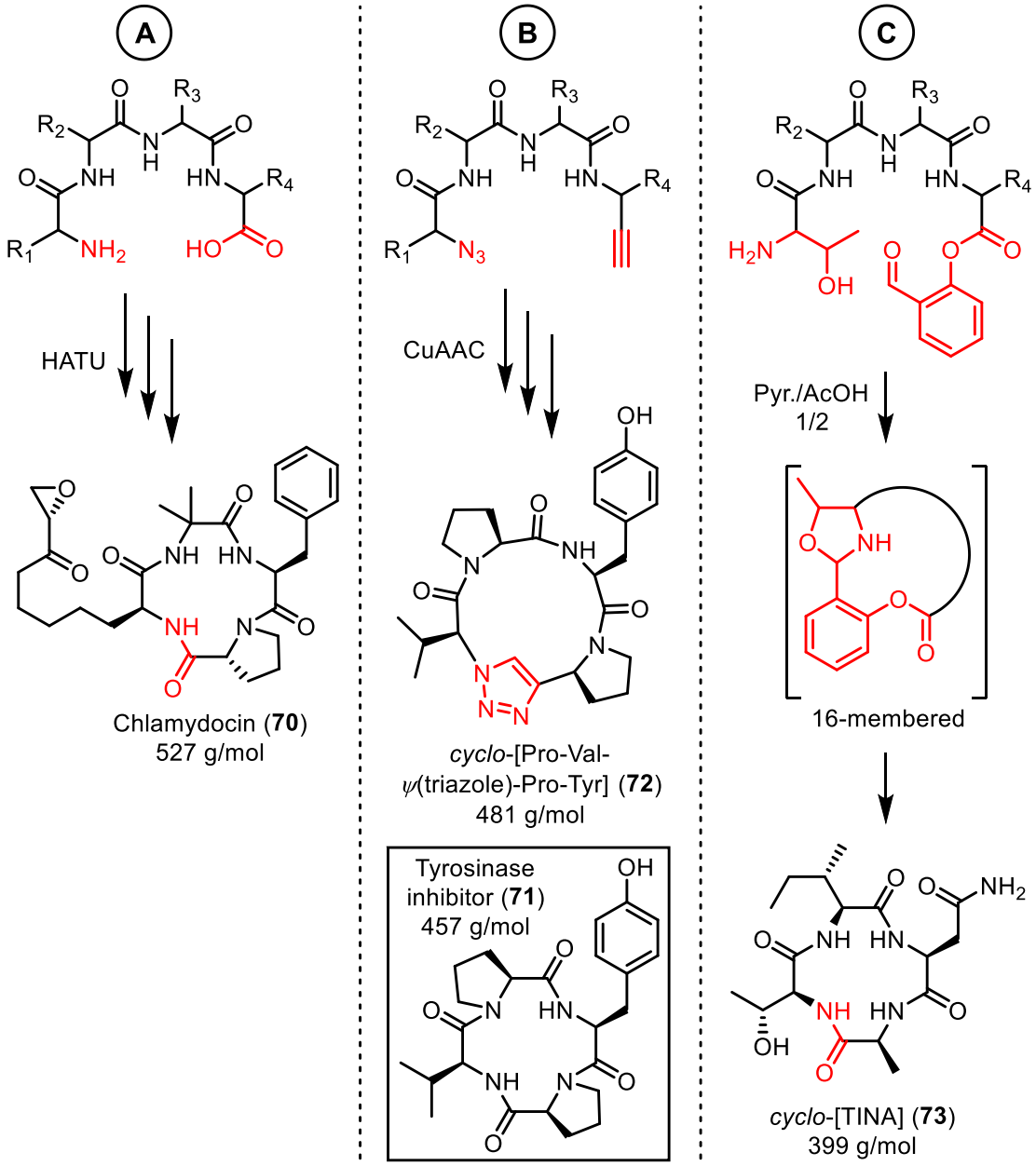
Collectively, the cyclic peptide KRAS PPIIs presented in **Figure 32** illustrate several critical insights. Despite their extended surface interactions, peptides **66**–**68** demonstrate that target specificity is not guaranteed. Moreover, high-affinity compounds like **65** often struggle with poor cell permeability and/or orally bioavailability. While CPP motifs offer a solution, their side effects limit their desirability, as evidenced by peptides **64**, **66**, and **67**. These findings define a key challenge in peptide drug design: engineering cyclic peptides that are small and lipophilic enough to cross membranes, yet large enough to maintain high binding affinity and specificity.

The development of LUNA18 has yielded a preliminary set of guidelines for peptides of comparable size. However, no such frameworks exist for smaller cyclic peptides that lie near the upper limit of the RO5. These peptides, which remain underexplored as KRAS PPIIs, may offer untapped potential. While the stapled peptide **69** offers a promising start, it also underscores the trade-off between size and binding affinity. Nonetheless, this does not imply that small PPIIs are inherently inferior in binding capability. For instance, MRTX-1133 (**2**, MW = 601 g/mol) successfully resides at the upper end of the RO5 range.

Consequently, the cyclic peptide KRAS PPIIs in **Figure 32** make a compelling case for exploring cyclic tri- and tetrapeptides as novel chemical probes for the RO5-RO4 boundary of the PPII chemical space. These intermediate scaffolds positioned between heavy *small molecules* and light PPIIs, potentially address KRAS in a completely novel fashion.

2.1.7 Synthesis Routes Towards Cyctetpep

Cyclic tetrapeptides have long been recognised as promising scaffolds for drug discovery, although on targets other than KRAS. The main barrier to their broader exploration is their challenging synthesis.[24] Due to amide-iminol tautomerism, these molecules experience significant ring strain, especially in 9–12 membered rings, making macrocyclization inherently difficult. Cyclisation yields are highly dependent on the peptide sequence, but they can be improved through various strategies, including: I) replacement of at least one amide bond by an ester bond, II) incorporation of D- and L-AAs and/or Gly, III) use of turn-inducing motifs (e.g., (pseudo)proline and *N*-methylation), IV) positioning the cyclisation site to be reactive and sterically accessible and V) choosing appropriate cyclisation reactions.[24] Proline exemplifies a turn-inducing motif that can enable reasonable yields in otherwise low-yielding tetrapeptide cyclisations. For instance, HATU-mediated cyclisation towards the histone deacetylase (HDAC) inhibitor chlamydocin A (**70**) proceeds with 56% yield, as shown in column A of **Scheme 25**.[202] However, proline's presence alone does not guarantee high yields; peptide molecular dynamics in solution, affected by sequence, solvent, and reagents, play a critical role in cyclisation behaviour.



Scheme 25: Published cyclisation reactions towards three cyctetpep.[202–204]

The tyrosinase inhibitor **71**, first isolated in 1993 from *Lactobacillus helveticus*, remains synthetically elusive despite extensive efforts.[205] To address this, Bock *et al*. reported the synthesis of the triazole analogue **72** *via* click cyclisation with a remarkable yield of 70% (**Scheme 25**, column B). The incorporation of a triazole moiety extends the ring to 13 carbon atoms, thereby reducing ring strain. Moreover, Cu^{I}-catalysed azide-alkyne cycloaddition (CuAAC) cyclisations likely involve multinuclear complexes, where Cu nuclei bridge the gap between the azide and alkyne termini, pre-organising them for efficient cyclisation.[206] Despite this success, click cyclisation requires two unnatural AAs, making it less desirable for large-scale or combinatorial synthesis.

Alternatively, peptides with *N*-terminal serine or threonine and a *C*-terminal salicylaldehyde ester can be cyclised through imine-induced ring contraction. WONG *et al*. successfully used this approach to generate a library of all-L cyctetpep, such as *cyclo*-[TINA] (**73**), with yields in the range 6–27% (**Scheme 25**, column C).[204] The authors hypothesise that the imine-induced ring contraction proceeds *via* a 16-membered intermediate, significantly reducing the activation barrier for the cyclisation. Computational calculations of the free energy changes during the reaction mechanism support this hypothesis.

Importantly, the three cyclisation examples shown in **Scheme 25** represent only a small subset of the numerous protocols developed over the past two decades.[207–209] Selecting the most suitable method for synthesising a given cyclic peptide remains largely a trial-and-error process. As such, building a comprehensive understanding of reliable macrocyclisation strategies is essential for efficiently unlocking the potential of this promising scaffold.

2.2 Aim

The second project of this thesis builds upon the SAR information accumulated in the project library, including the data generated in the first project. The principal objective was to develop a QSAR model capable of identifying novel KRAS PPIIs with greater accuracy and efficiency than conventional molecular docking methods. In parallel, the project aimed to establish the synthetic foundation necessary for the preparation of a derivative library based on a promising scaffold within our research group. To achieve this, the project was structured into four sequential phases:

I) Rapid and accurate prediction of the inhibitory potential of millions of prospective KRAS PPIIs sourced from specialised libraries.

II) Identification of an accessible scaffold exhibiting a privileged PPII shape and strong predicted activity.

III) Development of a robust and modular synthetic route to generate structurally diverse derivatives of the selected scaffold.

IV) Experimental evaluation of these derivatives.

A schematic overview of this project is shown in **Figure 33**.

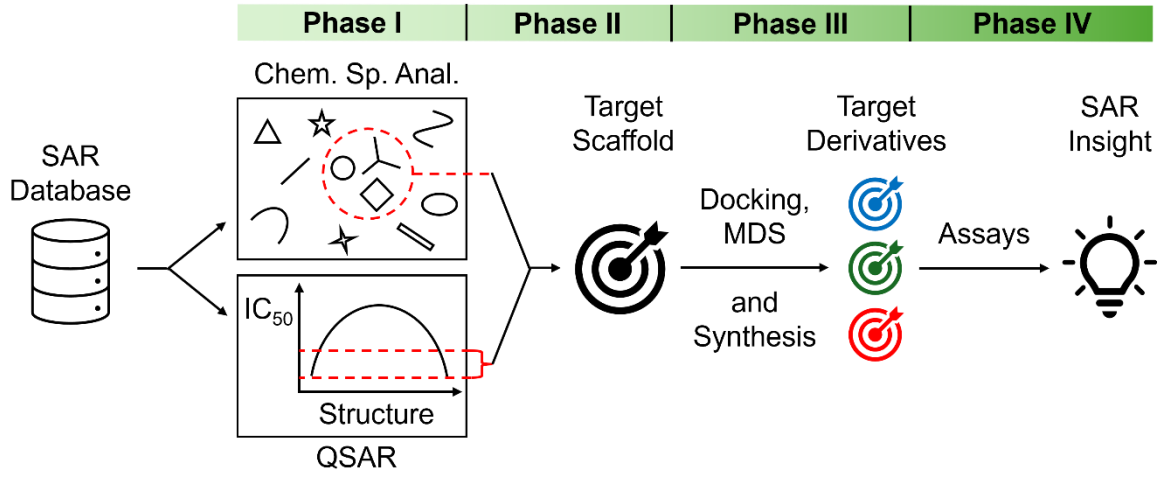


Figure 33: Schematic structure of the second project: I) Extraction of QSAR from project library and prediction of inhibitory potencies, II) Identification of target scaffold. III) Modular synthesis of target derivatives, IV) Testing of target derivatives.

The four phases were subdivided into a sequence of seven defined steps, which were executed in the following order:

1. Analyse the chemical space covered by the project library to enable meaningful comparison with structurally diverse PPII and cyclic peptide libraries.

2. Supplement the project library with curated external SAR data of known KRAS PPIIs, thereby increasing dataset diversity and robustness.
3. Construct a regression model with high predictive accuracy for estimating the inhibitory potency of candidate molecules, optimised for large-scale virtual screening.
4. Select a scaffold suitable for synthesis and identify target derivatives with favourable binding profiles using rigid receptor docking.
5. Evaluate KRAS-ligand interactions through MDS to gain insight into binding stability and interaction residues.
6. Synthesize the identified derivatives, employing a robust and modular approach to generate structural analogues efficiently.
7. Derive future scaffold optimisations based on insights from both *in silico* modelling and *in vitro* activity assessments.

2.3 Results and Discussion

2.3.1 Chemical Space Analysis

The chemical space analysis presented in the following section was conducted collaboratively by my colleague Sascha Koller and myself. The results are the outcome of an equally distributed joint effort. All subsequent *in silico* analyses, beginning with Section 2.3.2, were carried out independently by me. The structure and methodology of this section were informed by established guidelines for chemical space analyses of molecular libraries as reported in the literature.[170,210] The primary objective was to identify regions of chemical space not yet explored by the project library. The insights gained from this analysis are intended to guide future virtual screening campaigns and synthetic efforts, improving the strategic coverage of relevant chemical space.

2.3.1.1 Selection of Molecular Libraries

The first step of the chemical space analysis involved the selection of suitable reference libraries against which the project library could be compared. To ensure comprehensive benchmarking, three types of reference libraries were chosen: a maximally diverse library, a PPII-focused library, and a cyclic tetrapeptide (cyctetpep) library. Structurally diverse libraries of small molecules adhering to the RO5 are widely accessible; for this analysis, the DivSet provided by the company ChemDiv was selected as the diverse small molecule reference.

In contrast, the PPII chemical space remains relatively underexplored, with only a few curated libraries available.[211] Notably, the TIMBAL[18] and 2P2I[20] databases are either no longer accessible or too limited in size (the latter containing only 242 compounds). Larger and more relevant libraries such as Fr-PPIChem and DLiP-PPI would have been ideal candidates due to their size and DLiP-PPI's specific focus on the RO5/RO4 interface (see Section 2.1.4).[16,17] However, Fr-PPIChem is commercially restricted and DLiP-PPI cannot be downloaded from its web interface.

Three smaller but accessible PPII-focused libraries were therefore considered: iPPI[212], Enamine PPI[213], and Reinvent PPI[172]. The iPPI library is an open-source project containing published PPI modulators. The Enamine PPI library is the digital version of a commercial PPII library curated by the company Enamine. The Reinvent PPI library has been created entirely *in silico* and is not backed up by synthesised molecules.

Among them, the iPPI library, an open-source collection of published PPI modulators, was selected. This choice aligns with best practices for publicly funded research, particularly the FAIR principles: Findability, Accessibility, Interoperability, and Reproducibility.[214]

In addition, a custom click cyclic tetrapeptide (click cyctetpep) library was generated *in silico*. Using the Konstanz Information Miner (KNIME) software, a library of all permutations of 20 proteinogenic AAs in the click cyctetpep scaffold was created.[215] A detailed description of the workflow used for library generation is provided in the Experimental Section. A summary of the libraries included in this analysis is presented in **Table 7**.

Library	Entries	Downloaded/Created	Source
Project	1208	01.04.24	LDC
ChemDiv DivSet	50 000	14.05.24	Company
iPPI	2426	14.05.24	Open source
Click Cyctetpep	152 000	19.01.24	In-house

Table 7: Molecular libraries used for the chemical space analysis of the project library.

2.3.1.2 Structure Preparation

The second essential step in any data analysis process is the preparation of the input data, commonly referred to as *structure washing*. Inspired by two published structure washing pipelines, we designed a custom, user-friendly KNIME workflow specifically adapted to our input structures.[216,217] This workflow ensures compatibility and consistency across all molecular entries. A schematic overview of the structure preparation process is shown in **Figure 34**, which outlines the pipeline in three main steps.

Figure 34: Schematic representation of the library washing workflow developed in this project.

The first step involved the removal of inorganic compounds such as metal complexes, covalent inhibitors, i.e. Michael acceptors, unnecessary attributes like chirality, and structurally incorrect entries. Organic salts were converted to their corresponding free bases or acids. The second step focused on standardising the molecular representations. Since computational systems differentiate between resonance forms,

e.g. the two resonance structures of benzene, it is crucial to define a single representation for each molecule to ensure consistency across the dataset. In the third step, duplicate structures were eliminated, particularly those that arose from the removal of chirality and the unification of resonance forms. All molecular libraries used in this project underwent the same standardised structure-washing pipeline. A detailed description of the KNIME workflow employed can be found in the Experimental Section. The workflow itself can be found on GitHub.[218]

2.3.1.3 Comparison of Molecular Descriptors

The RO5 and RO4 comprise six molecular descriptors that are relevant for a ligand's binding affinity, solubility and membrane permeability. The descriptors, along with their respective thresholds, are summarised in **Table 1**. As an initial step in the chemical space analysis, the number of molecules in each library that comply with or violate the RO5 and RO4 criteria was determined. **Figure 35** illustrates the proportion of molecules that satisfy (green) or violate (red) at least one of the RO5 conditions (left) or RO4 conditions (right).

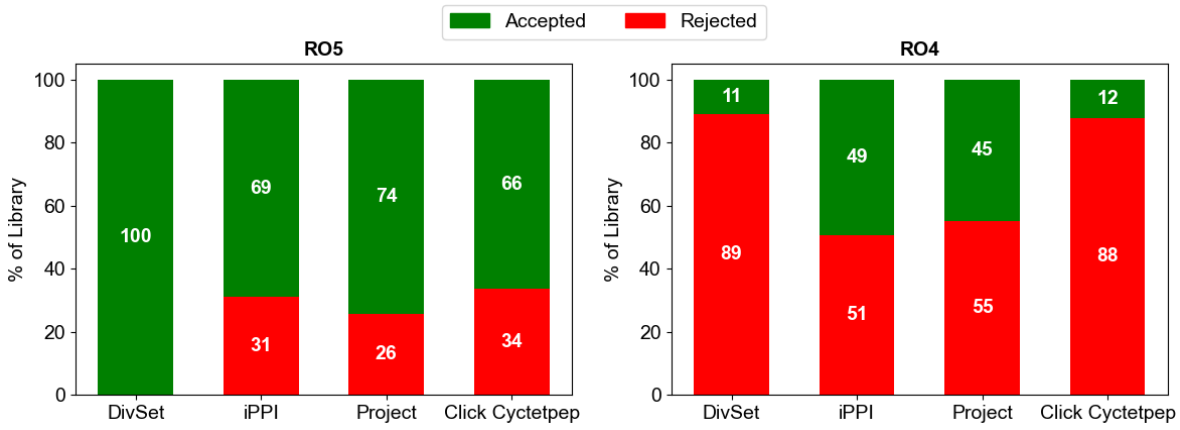


Figure 35: RO5 and RO4 compliance of the molecular libraries in percent, i.e. accepted (green) and rejected (red). Libraries from left to right: DivSet, iPPI, project and click cyctetpep.

As expected, the DivSet library exhibits full compliance with the RO5 criteria, as it was explicitly designed for this purpose. In contrast, the other three libraries show varying degrees of RO5 violations: 31% of the iPPI library, 26% of the project library, and 34% of the click cyctetpep library contain at least one RO5 violation. Regarding RO4 compliance, the iPPI and project libraries contain a higher proportion of RO4-compliant structures, at 49% and 45% respectively, compared to only 12% in the DivSet and 11% in the click cyctetpep libraries. The similarity in RO5 and RO4 compliance between the project and iPPI libraries is encouraging, as it supports the observed moderate to good

KRAS PPI inhibition for several scaffolds within the project library. In contrast, the lower-than-expected RO4 compliance of the click cyctetpep library is primarily attributable to their low SlogP values, which are log*P* values calculated according to a method published by WILDMAN and CRIPPEN.[219,220] **Figure 37** shows histograms of the SlogP and MW distributions for the four libraries: DivSet (blue), iPPI (green), project (orange) and click cyctetpep (red).

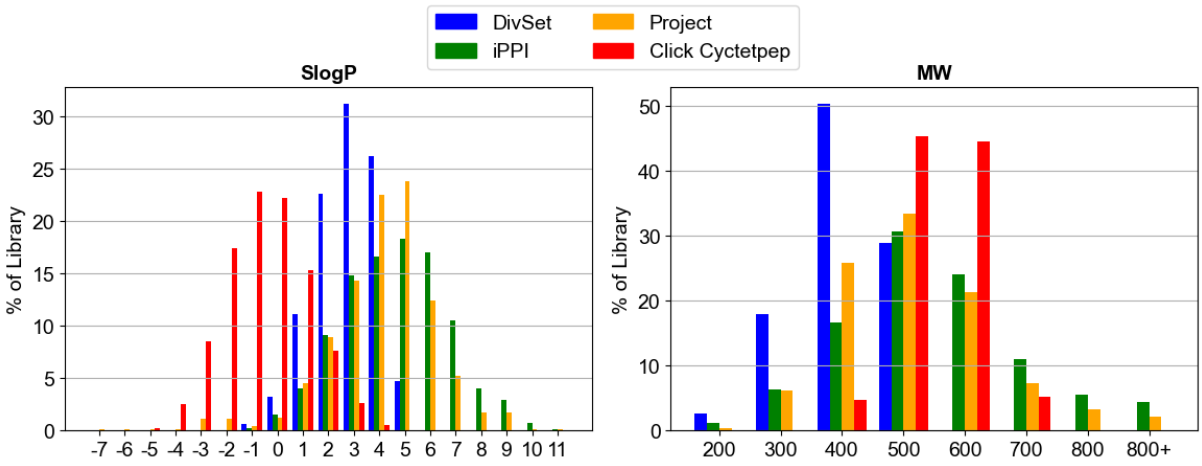


Figure 36: Histograms of SlogP (left) and MW (right) distributions of the four libraries DivSet (blue), iPPI (green), project (yellow) and click cyctetpep (red).

The iPPI and project libraries display SlogP distributions centred around 4.5, indicating a higher average lipophilicity compared to the DivSet library, which centres around an SlogP of 3. In contrast, the click cyctetpep library exhibits markedly lower SlogP values, clustering around 0, reflecting a considerably higher hydrophilicity than the other three libraries. Interestingly, despite their low lipophilicity, the click cyctetpep compounds fall well within the typical molecular weight range for PPIIs. This suggests a higher proportion of heteroatoms contributing to their overall molecular weight compared to the other libraries. This trend is further supported by the H-bond acceptor (HBA) and H-bond donor (HBD) distributions, as shown in (**Figure 37**).

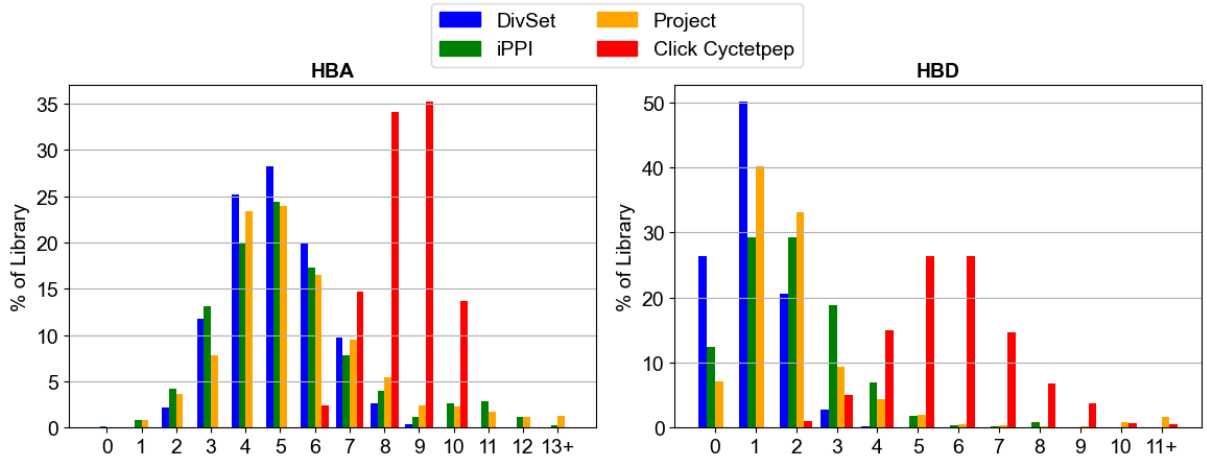


Figure 37: Histograms of HBA (left) and HBD (right) distributions of the four libraries DivSet (blue), iPPI (green), project (yellow) and click cyctetpep (red).

The majority of the click cyctetpep compounds contain 8 to 9 HBAs and 5 to 6 HBDs, whereas most molecules in the other three libraries typically feature 4 to 6 HBAs and only 1 to 2 HBDs. As a result, the click cyctetpep structures are significantly more polar than the average small molecule drug or PPII, with an approximate mean SlogP difference of ~4. Additionally, the click cyctetpep exhibit a higher number of rotatable bonds (NRB) and contain fewer cyclic motifs relative to the typical PPII compounds, as illustrated in **Figure 38**.

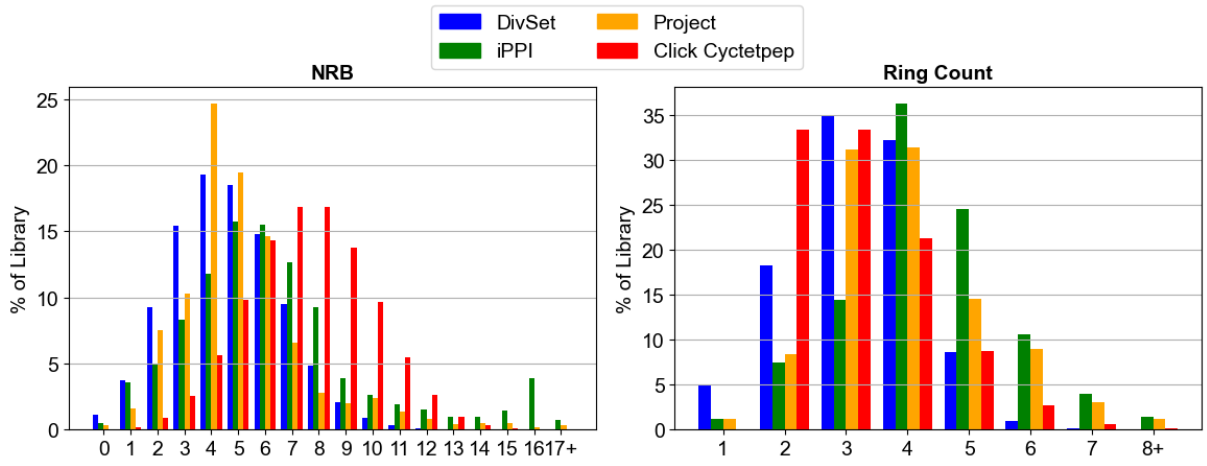


Figure 38: Histograms of NRB (left) and ring count (right) distributions of the four libraries DivSet (blue), iPPI (green), project (yellow) and click cyctetpep (red).

The amide-iminol tautomerism imparts rigidity to the macrocyclic cyctetpep backbone, implying that the elevated NRB observed in the click cyctetpep library, relative to other libraries, is primarily attributable to the side chains of the constituent AAs. Therefore, synthesising click cyctetpep exclusively from AAs bearing aromatic side chains could reduce the NRB while simultaneously increasing the number of ring systems.

Furthermore, the reduced number of heteroatoms in aromatic side chains would lower the counts of HBAs and HBDs, thereby shifting the SlogP values closer to those typically observed for PPIIs. Methylation of the four backbone nitrogen atoms would further enhance the lipophilicity of the click cyctetpep. Collectively, these modifications would enable the design of click cyctetpep with physicochemical properties aligned with PPIIs, while preserving the structural and pharmacokinetic advantages of macrocyclic drugs discussed in Section 2.1.5. In particular, the unique three-dimensional architecture of click cyctetpep positions them as a valuable addition to the chemical space represented in the project library.

2.3.1.4 Principal Moment of Inertia (PMI)

In 2003, SAUER and SCHWARZ introduced one of the most effective approaches for comparing molecular shape *in silico,* through the analysis of principal moments of inertia (PMIs).[221] Mechanically speaking, the moment of inertia quantifies the torque required to induce rotational acceleration around a given axis.[222] For instance, a linear molecule such as ethyne exhibits its smallest moment of inertia along the axis parallel to its triple bond, as it offers minimal resistance to rotation around this axis. Conversely, rotations around axes orthogonal to this bond require greater torque and thus yield larger moments of inertia.

PMI analysis was conducted for the four molecular libraries listed in **Table 7** using a KNIME-based workflow. A detailed description of this workflow is available in the Experimental Section. In the initial step, ten conformers were generated per structure, followed by geometry optimisation of each conformer. Subsequently, the PMIs for the three spatial dimensions ($I_{1\text{-}3}$) were calculated for each optimised conformer. To remove size dependency, the two smaller PMIs (I_1 and I_2) were normalised by dividing them by the largest PMI (I_3), yielding the dimensionless ratios $NPR_1 = I_1/I_3$ and $NPR_2 = I_2/I_3$. In the final step, these ratios were plotted in a two-dimensional space.

Figure 39 illustrates the resulting PMI plots for the reference molecules MRTX-1133 (left) and Ch-3 (right), with each data point representing one of the ten generated conformers.

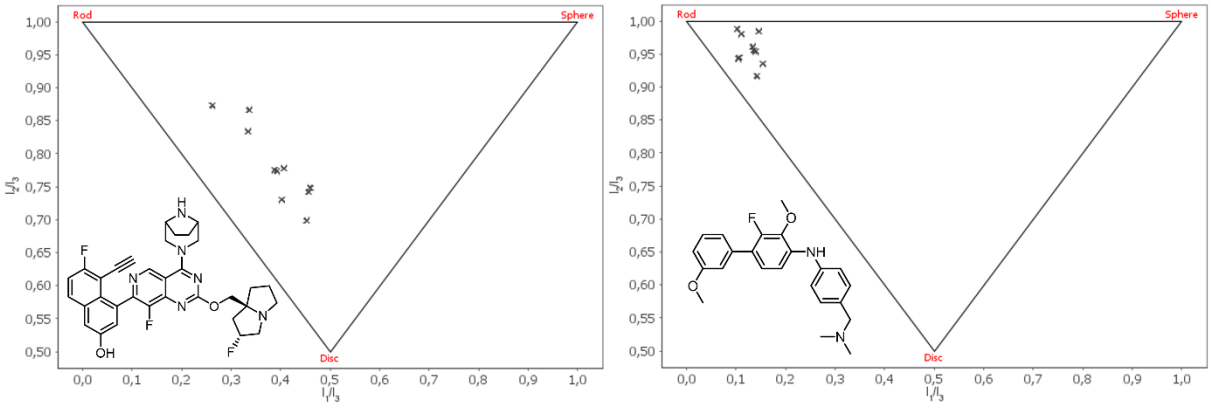


Figure 39: PMI scatter plots of MRTX-1133 (left) and Ch-3 (right) with ten conformers each.

The inclusion of multiple conformers per structure is essential, as it captures the breadth of molecular shapes that a compound may adopt due to Brownian motion in solution. Generating ten conformers per molecule ensures a more comprehensive representation of its shape flexibility. For example, MRTX-1133 is a rigid, planar, star-shaped scaffold that occupies a region in the PMI plot intermediate between an ideal rod (I_2/I_3 = 1 and I_1/I_3 = 0, top left) and an ideal disc (I_2/I_3 = I_1/I_3 = 0.5, bottom centre). In contrast, the linear and rigid molecule Ch-3 (8) exhibits conformers that cluster near the rod-like region of the plot, reflecting its limited shape adaptability.

By overlaying the PMI plots of all molecules within the four libraries, a general overview of their respective shape distributions can be obtained. **Figure 40** displays the PMI plots of the DivSet, iPPI, project, and click cyctetpep libraries (arranged from left to right and top to bottom). At first glance, it is evident that the DivSet library spans a broad range of shapes, including rod- and disc-like structures, with a modest presence of spherical conformers. In contrast, the iPPI and project libraries predominantly feature rod-shaped molecules, with the spherical region remaining largely unpopulated in all three of these libraries.

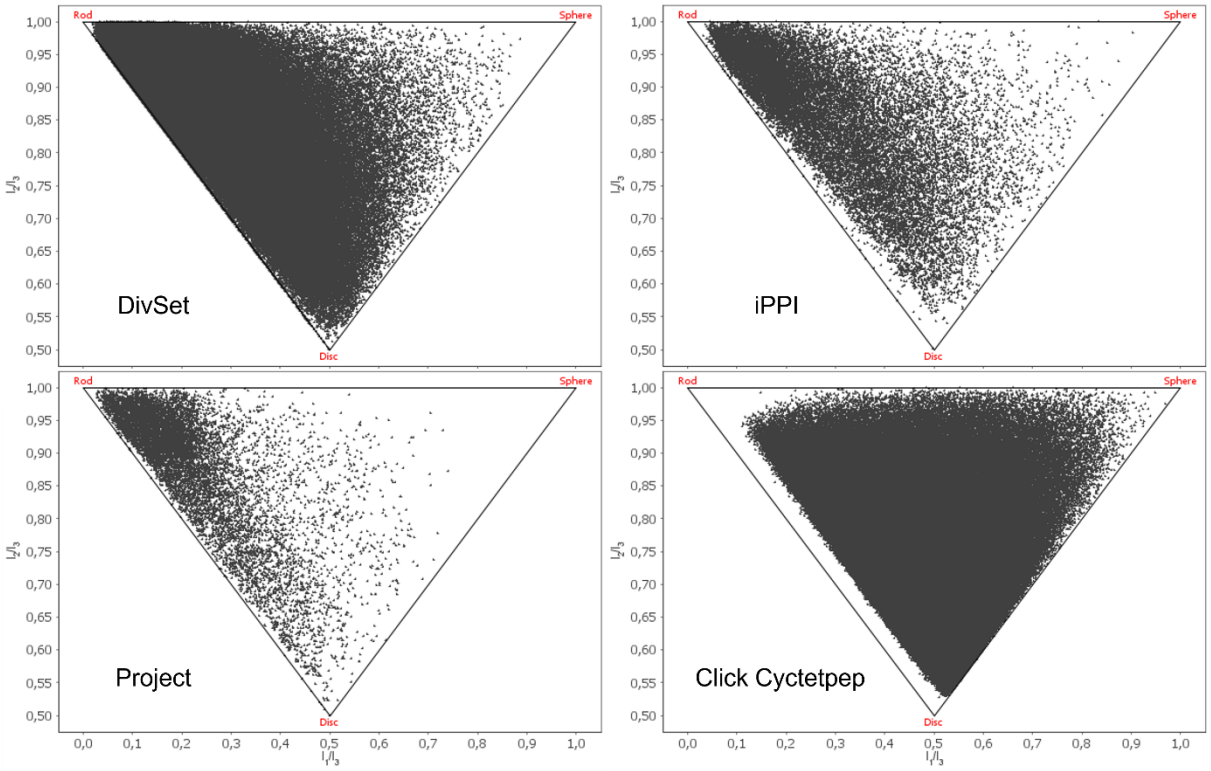


Figure 40: PMI scatter plots of the four libraries DivSet, iPPI, project and click cyctetpep with ten conformers per molecule (left to right and top to bottom).

Notably, the click cyctetpep library presents a complementary shape profile, characterised by a complete absence of strictly rod-shaped molecules. This is attributed to the rigid, disc-like macrocyclic backbone of cyctetpep structures, from which four side chains extend orthogonally. As a result, the library comprises a substantial proportion of disc-shaped and spherical conformers. Remarkably, even the structurally diverse DivSet library does not populate this region of shape space as thoroughly as the click cyctetpep library. This unique shape distribution provides the click cyctetpep library with a distinct advantage over the other libraries, enabling *in silico* screening within an underexplored region of chemical space.

In summary, the chemical space analysis positioned the project library reassuringly within the broader PPII space. However, it also revealed a notable deficiency in shape diversity, as the library is predominantly composed of rod-like molecules. In contrast, the click cyctetpep library was shown to occupy a complementary and underrepresented region of shape space, highlighting its value as an addition to the project library. Despite this advantage, its molecular descriptor profile suggests that, in its current form, it does not adequately span the desired interface between the RO5 and RO4 chemical spaces. Nevertheless, due to the prioritisation of establishing a

high-yielding synthetic route, the unmodified click cyctetpep library was retained for further development. The selection of target (click) cyctetpep derivatives is outlined in the following sections.

2.3.2 Python QSAR

In 2024, DUO *et al*. published a Python-based QSAR protocol for the identification novel SOS1 inhibitors.[223] The Python QSAR model developed in this project builds upon and significantly expands the methodology presented by DUO *et al*., adapting it to the specific task of efficiently identifying novel KRAS PPIIs. To enhance the model's predictive accuracy, several key modifications were introduced. Most notably, a molecular descriptor selection procedure was added to complement or even replace Extended Connectivity Fingerprints (ECFP). This enhancement substantially improves the robustness of the QSAR model and offers the potential to replace the high-dimensional fingerprint vector with a comparatively small set of molecular descriptors. This, in turn, allows for further optimisation of model efficiency by reducing the feature set without compromising performance.

It is important to note that preparation of the assay data used in this project was a collaborative effort between my colleague Sascha Koller and myself. All subsequent development and implementation of the Python QSAR model were conducted independently by me. KNIME workflows and Jupyter notebooks referenced in this section can be found on GitHub.[218]

2.3.2.1 ChEMBL Extension

Equation 2 in Section 2.1.3 of this project's introduction establishes that QSAR models relate structural features, encoded as molecular descriptors, to a quantitative measure of biological activity. For this study, the NE^{G12D} assay was selected as the primary activity readout for all compounds in the project library. This choice enabled the use of nearly the entire project library for model training. Prior to analysis, outliers and non-numerical activity values (e.g., entries listed as ">30 μM") were removed from the dataset. For each compound, the mean NE^{G12D} IC_{50} value was calculated when multiple measurements were available.

To further expand the training dataset, the project library was supplemented with publicly available data from the ChEMBL database.[224] Specifically, all compounds with

reported KRAS-related activity (Target ID: CHEMBL2189121) were extracted from ChEMBL version 34. Compounds identified as $KRAS^{G12C}$-specific inhibitors or lacking quantitative IC_{50} values were excluded to maintain relevance and data integrity.

The resulting IC_{50} values, ranging from 1 M to 1 nM, spanned several orders of magnitude, with the smallest IC_{50} values being associated with potent KRAS PPIIs, such as MRTX-1133 derivatives and cyclic peptides similar to LUNA18. Such variation poses challenges for machine learning due to skewed distributions and nonlinear scale effects. Therefore, all IC_{50} values (nM) were transformed using the negative base-10 logarithm to yield pChEMBL values, as shown in (**Equation 3**).

$$pChEMBL = -log_{10}(IC_{50})$$

Equation 3: Definition of the pChEMBL scale.

This transformation improves numerical stability, normalises the activity scale, and aligns with standard practices in cheminformatics and QSAR modelling. Furthermore, the pChEMBL scale, ranging from 0 to 9, offers a more practical and interpretable format for activity data than raw IC_{50} values. Following this transformation, all ChEMBL KRAS structures underwent the same structure standardisation process outlined in Section 2.3.1.2. The washed ChEMBL KRAS set was then merged with the project library, and any duplicate entries were removed.

To verify that the inclusion of ChEMBL compounds expanded the overall activity range, a 2D visualisation of both datasets was produced. This was achieved by applying t-distributed stochastic neighbour embedding (t-SNE) to the fingerprint matrix of each library.[225] **Figure 41** presents the t-SNE plots: the top plot shows only the project library, while the bottom plot includes both the project and ChEMBL KRAS datasets. Each molecule is shown as a single point, coloured by its pChEMBL value from 0 (dark blue) to 9 (bright yellow). While the x- and y-values themselves are arbitrary, the plots offer insight into structural relationships between molecules. More specifically, t-SNE embedding aims at plotting high-dimensional data in two or three dimensions, while preserving the relationship between the datapoints to a certain degree. Therefore, similar structures are represented as clusters, but the shape and distribution of the clusters is not reliably interpretable.

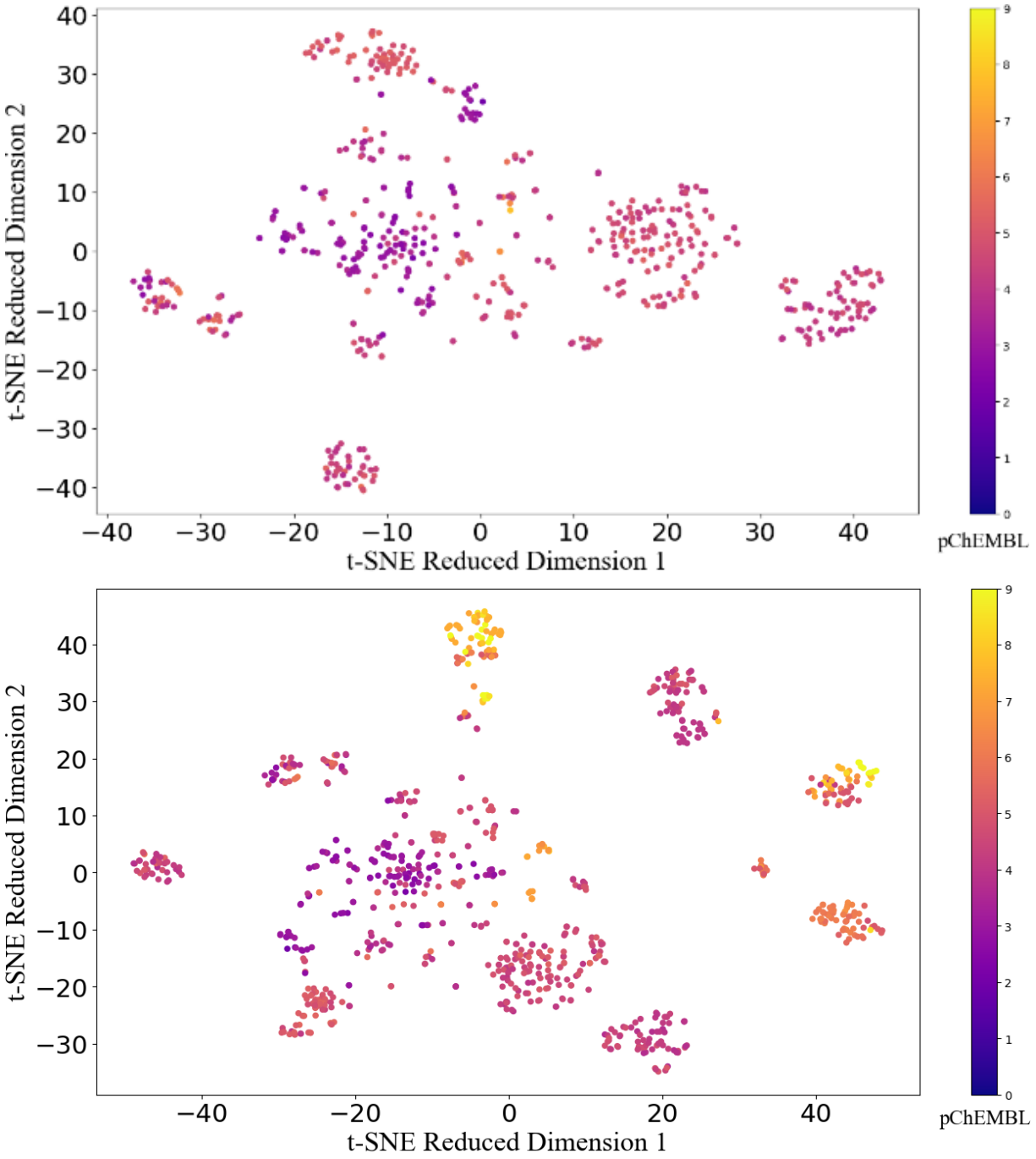


Figure 41: t-SNE plots of the project library (top) as well as the combination of project and ChEMBL KRAS libraries (bottom). Each structure is represented by one dot, coloured according to its pChEMBL value, i.e. 0 (dark blue) to 9 (bright yellow).

The main message to take away from **Figure 41** is that the combined library (bottom) has significantly more structures with pChEMBL values above 6 than the project library (top), i.e. the yellow cluster at -4/40 and the orange clusters at 45/-10 and 45/15. These clusters represent potent KRAS PPIIs, predominantly MRTX-1133 derivatives and cyclic peptides similar to LUNA18.

The t-SNE plots highlight that the ChEMBL KRAS library is a valuable extension of the project library. Integration of the ChEMBL KRAS dataset added approximately 50% more training examples (from 603 to 917) and extended the activity profile beyond the original pChEMBL range of 0–6. This broadened dataset improves the QSAR model's ability to recognise highly active scaffolds, including those in the pChEMBL range 6–9 that are particularly relevant for KRAS PPII discovery.

2.3.2.2 Feature Selection

With the input data carefully prepared, the next phase in building the QSAR model involved selecting the appropriate features. Each molecule in the dataset is represented by a specific set of features, and the quality of these features directly impacts the predictive performance of the model. Whereas Duo *et al*. relied solely on 512-bit ECFP fingerprints as features, this project introduced a hybrid approach.[223] Each molecule was characterised not only by its ECFP fingerprint but also by 20 additional, carefully selected molecular descriptors.

This modification represents an initial step towards a future goal of reducing the overall number of features to minimise predictive noise. In machine learning models, every feature not strongly correlated with the target variable contributes a degree of noise. Therefore, best practice recommends using as few features as possible, provided they are highly informative, while maintaining enough input diversity for accurate prediction.[226] While fingerprint vectors offer high specificity, they are not easily compressible without losing structural information. In contrast, molecular descriptors can be flexibly tailored to the modelling task and dataset.

To generate these descriptors, the Mordred software package was employed. Mordred provides one of the most comprehensive open-source collections of molecular descriptor algorithms currently available.[227] In this project, all 0D Mordred descriptors, requiring only 2D coordinates of the input molecules, were calculated for the extended project library, yielding approximately 1,600 features per molecule. A full list of all descriptors included in Mordred is available in the documentation.[228] Descriptors requiring 3D molecular coordinates were excluded to avoid the added complexity of conformer generation at this stage.

After descriptor calculation, any features containing missing or non-numerical values were discarded. The mutual information between each remaining descriptor and the

target activity variable (pChEMBL) was then computed. Mutual information, in simple terms, quantifies how much knowledge of one variable reduces uncertainty about another. It serves here as a measure of correlation between molecular features and bioactivity.[229] **Figure 42** shows a line plot of mutual information values plotted against the corresponding descriptor IDs.

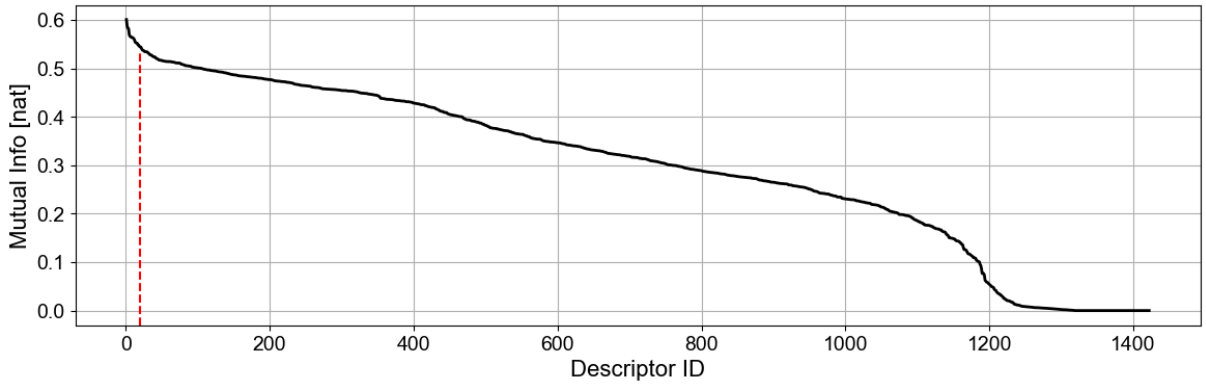


Figure 42: Mutual information of each descriptor plotted against their respective descriptor IDs. The descriptors with the largest mutual information (ID ≤ 20) were selected (up to red dashed line).

The plot exhibits a sharp exponential decline is observed across the top 50 descriptors, followed by a more gradual, approximately linear decrease extending through the subsequent 1,150 descriptors. This trend culminates in a second exponential drop, ultimately reaching a mutual information value near zero. Mutual information is expressed in natural units [nat], where a value of zero signifies no correlation between a descriptor and the activity variable, while increasing positive values reflect stronger correlations.

Empirical optimisation revealed that incorporating the top 20 descriptors, indicated by the red dashed line, alongside the ECFP significantly enhanced model performance. Although replacing ECFPs with a minimal, highly informative set of descriptors remains a long-term objective, the 20 selected descriptors were, in this case, employed in combination with the 512-bit ECFP vector. This integration yielded a total of 532 features per compound in the extended project library. The decision to retain the ECFP alongside the descriptors was guided by the resulting improvement in predictive performance. Importantly, the incorporation of a systematic descriptor selection process constitutes a substantive enhancement of the original protocol published by Duo et al.[223] The selected 20 descriptors are summarised in **Table 8**.

Descriptor	Definition	Interpretation
nN	Number of N atoms.	See *Definition*.[230]
nBondsD	Number of double bonds in non-kekulized molecular structure.	See *Definition*.[230]
NdO	Number of double-bonded O atoms.	See *Definition*.[230]
ATS8m	Moreau-Broto autocorrelation of lag 8 weighted by mass.	Distribution of atomic mass across molecular structure.[231]
ATS3v	Moreau-Broto autocorrelation of lag 3 weighted by vdW volume.	Distribution of vdW volume across molecular structure.[231]
ATS6i	Moreau-Broto autocorrelation of lag 6 weighted by ionisation potential.	Distribution of ionisation potential across molecular structure.[231]
BCUTse-1l	First lowest eigenvalue of Burden matrix weighted by Sanderson EN.	Describes how the atoms' Sanderson electronegativities are arranged and connected throughout a molecule.[232]
BCUTpe-1l	First lowest eigenvalue of Burden matrix weighted by Pauling EN.	Describes how the atoms' Pauling electronegativities are arranged and connected throughout a molecule.[232]
BCUTare-1l	First lowest eigenvalue of Burden matrix weighted by Allred-Rochow EN.	Describes how the atoms' Allred-Rochow electronegativities are arranged and connected throughout a molecule.[232]
Xc-3d	3-Ordered Chi cluster weighted by sigma electrons.	Describes topology of a molecule, while taking sigma electrons into account.[233]
SssNH	Sum of ssNH.	Measure of the polarity and steric accessibility of secondary amines in a molecule.[234]
SdO	Sum of dO.	Measure of the polarity and steric accessibility of double-bonded O atoms in a molecule.[234]
ETA_beta_s	Sigma contribution to valence electron mobile count.	Measure of the contribution of single bonds to a molecules' electron delocalisation.[235]
TIC1	1-Ordered neighbourhood total information content.	Measure of the complexity of a molecule.[236]
SMR_VSA3	MOE MR VSA Descriptor 3 ($1.82 \leq x < 2.24$).	Measure of the polarizability of a molecule.[219,237]
SMR_VSA6	MOE MR VSA Descriptor 6 ($2.75 \leq x < 3.05$).	Measure of the polarisability of a molecule.[219,237]
SlogP_VSA1	MOE logP VSA Descriptor 1 ($-\inf < x < -0.40$).	Measure of the lipophilicity of a molecule.[219,237]
MID_N	Molecular ID on N atoms.	Measure of the atomic environments of all N atoms in a molecule.[238]
piPC4	4-Ordered pi-path count (log scale).	Measure of the amount of branching in a molecule.[239]
MWC08	Walk count (leg-8).	Measure of the branching and size of a molecule.[240]

Table 8: The 20 Mordred descriptors with the highest mutual information content relative to the activity measure. Descriptor names and definitions were copied from the Mordred documentation.[228] Note that the *Interpretation* column is only provided as interpretation aid for the reader and does not necessarily capture the full meaning of each descriptor adequately.

2.3.2.3 Algorithm Selection

Following feature selection, the next step in the QSAR workflow was to determine the most suitable regression algorithm for modelling the dataset. This process involved partitioning the extended project library into a training set and a test set. Several regression models were trained on the training set and subsequently used to predict the pChEMBL values of compounds in the test set (data not previously encountered by the models). Model performance on the test set served as a measure of generalisability. The objective was to identify the regression algorithm that achieved the closest agreement between predicted and experimentally measured pChEMBL values. The overall model development workflow is illustrated in the right panel of **Figure 43**.

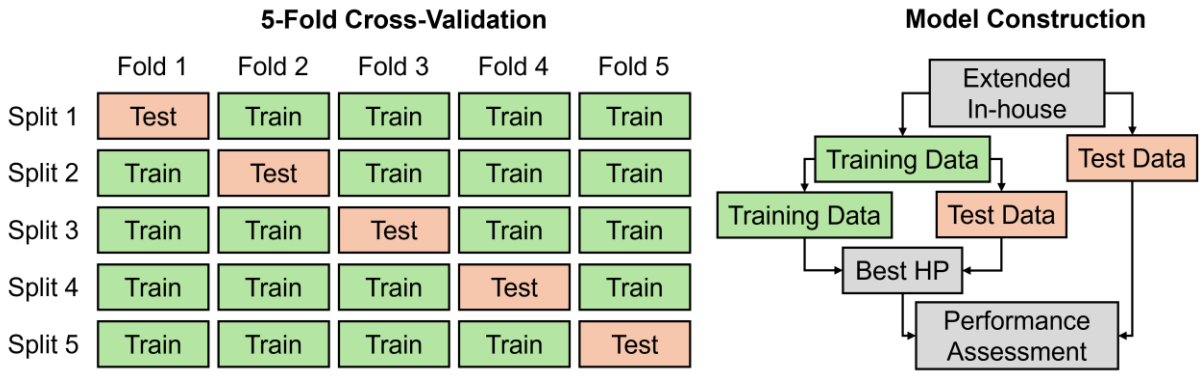


Figure 43: Visualisation of the five-fold cross-validation (CV) split (left) and of the model development workflow employed in this project (right). Each train-test split depicted on the right was performed as five-fold CV.

The complete QSAR modelling pipeline, from data preparation using the extended project library to model performance evaluation, was executed five times for each of the ten regression algorithms assessed. For each algorithm, this involved performing five-fold cross-validation (CV), as illustrated in the left panel of **Figure 43**. In five-fold CV, the dataset is partitioned into five equally sized subsets. Each fold uses a different subset as the test set while the remaining four subsets serve as the training set, ensuring that each molecule is used once for testing. This methodology mitigates biases arising from particularly challenging or lenient test splits and enables performance assessment across the entire dataset.

To account for the influence of hyperparameters (HPs), model parameters that are not learned during training but rather specified manually, a nested CV approach was implemented. Within each outer CV fold, an additional five-fold CV was used to

optimise the HPs, such as the number of trees, maximum tree depth, and minimum samples per leaf in the case of a random forest regressor (see Section 2.1.3). Due to computational constraints, a grid search over a defined subset of HP values was employed rather than an exhaustive search.

The best-performing HPs from each inner CV were then used to train the model on the outer fold's training set. These trained models subsequently predicted pChEMBL values for the respective test sets. Model performance was evaluated using the coefficient of determination (R^2) and root mean square error (RMSE). For each algorithm, the mean and standard deviation of these metrics across all five folds were computed to provide a robust measure of predictive accuracy and generalisability. A summary of the results is presented in **Table 9**

Algorithm	Train R^2	Test R^2	Train RMSE	Test RMSE
Decision Tree	0.975 (0.0118)	0.561 (0.0441)	0.219 (0.0546)	0.933 (0.0654)
Ada Boost	0.768 (0.0222)	0.685 (0.0445)	0.683 (0.0342)	0.787 (0.0197)
Elastic Net	0.784 (0.0790)	0.694 (0.0643)	0.651 (0.1369)	0.773 (0.0465)
Lasso	0.821 (0.0028)	0.728 (0.0292)	0.600 (0.0085)	0.734 (0.0459)
Ridge	0.918 (0.0030)	0.738 (0.0276)	0.407 (0.0081)	0.720 (0.0426)
Extra Trees	0.957 (0.0048)	0.749 (0.0498)	0.293 (0.0151)	0.701 (0.0647)
Gradient Boosting	0.881 (0.0407)	0.751 (0.0441)	0.483 (0.0831)	0.701 (0.0720)
K-Neighbours	0.977 (0.0508)	0.768 (0.0521)	0.092 (0.2068)	0.674 (0.0640)
SVR	0.994 (0.0008)	0.789 (0.0302)	0.106 (0.0072)	0.646 (0.0481)
Random Forest	0.992 (0.0105)	0.794 (0.0258)	0.091 (0.0932)	0.638 (0.0414)

Table 9: R^2 and RMSE for each regressor algorithm averaged across the five outer CV folds.

The selection of regression algorithms in this study was adopted from the work of Duo *et al*.[223] The algorithms were ranked in ascending order of their average R^2 scores obtained on the test sets. The R^2 metric quantifies the proportion of variance in the observed data that can be explained by the model and thus serves as an indicator of predictive strength. In the context of this project, R^2 was used to assess how accurately each model captured the relationship between molecular features and their corresponding pChEMBL values. An R^2 value of 0 indicates no correlation, whereas a value of 1 denotes perfect correlation. Complementarily, the root mean square error (RMSE) metric was used to quantify the average deviation between the predicted and measured pChEMBL values, providing a direct measure of prediction accuracy.

Given that the ultimate application of these models involves in silico screening of previously unseen compounds, particular emphasis was placed on test set

performance. The results, summarised in **Table 9**, show a clear trend: higher R^2 values are generally associated with lower RMSE values, suggesting that stronger correlations coincide with greater predictive precision.

Among the tested models, the random forest (RF) regressor exhibited the best performance, consistently achieving the highest R^2 and lowest RMSE values. On average, the RF model predicted IC_{50} values with an error of approximately fourfold the respective measured IC_{50} value (as per **Equation 3**). Furthermore, the low standard deviations in both R^2 and RMSE across the five folds indicate that the RF model's performance is stable and robust. Interestingly, these findings align with those reported by Duo et al., who also identified the RF regressor as the most effective algorithm for similar QSAR tasks.[223] This reinforces the suitability of the random forest approach for predicting pChEMBL values using ECFP fingerprints augmented with 20 additional molecular descriptors.

To gain further insights into the generalisation behaviour of the top-performing models, learning curves were generated for the three best regressors: RF, support vector (SV), and k-nearest neighbours (KNN). These are presented in **Figure 44**.

The learning curves were generated by training the optimised regression algorithms on incrementally larger subsets of the extended project library. The maximum training set comprised 80% of the total dataset, while the remaining 20% served as a test set. For each subset size, models were trained and subsequently used to predict pChEMBL values for both the training and test sets. The resulting predictions were used to compute R^2 and RMSE values, which were then plotted to visualise model performance across varying training set sizes.

In the generated plots, solid lines represent test set results, while dashed lines correspond to training set performance. Colour coding was applied as follows: blue for RF, red for SV and grey for KNN. These learning curves provide valuable insights into the learning dynamics and capacity of each model. Ideally, a learning curve exhibits a smooth hyperbolic shape that asymptotically approaches optimal values—R^2 approaching 1 and RMSE approaching 0, indicating improved generalisation with increasing data.

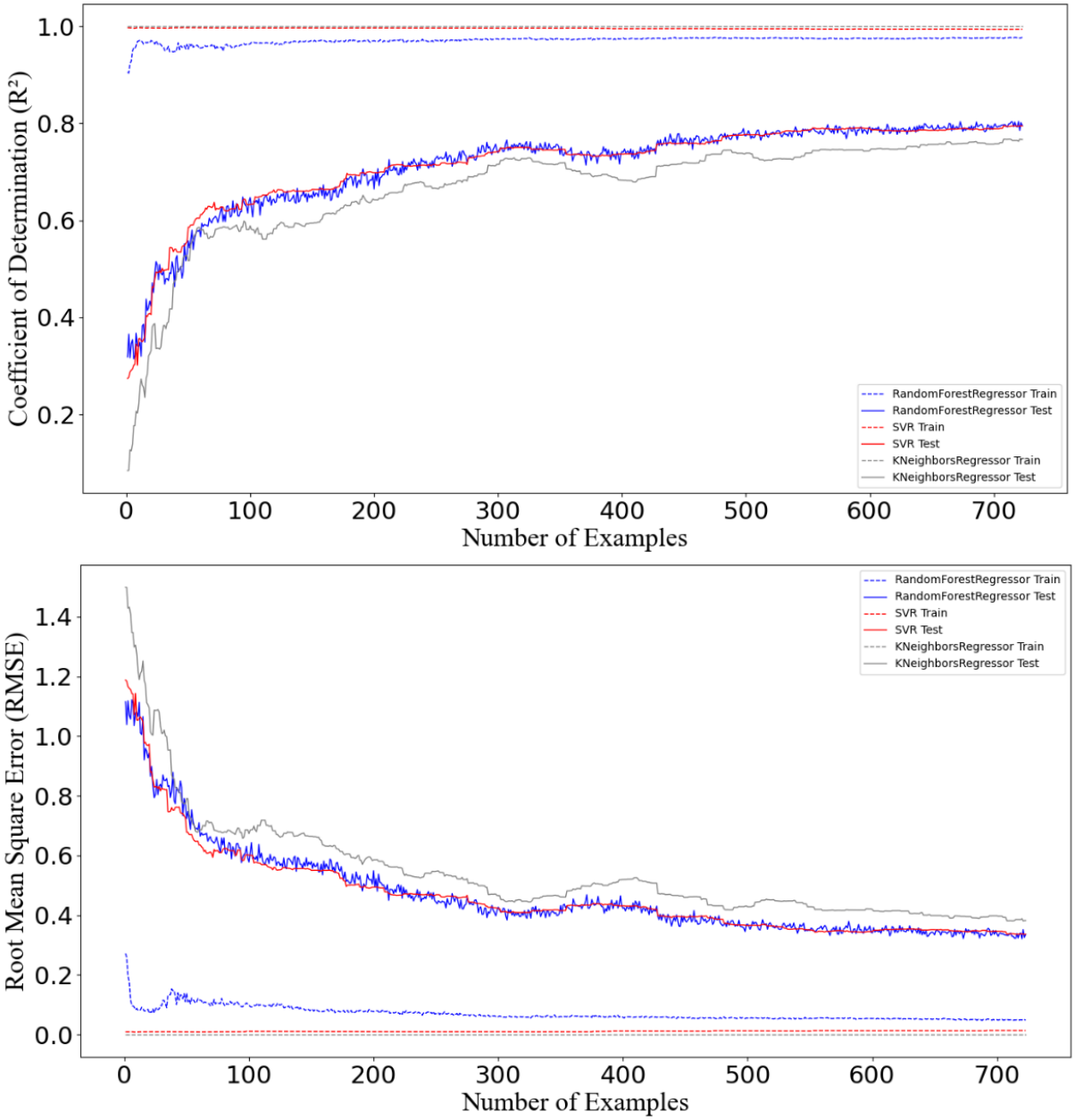


Figure 44: Learning curves of the optimised random forest, support vector and k-nearest neighbours regressor algorithms (blue, red and grey, respectively).

The training curves for the RF regressor suggest that approximately 300 training examples are sufficient for the model to reach near-optimal performance on the training data. Notably, the sharp increases in performance at the early stages underscore the importance of including at least 100 examples to establish a stable and minimally functional model. Beyond this point, all three algorithms showed consistent improvement, reaching R^2 values around 0.8 and RMSE values near 0.3 after training on approximately 700 examples.

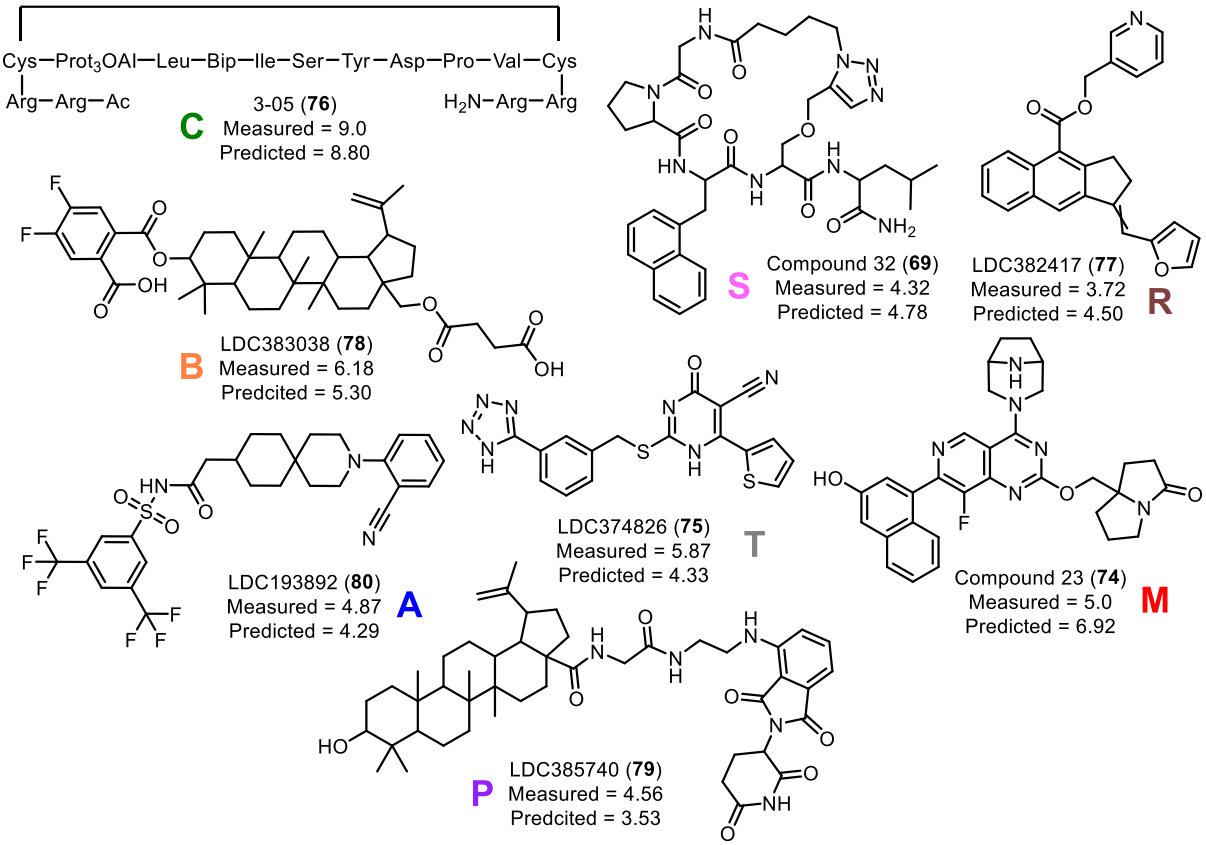
Importantly, none of the models demonstrated a continuous decline in performance beyond a certain training set size. This would be an indicator of *overfitting*; a phenomenon that severely compromises a model's generalisability. This consistency suggests that the temporary decrease in predictive accuracy between 350–400 examples, observed in all three learning curves, likely stems from specific QSAR characteristics of the compounds represented within this training subset.

Among the models, the KNN regressor displayed the most variability in its test set learning curve, characterised by noticeable fluctuations. This behaviour indicates a higher degree of model instability, irrespective of the training set size, despite achieving comparable performance metrics. In contrast, the RF and SV regressors showed more stable and consistent performance across both the training process (**Figure 44**) as well as across folds in (**Table 9**).

However, the SVR has an inherent sensitivity to the scale of input features, necessitating feature normalisation prior to training, typically to the range [0,1]. The RF regressor, by comparison, is scale-invariant and does not require such preprocessing, making it less prone to issues arising from feature heterogeneity. Considering its robust performance and lower sensitivity to data preprocessing requirements, the random forest regressor was identified as the most suitable and reliable model for the QSAR analysis in this project.

2.3.2.4 Structural Perspective of Prediction Accuracy

Prior to finalising the model, the structure-dependence of the RF regressor's predictions was evaluated through manual inspection. Aiming to maximise predictive performance, the model was trained on 90% of the extended project library. Hyperparameter optimisation was conducted through five-fold CV on this training set. The RF algorithm was then retrained using the identified optimal hyperparameters, and predictions were generated for the remaining 10% of the data, which served as the test set. Encouragingly, the model achieved strong predictive performance on the test set, with an average R^2 value of 0.79 and a RMSE of 0.29. A more detailed performance assessment across different compound classes is illustrated by the box plots shown in **Figure 45**.



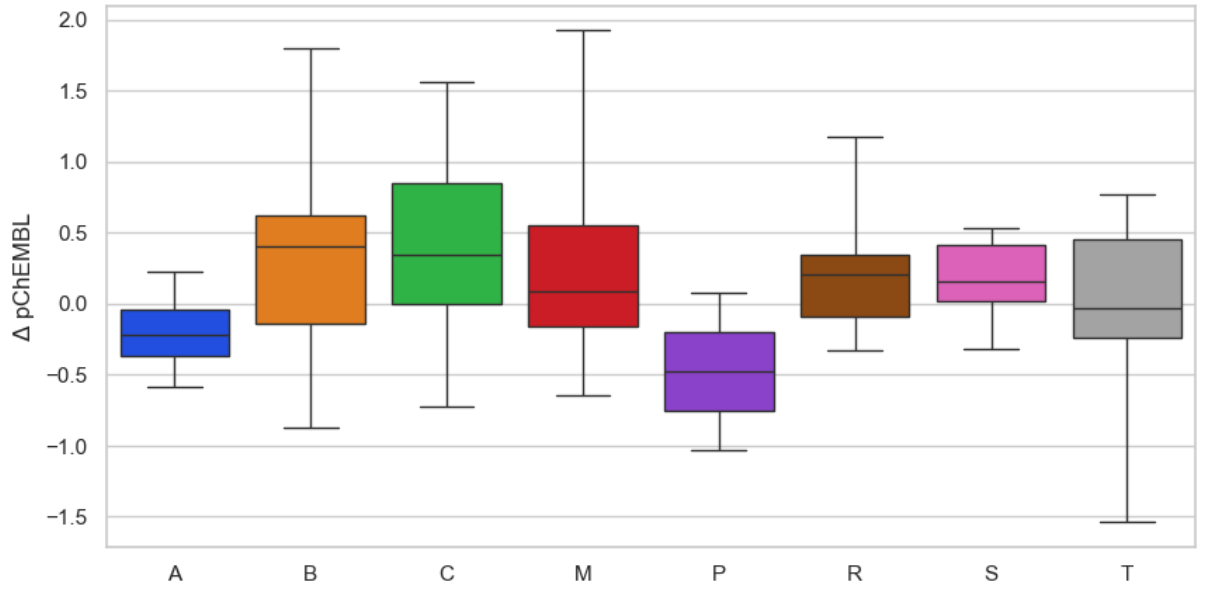


Figure 45: Top panel: Representative examples of the compound classes contained in the test set. Bottom panel: Box plots of the differences between predicted and measured pChEMBL values, sorted by compound class. The compound classes are: acylsulfonamides (**A**), betulinic acids (**B**), macrocyclic peptides (**C**), MRTX-1133 derivatives (**M**), PROTACs (**P**), small aromatic molecules (**R**), stapled peptides (**S**) and tetrazoles (**T**).

The box plots were created from the differences between predicted and measured pChEMBL values, defined as ΔpChEMBL in **Equation 4**.

$$\Delta pChEMBL = Predicted\ pChEMBL - Measured\ pChEMBL$$

Equation 4: Definition of ΔpChEMBL.

Each box plot visualises the distribution of ΔpChEMBL within a specific compound class. The central 50% of the data (interquartile range) is represented by coloured boxes, with the medians indicated by horizontal lines. Unlike conventional box plots, the whiskers extend to the full range of observed values, rather than the typical 1.5-fold interquartile range, to include all data points.

The median differences indicate that the RF model tends to overestimate the inhibitory activity in five of eight compound classes. In contrast, three classes, namely acylsulfonamides (**A**), PROTACs (**P**) and tetrazoles (**T**), exhibited a tendency for underestimated activity. Among all classes, acylsulfonamides (**A**) and stapled peptides (**S**) demonstrated the highest prediction accuracy, with ΔpChEMBL ranges of [-0.59, 0.23] and [-0.32, 0.53], respectively. Significantly wider ΔpChEMBL ranges were observed for the remaining classes, i.e. **B**:[-0.88, 1.80], **C**: [-0.72, 1.56], **M**: [-0.64, 1.92], **P**: [-1.03, 0.08], **R**: [-0.33, 1.07], **T**: [-1.54, 0.77].

It is important to note that 50% of the differences lie within the interquartile range, and extreme values such as -1.54 (**T**) and 1.92 (**M**) should not be viewed as representative of the overall model accuracy. However, these outliers warrant closer structural inspection.

The largest ΔpChEMBL was observed for derivative **74** from class **M**., a group comprising analogues of the potent $KRAS^{G12D}$ inhibitor MRTX-1133. Although derivative **74** differs structurally from MRTX-1133 only by the absence of two fluorine atoms and a terminal alkyne, and the addition of a ketone group, these minor changes drastically reduced its activity. Such a sharp activity drop suggests that subtle structural variations in high-affinity PPIIs like MRTX-1133 significantly impact potency. Improved predictions for derivative **74** may require a focused model trained on extensive SAR data for MRTX-1133 analogues, a level of detail beyond the scope of the general-purpose KRAS PPII model developed in this project.

Similarly, the significant underestimation of compound **75**, a tetrazole from class **T**, can be attributed to limited structural diversity within the training data. While most tetrazoles in class **T** feature an acidic tetrazole ring connected through a flexible linker to a ring system, compound **75** is unusually active despite the absence of the linker usually

required for activity. Consequently, the model underestimated the inhibitory potency of tetrazole **75**.

In contrast, a notably accurate prediction was achieved for the previously published 16-mer cyclic peptide 3-05 (**76**) in class **C**.[241] Its exceptional potency (pChEMBL value of 9.00) was adequately predicted by the RF model (pChEMBL = 8.80). Another example of high predictive accuracy is observed for compound 32 (**69**), which is part of the stapled peptide class (**S**) and was discussed in Section 2.1.6. Among all molecules in the extended project library, the members of class **S** exhibit the highest structural similarity to click cyctetpep. The relatively small prediction error for stapled peptide **69** (ΔpChEMBL = 0.46), and class **S** overall, suggests that a RF regressor trained on sufficient cyctetpep SAR data could yield a model well-suited for *in silico* screening of this compound class.

Class **R** is represented by the cyclopentyl quinoline (**77**) and comprises structurally diverse small molecules built around fused aromatic systems. Despite this structural variability, the RF model achieved commendable accuracy for this class, demonstrating its ability to generalise across chemically diverse scaffolds. Class **B**, in contrast, contains betulinic acid derivatives, such as compound **78**, differing in their ester substituents. The SAR of this class was explored in a recent publication from our research group.[49] These compounds are hypothesised to bind at the interface of the $KRAS^{G12D}$:SOS1 complex, chelating the Mg^{2+} cation through their acid moieties. This is a mode of action reminiscent of the biazole series described in Section 1.3.5.3. However, while the RF model has successfully learned the relevance of carboxylic acid groups for inhibition within this class, more diverse chemical modifications may be needed to improve model accuracy.

Notably, the RF model significantly underestimated the activity of PROTAC **79** (**P**) with a ΔpChEMBL of -1.03. Unlike other molecules in the dataset, PROTAC **79** is designed to induce proteasomal degradation of KRAS rather than inhibit the NE^{G12D}. Its unusually high potency in the NE^{G12D} assay, unexpected for a compound operating through a degradation mechanism, likely contributed to the poor prediction performance.

Finally, class **A** includes a comprehensive series of acylsulfonamides, such as compound **80**. These compounds typically feature a benzene ring with fluorinated

substituents on one side of the acylsulfonamide moiety and several rigidly connected rings on the other. The wealth of SAR data available for this class allowed the RF model to produce reliable prediction results.

In summary, the evaluation of prediction performance across compound classes indicates that the RF model provides sufficiently robust estimates for *in silico* screening aimed at identifying novel KRAS PPII scaffolds. While the model does not reliably predict the activity of individual derivatives, it is well-suited for scaffold prioritisation. Following scaffold selection, target derivatives can be identified through molecular docking of focused virtual libraries at the proposed binding site. To enable this next step, the final QSAR model was constructed by training the RF algorithm on the entire extended project library. The trained model was subsequently exported and applied to predict pChEMBL values across multiple screening libraries.

2.3.2.5 Activity Prediction

Structures for pChEMBL value prediction were sourced from a diverse array of chemical libraries. **Table 10** provides an overview of the employed libraries, sorted in descending order by the number of entries retained after washing. The *Entries* column denotes the total number of structures per library following this curation step. Libraries specifically focused on PPII scaffolds are highlighted in green.

Library	Entries	Downloaded/Created	Source
Enamine Screening	4 315 298	28.06.24	Commercial
ChEMBL 34	2 188 592	26.08.24	Open Source
ChemDiv BMS	299 919	06.08.24	Commercial
Click Cyctetpep	152 000	19.01.24	In-house
ChemDiv DivSet	49 854	14.05.24	Commercial
Enamine PPI	40 255	14.05.24	Commercial
Cyctetpep	40 110	04.12.23	Open source
ChemDiv KRAS	15 814	02.10.24	Commercial
Reinvent PPI	11 241	14.05.24	Open source
iPPI	2 345	14.05.24	Open source

Table 10: Molecular libraries used for pChEMBL prediction.

The measured pChEMBL values of the training data (extended project library) and the predicted pChEMBL values for each library in **Table 10** are summarised in the violin plots presented in **Figure 46**. It is important to note that the colours used in were assigned randomly and bear no relation to the colour scheme employed in the box plots of **Figure 45**.

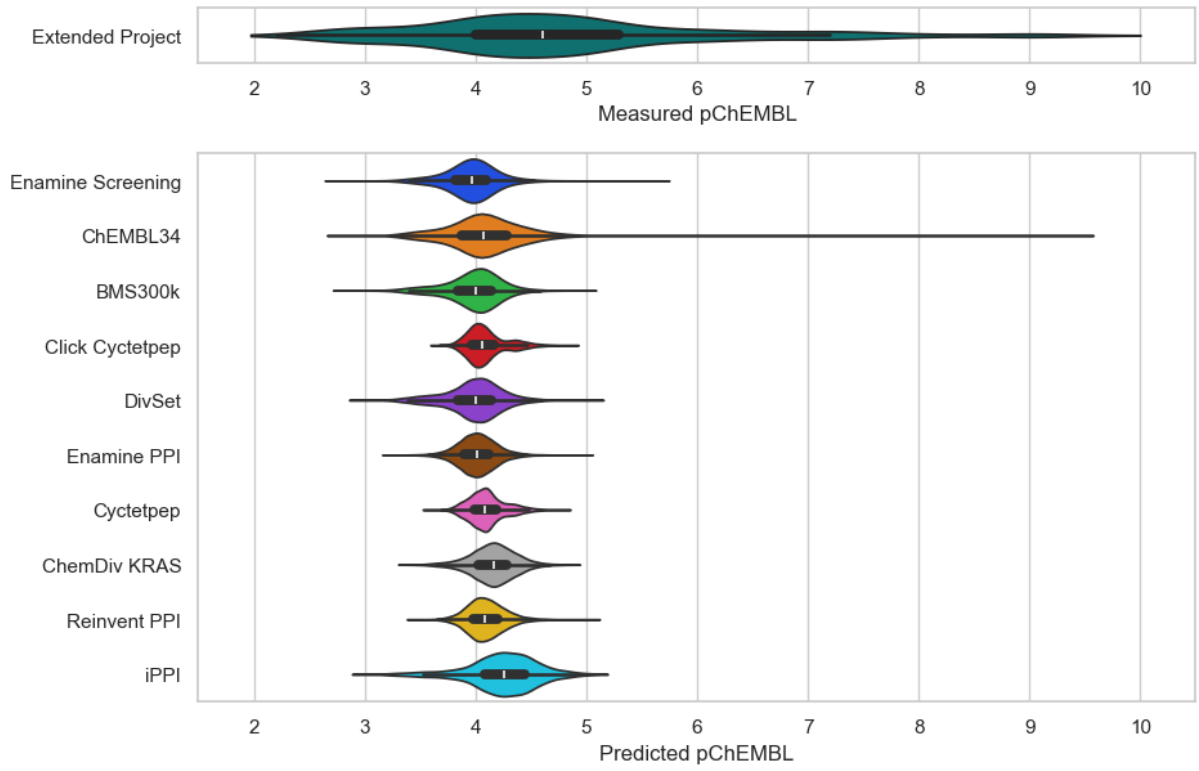


Figure 46: Violin plots of the predicted pChEMBL values for each of the screening libraries listed in **Table 10**.

The violin plots provide a visual representation of the distribution of predicted inhibitory activities across all compounds within each screening library. The width of each plot at a given pChEMBL value reflects the relative density of compounds predicted to exhibit that activity, thus providing insight into the inhibitory potential and diversity of each library. Embedded within each violin plot is a small box plot that highlights the interquartile range in black and the median predicted value in white. The full range of predicted values, including extreme outliers, is represented by the vertical extent of each violin.

Overall, the majority of analysed compounds received predicted pChEMBL values within the range of [3.5, 5.0], with medians clustering around 4.0. This distribution suggests a limited generalisability of the model, despite its promising predictive performance on the extended project library as discussed in Sections 2.3.2.3 and 2.3.2.4. Several factors may contribute to this behaviour:

1. Trained pChEMBL range: The model was exposed to relatively few examples outside the interquartile range of the training set (i.e., [4.0, 5.3]). As a result, it

may default to assigning lower pChEMBL predictions unless prompted otherwise by strongly learned structural features.

2. Descriptor limitations: The molecular descriptors employed may lack the discriminatory power needed to effectively distinguish between structurally diverse compounds in the screening libraries.
3. Chemical space mismatch: The chemical spaces covered by the training data and the screening libraries may exhibit limited overlap. Consequently, structurally diverse compounds from underrepresented regions may receive similar predictions due to insufficient model generalisation.
4. Overfitting: The model may have learned training set-specific artefacts (e.g., scaffold-based patterns or noise) rather than capturing broadly applicable SAR.

This observed behaviour underscores the importance of expanding the training dataset to include structurally diverse PPIIs with experimentally determined pChEMBL values outside the currently predominant predictive range, i.e. below 3.5 and above 5.5. Additionally, replacing the ECFP with a smaller set of carefully selected, highly informative molecular descriptors may improve the model's ability to generalise across novel chemical space.

Despite the aforementioned limitations, the violin plot corresponding to the iPPI library (comprising published PPIIs) exhibits the highest average predicted *p*ChEMBL values among all screened libraries. This observation supports the conclusion that the model is capable of distinguishing PPIIs from non-PPIIs with a reasonable degree of reliability. Moreover, it remains valid to infer that structural motifs associated with potent NE inhibition were successfully recognised in the highest-scoring compounds.

Consequently, the present QSAR model, while not optimised for predicting the precise inhibitory activity of individual compounds, remains suitable for identifying promising KRAS PPII scaffolds for future elaboration.

In the following paragraphs, one to two compounds per library will be examined in more detail as representative examples for the most promising KRAS PPII scaffolds. Selection criteria include: i) predicted pChEMBL value, ii) molecular flexibility, iii) overall molecular shape, iv) ease of derivatisation, and v) structural novelty with respect to the existing project library.

The selected compounds are illustrated in **Figure 47**. Notably, stereochemical information is omitted from these depictions, as stereoisomerism was not considered during the prediction process.

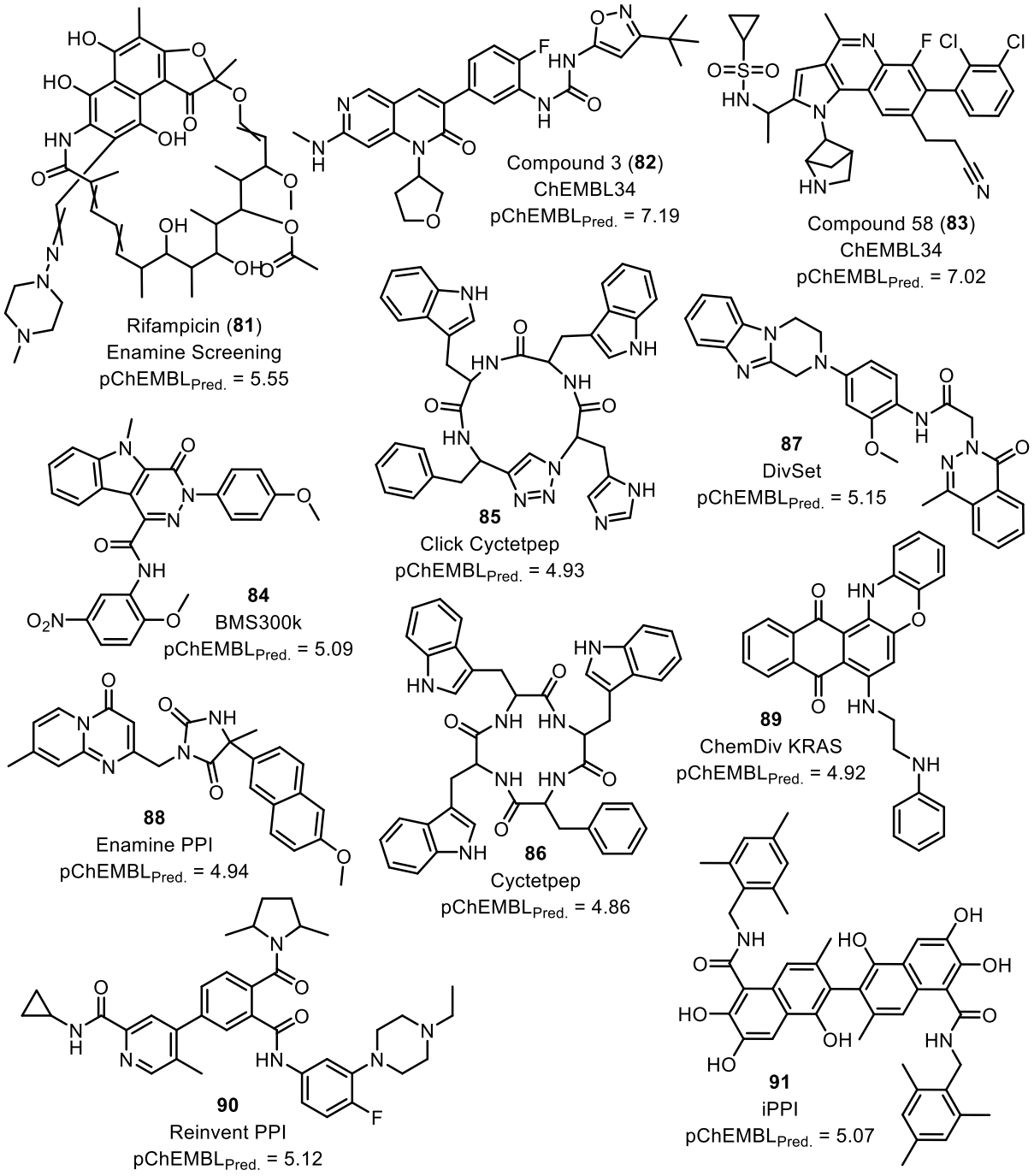


Figure 47: Hit structures, which resulted from the QSAR screening of the libraries in **Table 10**. All structures are depicted without stereo information.

According to the violin plot analysis, the Enamine screening library contains several highly promising scaffolds, with one notable example being rifampicin (**81**), a well-established anti-tuberculosis agent. Interestingly, rifampicin has also been reported to

exhibit antiproliferative activity against hepatic tumors.[242] However, to date, its inhibitory potential in the context of KRAS has not been evaluated. The predicted pChEMBL value suggests that rifampicin may inhibit NE^{G12D} in the low micromolar range. Crucially, rifampicin lacks structural analogues in the current project library, making it a valuable candidate for the expansion of scaffold diversity within the library.

The ChEMBL34 library yielded compounds with the highest predicted activities overall. However, the majority of these top-ranking molecules fall into two categories already represented in the project library: macrocyclic peptides similar to compound **76** and analogues of MRTX-1133 similar to **74** in **Figure 45**. This redundancy highlights a shortcoming in the current structure washing workflow, namely the incomplete removal of overlapping entries between the extended project library and ChEMBL34. Despite this duplication, the high predicted activity of these scaffolds still renders them valuable additions to the project library.

Two compounds stand out as particularly noteworthy. The pan-RAF inhibitor compound 3 (**82**) was reported to inhibit $BRAF^{WT}$ in HCT116 cells ($KRAS^{G13D}$) reducing cell viability with IC_{50} = 32 nM.[243] Its activity against KRAS has not yet been explicitly assessed, such as in a nucleotide exchange assay like NE^{G12D}. The second remarkable ChEMBL34 hit is compound 58 (**83**) with a reported inhibition of the $KRAS^{G12D/V}$ nucleotide exchange with $IC_{50} \leq 100$ nM.[244] Exact quantitative evaluation of its NE^{G12D} inhibition would provide valuable SAR data to the project library.

A structural analogue of compound **83** is found in the BMS300k hit **84**, which features a hybrid architecture combining a tricyclic fused ring system and a phthalazinone scaffold. This hybrid configuration presents a promising opportunity to explore the SAR of the phthalazinone scaffold more systematically by varying the three core ring systems.

The predicted pChEMBL values of the identified (click) cyctetpep hits **85** and **86** rank at the lower end of selection shown in **Figure 47** though the values fall only marginally short of the top-scoring compounds. Both derivatives feature aromatic and basic side chains, a combination that aligns with prior observations on the entropic benefits of aromatic residues in (click) cyctetpep structures, as discussed in Section 2.3.1.3. The presence of basic residues may reflect learned features from cell-penetrating peptide motifs prevalent in the macrocyclic peptides of the training set. However, the smaller

size and distinct architecture of cyctetpep compounds may render such cationic features less essential. Notably, the close structural resemblance between cyctetpep hits and their corresponding click analogues suggests that the introduction of the triazole linkage does not negatively impact predicted activity.

The DivSet library predominantly comprises small, linear molecules constructed from interconnected ring systems. Among these, compound **87** was identified as particularly promising, due to its structural similarity to compound 3 (**82**). At first glance, enamine PPI hit **88** may appear analogous to compound **87**; however, due to the presence of an asymmetric carbon in its imidazolidinedione ring, it assumes a distinctly different three-dimensional conformation. Its rigid connection to the naphthyl moiety is likely to confer an entropic advantage during binding. Similarly, the ChemDiv KRAS hit **89** features a rigid, boomerang-shaped ring system that is unprecedented within the extended project library. The terminal amines offer valuable points for functional extension in both two- and three-dimensional space, enhancing its utility as a scaffold.

Reinvent PPI hit **90** is capable of adopting a spatial conformation reminiscent of MRTX-1133, while retaining sufficient flexibility to explore a broader conformational landscape. This combination of shape adaptability and novelty regarding the project library supports its inclusion as a candidate scaffold for further derivatisation.

Finally, the iPPI hit **91** was reported to inhibit antiapoptotic Bcl-2 family proteins with EC_{50} = 3 µM in RS11846 cells.[245] It stands out through its polyphenolic core, which is a structural feature commonly found among frequent hitters.[246] Notably, such frequent hitter scaffolds were deliberately not excluded from the screening process, as strategic derivatisation may still yield valuable SAR insights. In solution, the two naphthyl rings adopt a staggered conformation, offering multiple functional handles for scaffold diversification in three dimensions.

In summary, this project successfully achieved the goal of designing an *in silico* screening pipeline tailored to the identification of promising PPIIs. The question of whether the identified scaffolds possess genuine activity against KRAS must ultimately be addressed through *in vitro* validation of the predicted hits depicted in **Figure 47**. Nonetheless, the incorporation of these scaffolds into the project library will significantly broaden the chemical space it encompasses. Given their favourable predicted activities, each compound in **Figure 47** represents a viable starting point for

subsequent SAR investigations, with an emphasis on structural diversity and PPI inhibition. Prioritisation of scaffolds for further development was guided primarily by considerations of modularity and shape. In this regard, the (click) cyctetpep scaffolds emerged as clear frontrunners. Their unique attributes, which make them especially attractive for scaffold-driven exploration, were outlined in detail in Sections 2.1.5 and 2.3.1.

2.3.2.6 Molecular Docking of Cyclic Tri- and Tetrapeptides

The development of synthetic strategies for peptide macrocycles was conducted in parallel with the chemical space analysis and the Python-based QSAR modelling described earlier. At the outset of this project, the (click) cyctetpep scaffold had not yet emerged as the most promising candidate, either from a synthetic or *in silico* perspective. Consequently, native cyclic tri- and tetrapeptides were initially considered viable targets. The molecular docking studies of exhaustive libraries of native cyctripep, native cyctetpep, and click cyctetpep scaffolds are presented in the following.

Multiple high-resolution crystal structures of KRAS with co-crystallized ligands are publicly available, the majority of which involve ligands binding to the P1 and P2 pockets. The P2 pocket is accessible exclusively in inactive $KRAS^{GDP}$, making it an especially attractive site for the design of PPIIs targeting this specific KRAS complex.[95] At the time of this study, MRTX-1133 was the most potent, non-covalent, small-molecule PPII targeting $KRAS^{G12D}$ *via* the P2 site. In our biochemical assays, MRTX-1133 showed potent activity, with $IC_{50}(NE^{G12D})$ = 0.0229±0.0160 µM and IC_{50}(CTG SNU-1) = 0.01 µM (single measurement). Structurally, MRTX-1133 is a star-shaped, high-molecular-weight small molecule, bearing a strong resemblance to the (click) cyctetpep and cyctripep scaffolds. Given this similarity, the crystal structure of $KRAS^{G12D,\ GDP}$:MRTX-1133 (PDB ID: 7RPZ) was selected for screening at the P2 site.

The P1 pocket, in contrast, is conserved across multiple RAS isoforms and mutants in their active and inactive states, making it a compelling target for the design of pan-RAS PPIIs.[95] Two distinct KRAS crystal structures were chosen to investigate this site: PDB IDs 6GJ8 and 6GQY. Structure 6GJ8 represents the complex $KRAS^{G12D,\ GPPCP}$:BI-2852 (**7**), in which ligand **7** exhibits a high binding affinity to the P1 pocket and inhibits the $KRAS^{G12D}$:SOS1 interaction with IC_{50} = 17.0±9.31 µM (**Table 5**). In contrast, Ch-3

showed no NE^{G12D} inhibition, despite clear evidence of binding to the P1 pocket on $KRAS^{Q61H, GPPNP}$ (PDB ID: 6GQY).

The click cyctetpep library was docked to the P1 pocket of both crystal structures (6GJ8 and 6GQY) with the dual objective of identifying promising NE^{G12D} inhibitors and assessing the sensitivity of docking results to the choice of crystal structure. A comprehensive description of the docking protocol can be found in the Experimental Section.

To establish robust docking parameters, the co-crystallised reference ligands **2**,**7**, and **8** were docked to their respective binding sites to establishing the ideal docking parameters for their respective crystal structures. **Figure 48** illustrates the superimposition of each co-crystallised ligand with its docked conformation, demonstrating the validity of the docking protocol.

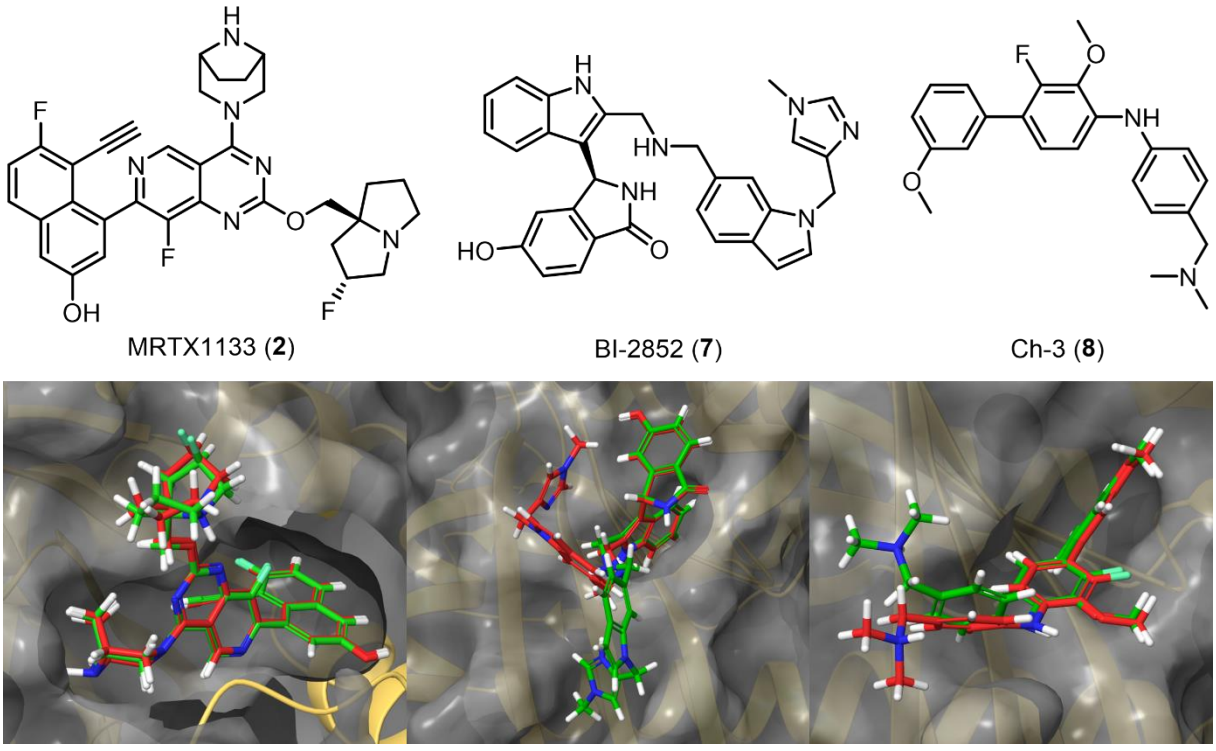


Figure 48: Top: Co-crystallised reference ligands of 7RPZ (**2**), 6GJ8 (**7**) and 6GQY (**8**) (left to right). Bottom: Crystallised conformation of references **2**, **7** and **8** superimposed with their respective docked poses (left to right). In the case of 7RPZ, part of the protein surface was removed to visualise MRTX-1133 (**2**) inside the pocket.

The crystallised conformations of BI-2852 and Ch-3 in 6GJ8 and 6GQY, respectively, were reproduced with sufficient accuracy using the docking settings employed in Section 1.3.5.2. In contrast, accurately simulating the binding pose of MRTX-1133 in 7RPZ required the implementation of an interaction constraint with Asp12. This

modification is well justified, as previous studies have reported that the ionic interaction between deprotonated Asp12 with the protonated, bridged piperazine moiety of MRTX-1133 is essential for its high affinity and selectivity for $KRAS^{G12D}$.[72] In addition, the pH range used during tautomer and protonation state generation in the LigPrep module had to be broadened to 7.0±2.0. The complete molecular docking pipeline is outlined in **Figure 49**.

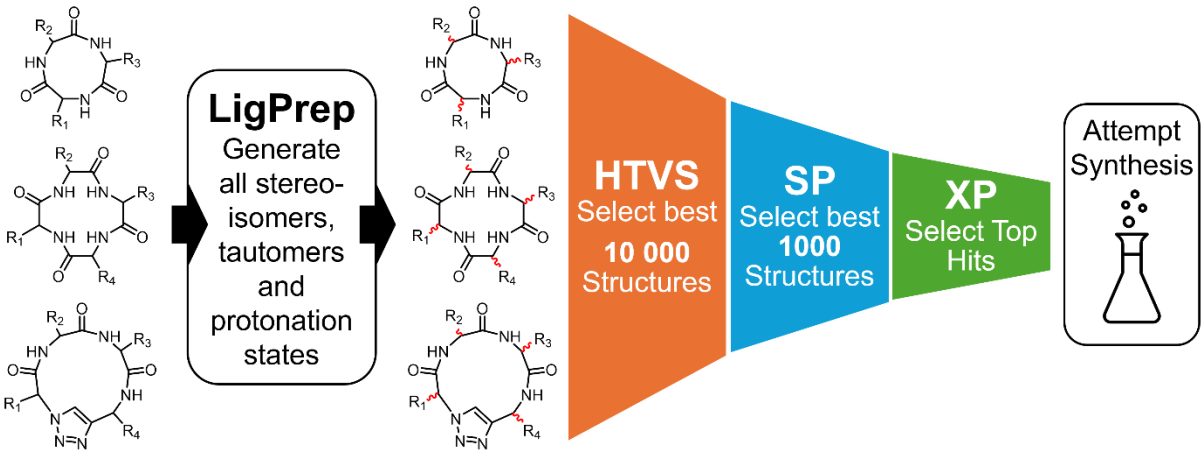


Figure 49: Schematic representation of the molecular docking pipeline employed for the selection of suitable target structures from the cyctripep and (click) cyctetpep libraries.

Ligand preparation was conducted using LigPrep (Schrödinger Maestro suite), generating all possible stereoisomers for the compounds in the cyctripep and (click) cyctetpep libraries. For each stereoisomer, all feasible tautomers and protonation states within the pH ranges 7.0±1.0 and 7.0±2.0 were generated. Before this process the library sizes were 2,680 (cyctripep), 40,110 (cyctetpep) and 152,000 structures, respectively. Notably, LigPrep produced several million distinct entries from the click cyctetpep library alone, highlighting the computational intensity and defining the practical upper limit of ligand set size on the available hardware.

Ligands prepared at pH = 7.0±1.0 were docked into the P1 pocket of 6GJ8 and 6GQY, while those prepared at pH = 7.0±2.0 were docked into the P2 pocket of 7RPZ. The docking workflow followed a tiered protocol. An initial round of high-throughput virtual screening (HTVS) was performed, from which the top 10,000 HTVS ligands (based on their docking score (DS)) were selected for further docking using the standard precision (SP) mode. Subsequently, the top 1,000 SP ligands were redocked using the extra precision (XP) mode. For each ligand in the XP step, up to ten docking poses were generated and retained.

Following docking, the reliability of the results was evaluated by comparing the structural similarity of the top 10 ligands in each XP docking output. Common structural features among top-scoring ligands included polar, acidic, and basic side chains in combination with at least one aromatic side chain, suggesting a preferred amino acid composition. However, the DS values of these ligands were distributed across a broad range, rather than being narrowly clustered, as would be expected in cases of high binding specificity. This dispersion indicates a degree of binding promiscuity and suggests that the modelled interactions may lack strong discriminative power.

The nine XP docking outputs of the cyctripep and (click) cyctetpep libraries, each docked into 6GJ8, 6GQY, and 7RPZ, are illustrated in **Figure 50**.

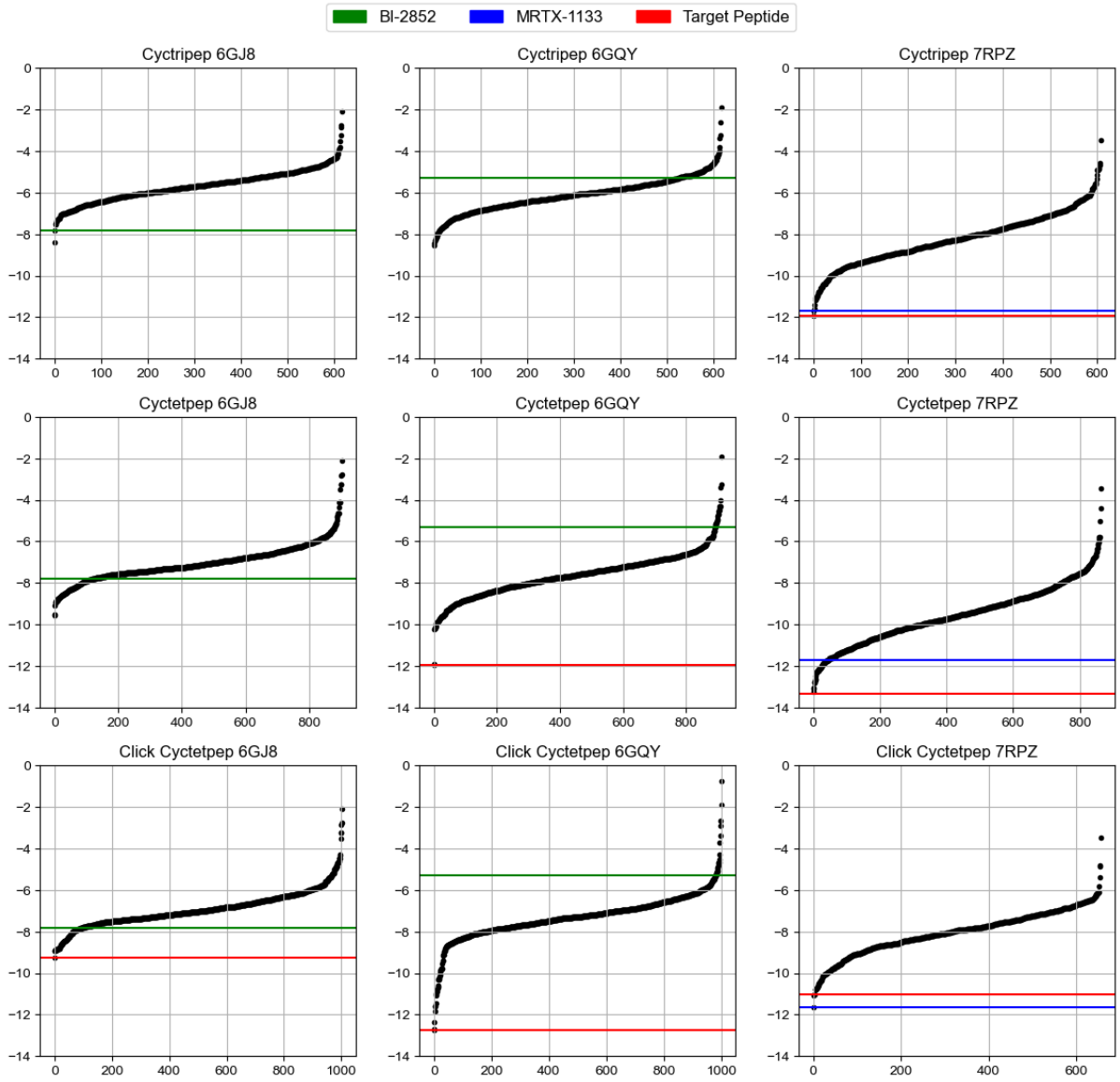


Figure 50: Scatter plots of the top DS [kcal/mol] of each ligand relative to the respective ligand ID. Each subplot contains the results of docking one library in XP mode to one crystal structure as indicated in the respective title. The ligand IDs are not conserved across the subplots, as sorting of the DS in ascending order was prioritised. Reference structures are indicted by the coloured, horizontal lines (see legend).

The scatter plots visualize the distribution of DS [kcal/mol] for all ligands subjected to XP docking. Each ligand is represented by a single black dot, and the ligands are arranged in ascending order of DS along the y-axis. It is important to note that due to this sorting approach, ligand identifiers are not preserved across subplots, which precludes direct comparison of individual compounds between docking runs.

A particularly noteworthy observation is the substantial reduction in the number of successfully docked structures when P2 in 7RPZ was chosen as docking receptor. significant portions of the 1000 input structures were not contained in the output. This loss of docking output is attributed to the combination of stringent interaction criteria imposed by XP mode and the steric constraints of the P2 binding pocket, which was originally induced by MRTX-1133. In practical terms, many ligands were unable to adopt geometries that fulfilled both the steric and interaction requirements necessary to be retained in the XP output for 7RPZ. This underscores the unique geometric and chemical demands of the P2 pocket in $KRAS^{G12D,\ GDP}$.

Green and blue horizontal lines mark the DS of the top-performing reference ligands in each docking run, i.e. BI-2852 and MRTX-1133. BI-2852, recognized as the highest-affinity ligand for the P1 site, consistently yielded more negative DS values than Ch-3, even when docking was performed against 6GQY, which is the crystal structure of the KRAS:Ch-3 complex.

The Ligands selected as targets for synthesis are indicated by red horizontal lines. Their structures and predicted binding poses are shown in **Figure 52**, along with the respective DS and predicted pChEMBL values. Notably, the latter was not calculated for cyctripep **92**, as the focus of this project had shifted exclusively towards cyclic tetrapeptides before the predictions were made.

Top Dock **92** (7RPZ)	Top Dock **93** (6GQY)	Top Dock **94** (7RPZ)
DS = -11.9 kcal/mol	DS = -11.9 kcal/mol	DS = -13.3 kcal/mol
$pChEMBL_{pred}$ = n.d.	$pChEMBL_{pred}$ = 4.41	$pChEMBL_{pred}$ = 4.39

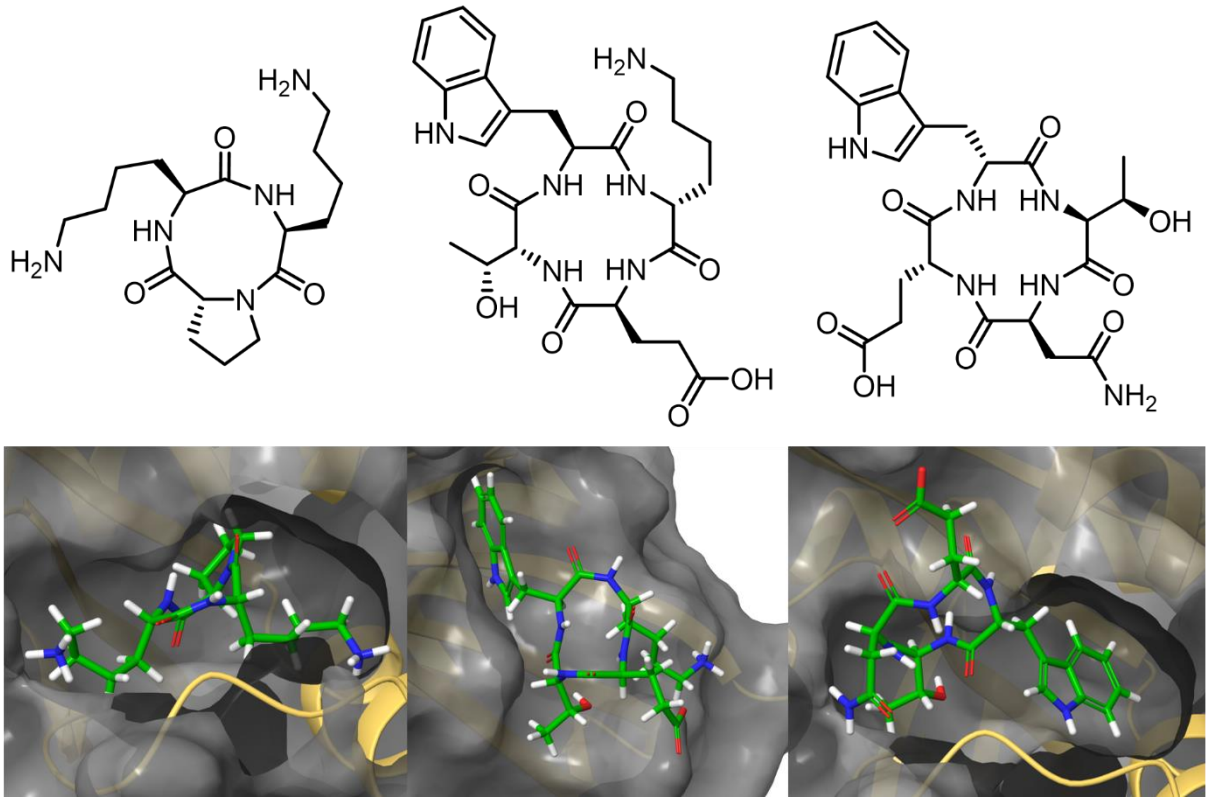


Figure 51: The ligands highlighted with green horizontal lines in **Figure 50** are depicted here with their docking poses below them. In the case of 7RPZ, part of the protein surface was removed to visualise the ligand inside the pocket.

Ligand selection was guided by combination of DS and anticipated synthetic accessibility. Despite the inherent challenges associated with the synthesis of cyctripep, compound **92** was selected due to precedent reports describing the successful synthesis of similar cyctripep analogues with acceptable yields.[247] Additionally, the incorporation of alternating *R*- and *S*-AAs was found to enhance macrocyclisation efficiency resulting in higher cyclisation yields, which identifies the cyctetpep analogues **93** and **94** as promising targets molecules.[24]

In the case of the click cyctetpep scaffold, derivatives **95**, **96**, and **97** were prioritised. The click reaction employed for macrocyclisation was anticipated satisfactory yields, even in the absence of turn-inducing residues (**Figure 52**). Furthermore, compounds **95** and **96** were selected for their chemically stable AA side chains, which minimise the risk of undesired side reactions and thereby support the robust optimisation of the

click-cyclisation protocol. Compound **97** was chosen to test the synthesis protocol's applicability to oxidation prone AAs such as Met.

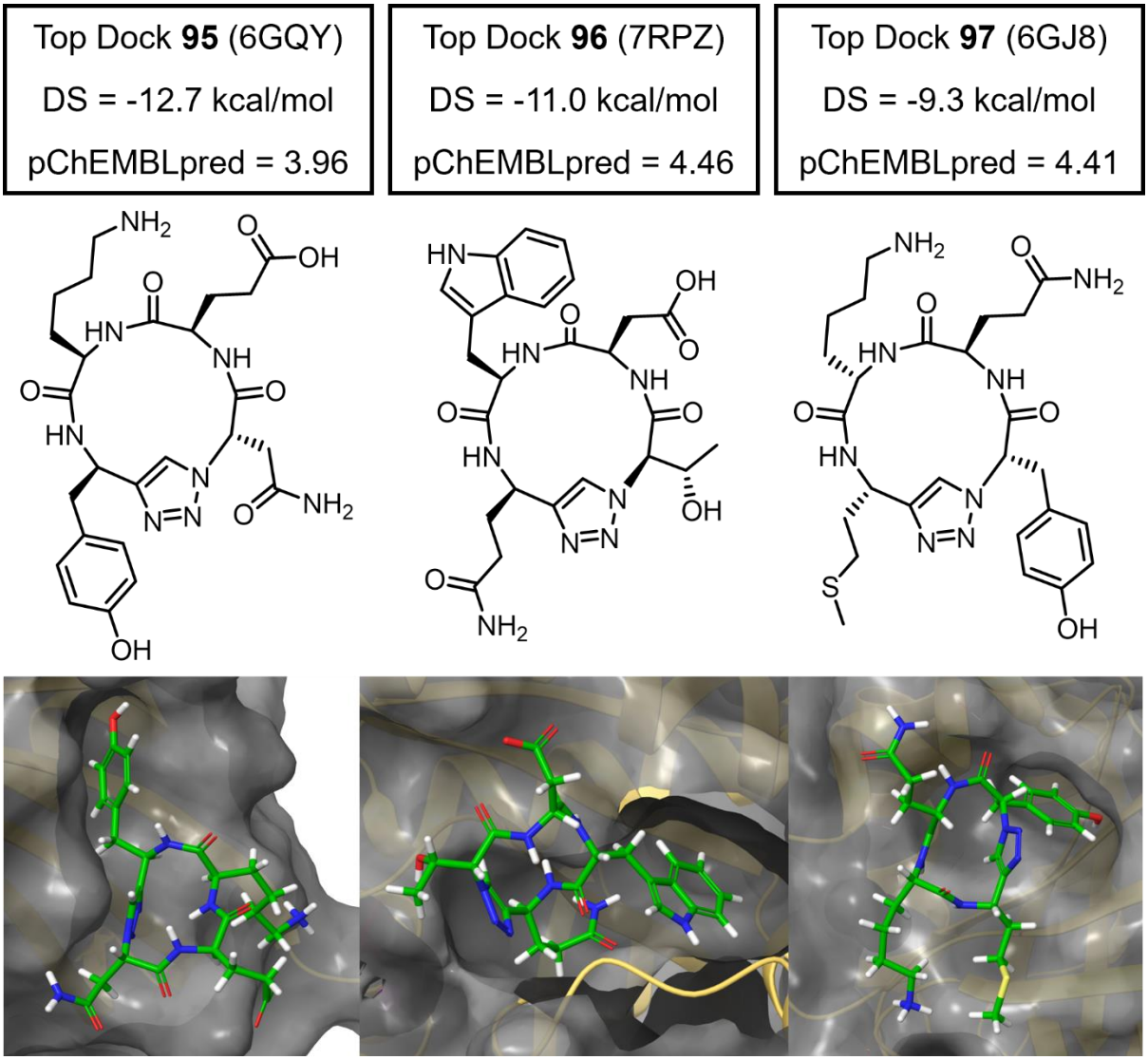


Figure 52: Continuation of **Figure 51**.

A noteworthy observation emerged during the evaluation of these target peptides: despite exhibiting excellent DS values, the compounds were only predicted to have moderate pChEMBL values. This discrepancy raises concerns about the reliability of rigid receptor docking as a sole metric for ranking macrocyclic peptides by their binding affinity and inhibitory potential toward KRAS. This concern is further substantiated by the previously noted wide variability in DS among the poses of each top-scoring peptide, indicative of potential limitations in docking specificity and scoring accuracy.

To address these concerns and gain a more comprehensive understanding of the ligand-protein interactions, molecular dynamics simulations (MDS) were conducted for

the top XP docking poses of click cyctetpep **95** and **96**. These simulations aimed to probe the dynamic behaviour of the complexes in solution, thereby providing a more nuanced assessment of binding stability, conformational flexibility, and key interaction motifs not fully captured in rigid receptor docking models.

2.3.2.7 Molecular Dynamics Simulations

The XP docking poses of the target cyctetpep derivatives **95** and **96** in **Figure 52** represent static models of protein-ligand complexes, optimised to reflect energetically favourable interactions. However, such static representations do not account for the conformational flexibility of proteins and ligands in solution, potentially leading to inaccurate predictions of actual binding modes. To overcome this limitation, MDS were employed, using the XP docking poses as starting structures. MDS simulate the motion of atoms over time, thereby offering dynamic insight into the stability and realism of the predicted binding poses. A binding pose that remains stable throughout the simulation is typically indicative of a viable interaction in solution, whereas significant positional drift or dissociation of the ligand from the binding site suggests an unstable or artefactual docking result.[248]

MDS were conducted for compounds **95** and **96** in complex with KRAS at P1 and P2 in the crystal structures 6GQY and 7RPZ, respectively (**Figure 52**). For comparison, simulations were also performed for the co-crystallised ligands Ch-3 (P1 in 6GQY) and MRTX-1133 (P2 in 7RPZ) using their experimentally resolved poses. Over a simulation duration of 500 ns, protein-ligand interactions were quantified, and interaction histograms were generated (**Figure 53**).

These histograms represent the interaction fractions, i.e. the proportion of simulation time a given interaction was maintained. Fractions in the range [0,1] denote single interaction persistence, while values >1 indicate multiple, concurrent interactions of the same type with the same residue.

A key observation from the histograms is that cyctetpep **95** and **96** engage with a substantially greater number of AA residues than their respective reference compounds, i.e. 7 *vs.* 13 for compound **95** compared to Ch-3 and 33 *vs.* 24 for compound **96** compared to MRTX-1133. The cyctetpep interact with same the residues as the reference molecules as well as with several more. Especially in the case of **95** these additional interactions are short-lived.

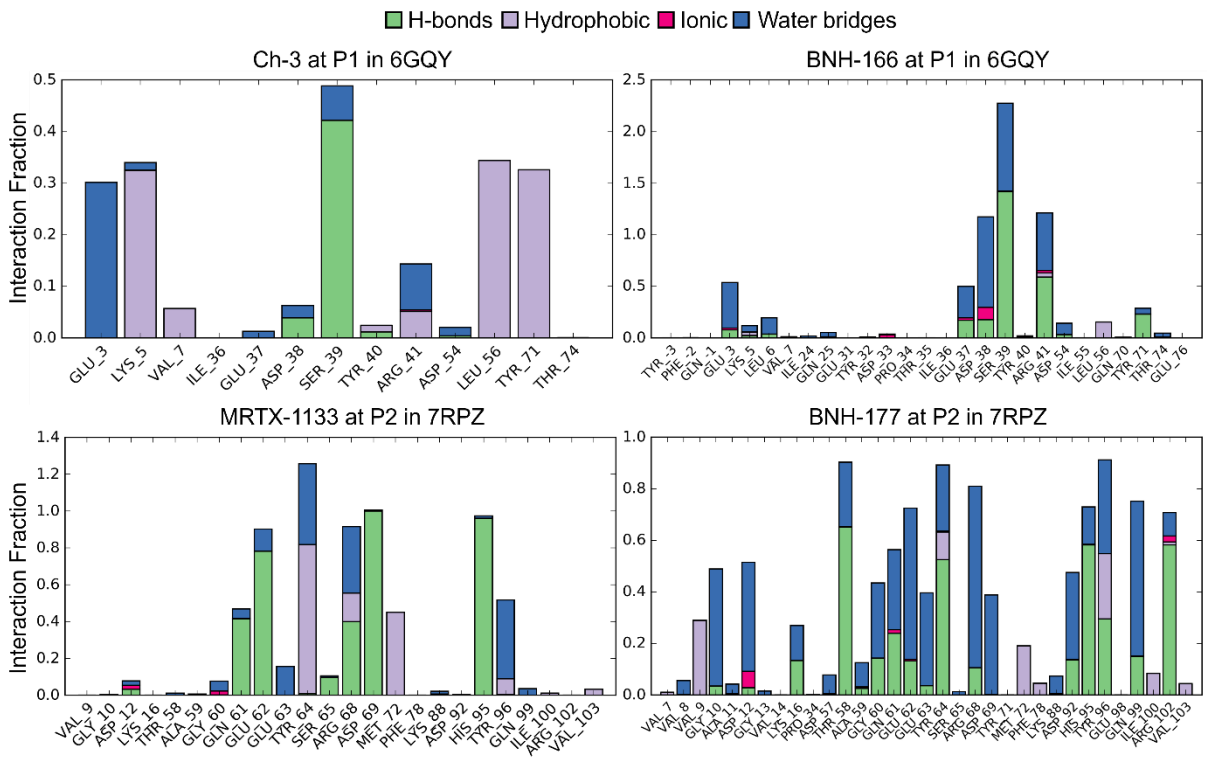


Figure 53: Interaction histograms calculated from the MDS results of Ch-3 (**8**) and BNH-166 (**95**) at P1 in 6GQY, as well as of MRTX-1133 (**2**) and BNH-177 (**96**) at P2 in 7RPZ. The interaction fractions represent the amount of time a specific residue-ligand interaction was detected. The interval [0,1] spans the complete simulation time frame of 500 ns. Values >1 indicate multiple, concurrent interactions of the same type with the same residue. Note that the y-axes' scales are not aligned between the histograms to allow for differentiation between interaction contributions of weak interactions in each histogram.

In addition to the number of interactions, the types of interactions differed notably. Whereas the reference compounds primarily exhibited H-bonds and hydrophobic contacts, the cyctetpep ligands formed a larger proportion of H-bonds and water-mediated bridges, with a marked increase in ionic interactions.

The interaction profiles summarised in **Figure 53** suggest that, in comparison to the reference ligands, the click cyctetpep **95** and **96** form less site-specific interactions with the KRAS protein targets. The reference compounds, Ch-3 and MRTX-1133, primarily engage in hydrophobic interactions and well-defined H-bonds with a limited subset of AA residues, indicative of site-specific and stable binding. Conversely, compounds **95** and **96** interact with a broader array of residues and solvent molecules, a pattern consistent with non-specific binding. This includes ligand pose variability and episodes of partial solvation, reflecting reduced conformational stability within the binding pocket.

These qualitative observations are substantiated through binding free energy (ΔG_{bind} in kJ/mol) analysis conducted over the 500 ns MDS. By calculating ΔG_{bind} for each snapshot along the simulation timeline and plotting these values against the snapshot IDs, we can identify decomplexation events, i.e. sudden losses in binding affinity indicative of ligand detachment or substantial pose rearrangement (**Figure 54**).

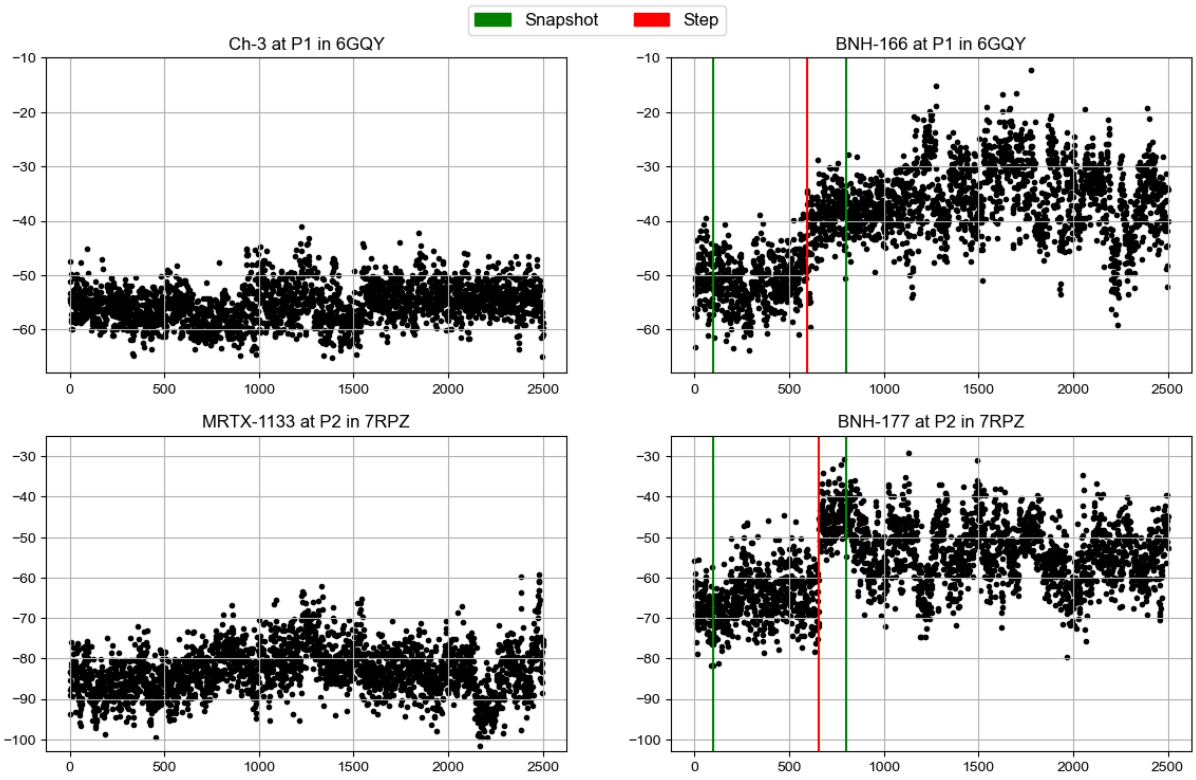


Figure 54: Scatter plots of ΔG_{bind} [kJ/mol] of simulated protein:ligand complexes plotted against the snapshot IDs across the simulation timeline. Exactly 2502 snapshots were taken in regular intervals of roughly 200 ps across the simulation timeframe of 500 ns. Red vertical lines indicate sharp increases in ΔG_{bind}. Green vertical lines identify the snapshots depicted in **Figure 55**.

The scatter plots of Ch-3 at P1 in 6GQY (top left) and MRTX-1133 at P2 in 7RPZ (bottom left) show stable ΔG_{bind} values in the ranges of approximately [-60,-50] kJ/mol and [-100,-70] kJ/mol, respectively. Crucially, no abrupt changes were observed throughout the full simulation timeframes, confirming their stable binding poses.

In contrast, compounds **95** at P1 in 6GQY (top right) and **96** at P2 in 7RPZ (bottom right) exhibited sudden increases in ΔG_{bind} (red vertical line), signalling instability of the binding poses. Comparison of the cyctetpep poses in the snapshots, located at the green vertical lines, highlight the changes in conformation (**Figure 55**).

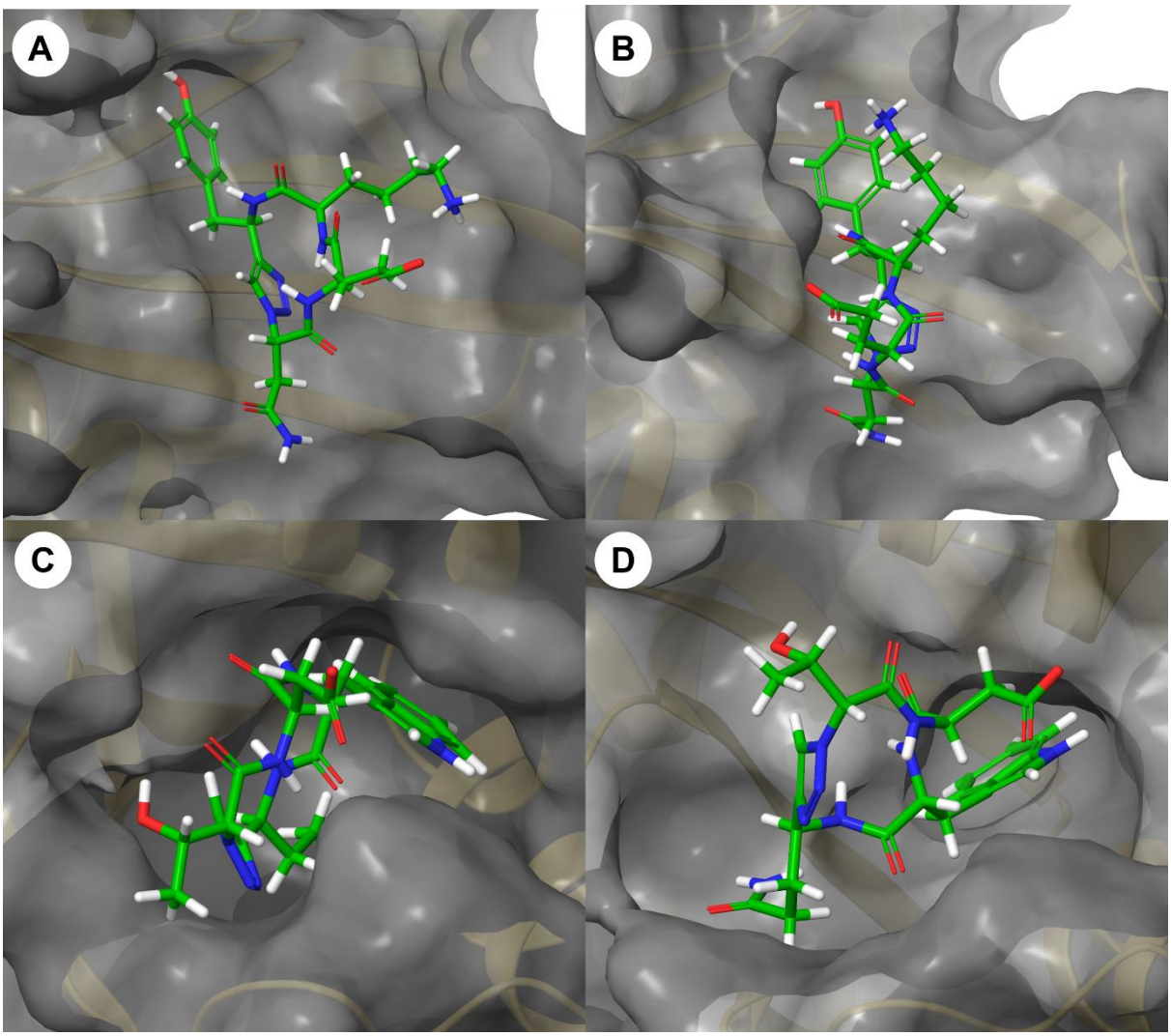


Figure 55: Snapshots of the simulated protein ligand complexes indicated with green vertical lines in the ΔG_{bind} scatter plots shown in **Figure 54**. The complexes: A) BNH-166 at P1 in 6GQY, B) BNH-166 at P1 in 6GQY, C) BNH-177 at P2 in 7RPZ and D) BNH-177 at P2 in 7RPZ.

The comparison between snapshots A and B for compound **95** reveals a migration of the macrocycle from one region of P1 to another, accompanied by detachment of the ε-amine group of Lys and a contraction of the P1 pocket, thereby diminishing hydrophobic interactions.

Even more pronounced changes were observed for compound **96** in snapshots C and D. Unlike P1, the P2 pocket is an induced binding site, formed in response to MRTX-1133.[95] The data show that cyctetpep **96** fails to stabilise P2 comparably, leading to a marked expansion of the pocket volume between snapshots C and D. This structural shift was accompanied by a 90° rotation of the indole side chain of Trp inside P2 and migration of the macrocycle from one region of the pocket to another.

In contrast to the drastic pose changes observed for cyctetpep **95** and **96**, the binding poses of the references MRTX-1133 and Ch-3 remained consistently stable throughout the entire simulation. These findings corroborate the interaction histograms shown in **Figure 53** and the ΔG_{bind} scatter plots in **Figure 54**.

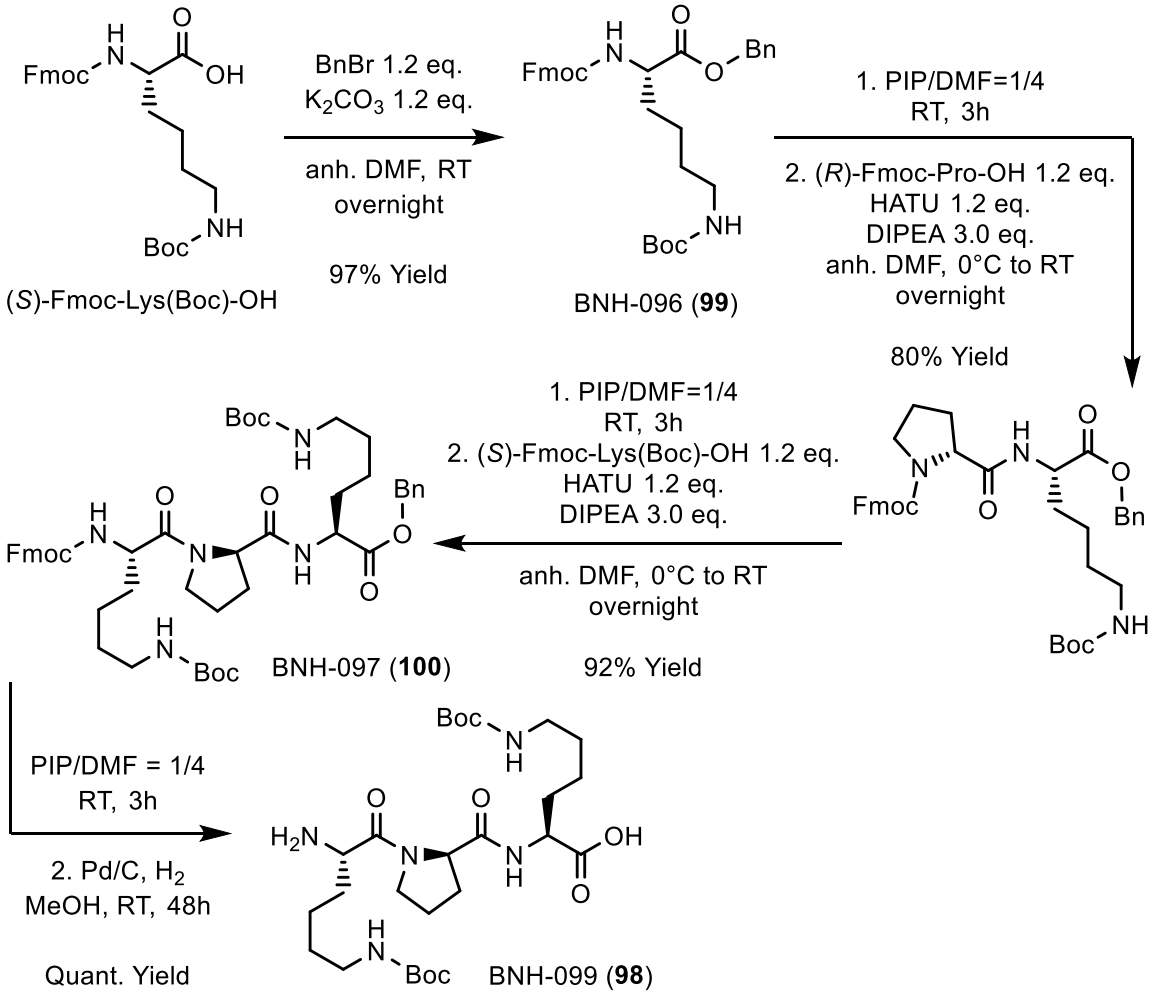
The MDS results for cyctetpep **95** at P1 in 6GQY and cyctetpep **96** at and P2 in 7RPZ indicate that both compounds exhibit weaker and less specific interactions with KRAS than initially suggested by their excellent DS of -12.7 kcal/mol (**95**) and -11.0 kcal/mol (**96**). These findings underscore the limitations of rigid receptor docking when used as the sole criterion for ranking macrocyclic peptides in terms of binding affinity and inhibitory potential toward KRAS. In contrast, the predicted pChEMBL values of click cyctetpep **95** and **96** correspond to moderate inhibitory potencies, with estimated IC_{50}(NE) values of 110 µM (**95**) and 34.7 µM (**96**), as calculated using **Equation 3**. These values align more closely with the transient binding behaviour and limited site specificity observed in the MDS.

Although the pChEMBL values indicate limited potency (**Figure 51** and **Figure 52**), the synthesis of the selected cyclic peptides was pursued for several compelling reasons. First, MRTX-1133, RMC-7977 and LUNA18 (**Figure 17**) constitute the majority of non-covalent KRAS PPIIs with sufficient potency for clinical application, underscoring the critical need for novel KRAS-targeting scaffolds. Second, the modular architecture of cyctripep and cyctetpep scaffolds enables rational design of ligands tailored to both P1 and P2 binding sites, offering a versatile platform for systematic optimisation of binding affinity and biological activity. Lastly, the chemical space defined by cyctripep and cyctetpep remains entirely unexplored in the context of KRAS inhibition, making their successful synthesis a valuable step toward expanding the toolbox for drug discovery targeting this historically challenging protein.

2.3.3 Cyclic Peptide Syntheses

2.3.3.1 Attempted Synthesis of Native Cyctripep **92**

The linear tripeptide BNH-099 (**98**) was selected as a precursor for the cyctripep **92**. Its peptide sequence, (*S*,*R*,*S*)-Lys(Boc)-Pro-Lys(Boc), features the turn-inducing element Pro at the centre, bringing its termini in close proximity and thereby facilitating cyclisation (**Scheme 26**).

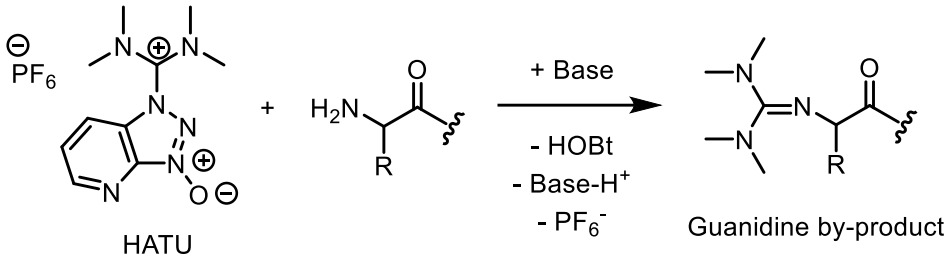


Scheme 26: Solution-phase synthesis of BNH-099 (**98**).

BNH-099 (**98**) was synthesised *via* solution-phase peptide coupling. The synthesis began with the benzylation of (*S*)-Fmoc-Lys(Boc) to afford BNH-096 (**99**) in excellent yield, followed by two sequences of Fmoc deprotection and HATU coupling.

Fmoc deprotection using piperidine in DMF required overnight evaporation of the deprotection reagents prior to peptide coupling. Furthermore, when employing uronium-based coupling reagents such as HATU, optimal yields and purity were

achieved by first mixing HATU with the carboxylic acid to be coupled. In this manner, the concentration of HATU in the reaction mixture is minimised by the time the amine is introduced. This pre-activation of the carboxylic acid is crucial, as uronium reagents like HATU can react with primary amines to form guanidine by-products, as shown in **Scheme 27**.[139]

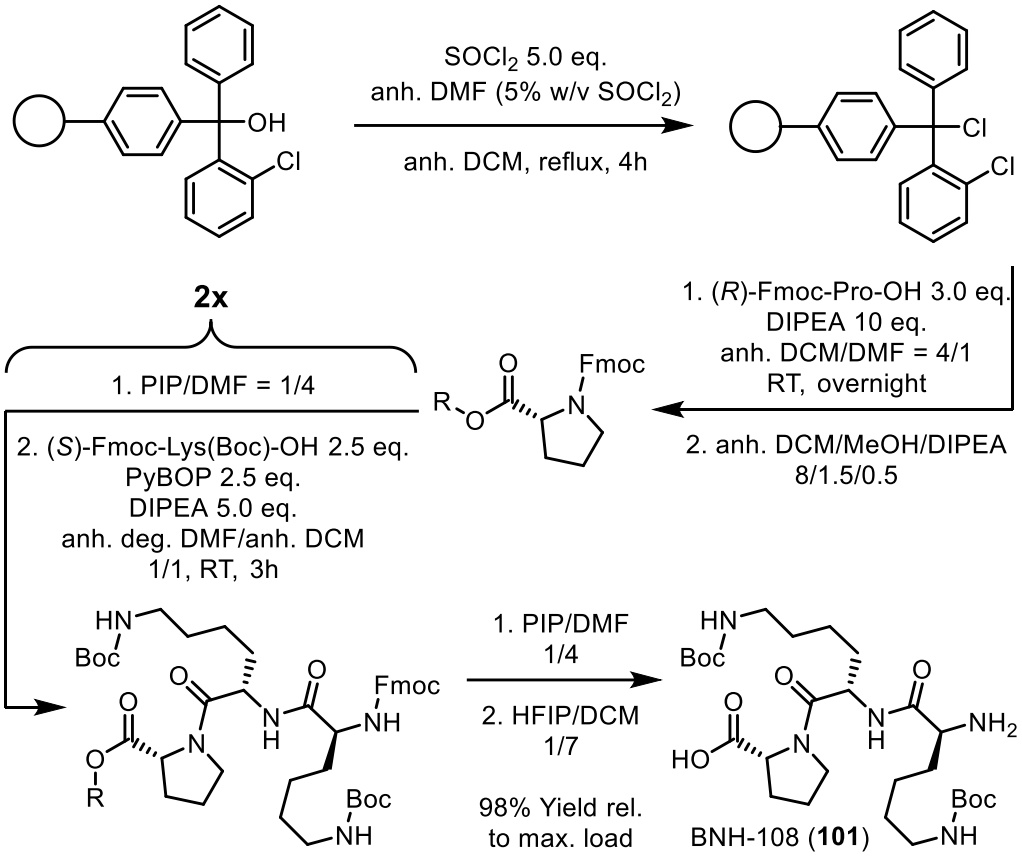


Scheme 27: Undesired side reaction between uronium-based coupling reagents, such as HATU, and primary amines. *N*-terminal guadinylation of peptides leads to truncated peptide sequences.[139]

The described pre-activation protocol was applied for both peptide couplings, affording the protected linear tripeptide BNH-097 (**100**). The deprotection-coupling sequence yielded the protected tripeptide **100** in 73% yield over four steps. In a subsequent two-step deprotection, the Fmoc group was removed with piperidine in DMF, followed by Pd-catalysed hydrogenation to remove the benzyl group. The linear tripeptide BNH-099 (**98**), featuring unprotected termini, was obtained in quantitative yield from tripeptide **100**, corresponding to an overall yield of 71% over seven steps.

Unfortunately, the solution phase synthesis of tripeptide **98** proved more laborious and time-consuming than initially anticipated. Streaking of intermediates on silica gel columns complicated the purification process. Additionally, the complete evaporation of the piperidine/DMF mixture required considerable time. Consequently, all subsequent linear peptide sequences were synthesised *via* the milder and more efficient solid-phase peptide synthesis (SPPS) using 2-chlorotrityl chloride (CTC) resin.

The linear tripeptide BNH-108 (**101**) was synthesised to assess the cyclisation efficiency of an alternative linear precursor of cyctripep **92** (**Scheme 28**). The first step involved resin activation by refluxing with $SOCl_2$ and anhydrous DMF in anhydrous DCM.[249] The activated, chlorinated resin should be used promptly, as it hydrolyses to the corresponding alcohol upon storage in the freezer for several weeks.



Scheme 28: Solid-phase peptide synthesis of BNH-108 (**101**). The specified equivalents and yield correspond to the maximum loading of the resin reported by the supplier.

The activated resin was loaded by shaking with excess (*R*)-Fmoc-Pro-OH in the presence of base overnight at room temperature. The specified equivalents correspond to the maximum loading capacity of the resin as reported by the supplier. Following loading with the first AA, unreacted trityl chloride groups on the resin were capped by conversion to the respective methyl ethers using MeOH and DIPEA in anhydrous DCM.

Subsequently, two (*S*)-Lys residues were appended to (*R*)-Pro moiety *via* two deprotection-coupling sequences. Each cycle involved Fmoc deprotection using piperidine in DMF, followed by coupling of (S)-Fmoc-Lys(Boc)-OH using PyBOP as the coupling reagent. Compared to HATU, PyBOP offers the advantage of reduced reactivity towards amines, thereby minimising the formation of undesired by-products.

The *N*-terminal Fmoc-group was removed with piperidine in DMF prior to cleavage of the protected tripeptide **101** from the resin. Notably, tripeptide **101** was obtained in excellent yield (98%) relative to the theoretical maximum loading of the resin. The outstanding purity of the peptide is partly attributed to the mild cleavage conditions

employed, i.e. hexafluoroisopropanol (HFIP) in DCM, which were identified as optimal following the screening of five different cleavage cocktails.

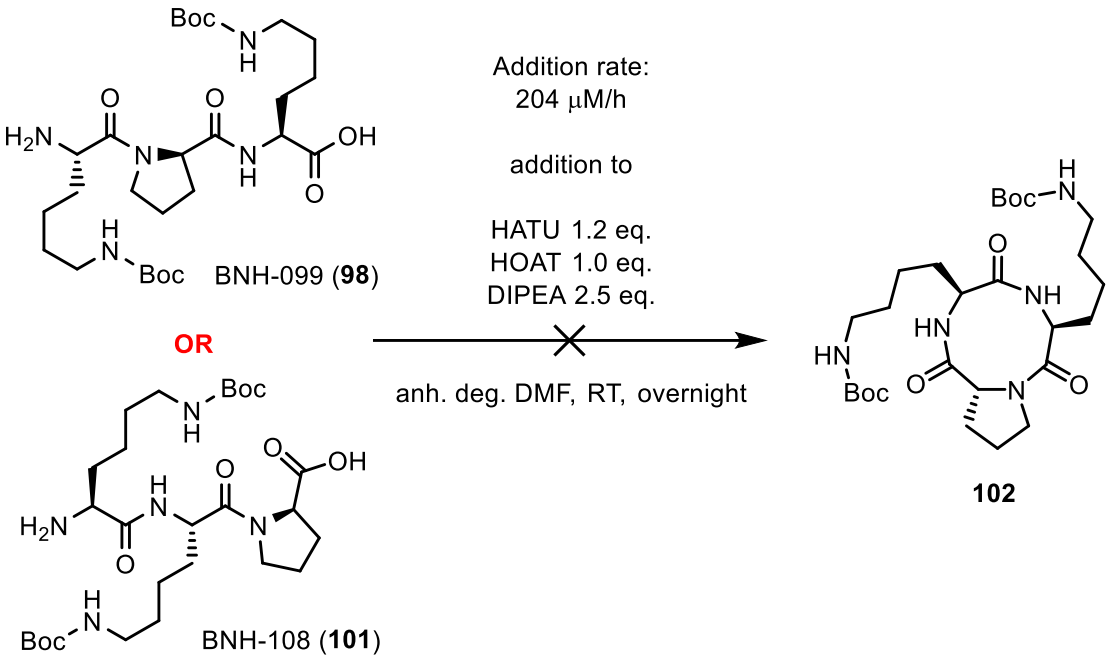
Reagents v/v	Comment
HFIP/DCM = 1/7	High purity
HFIP/TFE/DCM = 1/2/7	
AcOH/TFE/DCM = 1/1/8	
TFA/DCM = 1/99	Fragmentation observed
TFA/DCM = 2/98	
TFA/TIPS/H_2O = 95/2.5/2.5	Reference

Table 11: Tested cleavage cocktails for CTC resin.

TFA/TIPS/H_2O was employed as the reference deprotection cocktail, as it efficiently removed all protecting groups. All tested cleavage cocktails afforded the protected peptide as the major product. Notably, the TFA-containing cocktails generated several peptide fragments as minor by-products, while all HFIP-based cocktails yielded the cleaved peptide in good purity. Among these, the HFIP/DCM mixture was selected due to its ease of preparation and the superior product purity achieved compared to other formulations.

These findings are rationalised by the relatively high pKa of HFIP (9.3), which is comparable to that of the mild acid NH_4^+ (pKa = 9.2).[250,251] In contrast, other commonly used acids such as AcOH and TFA have significantly lower pKa values of 4.8 and 0.23, respectively.[251,252] In other words, HFIP is sufficiently acidic to cleave the peptide from the resin while preserving the integrity of the side chain protecting groups.

Cyclisation of the linear precursors **98** and **101** to yield the protected cyctripep **102** was attempted under high-dilution conditions. Each linear precursor was dissolved separately in anhydrous, degassed DMF (10 mL) and added with a syringe pump to a solution of HATU, HOAt, and DIPEA in anhydrous degassed DMF (50 mL), as illustrated in **Scheme 29**. This cyclisation strategy, previously applied successfully to aziridine-containing tetrapeptides, is known to promote macrocyclisation by minimising intermolecular side reactions.[253]



Scheme 29: Attempted cyclisation of the linear tripeptides **98** and **101** towards the protected cyctripep **102** in high-dilution conditions.

The addition rate of the linear tripeptides **98** and **101** to their respective reaction solutions is defined as:

$$Addition\ rate = \frac{total\ amount\ of\ substance\ [\mu mol]}{total\ volume\ of\ solvent\ [L] * \ total\ time\ of\ addition\ [h]}$$

Equation 5: Definition of the addition rates discussed in this project.

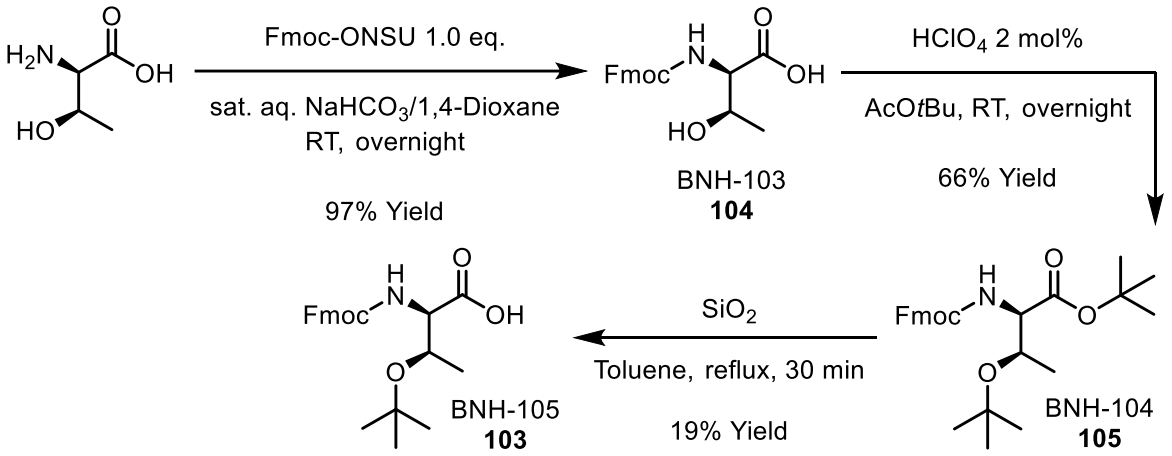
The calculated addition rate indicates that the average concentration of either tripeptide in the coupling solution could alternatively be achieved in 858 mL of solvent, had the linear precursors been introduced in a single portion. This comparison assumes that the cyclisation reaction proceeds rapidly relative to the time scale of addition. The intentionally low transient concentrations of precursors **98** and **101** in solution were intended to maximise the probability of intramolecular cyclisation and minimise the likelihood of intermolecular polymerisation.

However, no conversion to the protected cyctripep **102** was detected by LC/MS analysis. Instead, a complex mixture of products was observed, characterised by long retention times and high m/z values, indicative of oligomerisation or polymerisation. These findings suggest that the cyclisation reaction was too slow to outcompete the accumulation of linear precursors during the addition phase, thus favouring undesired intermolecular reactions.

Discouraged by these results, the focus of the project was redirected toward the synthesis of cyclic tetrapeptides. In contrast to the challenging cyclisation of tripeptides, a comparatively broad range of synthetic protocols has been reported for the preparation of this scaffold from linear tetrapeptides with diverse side chains.[24]

2.3.3.2 Attempted Synthesis of Native Cyctetpep **93** and **94**

The target cyctetpep **93** contains a (*R*,*R*)-Thr residue, necessitating the use of the protected AA (*R*,*R*)-Fmoc-Thr(*t*Bu)-OH (**103**) in the SPPS protocol. Compound **103** was synthesised from (*R*,*R*)-H_2N-Thr-OH *via* a three-step sequence involving selective protection of both the amine and the hydroxyl group, following a published protocol (**Scheme 30**).[254]



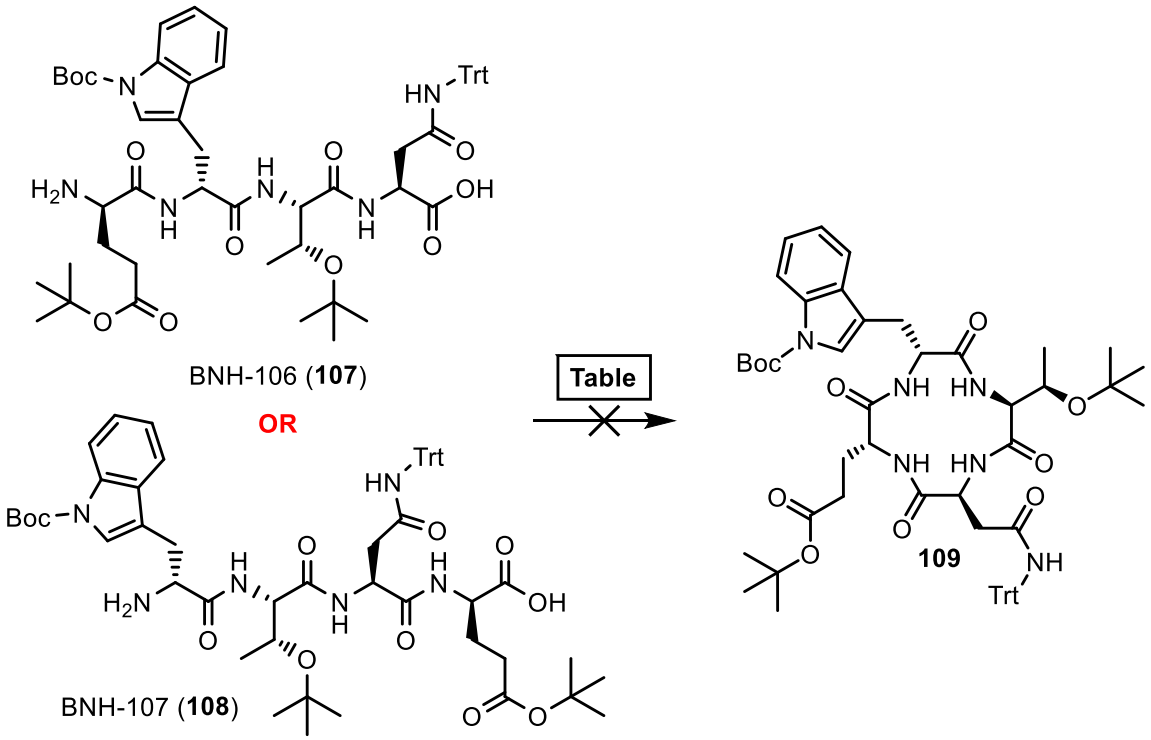
Scheme 30: Synthesis of (*R*,*R*)-Fmoc-Thr(*t*Bu)-OH (**103**) following a published protocol.[254]

Selective Fmoc protection of the α-amine was achieved using *N*-(9-Fluorenylmethoxy-carbonyloxy)succinimide (Fmoc-ONSU) in a basic medium, affording intermediate BNH-103 (**104**). Subsequent acid-catalysed alkylation of the alcohol and carboxylic acid moieties using *tert*-butyl acetate yielded BNH-104 (**105**) as the bis-*tert*-butyl protected intermediate. Final selective hydrolysis of the *tert*-butyl ester using silica in boiling toluene provided the desired protected Thr derivative **103** in 12% overall yield across three steps.

The protected Thr derivative **103** was employed in the SPPS of the linear tetrapeptide BNH-112 (**106**), utilising the same protocol as for the synthesis of protected tripeptide **102** (**Scheme 28**). Analogous procedures were used to synthesise the linear tetrapeptides BNH-106 (**107**) and BNH-107 (**108**), which were obtained with over 95%

purity, according to LC/MS analysis, without the need for further additional purification post-resin cleavage.

The protected tetrapeptides **106**, **107**, and **108** were used as precursors in the cyclisation reactions towards cyctetpep **93** and **94** (**Figure 52**). A broad range of macrocyclisation conditions were explored for the cyclisation of the linear precursors **107** and **108** towards the protected cyctetpep **109** (**Scheme 31**, **Table 12**).



Scheme 31: Attempted cyclisation of the linear tetrapeptides **107** and **108** towards the protected cyctetpep **109** employing the reaction conditions summarised in **Table 12**.

Coupling Reagent	Base	Solvent	Addition Rate [µM/h]
HATU 1.2 eq	DIPEA 8.0 eq.	DMF	120
HATU 1.5 eq.	DIPEA 4.0 eq.	DMF	120
HATU 1.5 eq.	DIPEA 4.0 eq.	DMF	198
HATU 3.0 eq.	DIPEA 4.0 eq.	DMF	120
HATU 3.0 eq.	DIPEA 4.0 eq.	DCM	120
PyBOP 4.0 eq.	DIPEA 4.0 eq.	DMF	120
DMTMM BF_4 3.0 eq.	DIPEA 4.0 eq.	DMF	120

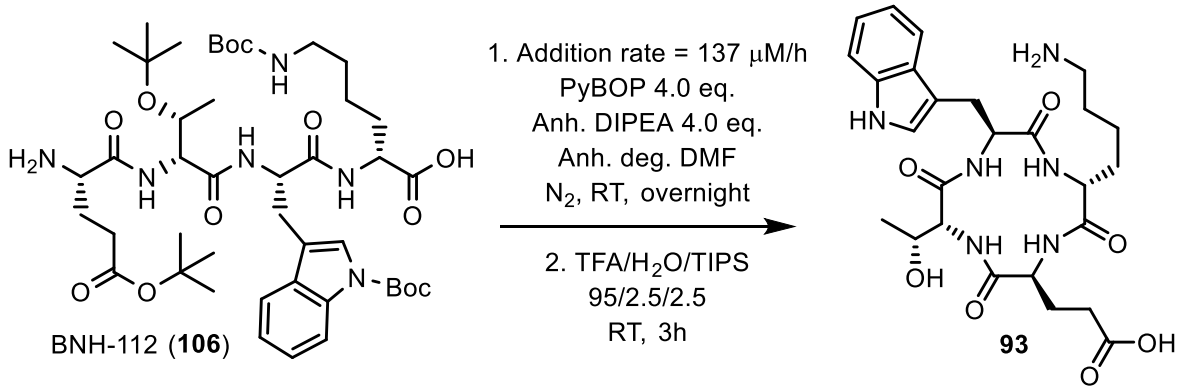
Table 12: Tested conditions for the cyclisation depicted in **Scheme 31**. The addition rates were calculated according to **Equation 5**.

Reference values for the reaction parameters were derived from previously published protocols for the cyclisation of linear tetrapeptides.[253,255] Both protocols employed moderate excesses of HATU and DIPEA, with comparable addition rates of

162 μM/h[253] and 107 μM/h[255]. In this study, the parameters systematically varied included the quantity of base and coupling agent, the identity of the coupling agent, the solvent system, and the peptide addition rate.

Despite these optimisations, no improvement in the conversion to cyctetpep **109** was observed by LC/MS analysis. Prolonged stirring for up to 48 hours post-addition had no effect on the yield either. In all cases, only trace amounts of the desired cyclic product were detected. The amount of cyctetpep **109** and the nature of side products varied slightly depending on the coupling reagent used. Notably, cyclisation employing 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium tetrafluoroborate (DMTMM BF_4) produced multiple broad peaks with similar m/z values (~985), indicative of undesirable epimerisation and peptide fragmentation.

The cleanest conversion was achieved with the coupling agent PyBOP, yielding a crude product mixture displaying a single broad peak with the expected m/z = 985, corresponding to the target macrocyclic product. Encouraged by this result, a preparative-scale cyclisation of linear tetrapeptide **106** was undertaken (**Scheme 32**).



Scheme 32: Successful cyclisation of linear tetrapeptide **106** towards the target cyctetpep **93**.

The linear tetrapeptide **106** was added to a solution of PyBOP and DIPEA in anhydrous, degassed DMF at an addition rate of 137 μM/h. Following cyclisation, the protected intermediate was purified by preparative HPLC, affording 9.4 mg of a white, amorphous solid. LC/MS analysis revealed the presence of significant impurities. Subsequent global deprotection yielded the target compound, which was isolated by HPLC as a white, amorphous solid. However, the isolated amount was insufficient for accurate quantification. Nevertheless, LC/MS and high-resolution mass spectrometry (HRMS) confirmed that the desired cyctetpep **93** was obtained with a purity of ≥ 95%.

Despite the high analytical purity, the extremely low isolated yield precluded further characterisation or biological evaluation. It is worth noting that cyctetpep **93** differs from target cyctetpep **94** (obtained from deprotection of **109**, **Scheme 31**) in its alternating absolute configurations at the α-carbon positions. Cyclic tetrapeptides with such stereochemical alternation have been reported to adopt flatter ring conformations and to be synthesised in higher yields, compared to analogues lacking this feature.[256]

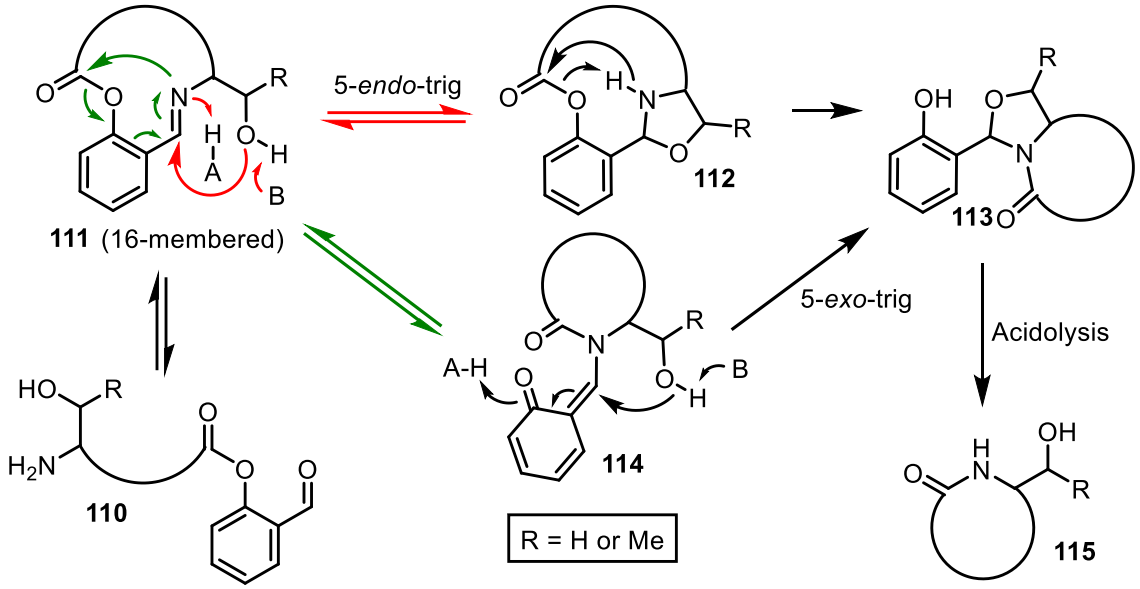
Given the poor yield observed for cyctetpep **93**, head-to-tail cyclisation of tetrapeptides *via* standard peptide coupling protocols was not further pursued. Alternative strategies that reduce the strain-induced activation barrier of tetrapeptide cyclisation have been described. One such approach is *imine-induced ring contraction*, which offers a promising avenue for improving the efficiency of cyclic tetrapeptide synthesis.[24]

2.3.3.3 Attempted Imine-Induced Ring Contraction Towards **94**

The synthesis of nine cyclic tetrapeptides through imine-induced ring contraction was published by WONG *et al*.[204] The authors demonstrated that even tetrapeptides composed entirely of L-AAs and lacking turn-inducing elements could be cyclised with acceptable yields of 7–27%. They proposed that the efficiency of this approach arises from the formation of a transient 16-membered cyclic intermediate, which is considerably less strained than the corresponding 12-membered cyclic tetrapeptides (**Scheme 33**).

The mechanism initiates from a tetrapeptide salicylaldehyde (SAL) ester **110** in which Ser or Thr occupies the *N*-terminal position. The first step involves imine formation between the *N*-terminal amine and the aldehyde of the SAL moiety towards the 16-membered intermediate **111**.

Two ring contraction pathways have been proposed for intermediates such as **111**, as illustrated by the red and green arrows. The red pathway entails a 5-*endo*-trig ring closure to generate intermediate **112**, a transformation that is disfavoured according to Baldwin's rules.[257] In this route, the peptide termini in intermediate **112** are brought into close proximity, enabling formation of the desired 12-membered macrocycle in intermediate **113**.



Scheme 33: Proposed reaction mechanism of the tetrapeptide head-to-tail cyclisation through imine-induced ring contraction. The published protocol works with tetrapeptide SAL esters with *N*-terminal Ser or Thr. A-H and B represent AcOH and pyridine, respectively.[204]

In contrast, the green pathway involves a pericyclic electron migration that leads to dearomatisation of the SAL benzene ring, resulting in intermediate **114**. This is followed by a 5-*exo*-trig ring closure to form intermediate **113**, which is a favourable pathway according to Baldwin's rules.[257] The final transformation comprises hydrolysis of the hemiaminal in intermediate **113**, thereby furnishing the native cyctetpep **115**.

After proposing the ring contraction mechanism, WONG *et al.* estimated the activation energies associated with various cyclisation processes, including those outlined in **Scheme 33**. Their computational results indicated that both the initial imine formation and the subsequent imine-induced ring contraction require approximately half the activation energy of direct head-to-tail cyclisation, thereby rationalising the improved efficiency of this synthetic strategy. The energies were estimated for cyclisation reactions with the Ser-Ala-Ala-Ala SAL ester derivatives shown in **Figure 56**.

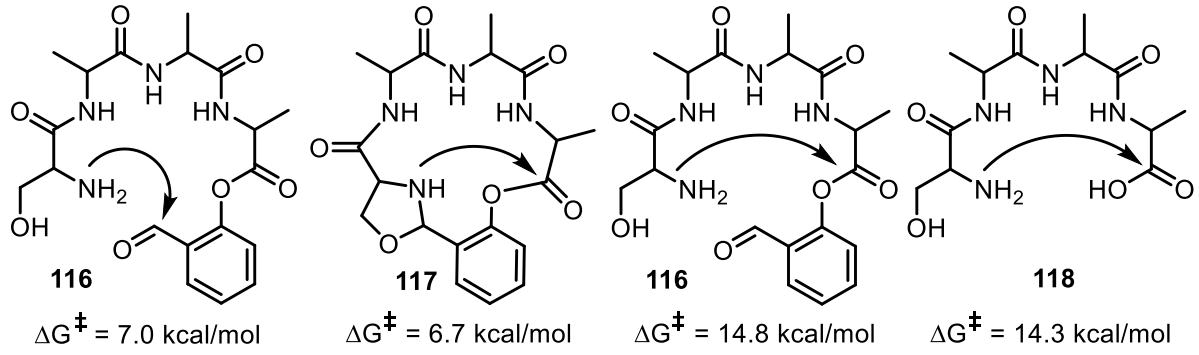
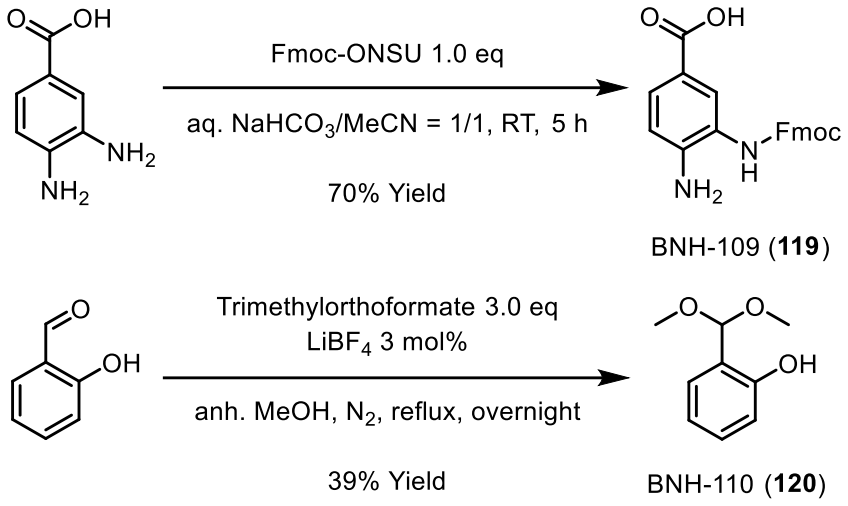


Figure 56: Comparison between cyclisation energies of native chemical ligation (SAL ester **116** and the resulting 16-membered intermediate **117**) and direct amide formation with ester **116** or acid **118**.[204]

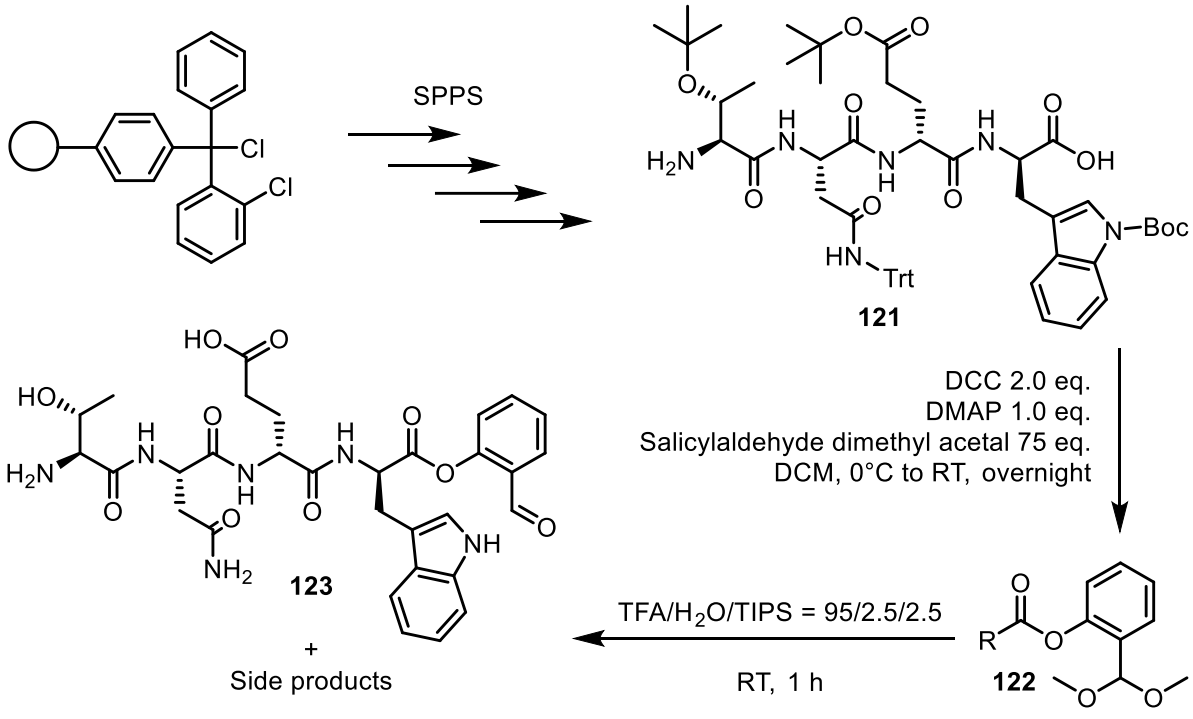
The imine formation step, commencing from SAL ester **116**, and the subsequent ring contraction of the resulting 16-membered intermediate **117** were calculated to exhibit activation barriers ($\Delta G^{‡}$) of 7.0 and 6.7 kcal/mol, respectively. In contrast, direct cyclisation of SAL ester **116** and peptide **118** was associated with significantly higher activation barriers of 14.8 and 14.3 kcal/mol, respectively.[204] Motivated by these favourable findings, the synthesis of cyctetpep **94** (**Figure 52**) was attempted through the imine-induced ring contraction strategy.

To this end, the requisite tetrapeptide SAL ester was synthesised either by Steglich esterification or by on-resin phenolysis.[258,259] These synthetic routes necessitated the intermediates BNH-109 (**119**) and BNH-110 (**120**). BNH-109 was obtained in good yield through Fmoc-protection of the *meta*-positioned amine in 3,4-diaminobenzoic acid using Fmoc-ONSU in basic medium (**Scheme 34**).[260]



Scheme 34: Synthesis of the precursors **119** and **120** through published procedures.[259,260]

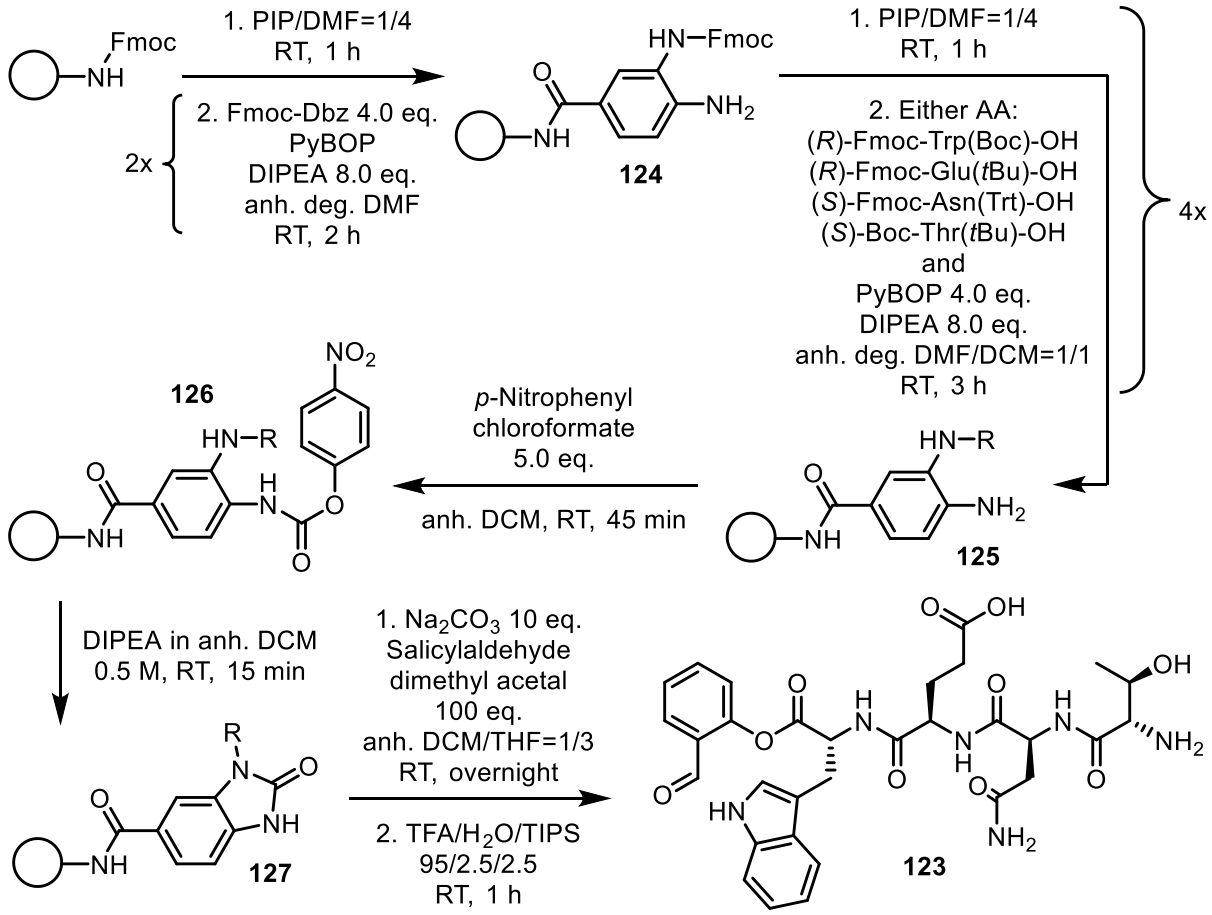
The dimethyl acetal **120** was synthesised from salicylaldehyde according to a published procedure involving trimethyl orthoformate and a catalytic amount the Lewis acid $LiBF_4$.[259] Subsequent distillation furnished the desired product in moderate yield. The Steglich esterification approach commenced with the synthesis of the protected tetrapeptide **121** through the SPPS procedure previously employed for the tripeptide **101** in **Scheme 28** (**Scheme 35**).



Scheme 35: Attempted synthesis of the SAL ester containing only small amounts of the desired peptide SAL ester **123** *via* Steglich esterification.

Subsequent Steglich esterification with a large excess of salicylaldehyde dimethyl acetal afforded >90% conversion to the corresponding ester **122**, as determined by LC/MS analysis. The reaction mixture was concentrated *in vacuo*, and the resulting residue was subjected to global deprotection using TFA/H_2O/TIPS. Precipitation of the product was achieved by the addition of cold MTBE, followed by centrifugation. However, LC/MS analysis of the precipitate revealed a complex mixture containing only small amounts of the desired peptide SAL ester **123**, alongside residual tetrapeptide **121** and various other side products.

Given the inefficacy of the Steglich esterification approach, the alternative on-resin phenolysis strategy was explored (**Scheme 36**).[258,260]

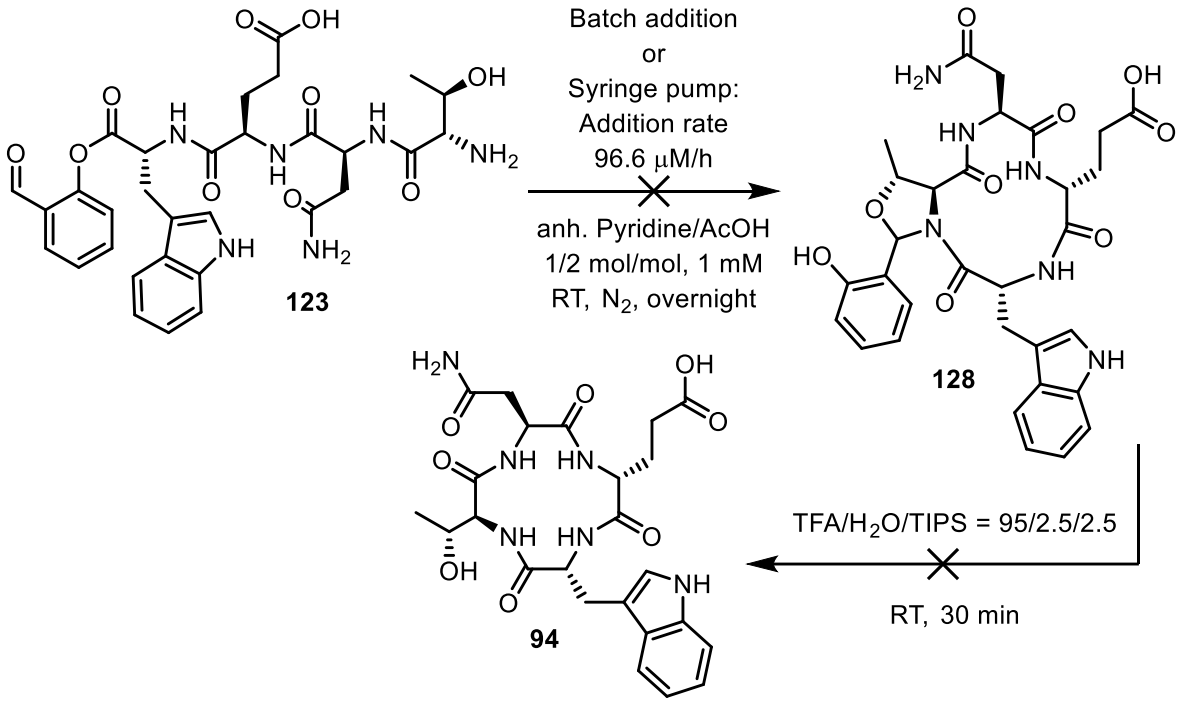


Scheme 36: On-resin phenolysis approach towards the unprotected tetrapeptide SAL ester containing only small amounts of the desired peptide SAL ester **123**.

The procedure commenced with deprotection of the Fmoc group from Rink amide MBHA resin, followed by loading with BNH-109 (**119**) under standard peptide coupling conditions. Complete loading of the resin was ensured by repeating the coupling step. Subsequently, the SPPS protocol previously described was employed to construct a tetrapeptide on the *meta*-positioned amine of the resin-bound Fmoc-Dbz (**124**). The resulting anchored tetrapeptide **125** was then subjected to carbamate formation on the *para*-positioned amine. Then resultant carbamate **126** underwent intramolecular cyclisation in the presence of DIPEA to yield the benzimidazolone derivative **127**. Phenolysis of the lactam linking the C-terminus of the tetrapeptide to the benzimidazolone nitrogen in intermediate **127** was achieved using a large excess of salicylaldehyde dimethyl acetal and Na_2CO_3. Following completion of the reaction, the mixture was filtered and the filtrate was concentrated under a stream of compressed

air. It is crucial to evaporate only the bulk of the DCM and THF, ensuring that the excess salicylaldehyde dimethyl acetal remains in the mixture. This excess is essential to minimise hydrolysis of the newly formed SAL ester during subsequent global deprotection under acidic conditions. The presence of salicylaldehyde dimethyl acetal favours transesterification with the acetal or aldehyde over hydrolysis by water.

According to LC/MS analysis, the desired SAL ester containing only small amounts of the desired peptide SAL ester **123** was obtained as the major product. The crude product was directly in the subsequent cyclisation step without further purification. For the cyclisation, the crude SAL ester was dissolved in a pyridine/AcOH mixture and stirred under N_2 at room temperature overnight.[204] The employed conditions were consistent with established protocols for Ser/Thr ligations (**Scheme 37**).[258,259,261]



Scheme 37: Attempted cyclisation of containing only small amounts of the desired peptide SAL ester **123** *via* imine-induced ring contraction.

A solution of SAL ester containing only small amounts of the desired peptide SAL ester **123** at a concentration of 1 mM was stirred overnight in buffer, resulting in complete consumption of the starting material. However, LC/MS analysis did not detect the formation of expected cyclic intermediate **128**. Instead, several side products were observed, with the free acid of the unprotected tetrapeptide emerging as the major product. This outcome suggests that hydrolysis of the SAL ester containing only small amounts of the desired peptide SAL ester **123** was the predominant side reaction,

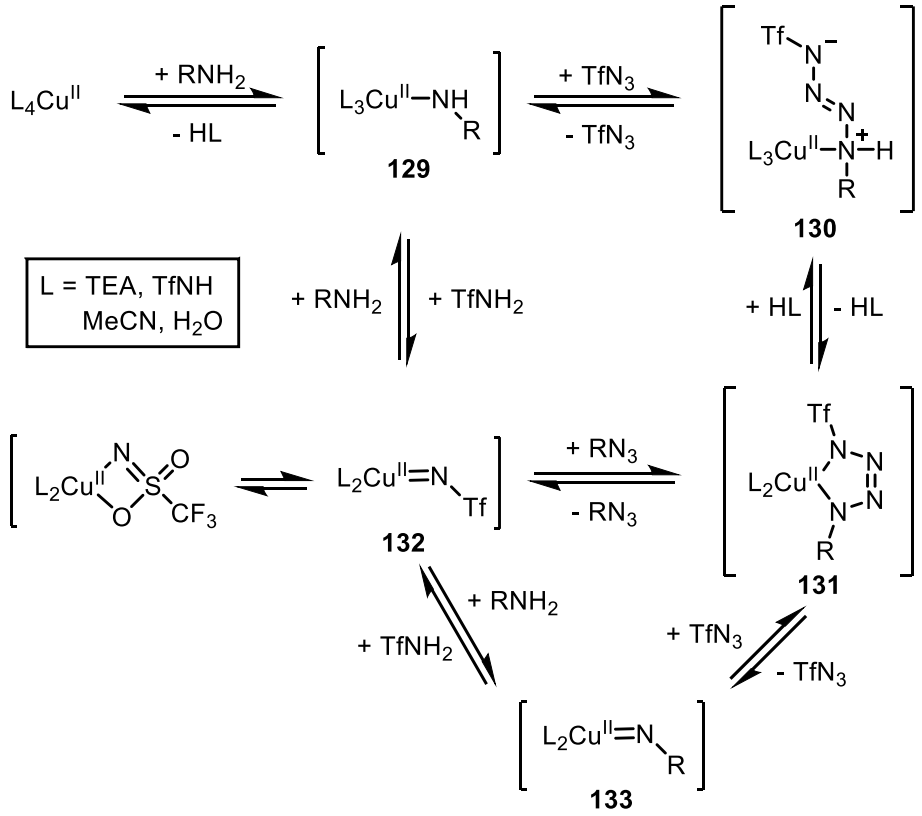
despite rigorous drying of laboratory glassware and reagents. In addition, undesired cyclodimerization may have contributed significantly to the consumption of the SAL ester, as such side reactions have been reported even at concentrations below 0.1 mM for tetrapeptide esters. In contrast, cyclisation of longer peptides (with more than four amino acids) has been reported to proceed cleanly even at concentrations as high as 10 mM.[261]

No improvement in yield was observed when the cyclisation of the ester containing only small amounts of the desired peptide SAL ester **123** was attempted under even higher dilution using a syringe pump. In this approach, a solution of the SAL ester in anhydrous DMSO (900 µL) was added over 10 hours to a pyridine/AcOH mixture with an addition rate of 96.6 µM/h. To rule out the possibility that the intermediate **128** was formed but was undetectable by LC/MS analysis, the reaction mixture was concentrated after overnight stirring and subsequently treated with TFA/H_2O/TIPS. Nevertheless, no formation of cyctetpep **94** was observed.

In light of these results, no further optimisation of the imine-induced cyclisation conditions was pursued. Instead, focus shifted to the more robust Cu-catalysed azide alkyne cycloaddition (CuAAC) approach.

2.3.3.4 Synthesis of α-Azido Acids

The α-azido acids required for the synthesis of compounds **95**-**97** were prepared from the corresponding AAs through a convenient two-step diazotransfer reaction. The reaction mechanism was initially proposed by FISCHER and ANSELME in the 1967 and later refined by NYFFELER *et al*.[262–264] In 2014, PANDIAKUMAR, SARMA and SAMUELSON substantiated the proposed mechanism through NMR studies employing isotope-labelled reagents (**Scheme 38**).[265]

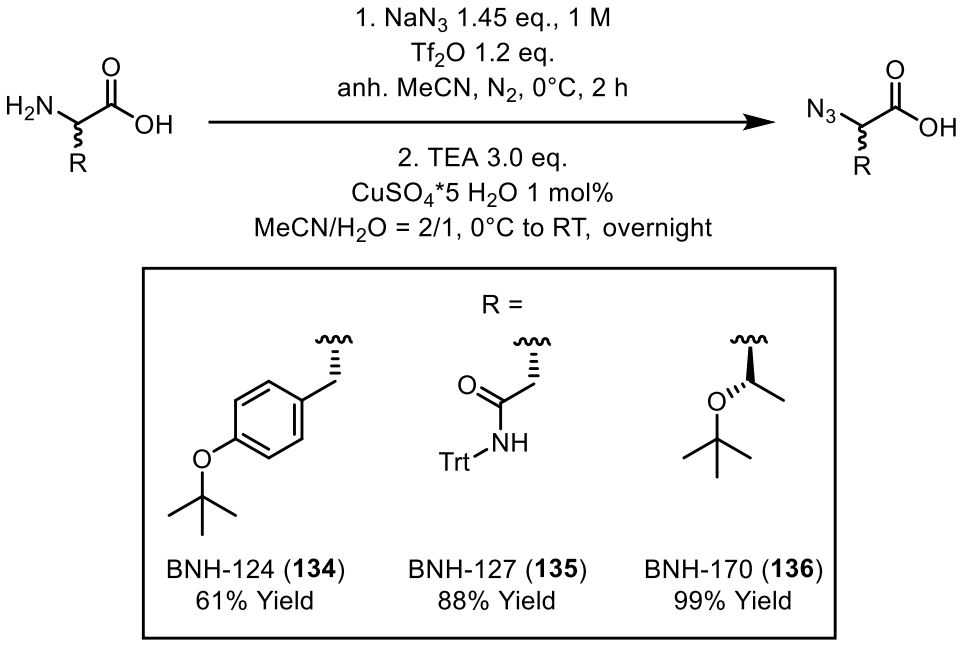


Scheme 38: Mechanism of Cu-catalysed diazotransfer from triflyl azide to primary amines. Modified from NYFFELER *et al.*[264]

In the first step of the mechanism, the α-amine of the AA coordinates to the metal centre (Cu^{II} or Zn^{II}), forming complex **129**. The amine then undergoes nucleophilic attack on the terminal nitrogen triflyl azide, generating complex **130**). Subsequent deprotonation of the amine nitrogen yields a cyclic tetrazene intermediate **131**, which undergoes a reverse [3+2] dipolar cycloaddition to afford the desired alkyl azide and a Cu^{II}-triflylamide complex **132**. The triflylamide ligand can then be displaced by another molecule of the AA starting material, thus perpetuating the catalytic cycle. The new round of the cycle can proceed either through the aforementioned pathway or through an alternative route involving the formation of a double bond between the starting

material amine and the Cu^{II} (complex **133**), followed by a [3+2] dipolar cycloaddition to produce the same tetrazene intermediate **131**.

Given the well-documented explosion risks associated with azide chemistry, particular caution was exercised in the selection of the reaction protocol. In particular, procedures that bring NaN_3 into contact with DCM were strictly avoided, as this may result in the formation of diazomethane, a highly volatile and explosive compound.[266] Accumulation of diazomethane during synthesis has been linked to catastrophic explosions.[267] Fortunately, safer alternatives have been developed.[268,269] Among these, a particularly convenient protocols involves the generation of triflyl azide from NaN_3 and triflyl anhydride in anhydrous MeCN. This method was employed for the synthesis of BNH-124 (**134**), BNH-127 (**135**), and BNH-170 (**136**), as shown in **Scheme 39**.



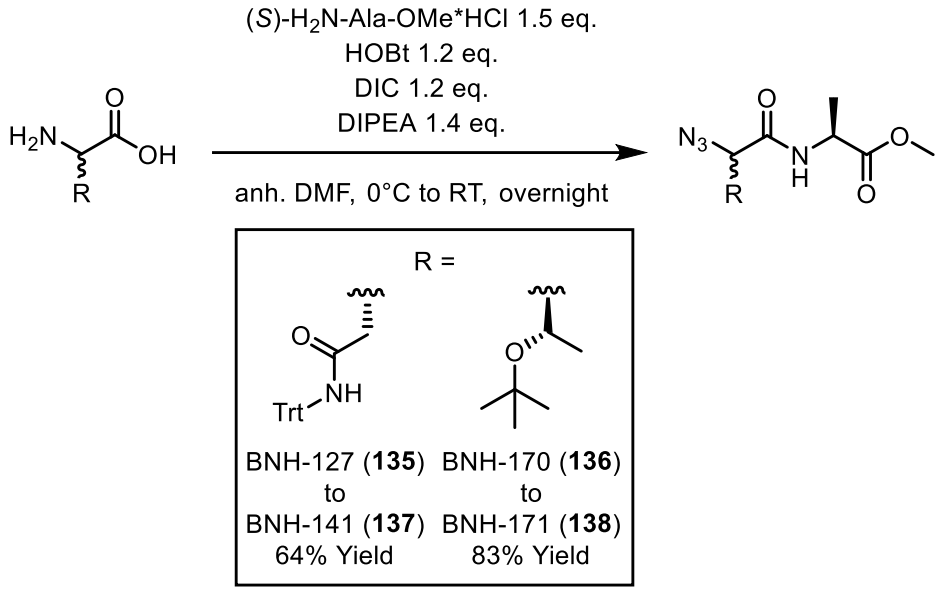
Scheme 39: Diazotransfer reactions towards the α-azido acids **134**–**136**.

In the first step, NaN_3 was suspended in anhydrous MeCN and the mixture was cooled to 0°C. The amount of NaN_3 was carefully calculated such that, if completely dissolved, the concentration would not exceed 1 M. This precaution was implemented to mitigate the risk of explosion associated with the subsequent formation of triflyl azide upon addition of triflyl anhydride (Tf_2O), which is known to be explosive at high concentrations.[270] After stirring the suspension for 2 hours, the triflyl azide solution was added to a precooled solution containing the respective AA, triethylamine (TEA) and a

catalytic amount of $CuSO_4$. The desired α-azido acids **134-136** were obtained in moderate to good yields.

Importantly, the yields were found to be highly dependent on the quality of the Tf_2O and the dryness of the MeCN. For optimal yields, freshly bought/prepared Tf_2O should be used, and rigorous exclusion of moisture must be ensured by using meticulously dried MeCN and oven-dried glassware.

The optical purity of the α-azido acids **135** and **136** was subsequently confirmed using a published method involving the coupling of the α-azido acids with (*S*)-H_2N-Ala-OMe to form the corresponding dipeptides (**Scheme 40**).[271]

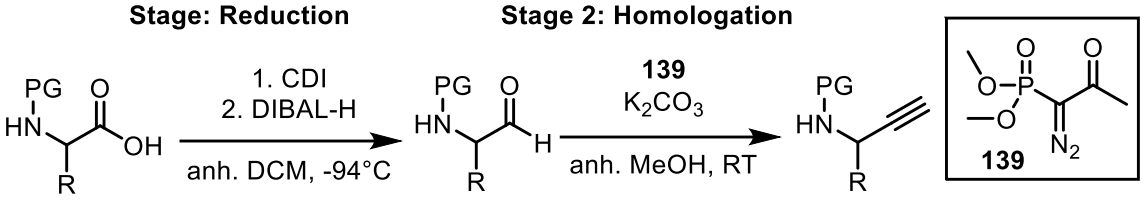


Scheme 40: Amide coupling reactions between the α-azido acids **135** or **136** and (*S*)-H_2N-Ala-OMe towards the dipeptides **137** and **138**.

Analysis of the dipeptides BNH-141 (**137**) and BNH-171 (**138**) by NMR spectroscopy revealed no evidence of epimerisation at the azido-substituted α-carbon. These findings ae consistent with previous literature reports indicating that racemisation does not occur under the employed diazotransfer reaction conditions.[266,269,272]

2.3.3.5 Synthesis of α-Amino Acetylenes

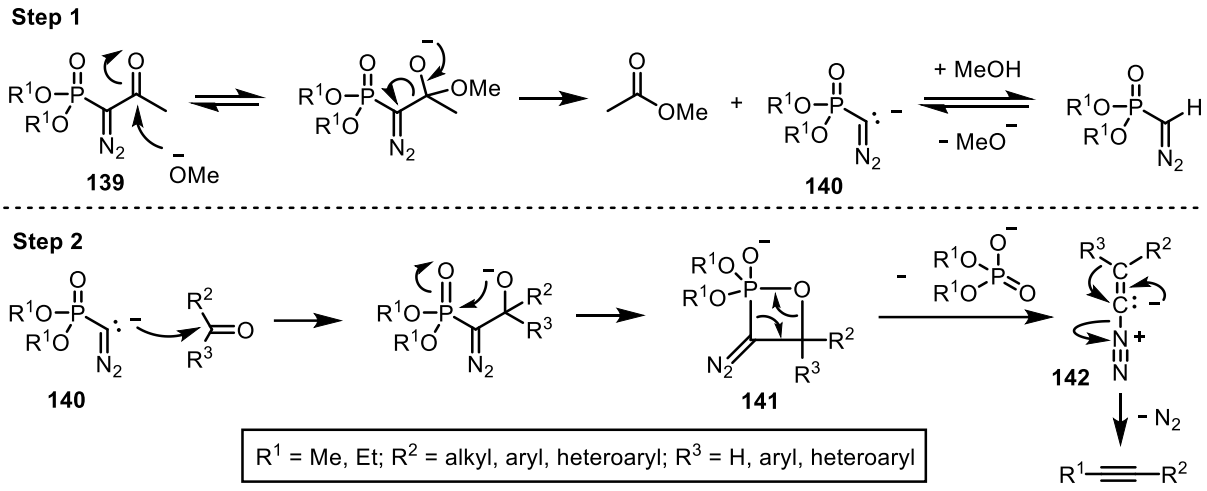
The α-amino acetylene building blocks used for the synthesis of click cyctetpep **95**-**97** in **Figure 52** were synthesised from the corresponding protected AAs through a mild, two-stage protocol. The first stage involved the reduction of the carboxylic acid to the corresponding aldehyde, followed by conversion to terminal alkyne in the second stage (**Scheme 41**).[272]



Scheme 41: Employed strategy for the conversion of protected AAs into the respective α-amino acetylenes through aldehyde intermediates.

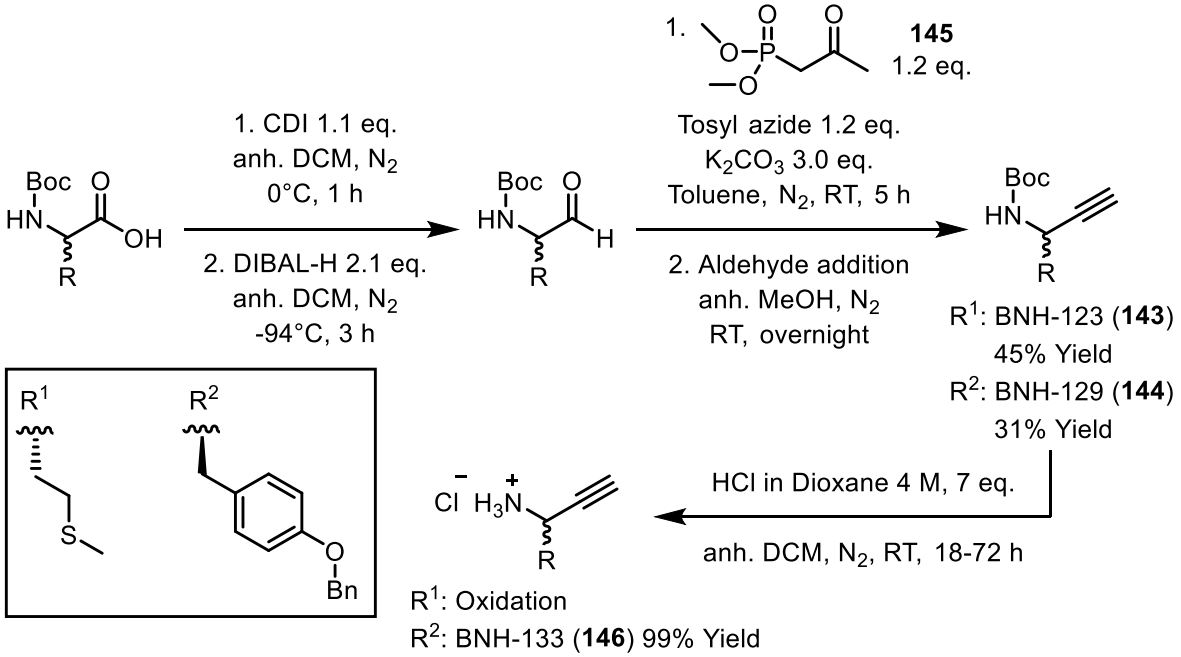
Preparation of the aldehyde can be achieved through activation of the carboxylic acid with carbonyl diimidazole (CDI) and subsequent reduction using DIBAL-H at -94°C.[273] Due to the susceptibility of the resulting α-amino aldehydes to epimerisation at the α-carbon under mildly acidic or basic conditions, e.g. during silica column chromatography, they should be used immediately after aqueous workup.[273] The second stage comprises a one-pot homologation of the aldehyde to the corresponding α-amino acetylene.[274] This transformation was achieved using the Ohira-Bestmann modification of the Seyferth-Gilbert homologation, which is widely used due to its broad functional group compatibility. In this reaction, the Ohira–Bestmann reagent **139** is deacylated by nucleophilic attack of MeO^-, which is formed *in situ* from K_2CO_3 in methanol solution (**Scheme 42**).[272]

In the second step, the diazo carbanion **140** engages in nucleophilic attack on the carbonyl group of the starting material, which may either be an aldehyde or a ketone bearing a broad range of substituents owing to the mildness of the reaction conditions. This addition yields the oxaphosphetane intermediate **141**, which subsequently decomposes to generate the diazoalkane **142**. The latter then undergoes α-elimination of N_2 and a concomitant 1,2-shift of one of the substituents to afford the terminal alkyne.[275,276]



Scheme 42: Reaction mechanism of the Ohira-Bestmann modification of the Seyferth-Gilbert homologation. Modified from KÜRTI and CZAKÓ.[277]

The α-amino acetylenes BNH-123 (**143**) and BNH-129 (**144**) were synthesised in accordance with the strategy outlined in **Scheme 41**. Boc and benzyl protecting groups were selected based on superior performance of the DIBAL-H reduction protocol when applied to Boc-protected AAs featuring side chains resistant to both acidic and basic conditions.[273] The synthetic route commenced with activation of the AA CDI at 0°C (**Scheme 43**).



Scheme 43: Reduction homologation sequence employed in the synthesis of **143**, **144** and **146**.

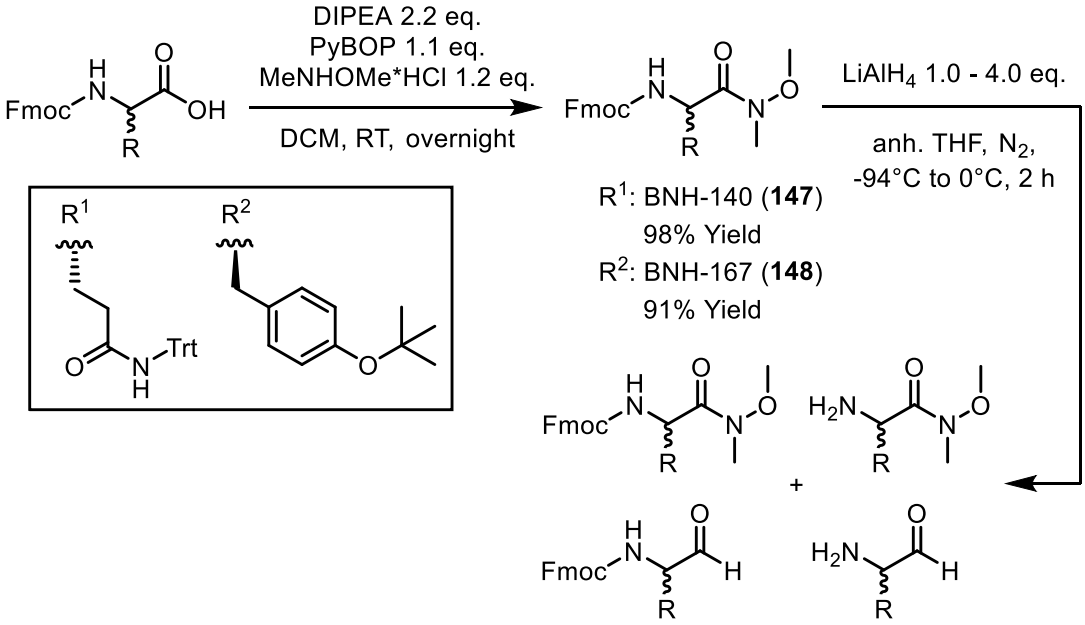
Optimal conversion to the aldehyde intermediate was achieved when freshly recrystallised CDI was used. The resulting acyl imidazole was subsequently reduced with DIBAL-H at -94°C. To minimise side reactions, DIBAL-H was added slowly over 2 hours through a syringe pump. Upon complete conversion, as monitored by TLC, the reaction was quenched with EtOAc and 25% aqueous tartaric acid. The use of tartaric acid, rather than its tartrate salt, facilitated rapid dissolution of the aluminium salts, thereby minimising epimerisation of the aldehyde. The aldehyde intermediates were obtained in sufficient purity following aqueous workup and were directly employed in the homologation reaction.

Concurrently, while the reduction was ongoing, the Ohira-Bestmann reagent was prepared from dimethyl-(2-oxopropyl)-phosphonate (**145**) and tosyl azide in the presence of K_2CO_3 in toluene.[274] No additional solvent was used, other than the toluene present in the commercial tosyl azide solution. After stirring at room temperature for 5 hours, the aldehyde was added in anhydrous MeOH to the reaction mixture containing the freshly prepared Ohira-Bestmann reagent. The homologation proceeded smoothly, furnishing the protected alkynes BNH-123 (**143**) and BNH-129 (**144**) in moderate yields over three steps.

Subsequent Boc deprotection of alkyne **144** under acidic conditions for 72 hours afforded the Tyr-derived alkyne BNH-133 (**146**) in excellent yield. However, analogous treatment of the Met-derived alkyne resulted in substantial oxidation of the side chain sulphur atom after only 18 hours, despite rigorous exclusion of oxygen. Notably, such oxidation has not been reported under comparable conditions.[278]

Encouraged by the successful synthesis of compound **146** , efforts were directed towards the preparation of the corresponding unprotected phenol derivative. Initial attempts using (*S*)-Fmoc-Tyr(tBu)-OH as the starting material, following the protocol in **Scheme 43**, provided the target aldehyde in low yield and poor purity. Subsequent homologation with the crude aldehyde failed to yield the desired alkyne. Employing excess DIBAL-H, as typically recommended for Fmoc-protected amino acids, also failed to improve the outcome.[273] The observed inefficiency is likely due to partial cleavage of the Fmoc group by the basic reducing conditions. A similar phenomenon has been reported during the reduction of Weinreb amides derived from Fmoc-protected AAs with $LiAlH_4$, with mitigation achieved by lowering the reaction

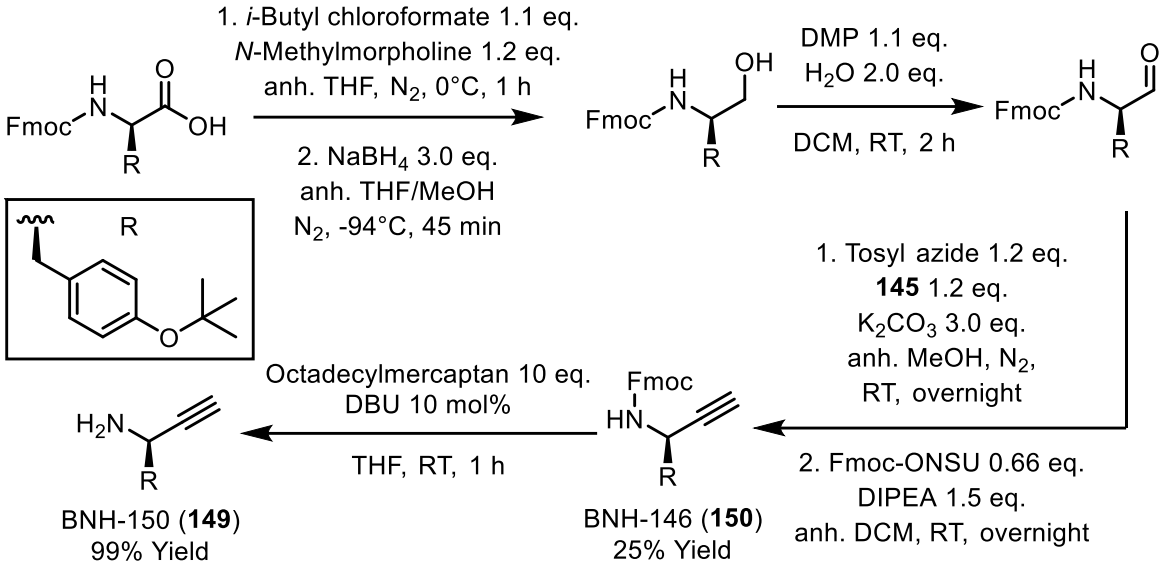
temperature.[279] Guided by this precedent, the Weinreb amides BNH-140 (**147**) and BNH-167 (**148**) were synthesised under standard coupling conditions (**Scheme 44**).



Scheme 44: Attempted synthesis of α-amino aldehydes by reduction of the Weinreb amides BNH-140 (**147**) and BNH-167 (**148**).

Stirring of the Weinreb amides with two equivalents of $LiAlH_4$ in anhydrous THF at −94°C for 1 hour failed to achieve complete consumption of the starting materials. Unexpectedly, the addition of a further two equivalents of $LiAlH_4$ had no appreciable effect on the conversion. Subsequent warming of the reaction mixture to 0°C and stirring for 30 minutes afforded a mixture of the protected and unprotected amides and aldehydes. Thes results contrast with previously reported aldehyde yields of 85% for (*S*)-BNH-140 and 50% for BNH-167 (**148**).[280,281]

Given the unsatisfactory conversion and the inherent fire hazard associated with $LiAlH_4$, an alternative reduction protocol using $NaBH_4$ was explored for the synthesis of BNH-150 (**149**) (**Scheme 45**). Activation of (*R*)-Fmoc-Tyr(*t*Bu)-OH was accomplished through formation of the mixed anhydride using *i*-butyl chloroformate. Subsequent reduction with $NaBH_4$ furnished the corresponding alcohol in >90% purity, as determined by LC/MS analysis. Even cleaner conversion may be achievable through activation with cyanuric chloride, as previously reported.[282]



Scheme 45: Synthesis of the α-amino acetylene **149** from (*R*)-Fmoc-Tyr(*t*Bu)-OH in a six-step reaction sequence involving $NaBH_4$ reduction, DMP oxidation, homologation and deprotection.

The crude alcohol was then oxidised using Dess-Martin periodinane (DMP) in the presence of two equivalents of H_2O, which served to accelerate the reaction, resulting in quantitative conversion to the corresponding aldehyde.[283] Notably, DMP has been shown to oxidise β-amino alcohols to α-amino aldehydes with the reduced epimerisation relative to Swern or TEMPO-based oxidations.[284]

Subsequent Ohira-Bestmann homologation yielded the Fmoc-protected alkyne BNH-146 (**150**) along with a small quantity of the corresponding free amine **149**. The latter was reprotected using Fmoc-ONSU to facilitate purification of alkyne **150** by silica gel chromatography. The four-step synthesis afforded the protected alkyne **150** with an overall yield of 25%.

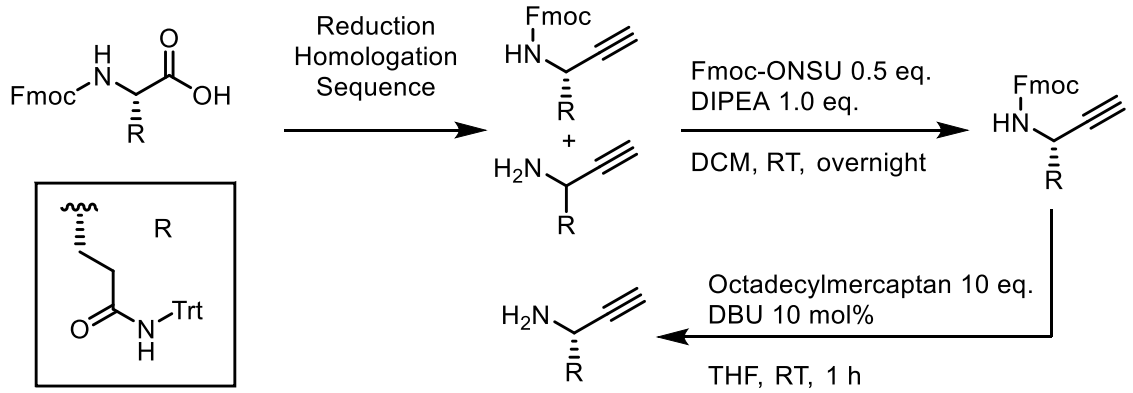
Removal of the Fmoc group under standard conditions using 20% piperidine in DMF proceeded with near-quantitative yield. However, the resulting piperidine-dibenzofulvene (DBF) adduct co-eluted with alkyne **149** on both normal- and reverse-phase columns. In response, seven deprotection cocktails were tested to identify a DBF scavenger that would yield an adduct a chromatographic profile amenable to separation from alkyne **149** (**Table 13**).

Cocktails	Results
20% Piperidine in DMF.	R_F(**149**) ≈ R_F(DBF-Adduct).
6% Piperazine in DMF.	R_F(**149**) ≈ R_F(DBF-Adduct).
50% Dicyclohexyl amine in DMF.	R_F(**149**) ≈ R_F(DBF-Adduct).
50% Ethanol amine in DMF.	Partial decomposition.
25% Benzylpiperazine in DMF.	R_F(**149**) ≈ R_F(DBF-Adduct).
15% Ethyl isonipecotate in DMF.	R_F(**149**) ≈ R_F(DBF-Adduct).
Octadecylmercaptan (10 eq.) DBU (10 Mol%) in THF.	Near quant. conversion. Excellent separation.

Table 13: Tested Fmoc-cleavage cocktails for the deprotection of alkyne **150** towards alkyne **149**.

The deprotection cocktails shaded in red were published by FIELDS.[285], while those shaded in blue were selected based on the ready availability of the corresponding bases. Neither group of cocktails produced sufficient differences in retention factor (R_F) values or product purity to enable effective separation. In contrast, the DBU-catalysed Fmoc deprotection in the presence of a large excess of octadecylmercaptan (shaded in green) achieved near quantitative conversion of alkyne **150** to alkyne **149**. This protocol enabled facile chromatographic removal of the octadecylmercaptan-DBF adduct.[286] An added advantage of this method is the elimination of potential piperidine-induced side reactions.[287,288]

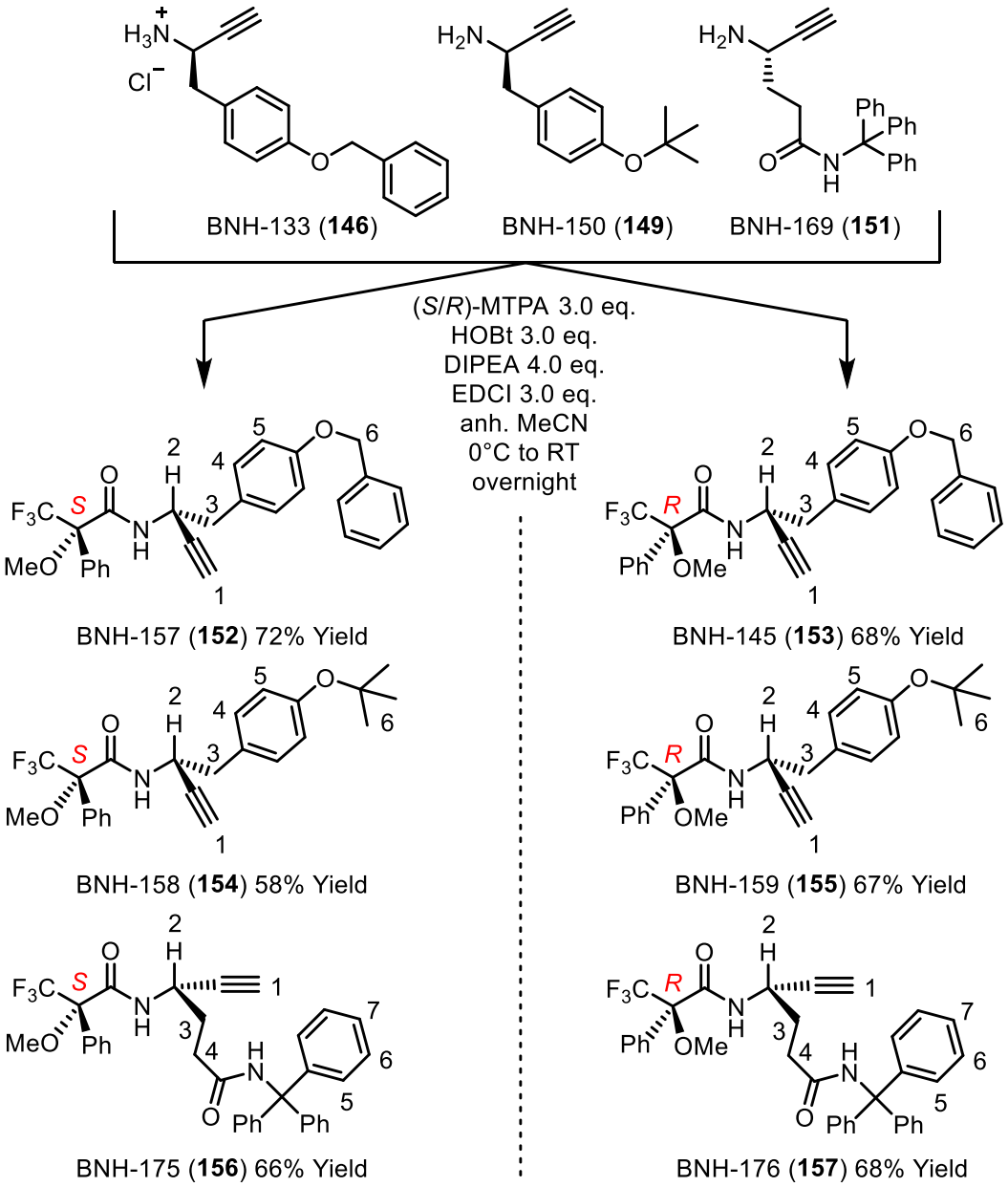
For the precursor (*S*)-Fmoc-Gln(Trt)-OH, reduction using CDI and DIBAL-H gave the corresponding aldehyde in sufficient yield and purity to allow successful homologation following the established sequence (**Scheme 43**), as depicted in **Scheme 46**.



BNH-169 (**151**) 10% Yield

Scheme 46: Synthesis of alkyne **151** through the described reduction homologation sequence with subsequent reprotection of the aldehyde and final DBU-catalysed Fmoc deprotection with octadecylmercaptan as DBF scavenger.

Reprotection of the partially deprotected alkyne mixture, followed by DBU-catalysed Fmoc removal, afforded the α-amino acetylene BNH-169 (**151**) in 10% yield over five steps. Throughout the syntheses of α-amino acetylenes **146**, **149**, and **151**, particular attention was paid to minimising epimerisation at the α-carbon. This was achieved by employing mild reaction conditions and limiting the storage time of the α-amino aldehyde intermediates. In order to assess the effectiveness of these precautions, the optical purity of the α-amino acetylenes was determined by coupling with (*R*)- and (*S*)-MTPA (Mosher's acid) through EDCI-mediated peptide coupling (**Scheme 47**).



Scheme 47: Synthesis of the Mosher amides **152**–**157** through peptide coupling of the amines **146**, **149** and **151** with (*R*)-and (*S*)-MTPA. The numbers have been assigned to the protons to facilitate comparison of their shifts in the respective diastereomers. The phenyl protons in amines **152** and **153** were omitted in this analysis, because they could not be assigned unambiguously.

Scheme 47 depicts the structures of the resulting Mosher amides BNH-157 (**152**), -145 (**153**), -158 (**154**), -159 (**155**), -175 (**156**) and -176 (**157**) alongside their assigned protons. The resulting (*S*)- and (*R*)-Mosher amides **152**–**157** were fully characterised and the absolute configurations were determined following the published protocol by HOYE, JEFFREY and SHAO.[289] Therein, the ^{1}H-NMR resonances were assigned and the shift differences $\Delta\delta^{SR}$ calculated from the proton shifts $\delta_{S/R}$ of the diastereomer pairs:

$$\Delta\delta^{SR} = \delta_S - \delta_R$$

Equation 6: Definition of the shift differences used for the determination of the absolute configuration at the α-carbon in the Mosher amides **152**–**157**. The (*S*)- and (*R*)-configurations refer to the α-carbon of the MTPA.

Amine	Proton	δ_S	δ_R	$\Delta\delta^{SR}$	
				ppm	Hz (600 MHz)
BNH-133 (**146**)	1	2.32	2.33	−0.01	−6
	2	5.03	5.08	−0.05	−30
	3	3.0	2.92	0.08	48
	4	6.95	6.85	0.1	60
	5	7.22	7.01	0.21	126
	6	5.07	5.04	0.03	18
	NH	7.01	6.78	0.23	138
BNH-143 (**149**)	1	2.30	2.32	−0.02	−8
	2	5.03	5.05	−0.02	−8
	3	3.25	2.92	0.33	132
	4	7.17	6.99	0.18	72
	5	6.94	6.85	0.09	36
	6	1.33	1.33	0	0
	NH	6.97	6.77	0.20	80
BNH-169 (**151**)	1	2.33	2.31	0.02	12
	2	4.77	4.78	−0.01	−6
	3	2.00	2.08	−0.08	−48
	4	2.36	2.45	−0.09	−54
	5	7.19	7.21	−0.02	−12
	6	7.29	7.29	0	0
	7	7.27	7.25	0.02	12
	NH-Trt	6.67	6.79	−-0.12	−72

Table 14: Summary of the ^{1}H shifts and their differences between diastereomer pairs resulting from coupling amines **146**, **149** and **151** with (*S*)- and (*R*)-MTPA. Orange: **152** *vs*. **153**, green: **154** *vs*. **155**, blue: **156** *vs*. **157**.

The absolute configuration of the α-carbons in the α-amino acetylenes was determined by analysis of the calculated $\Delta\delta^{SR}$ values for individual proton in each pair of (*S*/*R*)-Mosher amides. This stereochemical assignment is enabled by the characteristic solution-phase conformation adopted by Mosher amides (**Figure 57**).

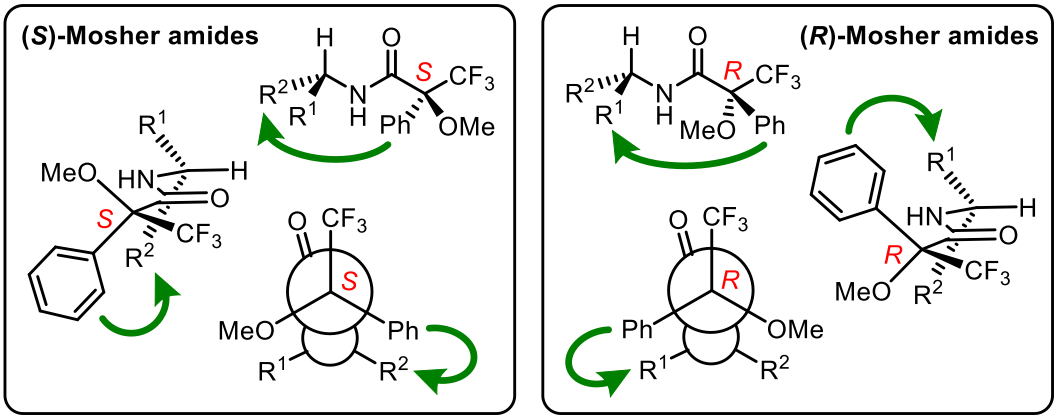


Figure 57: Common representations of Mosher amides with (*S*)- and (*R*)-configuration on Mosher's acid, left and right respectively. The anisotropic shielding caused by the phenyl moiety is represented as green arrows. Modified from HOYE, JEFFREY and SHAO.[289]

The α-proton, carbonyl, and CF_3 moieties assume a *syn*-coplanar conformation. While this conformation is not rigidly fixed, it dominates the spectroscopic profile of Mosher esters and amides.[289] In this geometry, each substituent on the α-carbon experiences a different chemical environment. Especially the phenyl substituent on MTPA provides anisotropic shielding to the moieties in close spatial proximity to it (green arrows).

Accordingly, in (*S*)-Mosher amides, protons associated with the R^2 substituent exhibit smaller chemical shifts compared to those on R^1, and the reverse is true for (*R*)-Mosher amides. The absolute configuration can therefore be reliably inferred from the $\Delta\delta^{SR}$ values: positive $\Delta\delta^{SR}$ values correspond to R^1 protons, while negative values correspond to R^2. Although protons situated within the *syn*-coplanar plane are generally excluded from interpretation, they are reported here for completeness.

The $\Delta\delta^{SR}$ values summarised in **Table 14** confirm the successful synthesis of the intended enantiomers: i.e. (*R*)-**146**, (*R*)-**149** and (*S*)-**151**. In Mosher amides **152** and **153** as well as **154** and **155**, the Tyr side chain and the alkyne moiety were assigned to R^1 and R^2, respectively. In contrast, in amides **156** and **157**, the Gln side chain and the alkyne moiety were assigned to R^2 and R^1, respectively.

It is important to note that the synthesised α-amino acetylenes were not obtained as enantiomerically pure compounds. Minor sets of ^{1}H-NMR signals with low integrals, corresponding to the opposite enantiomers, were detected in the spectra of the Mosher amides. The diastereomeric ratio (dr) and diastereomeric excess (de) of the major enantiomers were quantified by integration of the ^{19}F-NMR signals of the trifluoro-methyl group in each Mosher amide, using **Equation 7**. The resulting values are presented in **Table 15**.

$$de = \frac{Integral^{major} - Integral^{minor}}{Integral^{major} + Integral^{minor}} * 100\% \qquad dr = \frac{1 + de}{1 - de}$$

Equation 7: Definitions of the diastereomeric excess (de) and diastereomeric ratio (dr) calculated from the integrals in the ^{19}F-NMR spectra of the Mosher amides.

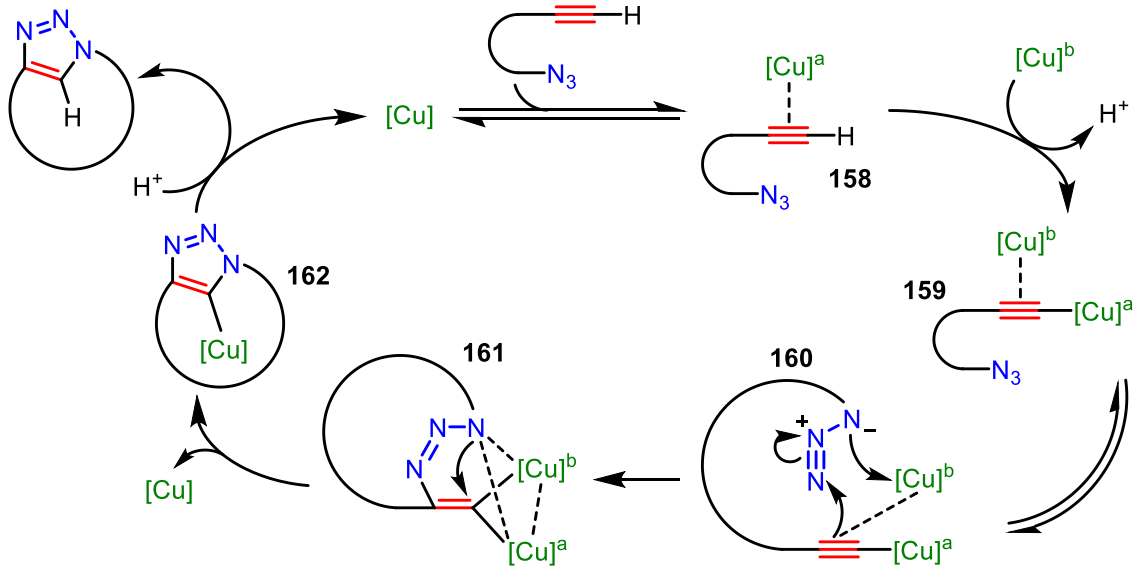
Amines	**146**		**149**		**151**	
Mosher amides	(*S*) **152**	(*R*) **153**	(*S*) **154**	(*R*) **155**	(*S*) **156**	(*R*) **157**
de in %	72	75	75	72	65	63
dr	6.1	7	7	6.1	4.6	4.41

Table 15: Summary of the diastereomeric excesses (orange) and diastereomeric rations (green) resulting from the ratios of the diastereomers' CF_3 group integrals. The absolute configuration of the MTPA in the Mosher amides is given in the *Mosher amides* column.

Notably, the desired (*R*)-enantiomers of the Tyr-derived alkynes **146** and **149** were obtained with higher selectivity than the (*S*)-enantiomer of the Gln-derived alkyne **151**. This discrepancy likely reflects the more efficient and cleaner reduction of the Tyr precursors to their corresponding α-aldehydes, relative to the Gln derivative. The reduced reactivity of the Gln substrate may be attributed to steric hindrance imposed by the bulky trityl protecting group on its side chain. In contrast, the reduction of Weinreb amides using DIBAL-H or $LiAlH_4$ has shown greater consistency across different amino acid substrates.[271,274] Overall, the α-amino acetylenes were obtained in sufficient purity to support subsequent assembly of click-cyclised tetrapeptides.

2.3.3.6 Synthesis of Click Cyctetpep

Several studies have reported the successful CuAAC-mediated cyclisation of tetrapeptides.[203,290–293] Yields of up to 70% have been achieved for the final click macrocyclisation step.[203] This efficiency has been attributed to the ability of Cu^{I} centres to preorganise the azide and alkyne termini of the linear tetrapeptide precursor, thereby facilitating intramolecular cyclisation. The proposed mechanism for this CuAAC-mediated macrocyclisation is illustrated in **Scheme 48**. The mechanism commences with coordination of Cu^{I} to the terminal alkyne of the linear precursor, affording complex **158**. This coordination lowers the pKa of the alkyne sufficiently to allow deprotonation, yielding a σ-bound Cu^{I}-acetylide intermediate that simultaneously engages a second Cu^{I} centre *via* π-complexation with the triple bond (**159**).



Scheme 48: Proposed mechanism of a CuAAC-mediated macrocyclization reaction. Modified from WORRELL, MALIK and FOKIN.[294]

In the second step, the second copper centre, denoted $[Cu]^{b}$, coordinates the azide of the linear precursor, thereby bringing the two termini into spatial proximity (**160**). Subsequent initiation of the intramolecular cycloaddition yields the dinuclear complex **161**, ich which two Cu^{I} atoms are σ-coordinated to the newly forming triazole. Ring closure to the 1,2,3-triazole is accompanied by the elimination of one Cu^{I} centre, affording the σ-complex **162**. Final protonation at the 4-position of the triazole delivers the desired macrocycle and liberates Cu^{I}.[294,295]

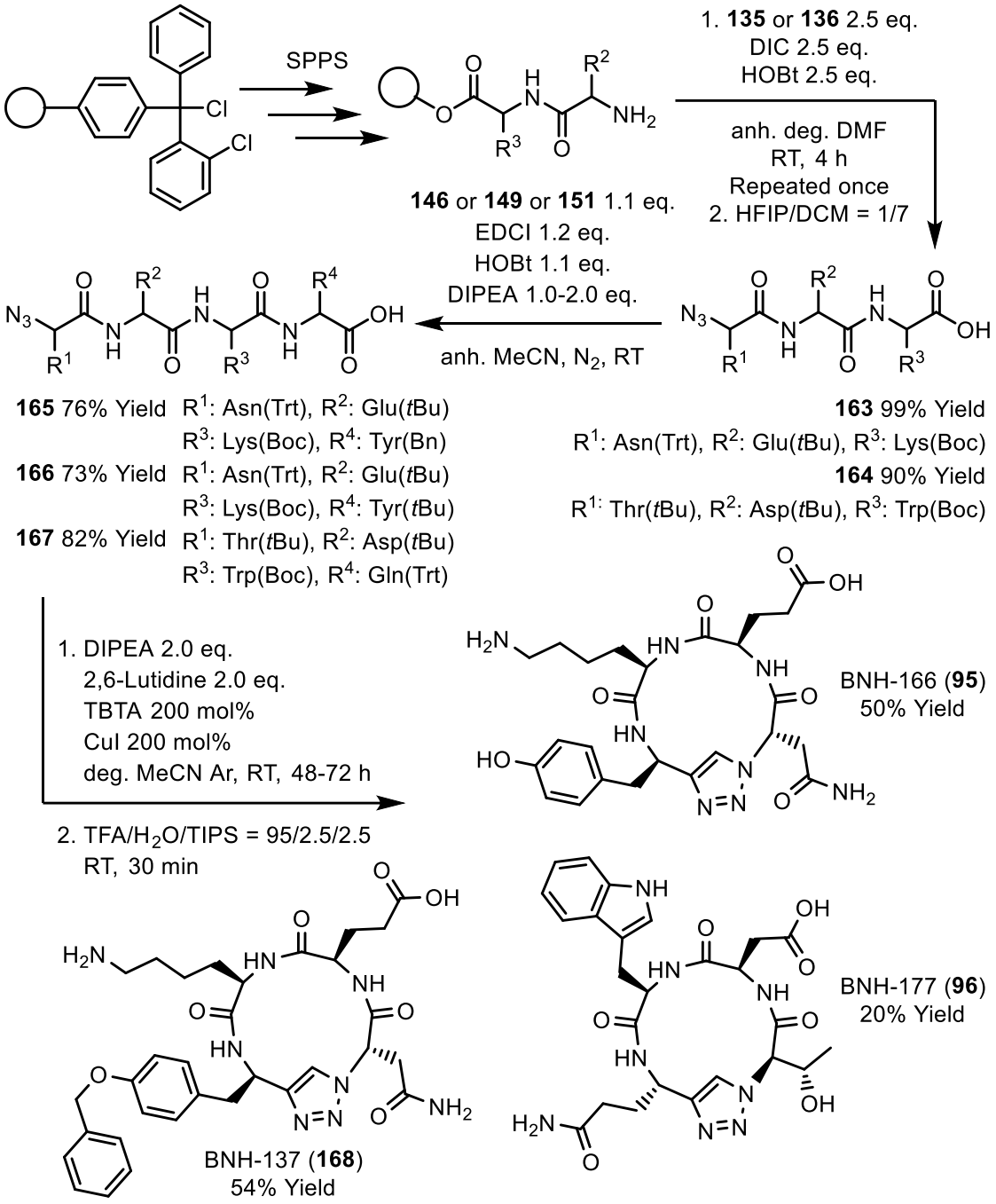
The equilibrium between intermediates **159** and **160** is presumably the key step responsible for the comparatively high cyclisation yields observed in CuAAC-mediated

macrocyclisations. Coordination of the azide to [Cu]^{b} results in the formation of a relatively large macrocyclic intermediate, a process that is particularly favourable when rigid tetrapeptide precursors are employed. The subsequent contraction of this macrocyclic intermediate *via* the 1,3-dipolar cycloaddition is expected to proceed with a significantly lower activation barrier than direct macrocyclisation in the absence of prior complexation. This mechanistic rationale bears resemblance to the imine-induced ring contraction strategy illustrated in **Scheme 33**.

The linear azide and alkyne tetrapeptides were synthesised through a combination of SPPS and solution-phase coupling. The second and third AA residues, counted from the N-terminus of the target linear tetrapeptide, were coupled on CTC resin following the protocol outlined in **Scheme 28** (**Scheme 49**). After *N*-terminal deprotection of the resin-bound dipeptide, the α-azido acid **135** or **136** was coupled. Notably, in contrast to standard SPPS protocols employing DIPEA, coupling of the α-azido acids was conducted in the absence of base. This modification is widely reported in the literature, albeit without explicit justification.[266,290,292] One plausible explanation is the electron-withdrawing nature of the azide group, which may render the α-proton of the α-azido acid prone to base-induced epimerisation upon formation of the active ester. Coupling was repeated once, until complete conversion was confirmed by a negative Kaiser test.

Upon successful coupling of the α-azido acid, the resulting tripeptide was cleaved from the resin using HFIP in DCM. The desired tripeptides **163** and **164** were obtained in yields of 99% and 90%, respectively, relative to the theoretical maximum loading of the CTC resin. Subsequent solution-phase coupling of tripeptide **163** with the α-amino acetylene **146** or **149**, and of tripeptide **164** with α-amino acetylene **151**, was carried out using EDCI in the presence of HOBt and DIPEA. One equivalent of DIPEA was employed to deprotonate the carboxylic acid of the tripeptide, and an additional equivalent to neutralise the HCl salt form of the α-amino acetylene **146**.

The resulting linear tetrapeptides **165**, **166**, and **167** were obtained in yields of 76%, 73% and 82%, respectively. All reaction by-products of the described EDCI coupling were conveniently removed through aqueous workup, affording the tetrapeptides in sufficient purity for the subsequent CuAAC-mediated macrocyclisation.



Scheme 49: Synthesis of the target click cyctetpep **95** and **96** as well as the benzyl-protected derivative **168** of target click cyctetpep **97**. A combination of SPPS, solution phase peptide coupling and CuAAC cyclisation was employed.

For the cyclisation step, 100 mg of linear tetrapeptide was dissolved in degassed MeCN to a final concentration of 0.2 mM. Scaling beyond 100 mg of peptide was avoided to prevent overloading of the HPLC column used in the final purification.

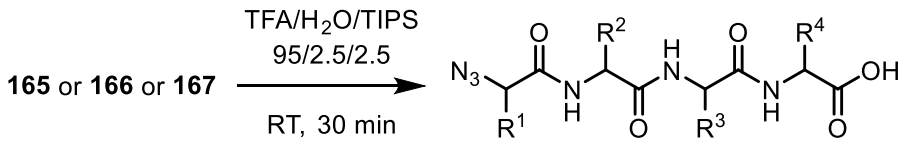
An excess of CuI catalyst and the ligand TBTA was employed to compensate for the high-dilution conditions required for efficient macrocyclisation of the linear tetrapeptides. TBTA has been reported to significantly enhance CuAAC-mediated

cyclisation yields in the synthesis of click cyctetpep.[293] This effect is presumably attributable to the ability of TBTA to stabilise the Cu^{I} centre by preventing oxidation, disproportionation, and aggregation *via* tetradentate coordination. Furthermore, the tertiary amine moiety of TBTA donates electron density to the Cu^{I} centre, thereby enhancing its catalytic activity in click reactions.[296]

Notably, cyclisation of linear tetrapeptide **167**, which features a sterically demanding Thr residue at the N-terminus, required 72 hours for complete conversion. In contrast, the corresponding linear precursors **165** and **166** were fully consumed after 48 hours under otherwise identical conditions. Following complete conversion, the reaction mixture was concentrated and globally deprotected using TFA/TIPS/H_2O. The resulting crude products were concentrated under a stream of compressed N_2 and purified by HPLC. Purification proved challenging due to the limited UV absorption and pronounced peak tailing and fronting of the click cyctetpep on both normal- and reverse-phase columns.

Nonetheless, the desired target click cyctetpep BNH-166 (**95**) and BNH-177 (**96**) were isolated in 50% and 20% yield, respectively. In addition, the benzyl-protected derivative **168** of the click cyctetpep **97** was obtained in 54% yield. These yields are consistent with literature reports and reflect the known sequence dependence of CuAAC macrocyclisation efficiency.[293]

As a final step, the linear tetrapeptides **165**-**167** were subjected to global deprotection to furnish the corresponding uncyclised analogues BNH-180 (**169**), BNH-179 (**170**), and BNH-178 (**171**). These compounds were prepared to investigate the impact of macrocyclisation on the biological activity of the cyclic tetrapeptides (**Scheme 50**).



BNH-180 (**169**) 38% Yield
R^1: Asn, R^2: Glu, R^3: Lys, R^4: Tyr(Bn)
BNH-179 (**170**) 50% Yield
R^1: Asn, R^2: Glu, R^3: Lys, R^4: Tyr
BNH-178 (**171**) 47% Yield
R^1: Thr, R^2: Asp, R^3: Trp, R^4: Gln

Scheme 50: Global deprotection of the linear tetrapeptides **165**–**167** towards the corresponding unprotected tetrapeptides **169**–**171**.

Analogous to the click cyctetpep, their unprotected linear counterparts were purified by HPLC. The moderate isolated yields are once again attributed to pronounced peak fronting and tailing observed on both normal- and reverse-phase columns. Additionally, despite near-quantitative conversion, as confirmed by LC/MS analysis, the modest deprotection yields suggest that the actual cyclisation efficiency may have exceeded the observed range of 20–54%. A substantial proportion of product appears to have been lost during chromatographic purification.

Finally, a notable disparity in solubility was observed between the (un)protected linear and cyclic tetrapeptides. Macrocyclisation significantly reduced the solubility of the peptides. The protected linear tetrapeptides were readily soluble in DCM and MeCN, whereas the corresponding protected cyclic analogues exhibited solubility only in DMSO. Following global deprotection, the linear peptides were soluble exclusively in DMSO, while the deprotected cyclic cyctetpep required warming to approximately 50°C in DMSO to achieve sufficient solubility.

2.3.3.7 Biochemical Data

At the time of writing, only three NE assay datapoints have been measured for the cyctetpep **95** and **96**. No NE inhibition was observed within the tested concentration range, up to 30 µM (**Table 16**). Their structures are shown in **Figure 58**.

ID	NE^{G12D} [µM]	NE^{WT} [µM]
95	>30	>30
96	>30	n.d.
2	0.0229±0.0160	0.0247 ± 0.0180
6	>300	>300
7	17.0 ± 9.31	33.1 ± 9.71
8	>300	>300

Table 16: Assay results of the click cyctetpep **95** and **96** (blue), as well as the references (white). All values are the result of multiple measurements.

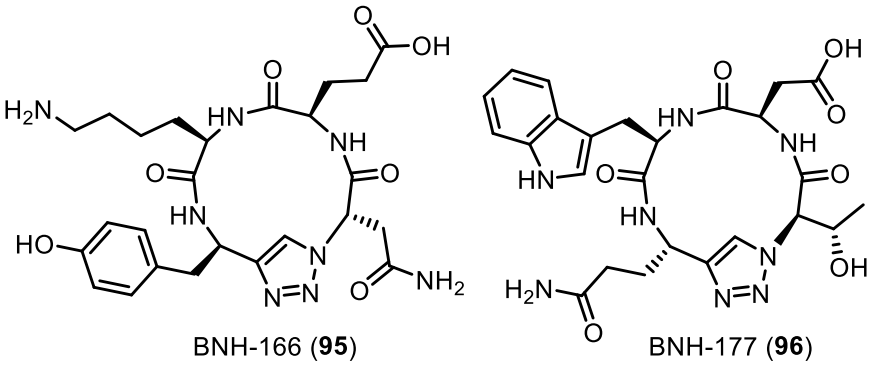


Figure 58: Structures of the click cyctetpep **95** and **96**.

Notably, the predicted IC_{50} values of the achiral structures of cyctetpep **95** and **96** were 110 µM and 34.7 µM, respectively. These predictions are in good agreement with the preliminary NE assay data; however, additional measurements at higher concentrations are required draw definitive conclusions. Unfortunately, the pronounced insolubility of the cyclic peptides precludes such follow-up experiments and might be a major contributing factor to the observed disappointing activity in the NE assay.

2.4 Summary and Outlook

In the second project of this thesis, the NE^{G12D} dataset from the project library was manually curated, and all molecules were processed using a standardized structure-washing protocol. This curated library was expanded to include 917 compounds through the integration of SAR data on KRAS PPIIs extracted from the ChEMBL34 database. Numerous molecules with pChEMBL values ≥ 6 were added, thereby broadening the SAR coverage of the library. Extended-connectivity fingerprints (ECFPs) were calculated for all structures and complemented by the 20 molecular descriptors with the highest mutual information relative to the assay data. Inclusion of these descriptors improved the predictive accuracy of the models.

The correlation between structural features and biological activity was evaluated using ten regression algorithms, with the random forest regressor yielding the best performance ($R^2 \approx 0.8$; RMSE ≈ 0.6). This construction of a robust QSAR model for KRAS-SOS PPIIs represents both a novel application and an innovative extension of the protocol reported by DUO et al.[223]

The trained random forest model was subsequently applied to predict $IC_{50}(NE^{G12D})$ values for over seven million compounds across ten screening libraries. Among these were five focused PPII libraries, including two newly generated in this project. Cyclic peptide libraries were constructed in KNIME from all 20 proteinogenic amino acids, encompassing all possible permutations within the cyclic peptide scaffolds. Eleven previously untested scaffolds were selected based on predicted IC_{50} values, molecular shape, and synthetic accessibility. In particular, the (click) cyclic tetrapeptide scaffolds emerged as highly promising.

Candidate cyclic peptides were prioritized for synthesis using rigid receptor docking. Although selected derivatives exhibited excellent docking scores, molecular dynamics simulations revealed limited binding affinity and specificity, more consistent with the predicted IC_{50} values in the low micromolar range. Nevertheless, these results represent encouraging leads for this novel KRAS PPII scaffold.

Synthesis was undertaken for the most promising *in silico* hits. Cyclisation experiments employing native tri- and tetrapeptides failed to yield the desired cyctripep and cyctetpep products in useful quantities, even under high-dilution conditions. Similar

outcomes were observed with the imine-induced ring contraction of tetrapeptide SAL esters. In contrast, CuAAC-mediated cyclisation emerged as a reliable and high-yielding method for pseudo-tetrapeptide macrocyclisation. Consequently, three promising click-cyclised tetrapeptide derivatives, i.e. cyctetpep **95**, **96** and **168**, were synthesised through an efficient, modular protocol.

The requisite α-azido acids and α-amino alkynes were prepared using convenient diazotransfer and Ohira-Bestmann homologation reactions. The optical purity of the synthesised unnatural AAs was confirmed through NMR analysis of their coupling products with Ala and Mosher's acid. These unnatural AAs were then assembled into pseudotetrapeptides using SPPS, followed by CuAAC-mediated cyclisation in a final high-yielding step.

Despite successful synthesis, the cyclic tetrapeptides **95** and **96** exhibited no inhibition of the SOS-catalysed nucleotide exchange on $KRAS^{G12D}$ within the tested concentration range, up to 30 µM. These findings are in good agreement with the predicted IC_{50} values of 110 µM and 34.7 µM for the corresponding achiral structures. However, due to the pronounced insolubility of compounds **95** and **96**, further activity measurements at higher concentrations were not feasible. To address this limitation, the Stoll group at the Ruhr University Bochum is currently investigating the interactions of **95** and **96** with KRAS in solution through chemical shift perturbation experiments.

Looking forward, three key areas for methodological improvement have been identified:

1. Library optimisation: Generate a click cyctetpep library with rigid, lipophilic side chains, i.e. substituted, aromatic rings. As discussed in Section 2.3, click cyctetpep are PPII-shaped; however, many members fall outside the Rule of Four (RO4). Enhancing compliance with the RO4 will increase overlap with the PPII chemical space.
2. Improved Docking Accuracy: Employ molecular dynamics simulations to generate multiple conformations of the target protein and perform ensemble docking against conserved binding sites, in particular P2 in PDB ID: 7RPZ. Incorporating receptor flexibility may overcome the limitations associated with rigid receptor docking.[297]

3. Feature Reduction in QSAR Model: Replace the ECFP with a reduced set of molecular descriptors. The implemented feature selection algorithm demonstrated potential to enhance accuracy by identifying and retaining only the most informative features. Substituting the 512-bit ECFP fingerprint with a smaller, high-quality descriptor set may further improve model performance.

In conclusion, this portion of the thesis describes the development of a novel pipeline for KRAS PPII screening. The synergy between efficient *in silico* prediction and high-yielding chemical synthesis enables rapid access to cyclic tetrapeptides, which represent a privileged, modular scaffold for PPI inhibition. Future improvements on the virtual screening strategy are expected to deliver a powerful and precise tool for exploring the KRAS PPII chemical space with unprecedented efficiency and accuracy.

3 Experimental

3.1 General Information

Synthesis: Chemicals were purchased from ABCR, Alfa Aesar, BLDpharm, Carbolution, Fisher Scientific, Merck, TCI, VWR. Ratios of solvents are given as volume ratios. Acetone, DCM, $CHCl_3$, EtOAc, EtOH, Hex, MeOH, MTBE and toluene were purchased from the local chemical storage. These solvents were reagent grade barrel goods. They were distilled with rotary evaporators immediately before use. Peroxides were removed from THF before use. This was achieved by passage through a column of activated alumina.[298] CDI was recrystallised from anh. THF.[273] Pyrrole was distilled at 5 mbar. HPLC grade MeCN was used. Reactions and workups were performed with dist. H_2O. Any chemicals and solvents not mentioned here were used as purchased without further purification. Reagent contents are given in weight %. Concentrations are given as molarities or % of weight per volume. Anhydrous solvents were stored over 3 or 4 Å molecular sieves. Molecular sieves were activated in a Heraeus Vacutherm vacuum oven with a Vacuubrand PC 2002 VARIO pump at 200°C and 5 mbar for 4 h. The molecular sieves were used immediately after activation. Reactions with anhydrous solvents were performed in flame-dried glassware, employing Schlenk techniques. Solvents were degassed with He before use in Pd-catalysed coupling reactions or SPPS. Microwave reactions were performed in a CEM Discover Lab Mate microwave oven. Reaction mixtures were concentrated with an IKA rotary evaporator (IKA HB10 basic heating bath with IKA RV10 basic rotor and Vacuubrand PC 2002 VARIO pump) at 40°C bath temperature. Aqueous solutions were lyophilised using a Zirbus GOT2000 lyophiliser.

Purification: HPLC grade MeCN was used. Dist. H_2O was purified with a SARTORIUS ARIUM® MINI water purifier before being used in preparative chromatography. TLC was performed with Macherey-Nagel pre-coated TLC sheets ALUGRAM® Xtra SIL G/UV. Spots were detected either by fluorescence quenching at 254 nm or by treatment with staining solutions followed by heat treatment with a heat gun. Employed stains and their use cases were: Ninhydrin for primary amines [Ninhydrin (1.50 g, 8.42 mmol) and glacial AcOH (3.00 mL, 52.4 mmol) in *i*-BuOH (100 mL)], bromocresol green for acids [bromocresol green (40.0 mg, 57.3 µmol) in EtOH (100 mL) and titrated to blue/green colour with aq. NaOH (0.1 M)] and $KMnO_4$ for everything else [$KMnO_4$

(1.50 g, 9.49 mmol), K_2CO_3 (10.0 g, 72.4 mmol) and aq. NaOH (10%, 1.25 mL) in H_2O (200 mL)]. Stain recipes were adapted from the literature.[299] The binary mixtures solvent+TFA and solvent+FA were always used with the ratio 1000/1. Normal phase chromatography was performed on a 1) Interchim PuriFlash XS 420 with cartriges PF-15SIHP-F0080/-40/-25/-12 or 2) using an Ismatec Reglo-Z gear pump connected to a Pharmacia Biotech fraction collector and cartridges packed with Macherey-Nagel silica gel 60M with particle size of 0.04-0.063 mm and pore size of 60 Å or 3) Interchim PuriFlash 4250 with HILIC column Macherey-Nagel 250/21 NUCLEODUR HILIC 5 µm. Reverse-phase chromatography was performed on an Interchim PuriFlash 4250 with 1) cartriges PF-15C18AQ-F0080/-40/-25/-12 or 2) HPLC column Macherey-Nagel VP250/21 NUCLEODUR 100-5 C18ec.

Analysis and Characterisation: HPLC grade MeCN was used. Dist. H_2O was purified with a SARTORIUS ARIUM® MINI water purifier before being used in analytical chromatography. The binary mixtures H_2O+TFA, H_2O+FA, MeCN+TFA and MeCN+FA were always used with the ratio 1000/1. **LC/MS** analyses were performed on a 1) Shimadzu Prominence-I, LC-2030C 3D Plus Liquid Chromatograph coupled with a LC/MS-2020 Liquid Chromatograph Mass Spectrometer operated with LabSolutions Version 5.109. or 2) Agilent Technologies 1100 Series Liquid chromatograph (Pump: BinPump G1312A, Autosampler: G1313A, UV-Detektor: VWD G1314A) coupled with a Bruker Daltonics micrOTOF mass spectrometer. LC/MS m/z ratios are given with their assigned molecular fragments and their relative intensities in percent. Liquid chromatograms with HILIC stationary phase were recorded on an Agilent Technologies 1220 Infinity II LC.

LC Methods:

Method 1: Chromatograph: Shimadzu, Column: Restek Raptor ARC-18, 1.8 µm, 90 Å, 50x2.1 mm, Temp.: 40°C, Flow: 0.4 mL/min, Inj. Vol.: 10 µL, Solvent A: H_2O+FA = 1000/1, Solvent B: MeCN+FA, Program: 95/5 to 5/95 over 13 min, 5/95 to 95/5 over 1 min, 95/5 over 2 min.

Method 2: Same as method 1 but with column: Macherey-Nagel EC NUCLEODUR C18 Gravity-SB, 1.8 µm, 110 Å, 50x2 mm.

Method 3: Chromatograph: Agilent Technologies 1100 Series LC, Column: MZ Analysetechnik PerfectSil Target ODS-3 HD, 5 µm, 100 Å, 100x4.6 mm,

Temp.: RT, Flow: 1.5 mL/min, Inj. Vol.: 5 µL, Solvent A: Aq. NH_4OAc (5.0 mM), Solvent B: MeCN, Program: 90/10 to 5/95 over 14 min, 5/95 over 10 min, 5/95 to 90/10 over 2 min, 90/10 over 4 min.

Method 4: Chromatograph: Shimadzu, Column: Restek Raptor ARC-18, 1.8 µm, 90 Å, 50x2.1 mm, Temp.: 40°C, Flow: 0.4 mL/min, Inj. Vol.: 10 µL, Solvent A: H_2O+FA, Solvent B: MeCN+FA, Program: 95/5 over 0.3 min, 95/5 to 5/95 over 6.7 min, 5/95 over 1 min, 5/95 to 95/5 over 1 min, 95/5 over 1 min.

Method 5: Chromatograph: Shimadzu, Column: Knauer Eurospher II, 2 µm, 100 Å, 100x2 mm, Temp.: 40°C, Flow: 0.3 mL/min, Inj. Vol.: 10 µL, Solvent A: H_2O+FA, Solvent B: MeCN+FA, Program: 95/5 to 5/95 over 11.8 min, 5/95 to 95/5 over 0.2 min, 95/5 over 4 min.

Method 6: Chromatograph: Agilent Technologies 1220 Infinity II LC, Column: Macherey-Nagel EC NUCLEODUR HILIC, 5 µm, 110 Å, 250x4 mm, Temp.: 40°C, Flow: 0.8 mL/min, Inj. Vol.: 10 µL, Solvent A: Aq. NH_4OAc (10 mM), Solvent B: MeCN, Program: 5/95 to 95/5 over 30 min, 95/5 over 5 min, 95/5 to 5/95 over 0.1 min, 95/5 over 5 min.

IR spectra were recorded on a 1) Bruker ALPHA Platinum-ATR with OPUS Version 7.5 or 2) Thermofisher Scientific NICOLET iS5 with iD7 ATR module and OMNIC 9.2.106. IR absorptions are given in cm^{-1}. Absorptions were assigned to functional group vibrations with Organikum 21. Auflage. and refer to valence vibrations if not stated otherwise.[300] **NMR** spectra were recorded at 25°C on a 1) Bruker Avance 400 MHz or 2) Bruker Avance III 600 MHz and postprocessed with TopSpin Version 4.1.3. Deuterated solvents were purchased from Deutero GmbH. Linear tetrapeptides and their respective cyclic counterparts were measured in extra pure DMSO-d6 (99.96%). NMR shifts are given in ppm relative to tetramethylsilane (0.00 ppm). The shifts were calibrated to traces of undeuterated solvent in the probed solutions. ^{1}H-NMR: $CHCl_3$ (s, 7.26 ppm), CD_3OD (quint, 3.31 ppm) and DMSO-d_6 (quint, 2.50 ppm). ^{13}C-NMR: (t, 77.06 ppm), (sept, 49.03 ppm) and (sept, 39.53 ppm).[301] Multiplicities of resonances were abbreviated: Singlet (s), doublet (d), triplet (t), quartet (q), quintet (quint), septet (sept), broad singlet (bs), doublet of doublets (dd), multiplet (m). Multiplet shifts are given as the range of the respective resonances. Coupling constants are given in Hertz (Hz). Assignment of the resonances was performed using 2D spectra: COSY, HSQC, HMBC, NOESY, TOCSY. The NMR data is given in the

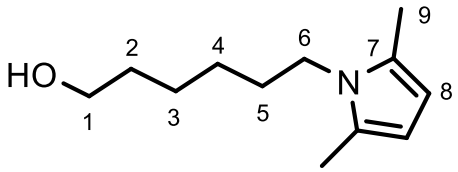
format “shift (multiplicity, coupling constant, number of assigned nuclei, assigned nuclei)”. **Optical rotation** values (α) were measured at 20°C in a A. Krüss Optronic P8000-T polarimeter. The values are given in $\frac{deg*cm^3}{g*dm}$ with concentration (c) given in $\frac{g}{100\ mL}$.

3.2 Project 1

3.2.1 Syntheses

1 (BNH-001) 6-(2,5-Dimethyl-1*H*-pyrrol-1-yl)hexan-1-ol

2,5-Hexanedione (571 mg, 5.00 mmol) and 6-amino-1-hexanol (604 mg, 5.00 mmol, 1.0 eq.) were dissolved in H_2O (2.5 mL) and the solution was refluxed for 20 min. The mixture was cooled to RT and extracted with EtOAc (4x3 mL) The pooled extracts were dried (Na_2SO_4) and conc. *in vacuo*. The crude product was purified chromatographically (RP, PF-15C18AQ-F0040, H_2O/MeCN=9/1 to 1/9 over 13 CV. The desired product was obtained as a light-brown oil (713 mg, 3.65 mmol, 73% yield). The product decomposes rapidly on TLC and in air. This procedure was adapted from the literature.[302] The alcohol **1** has been characterised previously with GC/MS only.[303]



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 5.76 (s, 2H, *H*8), 3.74-3.70 (m, 2H, *H*6), 3.64 (t, $^3J_{1\text{-}2}$=6.5 Hz, 2H, *H*1), 2.22 (s, 6H, *H*9), 1.67-1.53 (m, 4H, *H*2+5), 1.44-1.34 (m, 4H, *H*3+4).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 127.4 (*C*7), 105.1 (*C*8), 62.9 (*C*1), 43.7 (*C*6), 32.8 (*C*5), 31.1 (*C*2), 26.9, 25.7 (*C*3+4), 12.6 (*C*9).

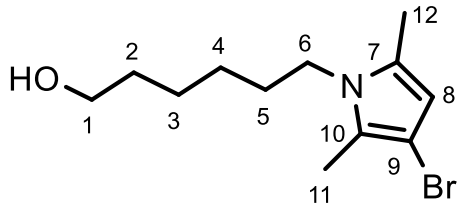
IR: 3344 (OH associated), 2929 (CH_2), 2857 (CH_2), 1407 (CH_3+CH_2 deform.).

LC/MS (Method 1, 220 nm, ESI+): t_R = 7.3 min, m/z: 196.1 [M+H$^+$] (100).

HRMS (ESI+): $C_{12}H_{22}NO^+$ [M+H$^+$] calc.: 196.1696 found: 196.1691.

19 (BNH-002) 6-(3-Bromo-2,5-dimethyl-1*H*-pyrrol-1-yl)hexan-1-ol

6-(2,5-Dimethyl-1*H*-pyrrol-1-yl)hexan-1-ol (**1**) (1.00 g, 5.12 mmol) was dissolved in THF at -94 C. Freshly recrystallised NBS (911 mg, 5.12 mmol, 1.0 eq.) was added in small portions over 30 min. The mixture was stirred for 3 h at -94°C. H_2O (15 mL) was added and the mixture extracted with DCM (4x15 mL). The organic extracts were pooled, washed with brine (3x30 mL), dried (Na_2SO_4) and conc. *in vacuo* with minimal temperature of the rotary evaporator bath. The obtained dark-brown oil was purified chromatographically (RP, PF-15C18AQ-F0040, H_2O/MeCN=9/1 to 1/9 over 13 CV: The desired product was obtained as a red-brown oil (407 mg, 1.48 mmol, 29% yield). The product decomposes rapidly on TLC and in contact with air. This procedure was adapted from the literature.[114]



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 5.82 (s, 1H, *H*8), 3.74-3.70 (m, 2H, *H*6), 3.64 (t, $^3J_{1\text{-}2}$=6.5 Hz, 2H, *H*1), 2.19-2.16 (m, 6H, *H*11+12), 1.64-1.54 (m, 4H, *H*2+5), 1.44-1.32 (m, 4H, *H*3+4).

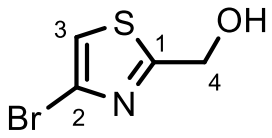
^{13}C-NMR (151 MHz, $CDCl_3$): δ = 127.4 (*C*12), 125.0 (*C7*), 107.9 (*C*8), 93.5 (*C*9), 62.9 (*C*1), 44.6 (*C*6), 32.7 (*C*2), 31.0 (*C*5), 26.8 (*C*4), 25.6 (*C*3), 12.4, 10.9 (*C*11+12).

LC/MS (Method 4, 254 nm, ESI+): t_R = 6.3 min, m/z: 276.0 [M(Br^{81})+H^+] (89), 274.0 [M(Br^{79})+H^+] (100).

20 (BNH-003) (4-Bromothiazol-2-yl)methanol

4-Bromo-2-formylthiazole (1.00 g, 5.21 mmol) was dissolved in MeOH (10 mL) and cooled with a water bath while $NaBH_4$ (374 mg, 9.89 mmol, 1.9 eq.) was added. The solution was stirred at RT for 2 h and was conc. *in vacuo* until a slurry was obtained. EtOAc (15 mL) and Hex (30 mL) were added and the resulting solution was passed through a short silica gel cake with EtOAc as eluent. The filtrate was conc. *in vacuo* to give the desired product as a colourless oil (1.00 g, 5.15 mmol, 99% yield). This

procedure was adapted from the literature.[115] The spectroscopic data agrees with the literature.



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.21 (s, 1H, *H*3), 4.94 (s, 2H, *H*4), 3.08 (bs, 1H, O*H*).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 127.9 (*C*1), 124.7 (*C*2), 117.2 (*C*3), 62.0 (*C*4).

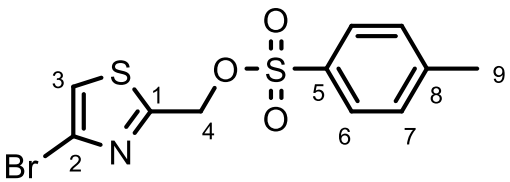
IR: 3262 (OH associated.), 3120 (=C-H), 1480 (ring vibr.), 1083 (C-O), 833 (=C-H deform.).

LC/MS (Method 1, 254 nm, ESI+): t_R = 2.8 min, m/z: 195.9 [M(Br^{81})+H^+] (100), 193.8 [M(Br^{79})+H^+] (86), 177.85 [M(Br^{81})-H_2O^-] (21), 175.85, [M(Br^{79})-H_2O^-] (16).

HRMS (ESI+): $C4H_5BrNOS^+$ [M+H^+] calc.: 193.9270, 195.9248 found: 193.9269, 195.9245.

21 (BNH-004) (4-Bromothiazol-2-yl)methyl 4-methylbenzenesulfonate

(4-Bromothiazol-2-yl)methanol (**20**) (806 mg, 4.15 mmol) was dissolved in anh. DCM (185 mL) at RT. *p*-Toluenesulfonyl chloride (950 mg, 4.98 mmol, 1.2 eq.) was added to the solution. Once the solution was clear again DIPEA (813 µL, 6.23 mmol, 1.5 eq.) was added dropwise. The mixture was stirred at RT for 2 h and conc. *in vacuo*. The mixture was washed with aq. $CuSO_4$ (10%, 2x20 mL), sat. $NaHCO_3$ (2x20 mL), dried (Na_2SO_4) and conc. *in vacuo* to give a pale-yellow oil. The crude product was purified chromatographically (NP, PF-15SIHP-F0040, EtOAc/Hex=2/8, R_F=0.42): The desired product was obtained as a white, amorphous solid (1.19 g, 3.44 mmol, 82% yield). This procedure was adapted from the literature.[304] The spectroscopic data agrees with the literature.



^{1}H-NMR (400 MHz, $CDCl_3$): δ = 7.83-7.78 (m, 2H, *H*6), 7.37-7.33 (m, 2H, *H*7), 7.25 (s, 1H, *H*3) 5.28 (s, 2H, *H*4), 2.45 (s, 3H, *H*9).

^{13}C-NMR (101 MHz, $CDCl_3$): δ = 163.7 (*C*1), 145.7 (*C*2), 132.5 (*C*5), 130.2 (*C*7), 128.2 (*C*6), 125.4 (*C*8), 119.1 (*C*3), 67.0 (*C*4), 21.8 (*C*9).

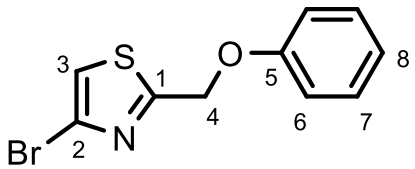
IR: 3115 (=C-H), 1362 (CH_3 deform.), 1170 (SO_2), 816 (=C-H deform.), 536 (C-Br).

LC/MS (Method 2, 254 nm, ESI+): t_R = 9.1 min, m/z: 371.9 [M(Br^{81})+Na^+] (5), 349.9 [M(Br^{81})+H^+] (100), 347.9 [M(Br^{79})+H^+] (92).

HRMS (ESI+): $C_{11}H_{10}BrNNaO_3S_2^+$ [M+Na^+] calc.: 369.9178, 371.9156 found: 369.9179, 371.9159.

22 (BNH-007) 4-Bromo-2-(phenoxymethyl)thiazole

Phenol (392 mg, 4.16 mmol, 1.2 eq.) was dissolved in anh. DMF (5 mL) and NaH (166 mg, 4.16 mmol, 1.2 eq.) was added at RT. On larger scales it is recommended to add the NaH at 0°C. The mixture was stirred at RT for 10 min. The tosylate **21** (1.21 g, 3.47 mmol) was added and the mixture was stirred at RT for 4 h. The reaction was quenched with H_2O (10 mL) and the resulting mixture was extracted with EtOAc (4x15 mL). The pooled organic extracts were washed with diluted brine (3x30 mL), dried (Na_2SO_4) and conc. *in vacuo*. The crude product was purified chromatographically (NP, PF-15SIHP-F0040, Hex, R_F(EtOAc/Hex=1/9)=0.38): The desired product was obtained as a white, crystalline solid (871 mg, 3.22 mmol, 92% yield). This procedure was adapted from the literature.[305]



^{1}H-NMR (400 MHz, $CDCl_3$): δ = 7.36-7.28 (m, 2H, *H*7), 7.25 (s, 1H, *H*3), 7.06-6.96 (m, 3H, *H*6+8), 5.35 (s, 2H, *H*4).

^{13}C-NMR (101 MHz, $CDCl_3$): δ = 168.7 (*C*1), 157.7 (*C*5), 129.9 (*C*7), 124.9 (*C*2), 122.2 (*C*8), 117.7 (*C*3), 115.1 (*C*6), 67.3 (*C*4).

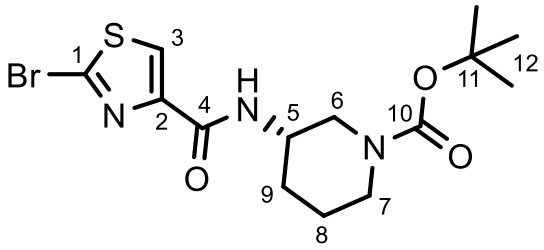
IR: 3060 (=C-H), 1241 (C-O-C), 749 (=C-H deform.), 507 (C-Br).

LC/MS (Method 3, 254 nm, ESI+): t_R = 11.0 min, m/z: 272.0 [M(Br81)+H$^+$] (100), 270.0 [M(Br79)+H$^+$] (95), 277.9 [M(Br^{81})-PhOH$^-$] (32), 275.9 [M(Br^{79})-PhOH$^-$] (32).

HRMS (ESI+): $C_{10}H_9BrNOS^+$ [M+H$^+$] calc.: 269.9583, 271.9562 found: 269.9584, 271.9566.

15 (BNH-012) *tert*-Butyl (*S*)-3-(2-bromothiazole-4-carboxamido)piperidine-1-carboxylate

2-Bromothiazole-4-carboxylic acid (400 mg, 1.92 mmol, 1.0 eq.), (*S*)-*tert*-butyl 3-aminopiperidine-1-carboxylate (385 mg, 1.92 mmol, 1.0 eq.) and HATU (731 mg, 1.92 mmol, 1.0 eq.) were dissolved in DMF (8 mL) and DIPEA (499 µL, 2.88 mmol, 1.5 eq.) was added. The mixture was stirred for 3.5 h at 80°C. H_2O (2 mL) was added and the mixture extracted with EtOAc (4x3 mL). The pooled organic extracts were washed with diluted brine (3x5 mL), dried (Na_2SO_4) and conc. *in vacuo* to give a yellow oil. The crude product was purified chromatographically (NP, PF-15SIHP-F0025, EtOAc/Hex=4/6, R_F=0.39): The desired product was obtained as a pale-yellow, amorphous solid (713 mg, 1.83 mmol, 95% yield). This procedure was adapted from the literature.[109]



^{1}H-NMR (400 MHz, $CDCl_3$): δ = 8.03 (s, 1H, *H*3), 7.32-7.21 (m, 1H, N*H*), 4.12-4.02 (m, 1H, *H*5), 3.67-3.60 (m, 1H, *H*6a), 3.44-3.35 (m, 3H, *H*6b+7), 1.94-1.85 (m, 1H, *H*9a), 1.77-1.68 (m, 2H, *H*8a+9b), 1.61-1.52 (m, 1H, *H*8b), 1.45 (s, 9H, *H*12).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 159.2 (*C*4), 155.0 (*C*10), 150.1 (*C*2), 135.9 (*C*1), 126.9 (*C*3), 80.1 (*C*11), 48.6 (*C*6), 45.5 (*C*5), 44.1 (*C*7), 29.9 (*C*9), 28.5 (*C*12), 22.6 (*C*8).

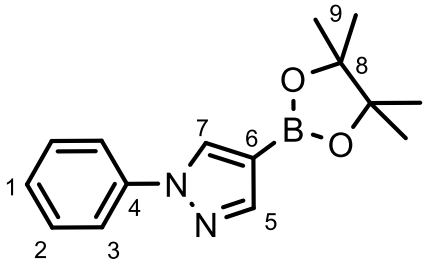
IR: 3318 (N-H), 1664 (C=O), 1417 (CH_3+CH_2 deform.), 1145 (C-O-C), 569 (C-Br).

LC/MS (Method 4, 254 nm, ESI+): t_R = 6.1 min, m/z: 391.9 [M(Br^{81})+H^+] (20), 389.9 [M(Br^{79})+H^+] (21), 335.8 [M(Br^{81})-$C_4H_7^-$] (100), 333.8 [M(Br^{79})-$C_4H_7^-$] (96), 291.9 [M(Br^{81})-$C_5H_7O_2^-$] (41), 289.9 [M(Br^{79})-$C_5H_7O_2^-$] (46).

HRMS (ESI+): $C_{14}H_{20}BrN_3NaO_3S^+$ [M+Na^+] calc.: 412.0301, 414.0281 found: 412.0301, 414.0281.

16 (BNH-014) 1-Phenyl-4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1*H*-pyrazole

4-Bromo-1-phenyl-1*H*-pyrazole (100 mg, 448 µmol), bis(pinacolato)diboron (125 mg, 493 µmol, 1.1 eq) and anh. KOAc (88.0 mg, 897 µmol, 2.0 eq.) were combined in anh. deg. dioxane (2 mL) and the mixture was Ar sparged, while placed in a sonicator for 5 min. Pd(dppf)Cl_2 (16.4 mg, 22.4 µmol, 5 mol%) was added and the mixture was stirred at 80°C for 18 h under Argon. The mixture was cooled to RT and filtered through a short celite cake with DCM as eluent. The organic phase was conc. *in vacuo* to give a dark-brown, oily residue. The crude product was purified chromatographically (NP, PF-15SIHP-F0012, EtOAc/Hex=1/9, R_F=0.28) to give the desired product as a white, amorphous solid (105 mg, 388 µmol, 87% yield). This procedure was adapted from the literature.[306] The spectroscopic data agrees with the literature.



^{1}H-NMR (400 MHz, $CDCl_3$): δ = 8.23 (s, 1H, *H*7), 7.97 (s, 1H, *H*5), 7.73-7.68 (m, 2H, *H*3), 7.48-7.40 (m, 2H, *H*2), 7.32-7.24 (m, 1H, *H*1), 1.35 (s, 12H, *H*9).

^{13}C-NMR (101 MHz, $CDCl_3$): δ = 146.9 (*C*5), 140.0 (*C*4), 133.8 (*C*7), 129.6 (*C*2), 126.8 (*C*1), 119.5 (*C*3), 109.0 (*C*6), 83.7 (*C*8), 24.9 (*C*9).

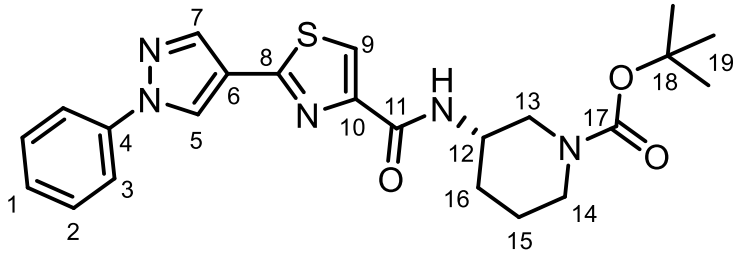
IR: 2977, 2930 (CH_3), 1131 (C-O), 755 (=C-H deform.).

LC/MS (Method 3, 254 nm, ESI+): t_R = 11.4 min, m/z: 271.2 [M+H^+] (100).

HRMS (ESI+): $C_{15}H_{20}BN_2O_2^+$ [M+H^+] calc.: 271.1615 found: 271.1615.

17 (BNH-016) *tert*-Butyl (*S*)-3-(2-(1-phenyl-1*H*-pyrazol-4-yl)thiazole-4-carboxamido)piperidine-1-carboxylate

The boronic acid ester **16** (100 mg, 370 µmol, 1.0 eq.), the bromide **15** (144 mg, 370 µmol, 1.0 eq.) and K_3PO_4 (157 mg, 740 µmol, 2.0 eq.) were combined in H_2O/dioxane (1/9, 2 mL). The mixture was Ar sparged, while placed in a sonicator for 5 min. Pd(dppf)Cl_2 (27.1 mg, 37.0 µmol, 10 mol%) was added and the mixture was stirred at 80°C for 48 h under Ar. The mixture was cooled to RT and filtered through a short celite cake with DCM as eluent. The organic phase was conc. *in vacuo* to give a dark-brown, oily residue. The crude product was purified chromatographically (RP, PF-15C18AQ-F0012, H_2O/MeCN=9/1 to 1/9 over 13 CV) to give the desired product as a pale-yellow, amorphous solid (101 mg, 222 µmol, 59% yield).



^{1}H-NMR (400 MHz, $CDCl_3$): δ = 8.46 (s, 1H, *H*9), 8.24 (s, 1H, N*H*), 8.11 (s, 1H, *H*5), 8.04 (s, 1H, *H*7), 7.78-7.72 (m, 2H, *H*3), 7.54-7.46 (m, 2H, *H*2), 7.40-7.33 (m, 1H, *H*1), 4.19-4.09 (m, 1H, *H*12), 3.77-3.67 (m, 1H, *H*13a), 3.53-3.34 (m, 3H, *H*13b+14), 2.03-1.89 (m, 2H, N*H*+*H*16a), 1.84-1.70 (m, 2H, *H*16b+15a), 1.67-1.55 (m, 1H, *H*15b), 1.45 (s, 9H, H19).

^{13}C-NMR (101 MHz, $CDCl_3$): δ = 160.5 (*C*11), 160.0 (*C*10), 155.1 (*C*17), 151.1 (*C*4), 150.6 (*C*6), 139.5 (*C*5), 129.7 (*C*2), 127.5 (*C*1), 125.6 (*C*9), 121.8 (*C*7), 119.5 (*C*3), 118.8 (*C*8), 79.9 (*C*18), 48.7 (*C*13), 45.4 (*C*12), 44.0 (*C*14), 30.1 (*C*16), 28.5 (*C*19), 22.7 (*C*15).

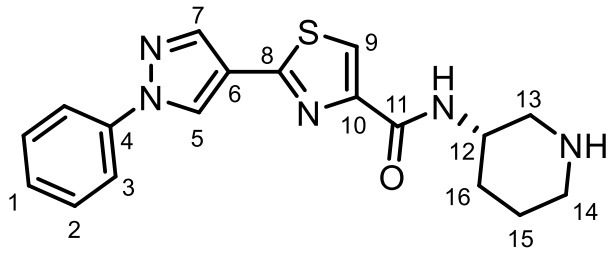
IR: 3399 (N-H), 1662 (C=O), 1146 (C-O-C), 756, 727 (=C-H deform.).

LC/MS (Method 2, 254 nm, ESI+): t_R = 10.2 min, m/z: 476.2 [M+Na$^+$] (6), 454.2 [M+H$^+$] (17), 398.1 [M-$C_4H_7^-$] (100), 354.1 [M-$C_5H_7O_2^-$] (16).

HRMS (ESI+): $C_{23}H_{27}N_5NaO_3S^+$ [M+Na$^+$] calc.: 476.1727 found: 476.1726.

14 (JES-248) (*S*)-2-(1-Phenyl-1*H*-pyrazol-4-yl)-*N*-(piperidin-3-yl)thiazole-4-carboxamide

The Boc-protected amine **17** (101 mg, 222 µmol) was dissolved in TFA/DCM (1/3, 4 mL) for 5 h at RT. The solution was conc. *in vacuo* and purified chromatographically (NP, PF-15SIHP-F0012, MeOH/DCM=1/9, R_F=0.41): The desired product was obtained as a pale-yellow, amorphous solid (68.3 mg, 193 µmol, 87% yield).



^{1}H-NMR (600 MHz, $CDCl_3$): 8.49 (s, 1H, *H*9), 8.09 (s, 1H, *H*5), 8.01 (s, 1H, *H*7), 7.94-7.85 (m, 1H, CON*H*), 7.79-7.70 (m, 2H, *H*3), 7.51-7.41 (m, 2H, *H*2), 7.36-7.29 (m, 1H, *H*1), 6.71 (s, 1H, N*H*), 4.38-4.26 (m, 1H, *H*12), 3.39-3.24 (m, 1H, *H*13a), 3.12-2.86 (m, 3H, *H*13b+14), 2.10-1.84 (m, 2H, *H*16a+15a), 1.84-1.67 (m, 2H, *H*16b+15b).

^{13}C-NMR (151 MHz, $CDCl_3$): 177.7 (*C*11), 160.8 (*C*8), 160.1 (*C*10), 150.3 (*C*4), 139.5 (*C*5), 129.7 (*C*3), 127.4 (*C*1), 125.7 (*C*9), 122.2 (*C*7), 119.5 (*C*2), 118.8 (*C*6), 48.5 (*C*13), 44.5 (*C*12), 44.4 (*C*14), 29.4 (*C*16), 21.9 (*C*15).

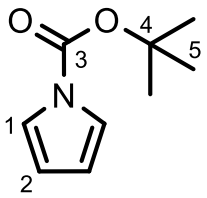
IR: 1652 (C=O), 1537 (N-H deform.), 1228 (C-N), 756, 725 (=C-H deform.).

LC/MS (Method 1, 254 nm, ESI+): t_R = 5.8 min, m/z: 354.0 [M+H$^+$] (100).

HRMS (ESI+): $C_{18}H_{20}N_5OS^+$ [M+H$^+$] calc.: 354.1383 found: 354.1386.

28 (BNH-020) *tert*-Butyl 1*H*-pyrrole-1-carboxylate

Pyrrole (freshly distilled, 2.00 g, 29.8 mmol), Boc_2O (7.79 g, 35.7 mmol, 1.2 eq.) and DMAP (500 mg, 4.09 mmol, 0.14 eq.) were dissolved in anh. MeCN. The mixture was stirred under Ar for 2 h at RT. The mixture was conc. *in vacuo* and purified chromatographically (NP, PF-15SIHP-F0012, Hex, R_F=0.17). Subsequent Kugelrohr distillation gave the desired product as a colourless oil (4.92 g, 29.4 mmol, 98% yield). This procedure was adapted from the literature.[122] The spectroscopic data agrees with the literature.



^{1}H-NMR (400 MHz, $CDCl_3$): δ = 7.27-7.23 (m, 2H, *H*1), 6.24-6.20 (m, 2H, *H*2), 1.61 (s, 9H, *H*5).

^{13}C-NMR (101 MHz, $CDCl_3$): δ = 149.0 (*C*3), 120.1 (*C*1), 111.9 (*C*2), 83.6 (*C*4), 28.1 (*C*5).

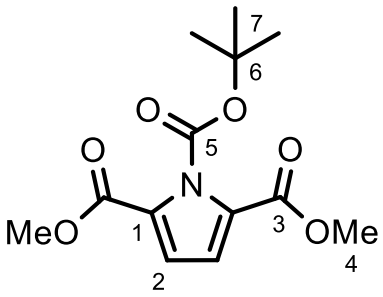
IR: 1740 (C=O), 1314 (C-N), 1149 (C-O-C), 950 (=C-H deform.).

LC/MS (Method 2, 254 nm, ESI+): t_R = 9.3 min, m/z: 168.0 [$M+H^+$] (97), 112 [$M-C_4H_7^-$] (61).

HRMS (APCI+): $C_9H_{14}NO_2^+$ [$M+H^+$] calc.: 168.1019 found: 168.1024.

29 (BNH-022) 1-(*tert*-Butyl) 2,5-dimethyl 1*H*-pyrrole-1,2,5-tricarboxylate

2,2,6,6-tetramethylpiperidine (6.47 mL, 44.9 mmol, 2.5 eq.) was dissolved in anh. THF (60 mL) at -94°C under N_2.*n*-BuLi (2.01 M, 22.3 mL, 44.9 mmol, 2.5 eq.) was added and the solution was stirred for 10 min. The Boc-pyrrole **28** (3.00 g, 17.9 mmol) was added and the mixture was stirred for 3 h at -94°C. The obtained mixture was added to methyl chloroformate (5.14 mL, 53.8 mmol, 3.0 eq.) and the dark-brown mixture was stirred at -94°C for 30 min under N_2. Sat. aq. NH_4Cl (10 mL) was added and the mixture was warmed to RT. The mixture was extracted with Et_2O (20 mL). The organic phase was washed with HCl (1 M, 2x30 mL) and brine (2x30 mL). The organic phase was dried (Na_2SO_4) and conc. *in vacuo* to give a red-brown solid. The crude product was purified chromatographically (NP, PF-15SIHP-F0040, EtOAc/Hex=1/9, R_F=0.25): The desired product was obtained as fine, off-white crystals (2.10 mg, 7.41 mmol, 41% yield). This procedure was adapted from the literature.[123] The spectroscopic data agrees with the literature.



^{1}H-NMR (400 MHz, $CDCl_3$): δ = 6.83 (s, 2H, *H*2), 3.86 (s, 6H, *H*4), 1.66 (s, 9H, *H*7).

^{13}C-NMR (101 MHz, $CDCl_3$): δ = 160.0 (*C*3), 149.0 (*C*5), 126.8 (*C*1), 115.9 (*C*2), 86.4 (*C*6), 52.1 (*C*4), 27.5 (*C*7).

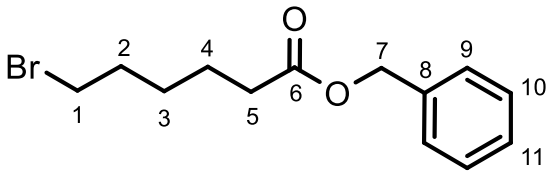
IR: 1775, 1722, 1705 (C=O), 1257 (C-N), 1155 (C-O-C), 745 (=C-H deform.).

LC/MS (Method 2, 254 nm, ESI+): t_R = 9.2 min, m/z: 306.0 [$M+H^+$] (17), 225.0 [$M-C_2H_2O_2^-$] (95), 184.0 [$M-C_5H_7O_2^-$] (100).

HRMS (ESI+): $C_{13}H_{17}NNaO_6^+$ [$M+Na^+$] calc.: 306.0948 found: 306.0951.

33 (BNH-023) Benzyl 6-bromohexanoate

Benzyl alcohol (485 µL, 4.68 mmol) and TEA (653 µL, 4.68 mmol, 1.0 eq.) were dissolved in anh. DCM (10 mL) at 0°C. 6-bromohexanoyl chloride (1.0 g, 4.68 mmol, 1.0 eq.) was added. The resulting white suspension was stirred at 0°C for 3 h. The mixture was washed with HCl (1 M, 2x5 mL) and sat. aq. $NaHCO_3$ (2x5 mL). The organic phase was dried (Na_2SO_4) and conc. *in vacuo* to give a pale-yellow oil. Subsequent Kugelrohr distillation gave the desired product as a colourless oil (1.25 g, 4.38 mmol, 93% yield). This procedure was adapted from the literature.[126] The spectroscopic data agrees with the literature.



^{1}H-NMR (400 MHz, $CDCl_3$): δ = 7.41-7.29 (m, 5H, *H*9-11), 5.12 (s, 2H, *H*7), 3.38 (t, $^{3}J_{1\text{-}2}$=6.8 Hz, 2H, *H*1), 2.38 (t, $^{3}J_{1\text{-}2}$=7.4 Hz, 2H, *H*5), 1.91-1.82 (m, 2H, *H*2), 1.73-1.63 (m, 2H, *H*4), 1.52-1.42 (m, 2H, *H*3).

^{13}C-NMR (101 MHz, $CDCl_3$): δ = 173.3 (*C*6), 136.2 (*C*8), 128.7 (*C*10), 128.3 (*C*9+11), 66.3 (*C*7), 34.2 (*C*5), 33.5 (*C*1), 32.5 (*C*2), 27.7 (*C*3), 24.2 (*C*4).

IR: 1731 (C=O), 1161 (C-O-C), 734 (=C-H deform.), 697 (C-Br).

LC/MS (Method 2, 254 nm, ESI+): t_R = 10.6 min, m/z: 132.0 [M-C_7H_7Br+NH_4^+] (100).

HRMS (ESI+): $C_{13}H_{17}BrNaO_2^+$ [M+Na^+] calc.: 307.0304 found: 307.0307.

30 (BNH-019 and BNH-025) Dimethyl 1*H*-pyrrole-2,5-dicarboxylate

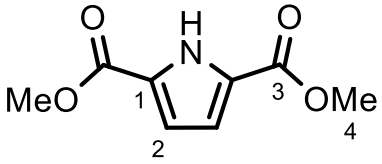
This diester was synthesised directly from pyrrole (synthesis a) as well as by deprotection of **29** (synthesis b).

Synthesis a:

Pyrrole (freshly distilled, 50.0 µL, 722 µmol), CCl_4 (544 µL, 5.62 mmol, 7.8 eq.) and $Fe(acac)_3$ (2.55 mg, 7.21 µmol, 1 mol%) were dissolved in MeOH (1.90 mL). The mixture was heated in a microwave tube (10.0 mL) for 6h with 150 W at 110°C under Ar. The resulting black suspension was filtered through a celite cake eluting with Et_2O/Hex=1/1. The filtrate was conc. *in vacuo* and purified chromatographically (NP, PF-15SIHP-F0012, $CHCl_3$, R_F=0.45): The desired product was obtained as a pale-yellow amorphous solid (23.9 mg, 130 µmol, 18% yield). This procedure was adapted from the literature.[307]

Synthesis b:

The Boc-protected diester **29** (500 mg, 1.77 mmol) was stirred in TFA/DCM (1/4, 15 mL) for 2 h at rt. The solution was conc. *in vacuo* and purified chromatographically (NP, PF-15SIHP-F0025, EtOAc/Hex=15/85, R_F($CHCl_3$)=0.45): The desired product was obtained as a white, amorphous solid (316 mg, 1.73 mmol, 97% yield). The spectroscopic data agrees with the literature.[307]



^{1}H-NMR (400 MHz, $CDCl_3$): δ = 9.77 (bs, 1H, N*H*), 6.87 (d, $^4J_{2\text{-}NH}$=2.6 Hz, 2H, *H*2), 3.89 (s, 6H, *H*4).

^{13}C-NMR (101 MHz, $CDCl_3$): δ = 160.9 (*C*3), 126.2 (*C*1), 115.7 (*C*2), 52.1 (*C*4).

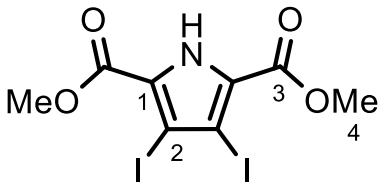
IR: 3291 (N-H), 1724, 1710 (C=O), 1257, 1245 (C-O-C), 758 (=C-H deform.).

LC/MS (Method 1, 254 nm, ESI+): t_R = 5.1 min, m/z: 184.0 [M+H$^+$] (100), 152.0 [M-MeO$^-$] (7).

HRMS (ESI+): $C_8H_9NNaO_4^+$ [M+Na$^+$] calc.: 206.0424 found: 206.0424.

31 (BNH-026) Dimethyl 3,4-diiodo-1*H*-pyrrole-2,5-dicarboxylate

The diester **30** (285 mg, 1.56 mmol) and NIS (770 mg, 3.42 mmol, 2.2 eq.) were dissolved in anh. DMF (10 mL). The mixture was stirred at 80°C for 4 h and then conc. *in vacuo*. The crude product was purified chromatographically (NP, PF-15SIHP-F0025, EtOAc/Hex=4/6, R_F=0.34): Off-white needles were obtained. Subsequent recrystallisation from DCM/Hex gave the desired product as white needles (598 mg, 1.37 mmol, 88% yield). This procedure was adapted from the literature.[125] The spectroscopic data agrees with the literature.



^{1}H-NMR (600 MHz, DMSO-d$_6$): δ = 13.12 (bs, 1H, N*H*), 3.82 (s, 6H, *H*4).

^{13}C-NMR (151 MHz, DMSO-d$_6$): δ = 158.7 (*C*3), 127.7 (*C*1), 86.7 (*C*2), 51.8 (*C*4).

IR: 3258 (N-H), 1715, 1692 (C=O), 1261 (C-O-C), 614 (C-I).

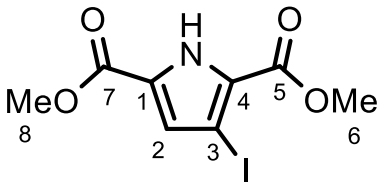
LC/MS (Method 1, 254 nm, ESI+): t_R = 7.6 min, m/z: 435.7 [M+H$^+$] (86), 403.7 [M-MeO$^-$] (11).

HRMS (ESI+): $C_8H_7I_2NNaO_4^+$ [M+Na$^+$] calc.: 457.8357 found: 457.8358.

32 (BNH-027) Dimethyl 3-iodo-1*H*-pyrrole-2,5-dicarboxylate

Powdered Zn (69.2 mg, 1.06 mmol, 1.15 eq.) and I_2 (23.3 mg, 91.7 mg, 0.1 eq.) were combined under Ar and stirred for 2 min at rt. DMA (1.5 mL) was added and the mixture was stirred for 2 min. The diiodide **31** (400 mg, 920 µmol) was added and the mixture was stirred at 120°C for 2.5 h under Ar. The mixture was conc. *in vacuo* and purified chromatographically (NP, PF-15SIHP-F0025, EtOAc/Hex=2/8, R_F=0.1): The desired

product was obtained as a white, amorphous solid (178 mg, 576 µmol, 62% yield). This procedure was adapted from the literature.[125] The spectroscopic data agrees with the literature.



^{1}H-NMR (600 MHz, DMSO-d_6): δ = 9.83 (bs, 1H, N*H*), 7.06 (d, $^{4}J_{2\text{-}NH}$=2.7 Hz, 1H, *H*2), 3.92 (s, 3H, *H*8), 3.89 (s, 3H, *H*6).

^{13}C-NMR (151 MHz, DMSO-d_6): δ = 160.0 (*C*5), 159.8 (*C*7), 126.9 (*C*4), 126.4 (*C*1), 124.6 (*C*2), 68.7 (*C*3), 52.4 (*C*6), 52.2 (*C*8).

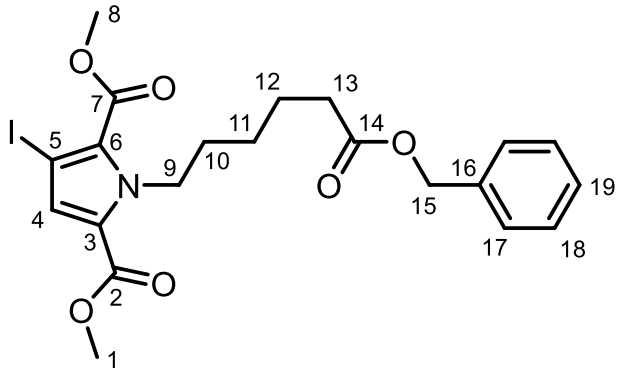
IR: 3272 (N-H), 1720, 1699 (C=O), 1262 (C-O-C), 618 (C-I).

LC/MS (Method 1, 254 nm, ESI+): t_R = 6.7 min, m/z: 309.8 [M+H$^+$] (100), 277.9 [M-MeO$^-$] (10).

HRMS (ESI+): $C_8H_8INNaO_4^+$ [M+Na$^+$] calc.: 331.9390 found: 331.9390.

34 (BNH-028) Dimethyl 1-(6-(benzyloxy)-6-oxohexyl)-3-iodo-1*H*-pyrrole-2,5-dicarboxylate

The iodopyrrole **32** (50.0 mg, 162 µmol), K_2CO_3 (112 mg, 809 µmol, 5.0 eq.) and benzyl 6-bromohexanoate **33** (46.0 µL, 221 µmol, 1.4 eq.) were sonicated in DMF (1.0 mL) for 5 min at RT and then stirred for 6.5 h at 70°C under Ar. Sat. aq. $NaHCO_3$ (2 mL) was added and the resulting suspension was washed with EtOAc (4x2 mL). The combined organic phases were washed with H_2O (2x5 mL), dried (Na_2SO_4) and conc. *in vacuo*. The crude product was purified chromatographically (NP, PF-15SIHP-F0025, EtOAc/Hex=1/9 R_F=0.2) and (RP, HPLC, aq. NH_4OAc (20 mM)/MeCN=95/5 to 5/95 over 13 CV): The desired product was obtained as a colourless oil (64.1 mg, 125 µmol, 77% yield).



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.38-7.29 (m, 5H, *H*17-19), 7.13 (s, 1H, *H*4), 5.11 (s, 2H, *H*15), 4.78-4.71 (m, 2H, *H*9), 3.88 (s, 3H, *H*1), 3.82 (s, 3H, *H*8), 2.36 (t, $^{3}J_{12\text{-}13}$=7.5 Hz, 2H, *H*13), 1.78-1.64 (m, 4H, *H*10+12), 1.41-1.33 (m, 2H, *H*11).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 173.4 (*C*14), 160.5 (*C*2), 160.2 (*C*7), 136.2 (*C*16), 128.7, 128.6, 128.4, 128.3, 127.9 (*C*3+6+17-19), 126.6 (*C*4), 68.6 (*C*5), 66.2 (*C*15), 51.9 (*C*1), 51.7 (*C*8), 47.7 (*C*9), 34.3 (*C*13), 31.5 (*C*10), 26.2 (*C*11), 24.6 (*C*12).

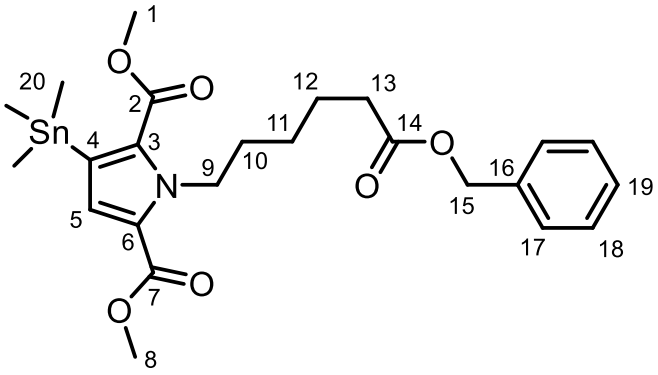
IR: v = 2861 (O-CH_3), 1713, 1699 (C=O), 981 (C-I).

LC/MS (Method 2, 254 nm, ESI+): t_R = 12.3 min, m/z: 536.1 [M+Na^+] (100), 514.2 [M+H^+] (26), 482.1 [M-MeO^-] (17).

HRMS (ESI+): $C_{21}H_{24}INNaO_6^+$ [M+Na^+] calc.: 536.0541 found: 536.0542.

35 (BNH-030) Dimethyl 1-(6-(benzyloxy)-6-oxohexyl)-3-(trimethylstannyl)-1*H*-pyrrole-2,5-dicarboxylate

The iododopyrrole **34** (300 mg, 584 µmol) and Hexamethyldistannane (287 mg, 877 µmol, 1.5 eq.) were dissolved in anh. deg. toluene (3.0 mL). The mixture was Ar sparged, while being sonicated for 5 min. $Pd(PPh_3)_4$ (33.8 mg, 29.2 µmol, 5 mol%) was added and the mixture was stirred for 17 h at 100°C under Ar. The mixture was left to cool to RT and was filtered through celite and the filter cake washed with DCM. The crude product was purified chromatographically (RP, PF-15C18AQ-F0025, aq. NH_4OAc (20 mM)/MeOH=95/5 to 5/95 over 13 CV) and (RP, PF-15C18AQ-F0040, aq. NH_4OAc (20 mM)/MeCN=1/1 to 9/1 over 13 CV): The desired product was obtained as a colourless oil (160 mg, 291 µmol, 49% yield). This procedure was adapted from the literature.[124]



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.39-7.29 (m, 5H, *H*17-19), 6.97 (s, 1H, *H*5), 5.11 (s, 2H, *H*15), 4.82-4.76 (m, 2H, *H*9), 3.84 (s, 3H, *H*1), 3.82 (s, 3H, *H*8), 2.38 (t, $^{3}J_{12\text{-}13}$=7.6 Hz, 2H, *H*13), 1.82-1.75 (m, 2H, *H*10), 1.75-1.68 (m, 2H, *H*12), 1.46-1.38 (m, 2H, *H*11), 0.24 (s, 9H, *H*20).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 173.6 (*C*14), 161.4 (*C*2), 161.3 (*C*7), 136.3 (*C*16), 132.1 (*C*6), 128.7 (*C*18), 128.30 (*C*3), 128.25 (*C*17), 127.7 (*C*19), 124.4 (*C*5), 66.2 (*C*15), 51.6 (*C*1), 51.4 (*C*8), 47.0 (*C*9), 34.4 (*C*13), 31.5 (*C*10), 26.4 (*C*11), 24.7 (*C*12), -8.29 (*C*20).

R: v = 2863 (O-CH_3), 1724, 1703 (C=O).

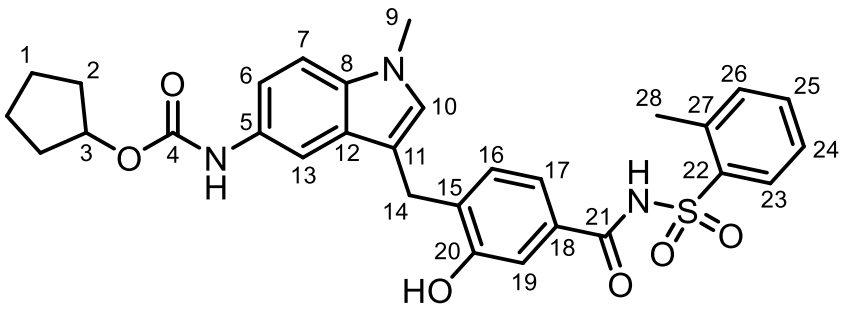
LC/MS (Method 3, 220 nm, ESI+): t_R = 15.9 min, m/z: 552.1 [M+H^+] (100).

HRMS (ESI+): $C_{24}H_{33}NNaO_6Sn^+$ [M+Na^+] calc.: 574.1222 found: 574.1224.

41 (BNH-031) Cyclopentyl (3-(2-hydroxy-4-((o-tolylsulfonyl)carbamoyl)benzyl)-1-methyl-1*H*-indol-5-yl)carbamate

Zafirlukast (100 mg, 174 µmol) was dissolved in anh. DCM (1.0 mL). A solution of BBr_3 in DCM (1 M, 869 µL, 869 µmol, 5.0 eq.) was added and the mixture was stirred for 17 h at RT. The reaction was quenched by addition of sat. $NaHCO_3$ (4 mL) and the DCM was removed *in vacuo*. H_2O (10 mL) was added and the mixture sonicated until all precipitate was suspended. The mixture was filtered and the filter cake was washed with H_2O (5 mL). The crude 4-((5-amino-1-methyl-1H-indol-3-yl)methyl)-3-hydroxy-N-(o-tolylsulfonyl)benzamide (44.2 mg, 98.3 µmol) was dried under high vacuum overnight. The off-white solid was dissolved in anh. DMF (1.0 mL). *N*-methylmorpholine (13.0 µL, 118 µmol, 1.2 eq.) and cyclopentyl chloroformate (17.5 mg, 118 µmol, 1.2 eq.) were added and the mixture was stirred for 6.5 h at RT.

Sat. aq. $NaHCO_3$ was added and the mixture was extracted with EtOAc (4x3 mL). The combined organic extracts were dried (Na_2SO_4) and conc. *in vacuo*. The crude product was purified chromatographically (RP, PF-15C18AQ-F0012, H_2O/MeCN=9/1 to 1/9 over 13 CV): The desired product was obtained as a white, amorphous solid (25.7 mg, 45.8 µmol, 26% yield over two steps). The carbamate formation procedure was adapted from the literature.[137]



^{1}H-NMR (400 MHz, $CDCl_3$): δ = 8.20-8.13 (m, 1H, *H*23), 7.53 (bs, 1H, O*H*), 7.42-7.35 (m, 1H, *H*25), 7.29-7.21 (m, 2H, *H*6+24), 7.20-7.14 (m, 1H, *H*26), 7.14-7.04 (m, 3H, *H*7, *H*13, *H*17), 7.03-6.96 (m, 1H, *H*19), 6.77-6.67 (m, 2H, *H*10+16), 5.19-5.10 (m, 1H, *H*3), 3.92 (s, 2H, *H*14), 3.58 (s, 3H, *H*9), 2.59 (s, 3H, *H*28), 1.89-1.75 (m, 2H, *H*2a), 1.75-1.60 (m, 4H, *H*2b+1a), 1.60-1.46 (m, 2H, *H*1b).

^{13}C-NMR (101 MHz, $CDCl_3$): δ = 165.9 (*C*4), 154.9 (*C*21), 154.5 (*C*20), 137.8 (*C*27), 137.1 (*C*22), 134.5 (*C*8), 133.9 (*C*12), 133.8 (*C*25), 132.5 (*C*26), 131.3 (*C*23), 130.5 (*C*19), 129.9 (*C*18), 128.4 (*C*16), 127.9 (*C*10), 126.4 (*C*24), 120.1 (*C*17), 115.8 (*C*13), 115.0 (*C*6), 111.8 (*C*11), 110.4 (*C*15), 109.6 (*C*7), 78.1 (*C*3), 32.9 (*C*2), 32.8 (*C*9), 25.7 (*C*14), 23.8 (*C*1), 20.4 (*C*28).

IR: v = 3240 (O-H), 2871 (N-CH_3), 1680 (C=O), 1426 (SO_2), 1224 (C-OH).

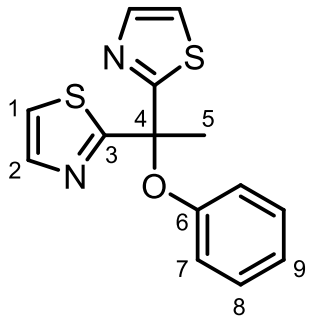
LC/MS (Method 3, 220 nm, ESI+): t_R = 7.5 min, m/z: 579.2 [M+NH_4^+] (100), 562.2 [M+H^+] (38).

HRMS (ESI+): $C_{30}H_{31}N_3NaO_6S^+$ [M+Na^+] calc.: 584.1826 found: 584.1828.

23 (BNH-032) 2,2'-(1-Phenoxyethane-1,1-diyl)dithiazole

The bromide **22** (100 mg, 370 µmol) and hexamethyldistannane (182 mg, 555 µmol, 1.5 eq.) were dissolved in anh. deg. toluene (2.0 mL). The solution was Ar sparged for 5 min while being sonicated. $Pd(PPh_3)_4$ (42.8 mg, 37.0 µmol, 10 mol%) was added and

the mixture stirred overnight at 80°C under Ar. The reaction mixture was filtered through celite with DCM as eluent. The filtrate was conc. *in vacuo*. The crude product was purified chromatographically (NP, PF-15SIHP-F0012, EtOAc/Hex=5/95) and (RP, PF-15SIHP-F0012, H_2O/MeCN=9/1 to 1/9 over 13 CV): The product was obtained as a white solid (32 mg, 111 µmol, 30% yield).



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.72-7.68 (m, 2H, *H*2+2'), 7.33-7.25 (m, 4H, *H*7, *H*8), 7.25-7.21 (m, 1H, *H*9), 7.00-6.96 (m, 2H, *H*1+1'), 2.44 (s, 3H, *H*5).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 149.8 (*C*3+3'), 145.4(*C*4), 132.6 (*C*6), 129.8, 129.7 (*C*7+8), 128.6 (*C*2+2'), 127.2 (*C*9), 122.5 (*C*1+1'), 21.8 (*C*5).

IR: v = 1374 (CH_3 deform.), 1170, 1146 (C-O-C).

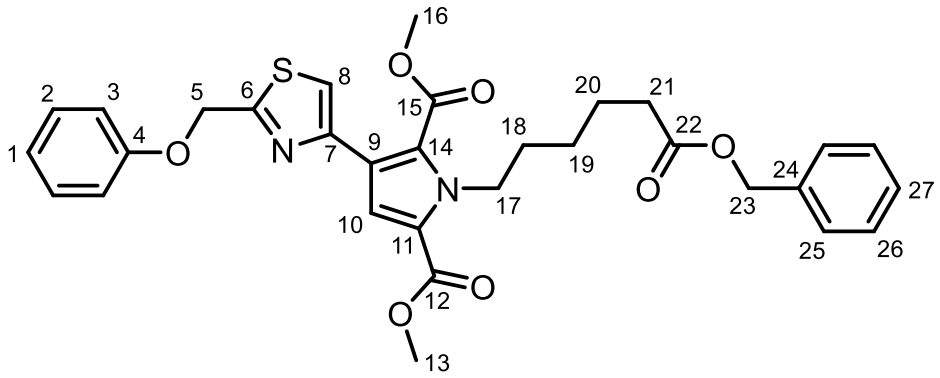
LC/MS (Method 3, 220 nm, ESI+): t_R = 11.1 min, m/z: not found.

HRMS (ESI+): $C_{14}H_{13}N_2OS_2^+$ [M+H$^+$] calc.: 289.0464 found: 289.0466.

36 (BNH-035) Dimethyl 1-(6-(benzyloxy)-6-oxohexyl)-3-(2-(phenoxymethyl)-thiazol-4-yl)-1*H*-pyrrole-2,5-dicarboxylate

The stannane **35** (160 mg, 291 µmol, 1.0 eq.), the bromide **22** (78.6 mg, 291 µmol, 1.0 eq.) and CsF (88.3 mg, 582 µmol, 2.0 eq.) were dissolved in anh. deg. DMF (2.0 mL). The mixture was Ar sparged, while being sonicated for 5 min. $Pd(PPh_3)_4$ (16.9 mg, 14.5 µmol, 5 mol%) and CuI (5.54 mg, 29.1 µmol, 10 mol%) were added and the mixture was stirred for 22 h at 50°C under Ar. The mixture was left to cool to RT and was filtered through celite. The filter cake was washed with DCM and the filtrate was conc. *in vacuo*. The crude product was purified chromatographically (NP, PF-15SIHP-F0025, EtOAc/Hex=15/85, R_F(EtOAc/Hex=2/8)=0.19) and (RP, HPLC, aq. NH_4OAc (20 mM)/MeCN=95/5 to 5/95 over 13 CV): The desired product was

isolated as a white, amorphous solid (93.3 mg, 162 µmol, 55% yield). This procedure was adapted from the literature.[127]



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.40 (s, 1H, *H*8), 7.38-7.29 (m, 7H, *H*2+25-27), 7.23 (s, 1H, *H*10), 7.05-7.98 (m, 3H, *H*1+3), 5.38 (s, 2H, *H*5), 5.12 (s, 2H, *H*23), 4.71-4.65 (m, 2H, *H*17), 3.85 (s, 3H, *H*16), 3.77 (s, 3H, *H*13), 2.37 (t, $^{3}J_{20\text{-}21}$=7.6 Hz, 2H, *H*21), 1.85-1.76 (m, 2H, *H*18), 1.75-1.67 (m, 2H, *H*20), 1.45-1.36 (m, 2H, *H*19).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 173.5 (*C*22), 165.8 (*C*6), 161.9 (*C*12), 161.0 (*C*15), 158.1 (*C*4), 149.2 (*C*7), 136.3 (*C*24), 129.8 (*C*2), 128.7 (*C*26), 128.3 (*C*25+27), 125.6 (*C*14), 125.3 (*C*11), 123.9 (*C*9), 121.9 (*C*1), 118.6 (*C*10), 116.2 (*C*8), 115.1 (*C*3), 67.5 (*C*5), 66.2 (*C*23), 51.9 (*C*13), 51.7 (*C*16), 47.0 (*C*17), 34.3 (*C*21), 31.5 (*C*18), 26.3 (*C*19), 24.6 (*C*20).

IR: v = 2863 (O-CH_3), 1722, 1704 (C=O), 1215, 1158 (C-O-C), 751, 691 (=C-H deform.).

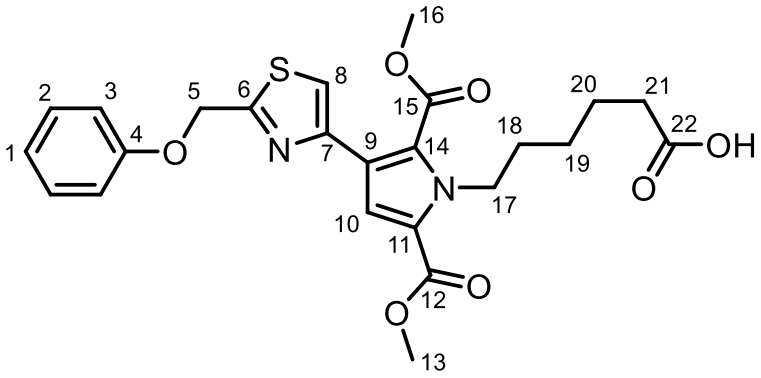
LC/MS (Method 2, 254 nm, ESI+): t_R = 11.5 min, m/z: not found.

HRMS (ESI+): $C_{31}H_{32}N_2NaO_7S^+$ [M+Na^+] calc.: 599.1822 found: 599.1822.

37 (BNH-039) 6-(2,5-Bis(methoxycarbonyl)-3-(2-(phenoxymethyl)thiazol-4-yl)-1*H*-pyrrol-1-yl)hexanoic acid

The benzyl ester **36** (80.5 mg, 140 µmol) was dissolved in MeOH (15 mL). The solution was Ar sparged for 5 min. Pd/C (10% with 50% H_2O, 164 mg, 77.0 µmol, 50 mol%) was added. The suspension was H_2 sparged for 5 min and stirred for 7 h at RT under H_2. The mixture was Ar sparged for 5 min and filtered through celite with DCM as eluent. The washed filter cake was purified chromatographically (RP, HPLC, H_2O+TFA/MeCN+TFA=95/5 to 5/95 over 13 CV): The desired product was isolated

(14.1 mg) as well as the starting material (25.2 mg). The starting material was reacted and purified again under the same conditions. The desired product was isolated as a white, amorphous solid (27.8 mg, 57.1 µmol, 41% yield overall).



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.38 (s, 1H, *H*8), 7.34-7.29 (m, 2H, *H*2), 7.21 (s, 1H, *H*10), 7.04-6.98 (m, 3H, *H*1+3), 5.39 (s, 2H, *H*5), 4.74-4.68 (m, 2H, *H*17), 3.85 (s, 3H, *H*16), 3.77 (s, 3H, *H*13), 2.37 (t, $^{3}J_{20\text{-}21}$=7.4 Hz, 2H, *H*21), 1.85-1.79 (m, 2H, *H*18), 1.73-1.66 (m, 2H, *H*20), 1.46-1.39 (m, 2H, *H*19).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 178.8 (*C*22), 166.1 (*C*6), 161.9 (*C*12), 161.1 (*C*15), 158.0 (*C*4), 149.2 (*C*7), 129.8 (*C*2), 125.6 (*C*14), 125.4 (*C*11), 123.8 (*C*9), 121.9 (*C*1), 118.7 (*C*10), 116.3 (*C*8), 115.1 (*C*3), 67.4 (*C*5), 51.9 (*C*13), 51.8 (*C*16), 47.0 (*C*17), 33.9 (*C*21), 31.4 (*C*18), 26.2 (*C*19), 24.4 (*C*20).

IR: v = 2861 (O-CH_3), 1703 (C=O), 1214, 1163 (C-O-C), 752, 690 (=C-H deform.).

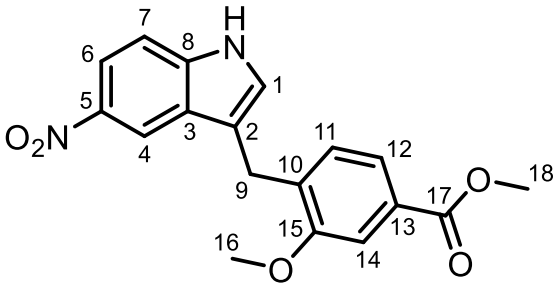
LC/MS (Method 1, 254 nm, ESI-): t_R = 9.9 min, m/z: 485.1 [M-H$^+$] (100).

HRMS (ESI+): $C_{24}H_{26}N_2NaO_7S^+$ [M+Na$^+$] calc.: 509.1353 found: 509.1354.

51 (BNH-041) Methyl 3-methoxy-4-((5-nitro-1*H*-indol-3-yl)methyl)benzoate

5-Nitroindole (3.00 g, 18.5 mmol, 1.0 eq.) and methyl 4-(bromomethyl)-3-methoxy-benzoate (4.79 g, 18.5 mmol, 1.0 eq.) were dissolved in anh. 1,4-dioxane (20 mL). Ag_2O (4.29 g, 18.5 mmol, 1.0 eq.) was added to the solution. The suspension was stirred for 17 h at 60°C. The reaction mixture was conc. *in vacuo* and filtered through celite with EtOAc as eluent. The filtrate was purified chromatographically (NP, PF-15SIHP-F0080, EtOAc/Hex=1/9 to 3/7, R_F($CHCl_3$)=0.21): The desired product was isolated as well as the starting materials 4-(bromomethyl)-3-methoxybenzoate (R_F(EtOAc/Hex=1/9)=3.35) and 5-nitroindole (R_F($CHCl_3$)=0.39). The starting materials

were reacted and purified again under the same conditions. The desired product was obtained as a yellow, amorphous solid (4.88 g, 14.4 mmol, 77% yield overall). This procedure was adapted from the literature.[137] The spectroscopic data agrees with the literature.



^{1}H-NMR (400 MHz, $CDCl_3$): δ = 8.72-8.64 (m, 1H, N*H*), 8.59-8.56 (m, 1H, *H*4), 8.10-8.04 (m, 1H, *H*6), 7.57-7.52 (m, 2H, *H*12 and *H*14), 7.40-7.34 (m, 1H, *H*7), 7.18-7.14 (m, 1H, *H*11), 7-13-7.09 (m, 1H, *H*1), 4.14 (s, 2H, *H*9), 3.93 (s, 3H, *H*18), 3.90 (s, 3H, *H*16).

^{13}C-NMR (101 MHz, $CDCl_3$): δ = 167.3 (*C*17), 157.1 (*C*15), 141.6 (*C*5), 139.5 (*C*8), 134.5 (*C*10), 129.8 (*C*11), 129.6 (*C*13), 127.0 (*C*3), 125.6 (*C*1), 122.2 (*C*12), 117.8 (*C*6), 117.4 (*C*2), 116.8 (*C*4), 111.3 (*C*7), 111.2 (*C*14), 55.7 (*C*18), 52.3 (*C*16), 25.4 (*C*9).

IR: v = 3363 (N-H), 2838 (O-CH_3), 1703 (C=O), 1379 (NO_2), 1226 (C-O-C), 717, 735 (=C-H deform.).

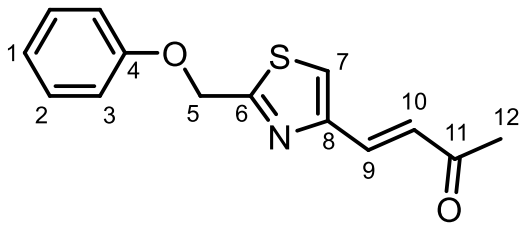
LC/MS (Method 2, 254 nm, ESI+): t_R = 9.6 min, m/z: 341.1 [M+H^+] (14), 309.0 [M-MeO^-] (100).

HRMS (ESI+): $C_{18}H_{16}N_2NaO_5^+$ [M+Na^+] calc.: 363.0951 found: 363.0946.

25 (BNH-042) (*E*)-4-(2-(Phenoxymethyl)thiazol-4-yl)but-3-en-2-one

The bromide **22** (50.0 mg, 185 µmol), 2-(di-*tert*-butyl-phosphino)-1-phenyl-1*H*-pyrrol (6.38 mg, 22.2 µmol, 12 mol%), methyl vinyl ketone (23.1 µL, 287 µmol, 1.5 eq.), TBAC (51.4 mg, 185 µmol, 1.0 eq.) and $NaHCO_3$ (17.1 mg, 204 µmol, 1.1 eq.) were dissolved in anh. DMF. The mixture was Ar sparged for 5 min, while being sonicated. $Pd_2(dba)_3$ (5.10 mg, 5.57 µmol, 3 mol%) was added and the mixture was stirred for 4 h at 80°C under Ar in a microwave reactor. The mixture was diluted with brine (10 mL)

and extracted with EtOAc (3x10 mL). The combined organic extracts were washed with H_2O (20 mL), dried (Na_2SO_4) and conc. *in vacuo*. The crude product was purified chromatographically (RP, PF-15C18AQ-F0012, aq. NH_4OAc (20 mM)/MeCN=9/1 to 1/9 over 13 CV) and (RP, HPLC, aq. NH_4OAc (20 mM)/MeCN=95/5 to 5/95 over 13 CV): The desired product was obtained as a white, amorphous solid (16.6 mg, 64.0 µmol, 35% yield). This procedure was adapted from the literature.[117]



^{1}H-NMR (400 MHz, $CDCl_3$): δ = 7.50 (s, 1H, *H*7), 7.46 (d, $^3J_{9\text{-}10}$=15.7 Hz, 1H, *H*9), 7.36-7.29 (m, 2H, *H*3), 7.03 (d, $^3J_{9\text{-}10}$=15.7 Hz, 1H, *H*10), 7.05-6.99 (m, 3H, *H*1+2), 5.37 (s, 2H, *H*5), 2.37 (s, 3H, *H*12).

^{13}C-NMR (101 MHz, $CDCl_3$): δ = 198.5 (*C*11), 168.7 (*C*6), 157.8 (*C*4), 151.9 (*C*8), 134.5 (*C*9), 129.8 (*C*3), 128.8 (*C*10), 122.8 (*C*7), 122.1 (*C*1), 115.0 (*C*2), 67.4 (*C*5), 28.5 (*C*12).

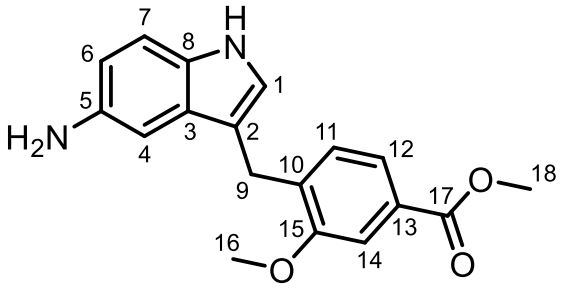
IR: v = 1660 (C=O), 1243 (C-O-C), 744, 689 (=C-H deform.).

LC/MS (Method 2, 254 nm, ESI+): t_R = 8.4 min, m/z: 301.1 [M+MeCN+H$^+$] (6), 260.0 [M+H$^+$] (100).

HRMS (ESI+): $C_{14}H_{14}NO_2S^+$ [M+H$^+$] calc.: 260.0740 found: 260.0742.

52 (BNH-043) Methyl 4-((5-amino-1*H*-indol-3-yl)methyl)-3-methoxybenzoate

The nitro ester **51** (50.0 mg, 147 µmol) was dissolved in MeOH/THF (2/1, 1.5 mL) The solution was N_2 sparged for 1 min. Pd/C (10% with 50% H_2O, 15.6 mg, 7.35 µmol, 5 mol%) was added. The mixture was stirred for 4 h at RT under H_2 and filtered through celite with DCM as eluent. The filtrate was conc. *in vacuo* and purified chromatographically (NP, PF-15SIHP-F0012, $CHCl_3$, R_F=0.13): The desired product was isolated as a white, amorphous solid, which turns brown in contact with air (41.9 mg, 135 µmol, 91% yield). This procedure was adapted from the literature.[308] The spectroscopic data agrees with the literature.



^{1}H-NMR (600 MHz, CD_3OD): δ = 7.57-7.53 (m, 1H, *H*12), 7.50-7.43 (m, 3H, *H*6+7+14), 7.22-7.18 (m, 1H, *H*2), 7.16-7.11 (m, 1H, *H*11), 7.10-7.06 (m, 1H, *H*4), 4.10 (s, 2H, *H*9), 3.91 (s, 3H, *H*18), 3.85 (s, 3H, *H*16).

^{13}C-NMR (151 MHz, CD_3OD): δ = 168.6 (*C*17), 158.6 (*C*15), 137.6 (*C*8), 136.8 (*C*5), 130.4 (*C*1), 130.4 (*C*13), 129.0 (*C*3), 127.0 (*C*11), 123.0 (*C*10), 122.9 (*C*6), 116.5 (*C*4), 114.7 (*C*2), 113.8 (*C*14), 113.7 (*C*12), 111.9 (*C*7), 56.0 (*C*18), 52.6 (*C*16), 26.0 (*C*9).

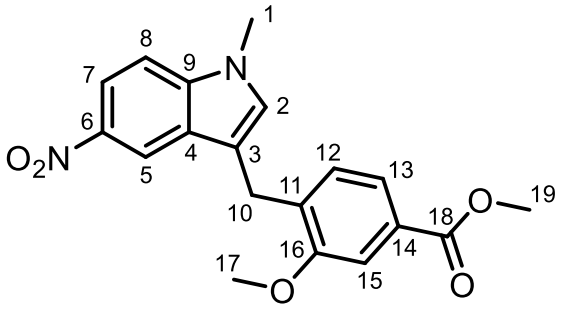
IR: v = 3006 (N-H), 1662 (C=O), 1178, 1128 (C-O-C).

LC/MS (Method 1, 254 nm, ESI+): t_R = 5.6 min, m/z: 352.1 [M+MeCN+H^+] (7), 311.0 [M+H^+] (100), 279.1 [M-MeO^-] (13).

HRMS (ESI+): $C_{18}H_{19}N_2O_3^+$ [M+H^+] calc.: 311.1390 found: 311.1389.

46 (BNH-044) Methyl 3-methoxy-4-((1-methyl-5-nitro-1*H*-indol-3-yl)methyl)benzoate

The nitro ester **51** (50.0 mg, 147 µmol) was dissolved in anh. THF under N_2. NaH (7.05 mg, 176 µmol, 1.2 eq.) was added. The bright-yellow solution turned dark red instantaneously. The mixture was stirred for 10 min at RT. MeI (11.0 µL, 176 µmol, 1.2 eq.) was added and the mixture was stirred for 3 h at RT. Aq. HCl (1 M, 5 mL) was added and the mixture was extracted with EtOAc (5x5 mL). The combined organic extracts were dried (Na_2SO_4) and conc. *in vacuo* to give an orange oil. The crude product was purified chromatographically (NP, PF-15SIHP-F0012, EtOAc/Hex=2/8, R_F=0.13) and (RP, HPLC, H_2O+TFA/MeCN+TFA=95/5 to 5/95 over 13 CV): The desired product was obtained as a bright-yellow, amorphous solid (27.2 mg, 76.8 µmol, 52% yield). This procedure was adapted from the literature.[137] The spectroscopic data agrees with the literature.



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 8.59-8.55 (m, 1H, *H*5), 8.12-8.01 (m, 1H, *H*7), 7.58-7.53 (m, 2H, *H*13+15), 7.28-7.24 (m, 1H, *H*8), 7.19-7.15 (m, 1H, *H*12), 6.92 (s, 1H, *H*2), 4.12 (s, 2H, *H*10), 3.95 (s, 3H, *H*19), 3.90 (s, 3H, *H*17), 3.78 (s, 3H, *H*1).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 167.1 (*C*18), 157.1 (*C*16), 141.4 (*C*6), 139.9 (*C*9), 134.5 (*C*11), 130.3 (*C*2), 129.8 (*C*12), 129.7 (*C*14), 127.3 (*C*4), 122.3 (*C*13), 117.4 (*C*7), 116.9 (*C*5), 116.5 (*C*3), 111.3 (*C*15), 109.2 (*C*8), 55.7 (*C*19), 52.2 (*C*17), 33.2 (*C*1), 25.3 (*C*10).

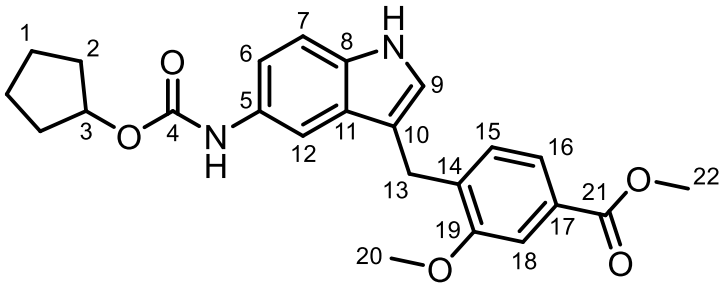
IR: v = 2836 (O-CH_3), 1705 (C=O), 1260 (NO_2).

LC/MS (Method 1, 254 nm, ESI+): t_R = 10.0 min, m/z: 377.0 [M+Na^+] (73), 355.0 [M+H^+] (64), 323.0 [M-MeO^-] (100).

HRMS (ESI+): $C_{19}H_{18}N_2NaO_5^+$ [M+Na^+] calc.: 377.1108 found: 377.1107.

53 (BNH-045) Methyl 4-((5-(((cyclopentyloxy)carbonyl)amino)-1*H*-indol-3-yl)methyl)-3-methoxybenzoate

The amino ester **52** (363 mg, 1.17 mmol) and *N*-methylmorpholine (129 µL, 1.17 mmol, 1.0 eq.) were dissolved in anh. DCM (5 mL) under N_2. Cyclopentyl chloroformate (145 µL, 1.17 mmol, 1.0 eq.) was added dropwise. The mixture was stirred for 3 h at RT. Aq. HCl (1 M, 15 mL) was added and the mixture was extracted with EtOAc (3x15 ml). The combined organic extracts were dried (Na_2SO_4) and conc. *in vacuo* to give a light-brown oil. The crude product was purified chromatographically (NP, PF-15SIHP-F0040, $CHCl_3$, R_F=0.31) and (RP, HPLC, aq. NH_4OAc (20 mM)/MeCN=95/5 to 5/95 over 13 CV): The desired product was obtained as a light-brown, amorphous solid (483 mg, 1.14 mmol, 97% yield). This procedure was adapted from literature.[137] The spectroscopic data agrees with the literature.[134]



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 8.01 (s, 1H, indole N*H*), 7.61-7.48 (m, 3H, *H*10+12+16), 7.27-7.23 (m, 1H, *H*6), 7.18-7.07 (m, 2H, *H*7+15), 6.91 (s, 1H, *H*9), 6.51 (s, 1H, CON*H*), 5.24-5.17 (m, 1H, *H*3), 4.08 (s, 2H, *H*13), 3.91 (s, 3H, *H*22), 3.89 (s, 3H, *H*20), 1.93-1.83 (m, 2H, *H*2a), 1.81-1.66 (m, 4H, *H*1a+2b), 1.64-1.54 (m, 2H, *H*1b).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 167.4 (*C*4), 157.3 (*C*21), 135.4 (*C*5), 133.5 (*C*8), 130.5 (*C*8), 129.7 (*C*15), 129.2 (*C*14), 127.9 (*C*11), 123.7 (*C*9), 122.2 (*C*16), 114.3 (*C*7), 111.4 (*C*6), 111.0 (*C*12+18), 77.9 (*C*3), 55.7 (*C*22), 52.2 (*C*20), 33.0 (*C*2), 25.5 (*C*13), 23.8 (*C*1).

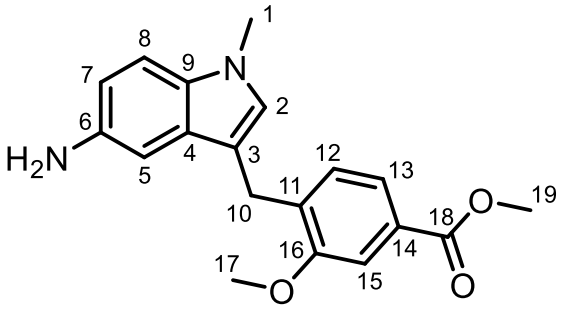
IR: v = 3327 (N-H), 2872 (O-CH_3), 1697 (C=O), 1224 (C-O-C), 760 (=C-H deform.).

LC/MS (Method 1, 254 nm, ESI+): t_R = 9.8 min, m/z: 445.1 [M+Na^+] (87), 423.1 [M+H^+] (32), 391.1 [M-MeO^-] (5), 355.0 [M-$C_5H_7^-$] (100), 311.1 [M-$C_6H_7O_2^-$] (56).

HRMS (ESI+): $C_{24}H_{26}N_2NaO_5^+$ [M+Na^+] calc.: 445.1734 found: 445.1731.

48 (BNH-047) Methyl 4-((5-amino-1-methyl-1*H*-indol-3-yl)methyl)-3-methoxybenzoate

The nitro ester **46** (50.0 mg, 141 µmol) was dissolved in MeOH/THF (2/1, 1.5 mL). The solution was N_2 sparged for 1 min. Pd/C (10% with 50% H_2O, 15.0 mg, 7.06 µmol, 5 mol%) was added. The mixture was stirred for 3 h at RT under H_2 and filtered through celite with DCM as eluent. The filtrate was conc. *in vacuo* and purified chromatographically (NP, PF-15SIHP-F0012, $CHCl_3$, R_F=0.17), (RP, HPLC, H_2O+TFA/MeCN+TFA=95/5 to 5/95 over 13 CV): The desired product was obtained as an off-white, amorphous solid (39.0 mg, 120 µmol, 85% yield). This procedure was adapted from the literature.[137] The spectroscopic data agrees with the literature.



^{1}H-NMR (600 MHz, CD_3OD): δ = 7.56-7.54 (m, 1H, *H*13), 7.50-7.45 (m, 3H, *H*7+8+15), 7.17-7.12 (m, 2H, *H*5+12), 7.12-7.10 (m, 1H, *H*2), 4.09 (s, 2H, *H*10), 3.91 (s, 3H, *H*19), 3.86 (s, 3H, *H*17), 3.77 (s, 3H, *H*1).

^{13}C-NMR (151 MHz, CD_3OD): δ = 168.6 (*C*18), 158.6 (*C*16), 138.1 (*C*9), 136.6 (*C*6), 131.4 (*C*2), 130.7 (*C*12), 130.5 (*C*14), 129.5 (*C*4), 123.2 (*C*11), 122.9 (*C*7), 116.6 (*C*5), 114.3 (*C*3), 114.1 (*C*15), 111.9 (*C*13), 111.8 (*C*8), 56.0 (*C*19), 52.6 (*C*17), 32.9 (*C*1), 25.9 (*C*10).

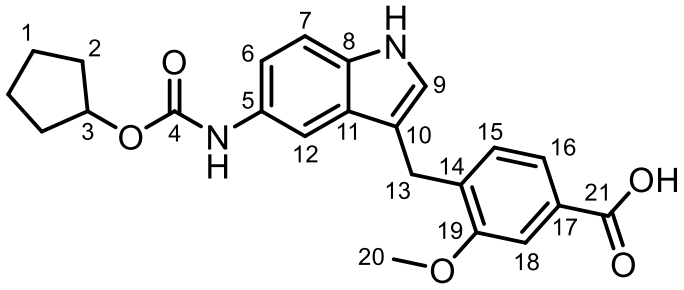
IR: v = 2960 (CH_3), 2923 (CH_2), 1701 (C=O), 1210, 1187 (C-O-C).

LC/MS (Method 1, 254 nm, ESI+): t_R = 6.1 min, m/z: 325.1 [$M+H^+$] (100), 293.0 [$M-MeO^-$] (8).

HRMS (ESI+): $C_{19}H_{21}N_2O_3^+$ [$M+H^+$] calc.: 325.1547 found: 325.1543.

54 (BNH-048) 4-((5-(((Cyclopentyloxy)carbonyl)amino)-1*H*-indol-3-yl)methyl)-3-methoxybenzoic acid

The carbamate ester **53** (50.0 mg, 118 µmol) was dissolved in H_2O/MeOH/THF (3/2/2, 1.75 mL). LiOH (14.2 mg, 592 µmol, 5.0 eq.) was added. The mixture was sonicated until the LiOH was well suspended and was then stirred for 19 h at RT. The organic solvents were removed *in vacuo*. The reaction was quenched with aq. HCl (1 M, 2 mL). White precipitate formed. The mixture was extracted with EtOAc (3x2 mL). The combined organic extracts were dried (Na_2SO_4) and conc. *in vacuo* to give a red-brown oil. The crude product was purified chromatographically (RP, HPLC, aq. NH_4OAc (20 mM)/MeCN=95/5 to 5/95 over 13 CV): The desired product was obtained as an off-white, amorphous solid (44.0 mg, 108 µmol, 91% yield). This procedure was adapted from the literature.[137] The spectroscopic data agrees with the literature.[134]



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 8.04 (bs, 1H, N*H*), 7.61-7.52 (m, 3H, *H*12+16+18), 7.25-7.21 (m, 1H, *H*6), 7.16-7.09 (m, 2H, *H*7+15), 6.92-6.87 (m, 1H, *H*9), 6.56 (bs, 1H, CON*H*), 5.25-5.17 (m, 1H, *H*3), 4.08 (s, 2H, *H*13), 3.91 (s, 3H, *H*16), 1.93-1.83 (m, 2H, *H*2a), 1.80-1.67 (m, 4H, *H*2b+1a), 1.63-1.55 (m, 2H, *H*1b).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 171.7 (*C*4), 157.3 (*C*21), 136.4 (*C*5), 133.6 (*C*8), 130.4 (*C*10), 129.8 (*C*15), 128.4 (*C*14), 127.9 (*C*11), 123.8 (*C*9), 122.9 (*C*16), 114.1 (*C*7), 111.4 (*C*6), 111.3 (*C*12+18), 55.7 (*C*20), 32.9 (*C*2), 25.5 (*C*13), 23.8 (*C*1).

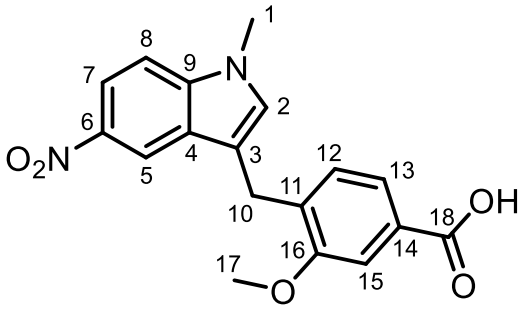
IR: v = 3322 (N-H), 2961 (CH_3), 2872 (O-CH_3), 1682 (C=O), 1412 (CH_2+CH_3 deform.), 1215 (C-O-C), 728 (=C-H deform.).

LC/MS (Method 1, 254 nm, ESI+): t_R = 8.6 min, m/z: 431.1 [M+Na^+] (36), 409.1 [M+H^+] (28), 341.0 [M-$C_5H_7^-$] (100), 323.0 [M-$C_5H_9O^-$] (40), 297.0 [M-$C_6H_7O_2^-$] (6).

HRMS (ESI+): $C_{23}H_{24}N_2NaO_5^+$ [M+Na^+] calc.: 431.1577 found: 431.1577.

47 (BNH-049) 3-Methoxy-4-((1-methyl-5-nitro-1*H*-indol-3-yl)methyl)benzoic acid

The nitro ester **46** (150 mg, 423 µmol) was dissolved in H_2O/MeOH/THF (3/3/5, 5.5 mL). LiOH (50.7 mg, 2.12 mmol, 5.0 eq.) was added. The mixture was sonicated until the LiOH was well suspended. The suspension was stirred for 5.5 h at RT. The reaction was quenched with aq. HCl (1 M, 5 mL). A yellow precipitate formed. The mixture was extracted with EtOAc (3x10 mL). The combined organic extracts were dried (Na_2SO_4) and conc. *in vacuo.* The crude product was purified chromatographically (RP, HPLC, H_2O+TFA/MeCN+TFA=95/5 to 5/95 over 13 CV): The desired product was obtained as a yellow, amorphous solid (143 mg, 419 µmol, 99% yield). This procedure was adapted from the literature.[137] The spectroscopic data agrees with the literature.[136]



^{1}H-NMR (600 MHz, CD_3OD): δ = 12.85 (bs, 1H, COO*H*), 8.52-8.48 (m, 1H, *H*5), 8.04-7.99 (m, 1H, *H*7), 7.61-7.56 (m, 1H, *H*8), 7.51-7.48 (m, 1H, *H*2), 7.48-7.44 (m, 1H, *H*13), 7.37-7.34 (m, 1H, *H*15), 7.27-7.23 (m, 1H, *H*12), 4.10 (s, 2H, *H*10), 3.91 (s, 3H, *H*17), 3.81 (s, 3H, *H*1).

^{13}C-NMR (151 MHz, CD_3OD): δ = 167.1 (*C*18), 156.6 (*C*16), 140.3 (*C*6), 139.4 (*C*9), 134.1 (*C*11), 131.5 (*C*15), 130.1 (*C*3), 129.7 (*C*12), 126.5 (*C*4), 121.7 (*C*13), 116.3 (*C*7), 115.9 (*C*5), 115.2 (*C*14), 110.9 (*C*2), 110.3 (*C*8), 55.4 (*C*17), 32.8 (*C*1), 24.4 (*C*10).

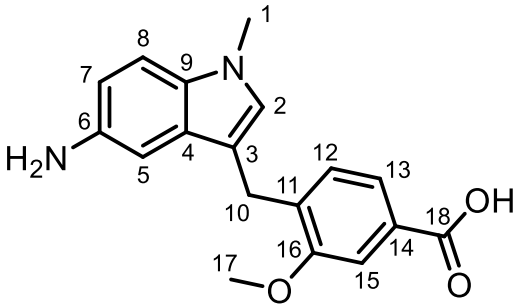
IR: v = 2924, 2835 (CH_2), 1684 (C=O), 1337 (NO_2), 1297, 1274 (C-O-C), 760, 740 (=C-H).

LC/MS (Method 1, 254 nm, ESI+): t_R = 8.7 min, m/z: 363.0 [M+Na^+] (48), 341.0 [M+H^+] (100), 323.0 [M-HO^-] (93).

HRMS (ESI+): $C_{18}H_{16}N_2NaO_5^+$ [M+Na^+] calc.: 363.0951 found: 363.0954.

49 (BNH-050) 4-((5-Amino-1-methyl-1*H*-indol-3-yl)methyl)-3-methoxybenzoic acid

The nitro acid **47** (50.0 mg, 147 µmol) was dissolved in MeOH/THF (2/1, 1.5 mL). The solution was N_2 sparged for 1 min. Pd/C (10% with 50% H_2O, 15.6 mg, 7.35 µmol, 5 mol%) was added. The suspension was H_2 sparged for 1 min and stirred for 4 h at RT under H_2. The mixture was filtered through celite with DCM and MeOH as eluents sequentially. The filtrate was conc. *in vacuo* and purified chromatographically (RP, HPLC, aq. NH_4OAc (20 mM)/MeCN=95/5 to 5/95 over 13 CV): The desired product was obtained as a white, amorphous solid (41.0 mg, 132 µmol, 89% yield). This procedure was adapted from the literature.[137]



^{1}H-NMR (400 MHz, DMSO-d_6): δ = 7.49-7.46 (m, 1H, *H*13), 7.45-7.40 (m, 1H, *H*15), 7.11-7.05 (m, 2H, *H*7+8), 6.88 (s, 1H, *H*2), 6.57-6.50 (m, 2H, *H*5+12), 3.90 (s, 3H, *H*17), 3.89 (s, 2H, *H*10), 3.62 (s, 3H, *H*1).

^{13}C-NMR (101 MHz, DMSO-d_6): δ = 167.4 (*C*18), 156.7 (*C*16), 140.8 (*C*6), 134.9 (*C*9), 130.9 (*C*11), 129.7 (*C*4), 129.3 (*C*7), 128.3 (*C*14), 127.7 (*C*2), 121.5 (*C*15), 111.8 (*C*12), 110.6 (*C*13), 109.8 (*C*8), 109.5 (*C*3), 102.0 (*C*5), 55.4 (*C*17), 32.2 (*C*1), 24.8 (*C*10).

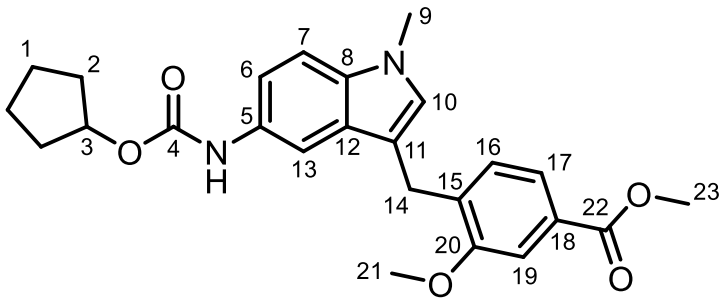
IR: v = 1724 (C=O), 1377 (C=O), 1246 (C-O-C), 779 (=C-H deform.).

LC/MS (Method 1, 254 nm, ESI+): t_R = 5.3 min, m/z: 311.1 [M+H$^+$] (100).

HRMS (ESI+): $C_{18}H_{19}N_2O_3^+$ [M+H$^+$] calc.: 311.1390 found: 311.1390.

44 (BNH-051) Methyl 4-((5-(((cyclopentyloxy)carbonyl)amino)-1-methyl-1*H*-indol-3-yl)methyl)-3-methoxybenzoate

The amino ester **48** (136 mg, 420 µmol) and *N*-methylmorpholine (50.8 µL, 462 µmol, 1.1 eq.) were dissolved in anh. DCM (2.0 mL) under N_2. Cyclopentyl chloroformate (57 µL, 462 µmol, 1.1 eq.) was added dropwise. The mixture was stirred for 2 h at RT. Aq. HCl (1 M, 15 mL) was added and the mixture was extracted with EtOAc (3x15 ml). The combined organic extracts were dried (Na_2SO_4) and conc. *in vacuo* to give a light-brown oil. The crude product was purified chromatographically (NP, PF-15SIHP-F0025, EtOAc/Hex=2/8, R_F=0.20): The desired product was obtained as a pale-yellow, amorphous solid (174 mg, 399 µmol, 95% yield). This procedure was adapted from the literature.[137] The spectroscopic data agrees with the literature.



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.63-7.49 (m, 3H, *H*13+17+19), 7.22-7.11 (m, 3H, *H*6+7+16), 6.75 (s, 1H, *H*10), 6.60 (bs, 1H, CON*H*), 5.24-5.17 (m, 1H, *H*3), 4.07 (s, 2H, *H*14), 3.92 (s, 3H, *H*23), 3.90 (s, 3H, *H*21), 3.69 (s, 3H, *H*9), 1.92-1.84 (m, 2H, *H*2a), 1.81-1.67 (m, 4H, *H*2b+1a), 1.65-1.55 (m, 2H, *H*1b).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 167.3 (*C*22), 157.2 (*C*4), 154.3 (*C*20), 135.6 (*C*5), 134.3 (*C*8), 130.2 (*C*11), 129.7 (*C*16), 129.1 (*C*6), 128.3 (*C*12), 128.2 (*C*13), 122.1 (*C*17), 115.3 (*C*18), 112.6 (*C*10), 110.9 (*C*19), 109.8 (*C*15), 109.4 (*C*7), 77.7 (*C*3), 55.7 (*C*23), 52.1 (*C*21), 32.9 (*C*2), 32.8 (*C*9), 25.3 (*C*14), 23.8 (*C*1).

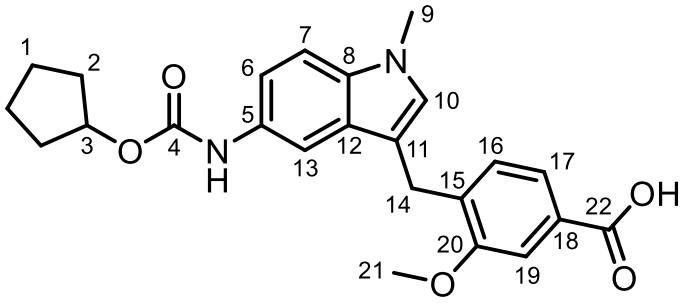
IR: v = 3326 (N-H), 1699 (C=O), 1227 (C-O-C), 760 (=C-H deform.).

LC/MS (Method 1, 254 nm, ESI+): t_R = 10.5 min, m/z: 459.1 [M+Na^+] (44), 437.1 [M+H^+] (100), 369.0 [M-$C_5H_7^-$] (89), 325.1 [M-$C_6H_7O_2^-$] (10).

HRMS (ESI+): $C_{25}H_{28}N_2NaO_5^+$ [M+Na^+] calc.: 459.1890 found: 459.1890.

43 (BNH-053) 4-((5-(((Cyclopentyloxy)carbonyl)amino)-1-methyl-1*H*-indol-3-yl)methyl)-3-methoxybenzoic acid

The carbamate ester **44** (72.6 mg, 166 µmol) was dissolved in H_2O/MeOH/THF (1/1/1, 3 mL). LiOH (19.9 mg, 832 µmol, 5.0 eq.) was added. The mixture was sonicated until the LiOH was well suspended. The suspension was stirred for 23 h at RT. Aq. HCl (1 M, 3 mL) was added. The mixture was extracted with EtOAc (3x5 mL). The combined organic extracts were dried (Na_2SO_4) and conc. *in vacuo*. The crude product was purified chromatographically (RP, HPLC, aq. NH_4OAc (20 mM)/MeCN=95/5 to 5/95 over 13 CV): The desired product was obtained as a white, amorphous solid (69.6 mg, 165 µmol, 99% yield). This procedure was adapted from the literature.[137] The spectroscopic data agrees with the literature.[135]



^{1}H-NMR (400 MHz, DMSO-d_6): δ = 9.24 (bs, 1H, N*H*), 7.64 (bs, 1H, *H*10), 7.50 (s, 1H, *H*19), 7.48-7.42 (m, 1H, *H*17), 7.30-7.22 (m, 1H, *H*6), 7.22-7.14 (m, 1H, *H*7), 7.14-7.08 (m, 1H, *H*16), 7.02 (s, 1H, *H*13), 5.14-5.01 (m, 1H, *H*3), 3.97 (s, 2H, *H*14), 3.92 (s, 3H, *H*21), 3.68 (s, 3H, *H*9), 1.94-1.76 (m, 2H, *H*2a), 1.76-1.61 (m, 4H, *H*2b+1a), 1.61-1.47 (m, 2H, *H*1b).

^{13}C-NMR (101 MHz, DMSO-d_6): δ = 167.4 (*C*22), 156.7 (*C*20), 153.7 (*C*4), 134.7 (*C*15), 133.3 (*C*5), 131.1 (*C*12), 130.0 (*C*18), 129.3 (*C*16), 128.5 (*C*13), 127.3 (*C*8), 121.6 (*C*17), 114.4 (*C*7), 111.2 (*C*11), 110.7 (*C*19), 109.5 (*C*6), 108.2 (*C*10), 76.3 (*C*3), 55.5 (*C*21), 32.4 (*C*9), 32.3 (*C*2), 24.7 (*C*14), 23.3 (*C*1).

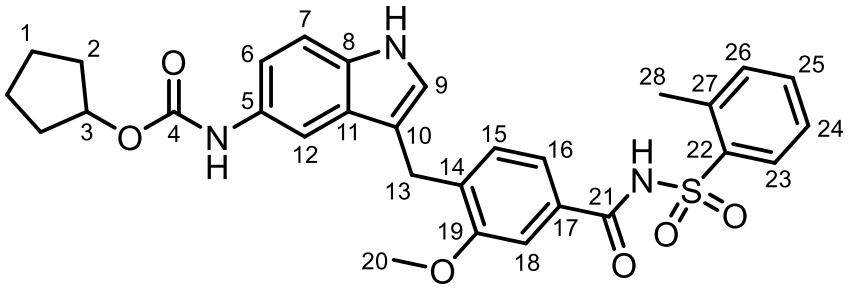
IR: v = 2954, 2871 (CH_3), 1689 (C=O), 1221 (C-O-C), 765 (=C-H deform.).

LC/MS (Method 1, 254 nm, ESI+): t_R = 9.3 min, m/z: 445.1 [M+Na^+] (23), 423.1 [M+H^+] (100), 355.1 [M-$C_5H_7^-$] (86), 337.0 [M-$C_5H_9O^-$] (14).

HRMS (ESI+): $C_{24}H_{26}N_2NaO_5^+$ [M+Na^+] calc.: 445.1734 found: 445.1734.

42 (BNH-054) Cyclopentyl (3-(2-methoxy-4-((o-tolylsulfonyl)carbamoyl)benzyl)-1*H*-indol-5-yl)carbamate

The carbamate acid **54** (56.0 mg, 137 µmol), 2-methylbenzenesulfonamide (47.0 mg, 274 µmol, 2.0 eq.), DMAP (16.8 mg, 137 µmol, 1.0 eq.), DIPEA (51.3 µL, 302 µmol, 2.2 eq.) and PyBOP (78.8 mg, 151 µmol, 1.1 eq.) were dissolved in anh. DCM (1.0 mL). The mixture was stirred for 48 h at RT under N_2. The mixture was conc. *in vacuo* and purified chromatographically (RP, PF-15C18AQ-F0012, H_2O/MeOH=9/1 to 9/1) and (RP, HPLC, aq. NH_4OAc (20 mM)/MeCN=95/5 to 5/95 over 13 CV): The desired product was obtained as a white, amorphous solid (53.0 mg, 94.4 µmol, 68% yield). This procedure was adapted from the literature.[137]



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 9.51 (bs, 1H, SO_2NH), 8.27-8.22 (m, 1H, *H*23), 8.21 (bs, 1H, N*H*), 7.57-7.45 (m, 2H, CON*H*+*H*6), 7.39-7.34 (m, 1H, *H*25), 7.29-7.24 (m, 2H, *H*24+26), 7.22-7.17 (m, 1H, *H*12), 7.11-7.00 (m, 2H, *H*16+18), 6.98-6.94 (m, 1H, *H*15), 6.91-6.87 (m, 1H, *H*7), 6.61-6.54 (m, 1H, *H*9), 5.23-5.15 (m, 1H, *H*3), 3.99 (s, 2H, *H*13), 3.80 (s, 3H, *H*20), 2.67 (s, 3H, *H*28), 1.93-1.82 (m, 2H, *H*2a), 1.82-1.65 (m, 4H, *H*2b+1a), 1.65-1.52 (m, 2H, *H*1b).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 164.8 (*C*21), 157.7 (*C*19), 154.6 (*C*4), 137.8 (*C*22), 137.0 (*C*27), 136.3 (*C*17), 134.0 (*C*6), 133.6 (*C*5), 132.5 (*C*24), 131.7 (*C*23), 130.3 (*C*10), 130.1 (*C*9), 129.8 (*C*15), 127.8 (*C*8), 126.5 (*C*25), 123.9 (*C*7), 119.6 (*C*16), 116.0 (*C*18), 113.7 (*C*11), 111.5 (*C*12), 110.2 (*C*14), 109.8 (*C*26), 78.0 (*C*3), 55.7 (*C*20), 32.9 (*C*2), 25.4 (*C*13), 23.8 (*C*1), 20.5 (*C*28).

IR: v = 3373 (N-H), 2960 (CH_3), 1680 (C=O), 1454, 1423 (CH_3+CH_2 deform.), 1333 (SO_2), 1219, 1160 (C-O-C).

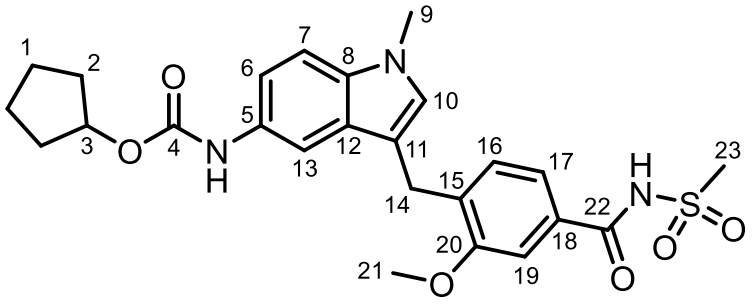
LC/MS (Method 1, 254 nm, ESI+): t_R = 10.0 min, m/z: 579.1 [M+NH_4^+] (27), 494.0 [M-$C_5H_7^-$] (100).

HRMS (ESI+): $C_{30}H_{31}N_3NaO_6S^+$ [M+Na^+] calc.: 584.1826 found: 584.1825.

45 (BNH-055) Cyclopentyl (3-(2-methoxy-4-((methylsulfonyl)carbamoyl)benzyl)-1-methyl-1*H*-indol-5-yl)carbamate

The carbamate acid **43** (73.3 mg, 174 µmol), EDCI HCl (34.9 mg, 182 µmol, 1.05 eq.) and DMAP (22.3 mg, 182 µmol, 1.05 eq.) were dissolved in anh. DCM (500 µL). The mixture was stirred for 1.5 h at RT under N_2. Methanesulfonamide (17.3 mg, 182 µmol, 1.05 eq.) was added. The mixture was stirred for 20 h at RT under N_2. Aq. HCl (1 M, 3 mL) was added and the mixture was extracted with EtOAc (3x5 mL). The combined organic extracts were dried (Na_2SO_4) and conc. *in vacuo* to give a purple oil. The crude

product was purified chromatographically (NP, PF-15SIHP-F0025, MeOH/DCM=2/98, R_F=0.11) and (RP, HPLC, aq. NH_4OAc (20 mM)/MeCN=95/5 to 5/95 over 13 CV): The desired product was obtained as a pale-pink, amorphous solid (35.0 mg, 70.1 µmol, 40% yield). This procedure was adapted from the literature.[137]



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.52 (bs, 1H, *H*6), 7.38-7.35 (m, 1H, *H*17), 7.23-7.18 (m, 1H, *H*19), 7.18-7.13 (m, 1H, *H*7), 7.11-7.02 (m, 2H, *H*13+16), 6.78-6.74 (m, 1H, *H*10), 6.65 (bs, 1H, CON*H*), 5.19-5.12 (m, 1H, *H*3), 4.01 (s, 2H, *H*14), 3.86 (s, 3H, *H*21), 3.67 (s, 3H, *H*9), 3.26 (s, 3H, *H*23), 1.90-1.78 (m, 2H, *H*2a), 1.78-1.63 (m, 4H, *H*2b+1a), 1.63-1.51 (m, 2H, *H*1b).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 166.0 (*C*22), 157.7 (*C*20), 154.5 (*C*4), 136.7 (*C*15), 134.4 (*C*8), 130.07, 130.03, 129.99 (*C*5+16+18), 128.5 (*C*10), 128.1 (*C*12),120.0 (*C*19), 115.5 (*C*13), 112.1 (*C*11), 110.1 (*C*6), 109.7 (*C*17), 109.6 (*C*7), 77.9 (*C*3), 55.8 (*C*21), 41.8 (*C*23), 32.92, 32.85 (*C*2+9), 25.4 (*C*14), 23.8 (*C*1).

IR: v = 2956, 2872 (CH_3), 1684 (C=O), 1434 (SO_2), 1222, 1153 (C-O-C), 757 (=C-H deform.).

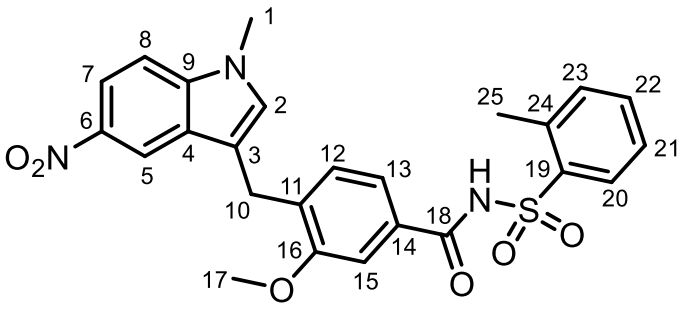
LC/MS (Method 1, 254 nm, ESI+): t_R = 9.2 min, m/z: 522.0 [M+Na^+] (9), 500.1 [M+H^+] (66), 432.0 [M-$C_5H_7^-$] (52), 388.0 [M-$C_6H_7O_2^-$] (100).

HRMS (ESI+): $C_{25}H_{29}N_3NaO_6S^+$ [M+Na^+] calc.: 522.1669 found: 522.1672.

38 (BNH-057) 3-Methoxy-4-((1-methyl-5-nitro-1*H*-indol-3-yl)methyl)-*N*-(*o*-tolylsulfonyl)benzamide

The nitro acid **47** (509 mg, 1.5 mmol), EDCI HCl (301 mg, 1.57 mmol, 1.05 eq.) and DMAP (192 mg, 1.57 mmol, 1.05 eq.) were dissolved in anh. DMF (10 mL). The solution was stirred for 1 h at RT under N_2. 2-Methylbenzenesulfonamide (269 mg, 1.57 mmol, 1.05 eq.) was added. The mixture was stirred for 18 h at RT under N_2. Aq.

HCl (1 M, 10 mL) was added. The mixture was extracted with EtOAc (3x15 mL), dried (Na_2SO_4) and conc. *in vacuo*. The crude product was purified chromatographically (NP, PF-15SIHP-F0080, MeOH/DCM=1/99, R_F(MeOH/DCM=2/98)=0.2): The desired product was isolated as well as the starting material **47**. The latter was subjected to the same procedure again to give the desired product as a bright-yellow, amorphous solid (671 mg, 1.36 mmol, 90% yield overall). This procedure was adapted from the literature.[137] The spectroscopic data agrees with the literature.[136]



^{1}H-NMR (600 MHz, DMSO-d_6): δ = 8.54-8.46 (m, 1H, *H*5), 8.06-7.96 (m, 2H, *H*7+20), 7.59-7.54 (m, 2H, *H*8+22), 7.54-7.50 (m, 1H, *H*13), 7.46-7.40 (m, 2H, *H*15+21), 7.40-7.36 (m, 1H, *H*23), 7.36-7.32 (m, 1H, *H*2), 7.29-7.22 (m, 2H, *H*12+SO_2N*H*), 4.08 (s, 2H, *H*10), 3.92 (s, 3H, *H*17), 3.79 (s, 3H, *H*1), 2.60 (s, 3H, *H*25).

^{13}C-NMR (151 MHz, DMSO-d_6): δ = 164.8 (*C*18), 156.6 (*C*16), 140.3 (*C*6), 139.4 (*C*9), 137.7 (*C*24), 136.8 (*C*19), 134.6 (*C*14), 133.4 (*C*22), 132.3 (*C*23), 131.5 (*C*2), 130.6 (*C*3), 130.4 (*C*20), 129.7 (*C*12), 126.4 (*C*4), 126.2 (*C*21), 120.8 (*C*15), 116.3 (*C*7), 115.9 (*C*5), 115.1 (*C*11), 110.32 (*C*13), 110.27 (*C*8), 55.7 (*C*17), 32.8 (*C*1), 24.4 (*C*10), 19.5 (*C*25).

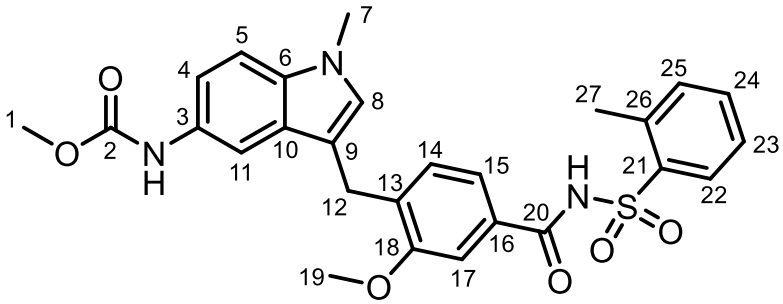
IR: v = 3310 (N-H), 1702 (C=O), 1401 (SO_2), 758 (=C-H deform.).

LC/MS (Method 1, 254 nm, ESI+): t_R = 10.0 min, m/z: 516.0 [M+Na^+] (69), 511.0 [M+NH_4^+] (37), 494.0 [M+H^+] (100), 323.0 [M-$C_7H_8NO_2S^-$] (91).

HRMS (ESI+): $C_{25}H_{23}N_3NaO_6S^+$ [M+Na^+] calc.: 516.1200 found: 516.1197.

40 (BNH-059) Methyl (3-(2-methoxy-4-((*o*-tolylsulfonyl)carbamoyl)benzyl)-1-methyl-1*H*-indol-5-yl)carbamate

The nitro sulfonamide **38** (325 mg, 659 µmol) was dissolved in MeOH/THF (1/1, 20 mL). The solution was N_2 sparged for 1 min. Pd/C (10% with 50% H_2O, 70.1 mg, 32.9 µmol, 5 mol%) was added. The suspension was H_2 sparged for 1 min and stirred for 18 h at RT under H_2. The pale-yellow suspension was filtered through celite. The filter cake was rinsed with DCM. TLC of the filtrate showed complete conversion to the desired product (R_F(MeOH/DCM=5/95)=0.23). The product oxidises rapidly on silica and in air. It was used in the next reaction without further purification. The produced amine (88.7 mg, 191 µmol) and *N*-methylmorpholine (42.1 µL, 383 µmol, 2.0 eq.) were dissolved in anh. DMF (2.0 mL). Methyl chloroformate (29.6 µL, 383 µmol, 2.0 eq.) was added dropwise. The mixture was stirred for 1 h at RT under N_2. The reaction was quenched with aq. HCl (1 M, 10 mL) and extracted with EtOAc (3x10 mL). The combined organic extracts were dried (Na_2SO_4) and conc. *in vacuo* to give a dark-brown residue. The crude product was purified chromatographically (RP, HPLC, H_2O+TFA/MeCN+TFA=95/5 to 5/95 over 13 CV): The desired product was obtained as a white, amorphous solid (52.8 mg, 101 µmol, 15% yield over two steps). This procedure was adapted from the literature.[137] The spectroscopic data agrees with the literature.[133]



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 9.68-9.59 (bs, 1H, CON*H*), 8.29-8.21 (m, 1H, *H*22), 7.56-7.44 (m, 2H, *H*5+24), 7.42-7.35 (m, 1H, *H*23), 7.35-7.30 (m, 1H, *H*11), 7.30-7.25 (m, 2H, *H*25), 7.24-7.15 (m, 2H, *H*4+15), 7.14-7.02 (m, 2H, *H*14+17), 6.79-6.74 (m, 1H, *H*8), 6.11 (bs, 1H, SO_2N*H*), 4.01 (s, 2H, *H*12), 3.81 (s, 3H, *H*19), 3.75 (s, 3H, *H*1), 3.69 (s, 3H, *H*7), 2.68 (s, 3H, *H*27).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 164.7 (*C*2), 157.7 (*C*18), 137.8 (*C*26), 136.8 (*C*21), 136.5 (*C*6), 134.1 (*C*24), 132.6 (*C*25), 131.6 (*C*22), 130.0 (*C*14), 129.9 (*C*16), 128.5

(*C*8), 128.1 (*C*17), 126.5 (*C*23), 119.8 (*C*4), 109.8 (*C*11), 109.6 (*C*15), 55.7 (*C*19), 52.6 (*C*1), 32.8 (*C*7), 25.3 (*C*12), 20.5 (*C*27).

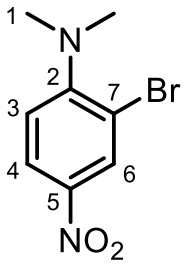
IR: v = 1693 (C=O), 1453 (CH_3 deform.), 1423 (SO_2), 1159 (C-O-C), 865 (=C-H deform.).

LC/MS (Method 1, 254 nm, ESI+): t_R = 9.2 min, m/z: 544.0 [M+Na^+] (38), 539.1 [M+NH_4^+] (21), 522.0 [M+H^+] (94), 351.0 [M-$C_7H_8NO_2S^-$] (28).

HRMS (ESI+): $C_{27}H_{27}N_3NaO_6S^+$ [M+Na^+] calc.: 544.1513 found: 544.1517.

57 (BNH-068) 2-Bromo-*N,N*-dimethyl-4-nitroaniline

N,N-Dimethyl-4-nitroaniline (1.0 g, 6.02 mmol) was dissolved in glacial AcOH (20 mL). Br_2 (310 µL, 6.02 mmol, 1.0 eq.) was added dropwise at RT. The mixture was stirred overnight at RT. The reaction was quenched with sat. aq. NaS_2O_3 (5 mL) and conc. *in vacuo*. Sat. aq. $NaHCO_3$ (100 mL) was added portionwise. The mixture was extracted with EtOAc (3x50 mL). The combined organic extracts were dried (Na_2SO_4) and conc. *in vacuo* to give a yellow solid. The solid was dissolved in MeOH at RT, filtered and recrystallised from hot MeOH. Fine, bright-yellow needles were obtained. The mother liquor was purified chromatographically (NP, PF-15SIHP-F0080, EtOAc/Hex=5/95): The desired product was obtained as bright-yellow needles (1.35 g, 5.50 mmol, 91% yield). The spectroscopic data agrees with the literature.[309]



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 8.45-8.41 (m, 1H, *H*6), 8.15-8.08 (m, 1H, *H*4), 7.06-7.00 (m, 1H, *H*3), 3.01 (s, 6H, *H*1).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 157.0 (*C*2), 141.5 (*C*5), 130.5 (*C*6), 123.9 (*C*4), 118.6 (*C*3), 114.8 (*C*7), 43.5 (*C*1).

IR: v = 1496 (-CH_3 deform.), 1311 (-NO_2), 1111 (C-N).

LC/MS (Method 1, 254 nm, ESI+): t_R = 8.6 min, m/z: 247.0 [M(Br^{81})+H^+] (100), 244.9

[M(Br^{79})+H^+] (94).

HRMS (ESI+): $C_8H_{10}BrN_2O_2^+$ [M+H^+] calc.: 244.9920 found: 244.9919.

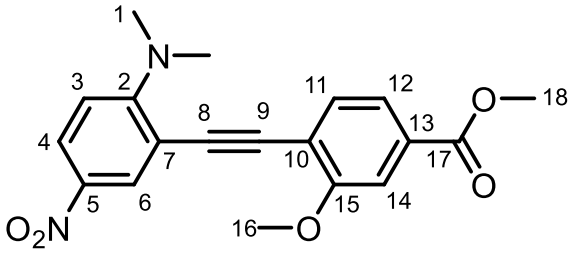
56 (BNH-082) Methyl 4-((2-(dimethylamino)-5-nitrophenyl)ethynyl)-3-methoxybenzoate

Sonogashira Coupling Towards TMS-Alkyne:

Methyl 4-iodo-3-methoxybenzoate (1.0 g, 3.42 mmol) and deg. TEA (716 µL, 5.14 mmol, 1.5 eq.) were dissolved in deg. DMF (10 mL). $PdCl_2(PPh_3)_2$ (120 mg, 171 µmol, 5 mol%) and CuI (33.0 mg, 171 µmol, 5 mol%) were added. The mixture was stirred for 10 min at RT under Ar. TMS-acetylene (617 µL, 4.45 mmol, 1.3 eq.) was added. The mixture was stirred for 2 h at RT under Ar. The reaction was quenched with H_2O (50 mL) and brine (20 mL). The mixture was extracted with EtOAc (3x50 mL). The combined organic extracts were dried (Na_2SO_4) and conc. *in vacuo*. The crude product was purified chromatographically (NP, PF-15SIHP-F0080, EtOAc/Hex=5/95, R_F (EtOAc/Hex=1/9)=0.44: The desired TMS-alkyne was isolated as an orange, amorphous solid (880 mg, 3.36 mmol, 98% yield). This procedure was adapted from the literature.[136]

Sonogashira Coupling Towards Internal Alkyne:

The bromoaniline **57** (573 mg, 2.34 mmol), TMS-alkyne (921 mg, 3.51 mmol, 1.5 eq.) and PPh_3 (61 mg, 234 µmol, 10 mol%) were dissolved in TEA/PEG200/H_2O (18/2/1, 10.5 mL). The solution was Ar sparged for 10 min, while being sonicated. CsF (711 mg, 4.68 mmol, 2.0 eq.), $PdCl_2(PPh_3)_2$ (82.1 mg, 117 µmol, 5 mol%) and CuI (44.6 mg, 234 µmol, 10 mol%) were added. The mixture was stirred overnight at 80°C under Ar. The reaction was quenched with H_2O (50 mL) and extracted with EtOAc (3x50 mL). The combined organic extracts were dried (Na_2SO_4) and conc. *in vacuo*. The crude product was purified chromatographically (NP, PF-15SIHP-F0080, AcOH/DCM/Hex=2/23/75): The desired product was isolated as a yellow, amorphous solid (806 mg, 2.27 mmol, 97% yield and 95% yield over two steps). This procedure was adapted from the literature.[142] The spectroscopic data agrees with the literature.[136]



^{1}H-NMR (400 MHz, $CDCl_3$): δ = 8.41-8.35 (m, 1H, *H*6), 8.09-8.02 (m, 1H, *H*4), 7.67-7.61 (m, 1H, *H*11), 7.59-7.55 (m, 1H, *H*14), 7.55-7.48 (m, 1H, *H*12), 6.91-6.83 (m, 1H, *H*3), 3.96 (s, 3H, *H*18), 3.94 (s, 3H, *H*16), 3.30 (s, 3H, *H*1).

^{13}C-NMR (101 MHz, $CDCl_3$): δ = 166.6 (*C*15), 160.0 (*C*17), 156.9 (*C*2), 139.0 (*C*5), 132.8 (*C*12), 131.8 (*C*6), 131.3 (*C*10), 125.4 (*C*4), 122.0 (*C*11), 117.0 (*C*13), 115.3 (*C*3), 111.4 (*C*14), 110.5 (*C*7), 94.1 (*C*8), 91.7 (*C*9), 56.1 (*C*18), 52.5 (*C*16), 43.0 (*C*1).

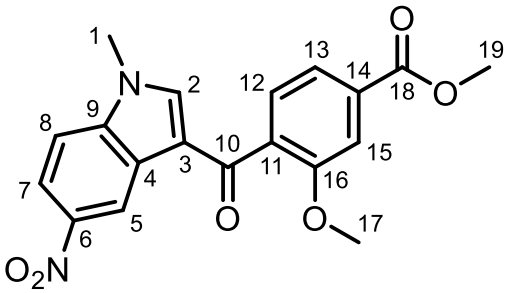
IR: v = 2204 (-C≡C-), 1722 (C=O), 1327 ($-NO_2$).

LC/MS (Method 1, 254 nm, ESI+): t_R = 10.5 min, m/z: 355.0 [$M+H^+$] (100).

HRMS (ESI+): $C_{19}H_{19}N_2O_5^+$ [$M+H^+$] calc.: 355.1288 found: 355.1289.

50 (BNH-081) Methyl 3-methoxy-4-(1-methyl-5-nitro-1*H*-indole-3-carbonyl)benzoate

The alkyne **56** (400 mg, 1.13 mmol) was dissolved in anh. DMSO (10 mL). TBAI (83.3 mg, 226 µmol, 20 mol%) and TBHP (70% in H_2O, 872 µL, 6.77 mmol, 6.0 eq.) were added sequentially. The mixture was stirred overnight at 80°C. The reaction was quenched with H_2O (30 mL). Brine was added (20 mL) and the mixture was extracted with DCM (3x30 mL). The combined organic extracts were dried (Na_2SO_4) and conc. *in vacuo* to give a dark-brown residue. The crude product was purified chromatographically (NP, PF-15SIHP-F0080, EtOH/DCM/Hex=2/28/70, R_F(EtOH/DCM/Hex=2/48/50)=0.28): The desired product was obtained as a pale-yellow solid (194 mg, 526 µmol, 46% yield). This procedure was adapted from the literature.[143] The spectroscopic data agrees with the literature.[136]



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 9.30-9.25 (m, 1H, *H*5), 8.24-8.18 (m, 1H, *H*7), 7.75-7.70 (m, 1H, *H*13), 7.70-7.66 (m, 1H, *H*15), 7.47-7.45 (m, 1H, *H*2), 7.45-7.42 (m, 1H, *H*12), 7.42-7.38 (m, 1H, *H*8), 3.97 (s, 3H, *H*19), 3.87 (s, 3H, *H*1), 3.85 (s, 3H, *H*17).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 188.8 (*C*10), 166.5 (*C*18), 156.8 (*C*16), 144.3 (*C*6), 140.9 (*C*2), 140.5 (*C*9), 134.4 (*C*11), 132.9 (*C*14), 128.7 (*C*12), 126.0 (*C*4), 122.0 (*C*13), 119.8 (*C*5), 119.3 (*C*7), 118.5 (*C*3), 112.7 (*C*15), 110.1 (*C*8), 56.2 (*C*17), 52.6 (*C*19), 34.2 (*C*1).

IR: v = 1723 (C=O), 1290 (-NO_2), 747 (=C-H deform.).

LC/MS (Method 1, 254 nm, ESI+): t_R = 8.3 min, m/z: 369.0 [M+H^+] (100).

HRMS (ESI+): $C_{19}H_{16}N_2NaO_6^+$ [M+Na^+] calc.: 391.0901 found: 391.0902.

3.2.2 Rigid Receptor Docking

This docking procedure was performed on our research group's workstation with Windows installed. Specifications about the used hard- and software are provided in Section 3.3.1. Screenshots of the default settings for each used software package are stored in the supplementary data. Listed below are only those settings, which differ from the default configuration. Molecular structures were built in MarvinSketch and loaded into Maestro. The LigPrep package was used to clean and minimise the structures. The Protein Preparation Wizard was used to import protein crystal structure 6ZL5 *via* its PDB ID. The force field OPLS3 was used in LigPrep and the Protein Preparation Wizard. The Glide package, i.e. receptor grid generation and ligand docking packages, was used for rigid receptor molecular docking. Suitable docking parameters were identified by preparing the crystallised ligand BI-2852 (**7**) with LigPrep and docking it to P1 in 6ZL5 in SP mode. Docking parameters were deemed suitable only if one of the docking poses was nearly identical with the crystallised pose. The optimal settings, which differed from the default configuration are listed in **Table 17**.

Package	Changed Settings
LigPrep	Generate possible states at target pH: 7.0 ± 1.0. Determine chiralities from 3D structure.
ProtPrep	Fill in missing side chains using Prime. Fill in missing loops using Prime. Delete water beyond 4Å from het groups. Generate het states using Epic: pH: 7.0 ± 1.0. Commit all alternate AA positions. Remove water with less than 2 H-bonds to non-waters.
Receptor Grid Generation	Crystallised ligand picked. Rotatable groups: All groups selected.

Table 17: Optimised settings, which differ from the default settings, used for the rigid receptor docking of the biazole and zafirlukast scaffolds to P1 in 6ZL5.

Once the optimal docking parameters were determined, the biazole and zafirlukast series, as well as the references Fesik-Ile (**6**), BI-2852 (**7**) and Ch-3 (**8**), were docked to P1 in 6ZL5 in SP mode. SP settings: “Write out at most: 10 poses per ligand” and “Number of poses per ligand to include: 100”. The output pose with the lowest docking score for each ligand is shown in the Appendix.

3.3 Project 2

3.3.1 General

The *in silico* studies described in Sections 3.3.4 and 3.3.5 were conducted by my colleague Sascha Koller and myself. The results are the outcome of an equally distributed joint effort. All subsequent *in silico* studies were carried out independently by me.

The following hardware was used:

a) Intel® Core™ i9-7960X CPU @2.80 GHz, 64.0 GB RAM, NVIDIA GeForce GT 1030, Windows 10

b) Intel® Core™ i9-7960X CPU @2.80 GHz, 62.5 GB RAM, NVIDIA GeForce GTX 1080, Ubuntu 16.04.5 LTS.

Importantly, molecular modelling results differ between Windows and Linux. Therefore, all *in silico* calculations, except molecular dynamics simulations, were performed in Windows on hardware a). Molecular dynamics simulations were performed in Linux on hardware b).

The following software was used:

MarvinSketch v23.17, KNIME v5.2.3, Anaconda v2.6.2: Two environments were created, namely 'my-rdkit-env' and 'mordred_env'. The RDKit environment had the RDKit v2024.09.4 and Jupyter notebook v7.4.0 installed. The Mordred environment had Mordred v1.2.0 and Jupyter notebook v7.4.0 installed as well as NumPy v1.19.5. The latest numpy version conflicted with the calculation of some descriptors. All molecular modelling tasks were conducted with the Schrödinger Maestro suite v11.5.011, MMshare v4.1.011, release 2018-1.

All KNIME workflows and Jupyter notebooks employed in this thesis can be found on GitHub.[218]

3.3.2 Cheminformatics Software

KNIME

KNIME (Konstanz Information Miner) is a free and open-source software ecosystem designed for data analysis. Its core component, the KNIME Analytics Platform, enables users to build complex data analysis pipelines without requiring programming knowledge.

Data processing in KNIME is done using *nodes*, modular software components with individual graphical user interfaces. Each node performs a specific transformation or operation on the input data and allows users to customize its behaviour through options like checkboxes and dropdown menus.

Nodes are connected via *edges*, which represent the flow of data between processing steps. These connected structures form *workflows*, which visually represent the entire data transformation pipeline.[310]

A basic example workflow is shown in **Figure 59**.

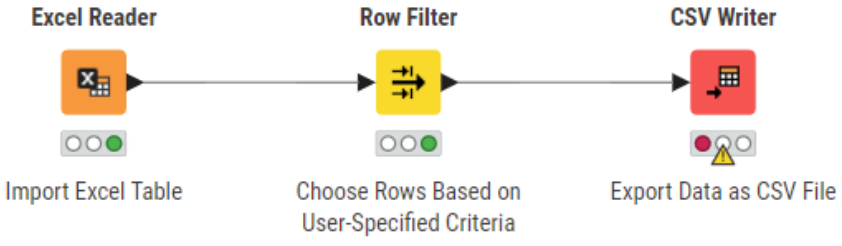


Figure 59: KNIME example workflow.

Here, the “Excel Reader” node imports data from an Excel file. The resulting table is passed to the “Row Filter” node, which selects the first ten rows. These filtered rows are then forwarded to the “CSV Writer” node. Upon execution of the latter, the user obtains a CSV file containing the selected rows with all columns included. Execution status, warnings, and errors are indicated by traffic light icons beneath each node. A vast selection of nodes is available from KNIME itself as well as from third-party providers, including RDKit and Vernalis.[311] Published, automated workflows are available as well.[216,312,313] KNIME has long been a valuable tool for many researchers in the life sciences.[314]

Jupyter Notebook

The Jupyter Notebook project is a framework for interactive computing.[315] A Jupyter Notebook file, commonly referred to as a *Jupyter notebook*, consist of so-called *cells*. Each cell contains code written in a programming language, i.e. Python. Cells can be executed individually and not necessarily in sequential order. Variables remain stored in memory for the whole duration of the Jupyter notebook session. Splitting a complete program into cells enables users to inspect changes in variables and view intermediate outputs. Cells can also include notes to clarify code snippets, and plot graphs directly within them.

The fields of Cheminformatics and Bioinformatics make extensive use of Jupyter notebooks. They allow scientists to build programming pipelines that align with the FAIR principles. An excellent resource for those interested in Computer-Aided Drug Design (CADD) has been created by the Volkamer group at Saarland University.[316] A wealth of CADD expertise has been distilled into nearly a dozen Jupyter notebooks.[317]

RDKit

RDKit is an open-source toolkit for cheminformatics.[318] It enables handling of molecules in Python, including structural representation, 2D/3D molecular operations, descriptor generation, and much more.

Mordred

Mordred is a free molecular descriptor calculation software developed by MORIWAKI *et al*.[227] While many similar free software packages exist, Mordred is superior to most of them. It can compute more than 1800 descriptors with significantly better performance.

Scikit-learn

Scikit-learn is an open-source software machine learning library for Python.[319] Most of the software packages used in the Jupyter notebooks of this project has been imported from Scikit-learn.

3.3.3 Generation of Cyclic Peptide Libraries

In KNIME the “MarvinSketch” node was used to draw 20 proteinogenic α-AAs, namely R, H, K, D, E, S, T, N, Q, C, G, P, A, V, I, L, M, F, Y, and W. Using these inputs, linear and cyclic (click) di-, tri- and tetrapeptides were generated in silico, including all possible permutations of the 20 AAs. The corresponding workflow is illustrated in **Figure 60**.

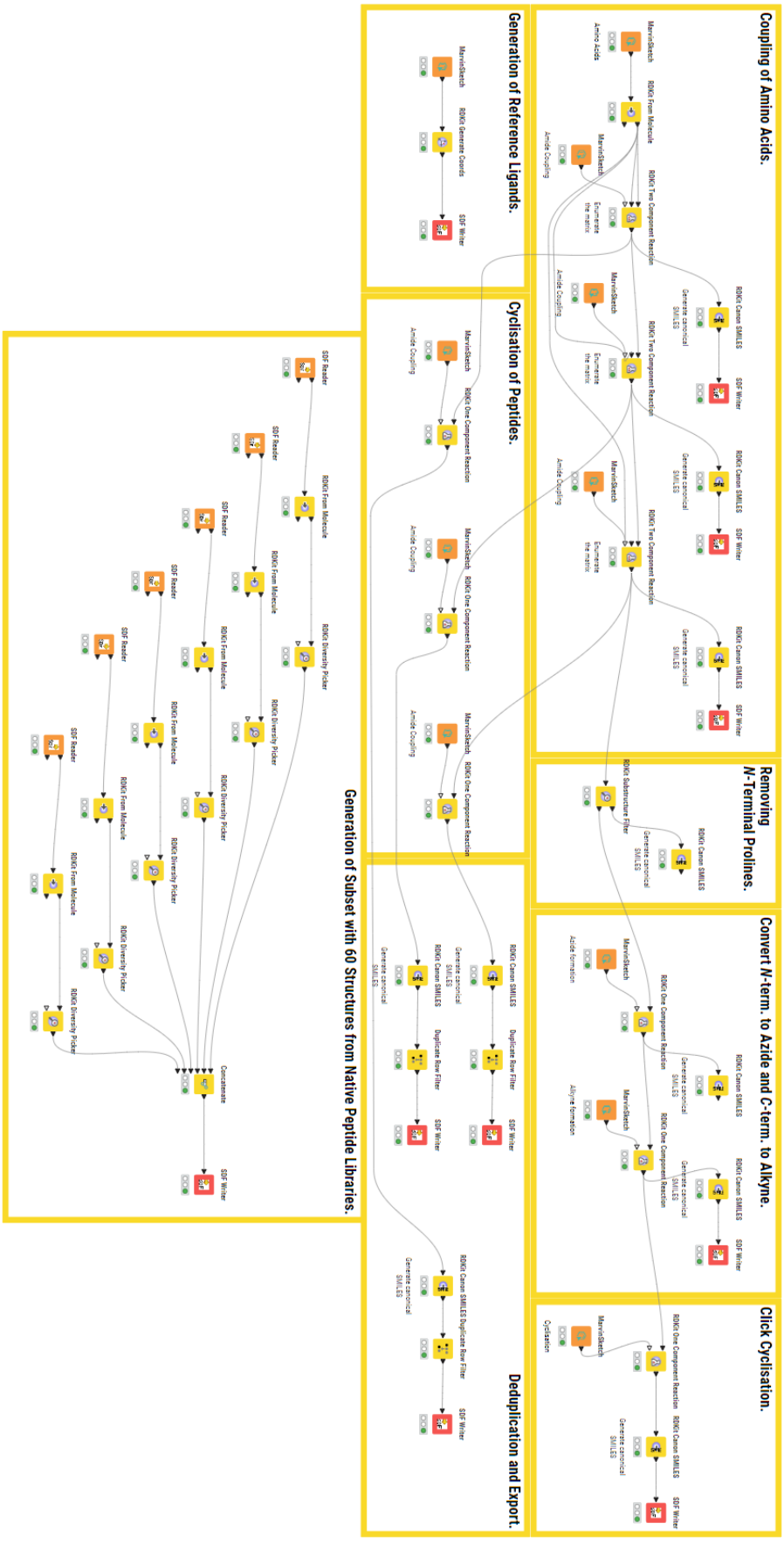
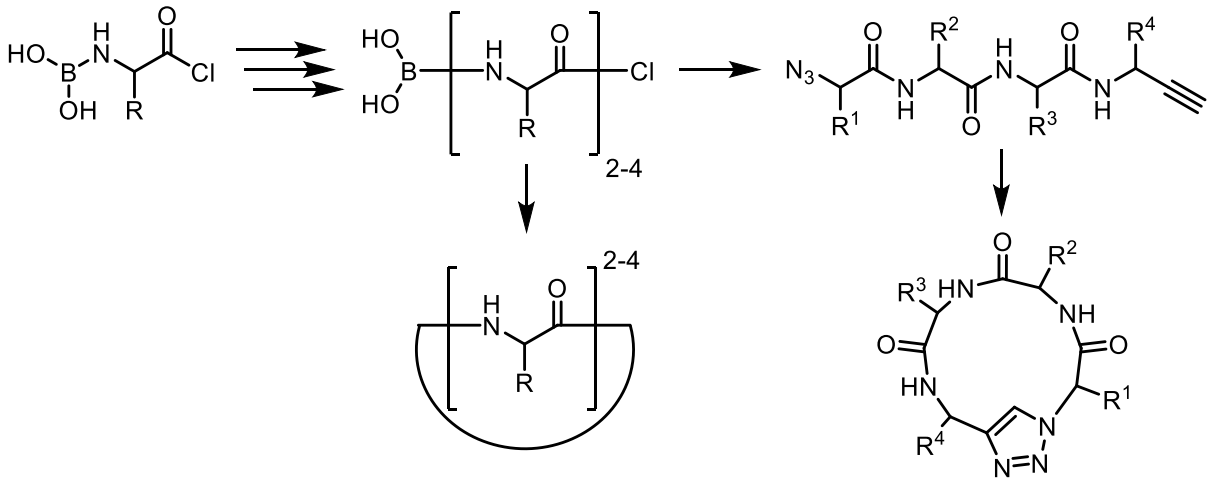


Figure 60: KNIME workflow used for the generation of achiral linear and cyclic (click) di-, tri-, and tetrapeptide libraries.

KNIME must be able to distinguish between the terminal carboxylic acid and amine groups and the functional moieties present in the AA side chains. The former can be modified *in silico* in ways that are not necessarily chemically feasible. In this case, they carboxylic acid chlorides and an aminoboronic acids were used. **Scheme 51** illustrates the conversion of the input AAs through the reaction steps performed *in silico*.



Scheme 51: Schematic representation of the way chemically unique peptide termini lead to selective head-to-tail cyclic peptides in the KNIME workflow shown in **Figure 60**.

Chirality was not defined, as all possible stereoisomers were later generated during ligand preparation in Maestro. Up to three peptide coupling reactions were carried out with the 20 AAs using the “RDKit Two Component Reaction” node, which requires RDKit molecules and a reaction SMARTS string. The latter was created using the “MarvinSketch” node.

Since the depictions of RDKit molecules are not readily interpretable, the “RDKit Canon SMILES” node was used to convert them into canonical SMILES, enabling the display of clean 2D structure images. The resulting structures were exported using the “SDF Writer” node.

The obtained linear di-, tri-, and tetrapeptides were then cyclized intramolecularly using the “RDKit One Component Reaction” node. Cyclization can result in multiple linear sequences forming the same cyclic peptide. For example, GAAA, AGAA, AAGA, and AAAG all yield the same cyclic structure upon cyclization. Such duplicates were removed using the “Duplicate Row Filter” node, which requires canonicalized SMILES as input to function correctly in this context.

A similar workflow was applied to generate click cyclic tetrapeptides. Starting from the linear tetrapeptides, *N*-terminal prolines were filtered out using the "RDKit Substructure Filter" node, as secondary amines are not easily converted to azides in the laboratory, which are essential for click coupling. The *N*-terminal aminoboronic acids were converted into azides using the "RDKit One Component Reaction" node, while the *C*-terminal acid chlorides were converted into terminal alkynes *via* the same node. These modified peptides were then cyclized intramolecularly to yield the click cyclic tetrapeptide library.

To limit computational time during test docking runs, a subset of 60 amino acids was selected using the "RDKit Diversity Picker" node. This node selects a subset of structurally diverse molecules based on the Tanimoto distance between molecular fingerprints.[320] From each of the six sets of native peptides (linear and cyclic di-, tri-, and tetrapeptides) a set of ten peptides was selected. Additionally, the reference structures depicted in **Figure 19** were drawn within the KNIME workflow and exported as a separate SDF file.

Experience has shown that handling molecules *in silico* is prone to error. Therefore, verification of the number of structures in the peptide libraries was essential. Linear peptide sequences have a distinct start (*N*-terminus) and end (*C*-terminus). Each permutation is unique. The number of possible permutations N of 20 AAs in a linear peptide of length n is calculated with **Equation 8**.

$$N_{20}(n) = 20^n$$

Equation 8: Calculation of all possible peptide sequences with 20 AAs in a linear peptide of length n.

Thus, the linear di-, tri- and tetrapeptide libraries contain 400, 8,000 and 160,000 unique structures, respectively.

In contrast, native cyclic peptides can be conceptualised as a necklace of coloured beads, where each colour represents one of the 20 AAs. The number of unique permutations N for a cyclic peptide of length n with 20 available AAs can be calculated with **Equation 9**.

$$N_{20}(n) = \frac{1}{n}\sum_{i=1}^{n} 20^{\gcd(i,n)}$$

Equation 9: Calculation of all possible peptide sequences with 20 AAs in a native cyclic peptide of length n.

Here, $gcd(i,n)$ denotes the greatest common divisor of the integers i and n.[321] Applying **Equation 9** yields 210, 2,680 and 40,110 unique structures for the native cyclic di-, tri- and tetrapeptide libraries, respectively.

For the click cyclic peptide library, no duplicate structures result from the cyclisation step due to the lack of rotational symmetry in their peptide backbone. The only constraint is that that proline cannot be present at the *N*-terminus, as its secondary amino group cannot be converted to an azide. Therefore, the 8,000 sequences containing *N*-terminal proline are excluded from the click cyctetpep library. As a result, both the linear and cyclic click tetrapeptide libraries both contain 152,000 unique structures.

3.3.4 Data Preparation

The nucleotide exchange assay values were selected for the following *in silico* studies. The assays conducted by LDC produce reliable results only in specific concentration ranges.[322] IC_{50} values falling outside of this range cannot be determined precisely. LDC denoted such measurements using expressions like “>3000” or “≤ 50 µM”. These range-based values and obvious outliers were excluded from further analysis. No such edits were made to the other libraries.

Standardisation of all libraries was performed in KNIME. The standardisation pipeline of the project library, shown in **Figure 61**, serves as a representative example. To improve visibility, the workflow is presented in four different sections (from top to bottom). The pipeline should be read from left to right, top to bottom. This workflow was inspired by the standardisation pipeline developed by FALCÓN-CANO *et al*.[217]

Since both the project library and the ChEMBL KRAS extension contained assay values. More preparatory steps were necessary, in comparison to the other libraries. After importing the data, the SMILES column was identified using the “Molecule Type Cast” node. The required columns were selected, atom valences were corrected, and inorganics removed. Any structures containing elements other than B, C, N, H, O, P, S, F, Cl, Br, or I were discarded.

The structured were then desalted and kekulized. Stereochemistry was removed and formal charges were standardised. The structures were normalised and aromatised. Subsequently, InChI codes and canonical SMILES were generated. Any entries with

missing InChIs and SMILES, as well as any duplicate entries, were removed. Compounds with missing assay values were also excluded.

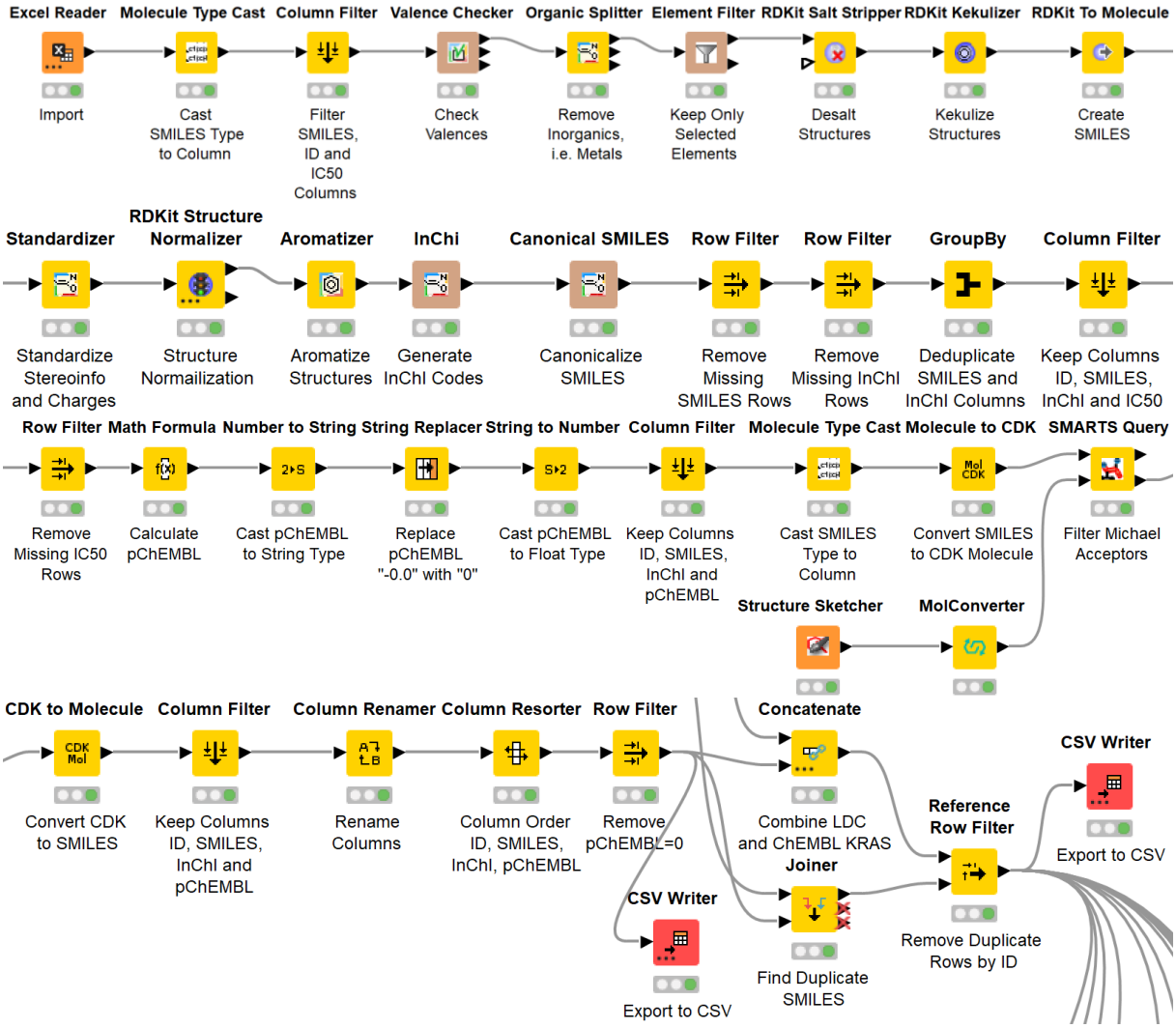


Figure 61: Data preparation and structure washing workflow.

The “Math Formula” node was used to convert IC_{50} values (nM) into dimensionless pChEMBL values using **Equation 3**. Ligands targeting $KRAS^{G12C}$ were removed using the “SMARTS Query” node, with a Michael acceptor substructure defined using the “Structure Sketcher” node. The table was then cleaned, and all rows with pChEMBL values equal to zero were deleted. The resulting table was exported and also concatenated, without duplicates, with the ChEMBL KRAS extension, which had been prepared in an analogous manner beforehand. This merged dataset was subsequently exported and subtracted from all other libraries using the “Substructure Filter” node. The other libraries were standardised through similar, albeit shorter pipelines.

3.3.5 Diversity Analysis

Molecular Descriptor Histogram Generation

The workflow snippet shown in **Figure 62** illustrates the visualisation of value distributions for key molecular descriptors, namely MW, SlogP, TPSA, and the numbers of rotatable bonds, HBDs, HBAs and rings. The pipeline should be read from left to right.

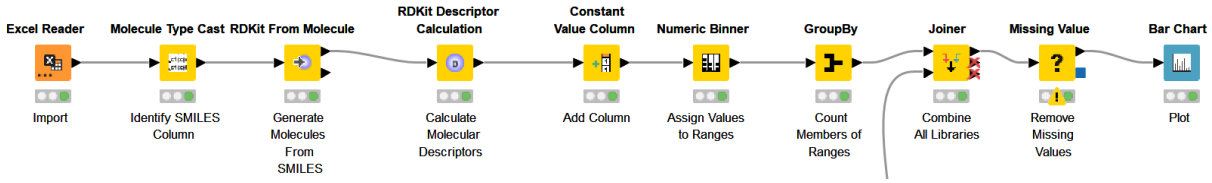


Figure 62: Molecular descriptor histogram generation workflow.

Initially, the library was imported, and the SMILES column was identified and converted to RDKit molecule objects. The selected molecular descriptors were then calculated and sorted into value ranges, so-called *bins*. The number of molecules in each bin was counted. This procedure was repeated for all libraries under investigation. Subsequently, the resulting descriptor tables were joined, missing values were removed, and the final aggregated table was used to generate the histograms.

RO5 and RO4 Compliance Plot Generation

Figure 63 shows a representative snippet of the workflow used to visualise the proportion of molecules in each library that comply with the RO5 and RO4 criteria. The pipeline should be read from left to right and top to bottom.

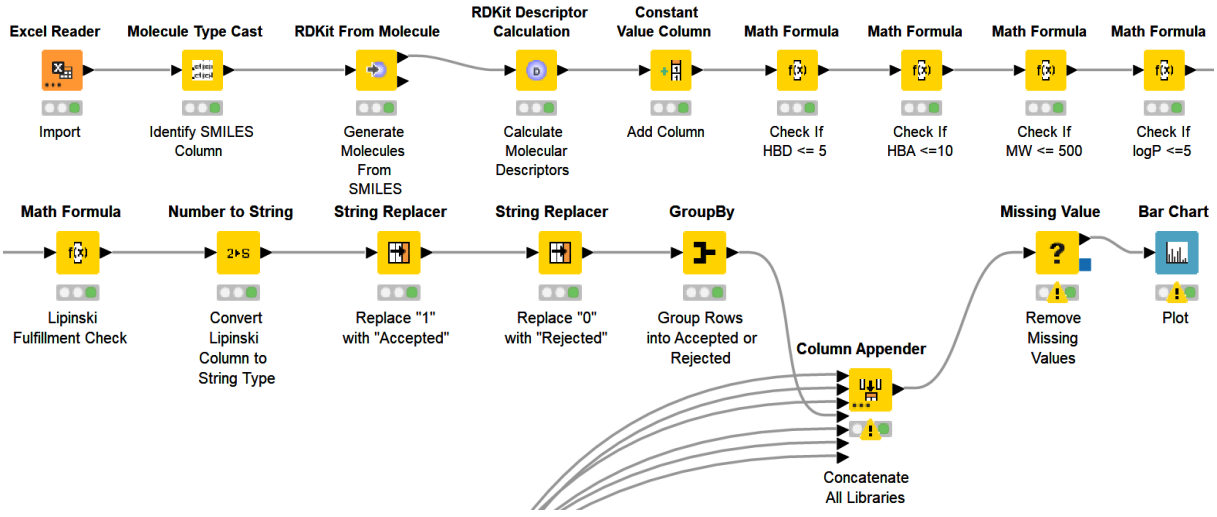


Figure 63: RO5 and RO4 compliance bar plot generation workflow.

Molecular descriptors were computed as detailed for the workflow shown in **Figure 62**. Each “Math Formula” node assessed whether a given descriptor value fell within the relevant threshold, appending a Boolean column with values of “1” for compliance and “0” for violation. A final “Math Formula” node combined these Booleans to determine overall RO5 compliance, assigning each molecule a status of “Accepted” or “Rejected.” Subsequently, the relative proportions of accepted and rejected molecules were calculated. The resulting tables were merged, any entries with missing values were excluded, and the final data were visualised as bar plots.

PMI Plot Generation

The workflow snippet shown in **Figure 64** demonstrates the procedure used to visualise the principal moments of inertia (PMIs) for all molecules within a given library. The pipeline should be read from left to right and top to bottom.

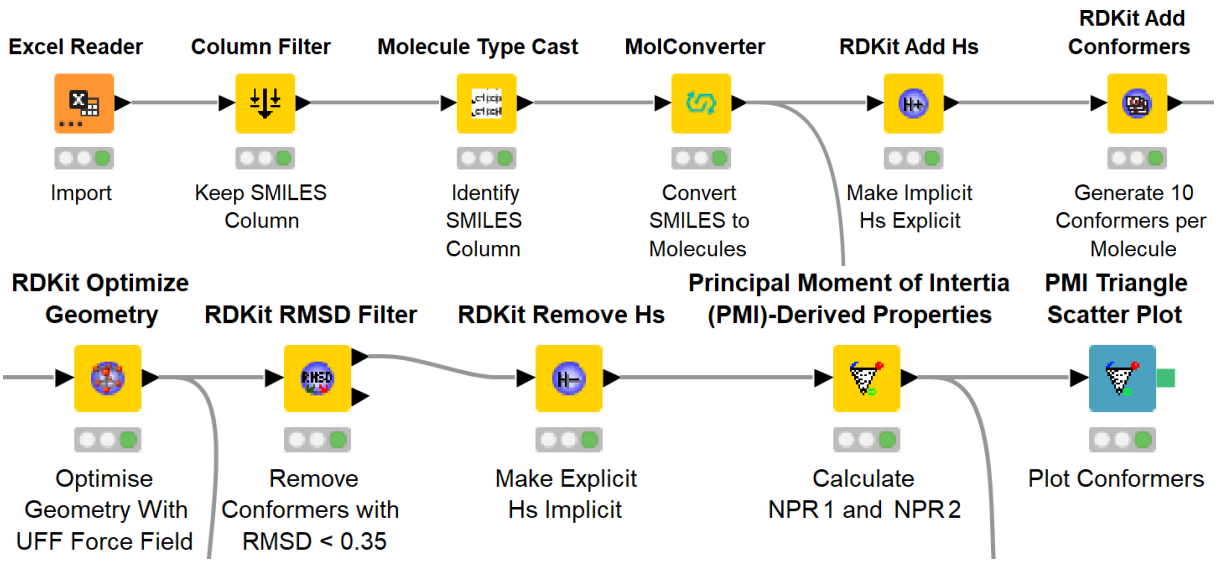


Figure 64: PMI plot generation workflow.

Molecular structures were imported from SMILES and converted as described previously. The “RDKit Add Hs” node was used to add explicit hydrogens, a necessary step for accurate 3D conformer generation. Ten conformers were generated per molecule, and their geometries optimised using the Universal Force Field (UFF), which was the only force field that consistently completed without errors. Redundant conformers were removed (defined as those with an RMSD below 0.35 relative to the initial conformer). Explicit hydrogens were then reverted to implicit form, and the normalised principal moment of inertia ratios (NPR_1 and NPR_2) were computed. The resulting data points were plotted on a PMI triangle scatter plot. This procedure was

repeated for each library and selected reference structures, namely MRTX-1133, BI-2852, BAY-293, and Ch-3.

3.3.6 Python QSAR

The QSAR model developed in this project builds upon the protocol published by DUO *et al*. in 2024.[223] The original source code was adapted to predict the inhibitory potency of molecules against the KRAS nucleotide exchange assay based on their molecular structure. The initial step focussed on selecting molecular descriptors. Using Mordred, all available 2D molecular descriptors were calculated for the combined project-ChEMBL library. Feature with missing or non-numerical values were excluded. From this processed set, the 20 best performing descriptors were selected; these are listed in **Table 8**.

In parallel, ECFP fingerprints (radius = 3, length = 512 bits) were generated for all molecules the project-ChEMBL library and the resulting data set was divided into training and test sets using an 80:20 split. A five-fold cross-validation hyperparameter search was conducted on the training set, with the examined hyperparameters summarised in **Table 18**.

Subsequently, regression models were trained on the training subset, and their performance was evaluated by predicting pChEMBL values for the test set. Performance metrics, including R^2, RMSE, and standard deviation, are reported in **Table 9**. The best-performing model, based on the random forest regressor algorithm, demonstrated satisfactory fit and predictive performance. This model was then retrained on the entire project-ChEMBL dataset following a second five-fold cross-validation hyperparameter optimisation. The final model was used to predict pChEMBL values for all libraries.

Algorithm	Tested Parameters
Decision Tree	criterion: squared_error, friedman_mse; min_samples_split: 2,3,4,5,6,7,8,9;
Ada Boost	learning_rate: 0.001, 0.01, 0.1, 1; loss: linear, square, exponential;
Elastic Net	alpha: 0.001, 0.01, 0.1, 1, 10; l1_ratio: 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8; max_iter: 100000;
Lasso	alpha: 0.001, 0.01, 0.1, 1, 10; selection: cyclic, random; max_iter: 100000;
Ridge	alpha: 0.001, 0.01, 0.1, 1, 10; solver: auto; max_iter: 100000;
Extra Trees	bootstrap: True, False; min_samples_split: 2,3,4,5,6,7,8,9;
Gradient Boosting	learning_rate: 0.001, 0.01, 0.1, 1; min_samples_split: 2, 3, 4, 5, 6, 7, 8, 9; loss: squared_error, absolute_error, huber, quantile; criterion: squared_error, friedman_mse;
K-Neighbours	n_neighbors: 2, 3, 4, 5, 6, 7, 8, 9, 10; weights: uniform, distance; algorithm: auto, ball_tree, kd_tree, brute; leaf_size: 20, 30, 40; p: 1, 2;
SVR	kernel: rbf, linear, poly, sigmoid; gamma: scale, auto; C: 0.001, 0.01, 0.1, 1, 10, 20, 50, 100
Random Forest	bootstrap: True, False; max_features: auto, log2, sqrt; min_samples_split: 2,3,4,5,6,7,8,9;

Table 18: Tested hyperparameters during the QSAR model construction. Best parameters for each model are highlighted in red.

3.3.7 Rigid Receptor Docking of Cyctetpep Library

The general docking procedure followed a protocol analogous to that described in Section 3.2.2. The force field OPLS-2005 was used instead. To determine appropriate docking parameters, the crystallised ligands were first prepared using LigPrep and subsequently docked into their respective binding sites within the original crystal structures, employing Glide in XP mode. Docking parameters were considered suitable only if at least one generated pose closely resembled the experimentally determined crystallographic pose. **Table 19** summarises the optimised settings that deviated from the default configuration.

Package	PDB ID	Changed Settings
LigPrep	6GJ8 and 6GQY	Crystallised ligands: Generate possible states at target pH: 7.0 ± 1.0. Determine Chiralities from 3D structure. Cyclic peptides: Generate possible states at target pH: 7.0 ± 1.0. Generate all combinations.
	7RPZ	Crystallised ligands: Determine Chiralities from 3D structure. Cyclic peptides: Generate all combinations.
ProtPrep	6GJ8, 6GQY and 7RPZ	In the case of 6GQY delete all chains except chain A. Fill in missing side chains using Prime. Fill in missing loops using Prime. Delete water beyond 4Å from het groups. Generate het states using Epic: pH: 7.0 ± 1.0. Commit all alternate AA positions. Remove water with less than 2 H-bonds to non-waters.
Receptor Grid Generation	6GJ8 and 6GQY	Crystallised ligand picked. Rotatable groups: All groups selected.
	7RPZ	Crystallised ligand picked. Rotatable groups: All groups selected. Carboxylate oxygen atoms of Asp12 picked as receptor atoms in 'H-bond/Metal' tab. 'Use Symmetry' box was ticked.
Ligand Docking	7RPZ	Constraints: Name: A: Asp12: OD2 (hbond). Receptor Constraint Type: H-bond. Ligand Feature: Donor including aromatic H + halogens (11 patterns).

Table 19: Optimised settings, which differ from the default settings, used for the rigid receptor docking of the cyctetpep library to P1 in 6GJ8 and 6GQY, as well as to P2 in 7RPZ.

Following optimisation of the docking parameters, the cyclic peptide libraries were docked against all three protein structures. Each docking was performed in three sequential precision modes: HTVS, SP, and XP. Crystallised ligands and well-characterised KRAS SOS PPIIs were included in each run as reference compounds. For HTVS and SP docking, the following settings were applied: "Write out at most: 1 pose per ligand" and "Number of poses per ligand to include: 10". For XP docking, the settings were adjusted to "Write out at most: 10 poses per ligand" and "Number of poses per ligand to include: 100". From the HTVS run, the 10,000 ligands with the most favourable docking scores were selected for subsequent SP docking. The top 1,000

ligands from the SP run were then advanced to the XP docking stage. All final XP-mode output structures are provided in the supplementary data.

3.3.8 Molecular Dynamics Simulations

Molecular dynamics simulations were conducted with PDB entries 7RPZ and 6GQY, using their respective crystallised ligands, MRTX-1133 and Ch-3, as well as the top-scoring docking poses of BNH-166 and BNH-177 identified through the screening pipeline outlined in Section 3.3.7. Protein preparation followed the procedure described in that section, while the ligands were used without further modification. All simulations employed the OPLS3 force field. **Table 20** summarises the optimised simulation parameters that differed from the default settings.

Package	PDB ID	Changed Settings
System Builder	7RPZ and 6GQY	Solvation: Minimize Volume. Show boundary box. Use Custom charges + Do not use. Ions: Recalculate → Neutralize by adding 9 Na+ ions. Add salt (NaCl, 0.15 M).
Molecular Dynamics		Simulation time (ns): total 500. Recording interval (ps): trajectory 200. Approximate number of frames: 2500.

Table 20: Optimised settings, which differ from the default settings, used for the molecular dynamics simulations with Ch-3 and BNH-166 at P1 in 6GQY, as well as with MRTX-1133 and BNH-177 at P2 in 7RPZ.

Protein-ligand interactions were analysed using the “Simulation Interaction Diagram” package. To assess binding stability and energetics, extensive MMGBSA (Molecular Mechanics Generalized Born Surface Area) calculations were performed across all frames of the simulation trajectories using the thermal_mmgbsa.py script. The command used was: $SCHRODINGER/run thermal_mmgbsa.py xx-out.cms -HOST ‘localhost:16’ -step_size 1 -NJOBS 16 -lig_asl “ASL”. The appropriate ligand ASL (atom selection language) expression was determined *via* the GUI of the “Simulation Interaction Diagram” package. For instance, the ASL for MRTX-1133 in the 7RPZ structure was identified as “mol.num 3”.

3.3.9 Syntheses

Procedure A: SPPS Towards Native Tri- and Tetrapeptides

Kaiser Test:

Three stock solution were prepared: a) KCN (1.63 mg, 253 µmol) was dissolved in H_2O (25.0 mL). An aliquot of this aq. KCN (0.001 M, 1 mL) was added to pyridine (49.0 mL) (freshly distilled from ninhydrin). The solution was mixed well. b) ninhydrin (1.00 g, 5.61 mmol) was dissolved in EtOH (20.0 mL). and c) phenol (16.0 g, 170 mmol) was dissolved in EtOH (20.0 mL). Three to six resin pearls were placed in an empty mass spectrometry vial (2 mL). Two drops of stock solution a, b and c were added. The mixture was heated for 5 min at 110°C. In the presence of primary amines, the mixture will take on a characteristic dark-blue colour. Faint hues of violet, red, yellow, green or brown were observed as well, but should not be interpreted as indicating the presence of primary amines. Secondary amines such as in proline cannot be identified with this test. This procedure was adapted from the literature.[323]

Activation of CTC resin

CTC resin with a maximum loading of 1.55mmol/g was used. All SPPS yields were calculated relative to maximum loading. The resin (5.00 g, 7.75 mmol) was suspended in anh. DCM (50 mL). $SOCl_2$ (2.81 mL, 38.8 mmol, 5.0 eq.) and anh. DMF (141 µ, 5% rel. to $SOCl_2$) were added. The suspension was refluxed for 4 h under N_2 and left to cool to RT. The suspension was filtered. The filter cake was washed with anh. DMF (3x10 mL) and anh. DCM (3x10 mL). The filter cake was dried *in vacuo* overnight. The activated resin was stored at 4°C for max. 1 month. This procedure was adapted from the literature.[249]

SPPS protocol on CTC resin

Fmoc-SPPS was employed. If not specified otherwise, 10 mL of solvent per gram of resin was used. Equivalents of reagents were calculated respective to the theoretical max. loading of the resin. The first AA (2 eq.) was loaded by shaking with CTC resin and anh. DIPEA (10 eq.) in anh. deg. DMF/anh. DCM (4/1) overnight at RT. The resin was washed with anh. DMF (2x) and shaken with anh. DCM/anh. MeOH/anh. DIPEA (16/3/1) for 10 min (2x). The resin was washed with DMF (3x) and DCM (3x) and dried in vacuum overnight. Fmoc removal was carried out by shaking the resin in piperidine/DMF (1/4) for 10 min at RT (2x). The resin was washed with DMF (3x) and

DCM (3x). Further AAs (2.5 eq.) were coupled with PyBOP (2.5 eq.) and DIPEA (5 eq.) in anh. deg. DMF/anh. DCM (1/1) for 3 h at RT. The resin was washed with DMF (3x) and DCM (3x). Completion of each coupling step was verified with a Kaiser test. After complete coupling, the Fmoc group was removed as described above. After the final Fmoc removal, the resin was washed with DCM (5x), iPrOH (2x), Hex (2x), DCM (2x), MeOH (2x) and DCM (2x). The peptide was cleaved off the resin by shaking in HFIP/DCM (1/7, 6 mL/g resin, 10x5 min). The combined filtrates were conc. under a stream of compressed air and finally dried *in vacuo* at RT(!). The obtained residue was dissolved in MTBE and conc. *in vacuo* at RT(!) (2x).

Procedure B: Reduction-Homologation Sequence

DIBAL-H Reduction:

Boc-/Fmoc-AA (1.0 eq.) was dissolved in anh. DCM. The solution was cooled to 0°C under N_2. Freshly recrystallised CDI (1.1 eq.) was added. The mixture was stirred for 60 min and cooled to -94°C. DIBAL-H (1.2 M in toluene, 2.1 eq.) was added dropwise over 110 min using a syringe pump. The mixture was stirred for 30 min. The reaction was quenched with EtOAc. The cooling bath was removed. Aq. tartaric acid (25%) was added, while stirring vigorously. The mixture was warmed to RT with a H_2O bath, while stirring vigorously for 15 min. Once complete dissolution of the aluminium salts was achieved, the layers were separated, and the aq. phase was extracted with EtOAc (2x). The combined organic extracts were washed with aq. HCl (1 M) and sat. $NaHCO_3$. The extracts were dried (Na_2SO_4) and concentrated *in vacuo*. The crude product was used immediately (!) without further purification. This procedure was adapted from the literature.[273]

Ohira-Bestmann Homologation:

The following equivalents are given relative to the aldehyde. K_2CO_3 (3.0 eq.) was flame dried in a Schlenk flask. Toluenesulfonyl azide (10% in toluene, 1.2 eq.) and dimethyl-(2-oxopropyl)-phosphonate (1.2 eq.) were added sequentially. The mixture was stirred for 5 h at RT under N_2. A solution of the crude aldehyde in anh. MeOH was added to the off-white, opaque mixture. The reaction mixture changed colour rapidly to pale-yellow and turned opaque shortly after. The mixture was stirred overnight at RT under N_2 and was conc. *in vacuo*. The obtained residue was dissolved in EtOAc, washed with

H_2O (2x), dried (Na_2SO_4), and conc. *in vacuo* to give a brown residue. This procedure was adapted from the literature.[274]

Procedure C: Diazotransfer Reaction

NaN_3 (1.45 eq.) was suspended in anh. MeCN (1 M relative to NaN_3), and the mixture was cooled to 0°C. Triflyl anhydride (1.2 eq.) was added dropwise, and the mixture was stirred for 2 h at 0°C under N_2, and further 30 min at RT under N_2. $CuSO_4$*5 H_2O (1 mol%), the AA (1.0 eq.), and TEA (3.0 eq.) were combined with MeCN/H_2O (2/1). The suspension turned clear upon addition of the base. The mixture was cooled to 0°C. The previously prepared triflyl azide solution was filtered and added dropwise at 0°C. The mixture turned from pale-blue to yellow to red to green. The mixture was stirred 30 min at 0°C. Cooling was removed and the mixture was stirred overnight at RT under N_2. The organic solvent was removed under a stream of compressed air. The aq. phase was diluted with H_2O and extracted with EtOAc (3x). The aq. phase was acidified to pH=1 with conc. HCl. The resulting mixture was extracted with EtOAc (3x). The combined organic extracts were dried (Na_2SO_4) and conc. *in vacuo*. Chromatographic purification provided the desired α-azido acid. This procedure was adapted from the literature.[269]

Procedure D: Synthesis of Click Cyclic Tetrapeptides

<u>Linear Pseudo Tetrapeptide:</u>

The azido tripeptides were synthesised following a modified SPPS procedure. The first two AAs were installed according to general procedure A. The α-azido acid (2.5 eq.) and HOBt (2.5 eq.) were dissolved in anh. deg. DMF and DIC (2.5 eq.) was added. The mixture was stirred for 15 min at RT, whereupon precipitate formed. The mixture was added to the resin and was shaken for 4 h at RT. This coupling step was repeated once. The resin was washed with DMF (3x) and DCM (3x). Completion of the coupling step was verified with a Kaiser test. The azido tripeptide was cleaved off the resin as described in general procedure A. The azido tripeptide (1.0 eq.), the α-amino acetylene (1.1 eq.), EDCI (1.2 eq.), HOBt*H_2O (1.1 eq.), and DIPEA (1.0 eq. for free base α-amino acetylene and 2.0 eq. for respective HCl salt) were dissolved in anh. MeCN and stirred overnight at RT under N_2. The mixture was conc. *in vacuo* and dissolved in EtOAc. The solution was washed with H_2O, sat. $NaHCO_3$ (20 mL), and $KHSO_4$ (1 M). The solution was dried (Na_2SO_4) and conc. *in vacuo*.

Cyclisation:

No more than 100 mg of pseudo-tetrapeptide were cyclised at a time. More compound would have overloaded our HPLC column after deprotection. The pseudo-tetrapeptide (100 mg, 1.0 eq.) was dissolved in MeCN (0.2 mM) and the solution was He sparged for 7 min and Ar sparged for 8 min. DIPEA (2.0 eq.), 2,6-lutidine (2.0 eq.), TBTA (2.0 eq.), and CuI (2.0 eq.) were added sequentially. The mixture was stirred for the indicated time at RT under Ar. The mixture was conc. *in vacuo* at 40°C. This cyclisation procedure was adapted from the literature.[293]

Deprotection:

The crude residue was stirred in TFA/TIPS/H_2O (95/2.5/2.5, 10 mL) for 30 min at RT. The mixture was conc. under a stream of compressed N_2. The residue was dissolved in MeCN/H_2O/DMSO (1/1/2, 2 mL) and purified chromatographically.

Procedure E: N_3-AA-Ala-OMe Synthesis

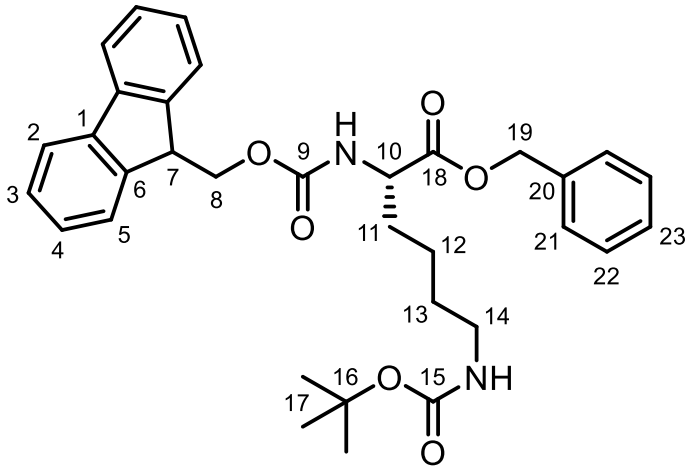
The α-azido acid (1.0 eq.) and HOBt*H_2O (1.2 eq.) were dissolved in anh. DMF at 0°C. DIC (1.2 eq.) was added, and the mixture was stirred for 20 min at 0°C. A solution of H_2N-Ala-OMe*HCl (1.5 eq.) and DIPEA (1.4 eq.) in anh. DMF was added. The mixture was stirred for 20 min at 0°C. Cooling was removed and the mixture was stirred overnight at RT. The mixture was conc. in high vacuum overnight at RT. Chromatographic purification provided the desired azido dipeptides.

Procedure F: Mosher Amide Synthesis

(*R*)- or (*S*)-α-Methoxy-α-trifluoromethylphenylacetic acid (3.0 eq.), HOBt*H_2O (3.0 eq.) and DIPEA (3.0 eq.) were dissolved in anh. MeCN at 0°C. The solution was stirred for 5 min and EDCI (3.0 eq.) was added. The solution was stirred for 20 min at 0°C. The α-amino alkyne (1.0 eq.) and DIPEA (1.0 eq.) were dissolved in anh. MeCN. The alkyne solution was added to the precooled solution. The mixture was stirred for 10 min at 0°C. Cooling was removed and stirring was continued overnight at RT. The mixture was conc. *in vacuo* and the obtained residue was dissolved in EtOAc. The solution was washed with H_2O, sat. $NaHCO_3$, and $KHSO_4$ (1 M). The solution was dried (Na_2SO_4) and conc. *in vacuo*. Chromatographic purification provided the desired Mosher amides.

99 (BNH-096) Benzyl N^2-(((9*H*-fluoren-9-yl)methoxy)carbonyl)-N^6-(*tert*-butoxycarbonyl)-L-lysinate

(*S*)-Fmoc-Lys(Boc)-OH (1.0 g, 2.13 mmol) was dissolved in anh. DMF (10 mL). K_2CO_3 (354 mg, 2.56 mmol, 1.2 eq.) and benzyl bromide (305 µL, 2.56 mmol, 1.2 eq.) were added, and the mixture was stirred overnight at RT. The reaction was quenched with H_2O (50 mL) and extracted with EtOAc (3x50 mL). The combined organic extracts were washed with H_2O (2x50 mL) and brine (50 mL). The solution was dried (Na_2SO_4) and conc. *in vacuo* to give a colourless oil. The crude product was purified chromatographically (NP, PF-15SIHP-F0080, EtOAc/Hex=2/8): The desired product was obtained as a white, amorphous solid (1.16 g, 2.07 mmol, 97% yield). This procedure was adapted from the literature.[324] The spectroscopic data agrees with the literature.



^{1}H-NMR (400 MHz, $CDCl_3$): δ = 7.80-7.73 (m, 2H, *H*2), 7.64-7.57 (m, 2H, *H*5), 7.44-7.27 (m, 9H, *H*3+4+21-23), 5.55-5.41 (m, 1H, α-N*H*), 5.22 (d, $^3J_{19ab}$=12.2 Hz, 1H, *H*19a), 5.15 (d, $^3J_{19ab}$=12.2 Hz, 1H, *H*19b), 4.55 (bs, 1H, ε-N*H*), 4.49-4.32 (m, 3H, *H*8+10), 4.22 (t, $^3J_{10\text{-}11}$=7.0 Hz, 1H, *H*7), 3.13-2.98 (m, 2H, *H*14), 1.94-1.80 (m, 1H, *H*11a), 1.79-1.62 (m, 1H, *H*11b), 1.44 (s, 9H, *H*17), 1.40-1.25 (m, 4H, *H*12+13).

^{13}C-NMR (101 MHz, $CDCl_3$): δ = 172.4 (*C*18), 156.2, 156.1 (*C*9+15), 143.8 (*C*6), 141.4 (*C*2), 135.4 (*C*20), 128.7 (*C*21), 128.6 (*C*22), 128.4 (*C*23), 127.8 (*C*3), 127.2 (*C*4), 125.2 (*C*5), 120.1 (*C*2), 79.3 (*C*16), 67.3, 67.1 (*C*8+19), 53.9 (*C*10), 47.2 (*C*7), 40.1 (*C*14), 32.2 (*C*11), 29.7 (*C*12), 28.5 (*C*17), 22.4 (*C*13).

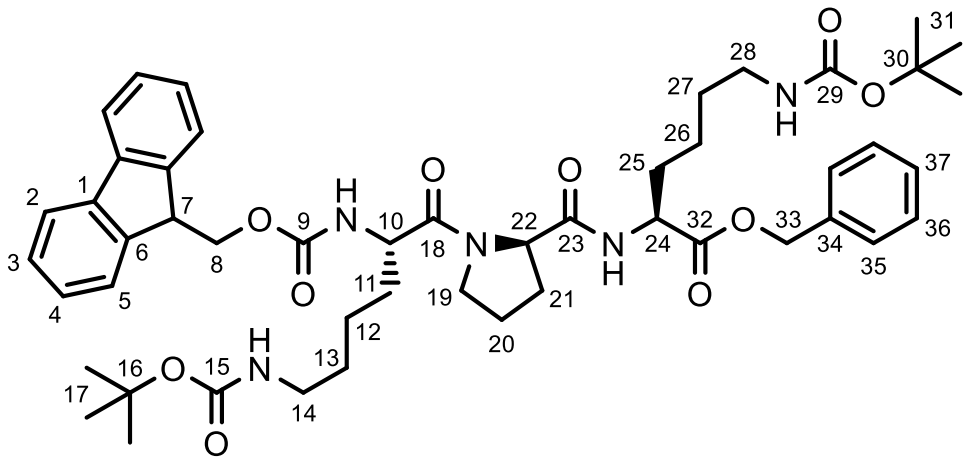
IR: v = 3350 (-N-H), 1684 (C=O), 1521 (-N-H deform.), 732 (=C-H deform.).

LC/MS (Method 1, 254 nm, ESI+): t_R = 11.3 min, m/z: 581.2 [M+Na^+] (52), 559.2 [M+H^+] (14), 503.2 [M-$C_4H_7^-$] (18), 459.2 [M-$C_5H_7O_2^-$] (100).

HRMS (ESI+): $C_{33}H_{38}N_2NaO_6^+$ [M+Na^+] calc.: 581.2622 found: 581.2621.

100 (BNH-097) Benzyl *N*^{2}-*N*^{2}-(((9*H*-fluoren-9-yl)methoxy)carbonyl)-*N*^{6}-(*tert*-butoxycarbonyl)-L-lysyl-D-prolyl-*N*^{6}-(*tert*-butoxycarbonyl)-L-lysinate

The protected lysine **99** (788 mg, 1.41 mmol) was dissolved in piperidine/DMF (1/4, 10 mL). The solution was stirred for 3 h at RT and was conc. in high vacuum overnight at RT. (*S*)-Fmoc-Pro-OH (571 mg, 1.69 mmol, 1.2 eq.), HATU (643 mg, 1.69 mmol, 1.2 eq.) and DIPEA (732 µL, 4.23 mmol, 3.0 eq.) were dissolved in anh. deg. DMF (5 mL). The mixture was stirred for 30 min at 0°C. A solution of (*S*)-H_2N-Lys(Boc)-OBn in anh. deg. DMF (5 mL) was added and the mixture was stirred overnight at RT. The reaction was quenched with H_2O (50 mL) and brine (20 mL), and was extracted with EtOAc (3x50 mL). The combined organic extracts were dried (Na_2SO_4) and conc. *in vacuo* to give a pale-yellow oil, which was purified chromatographically (NP, EtOAc/Hex=45/55, R_F (EtOAc/Hex=6/4)=0.49): The dipeptide (*S*,*S*)-Fmoc-Pro-Lys-OBn was obtained as a white, amorphous solid (655.79 g/mol, 744 mg, 1.13 mmol, 80% yield). The dipeptide was dissolved in piperidine/DMF (1/4, 10 mL). The solution was stirred for 3 h at RT and was conc. in high vacuum overnight at RT. (*S*)-Fmoc-Lys(Boc)-OH (637 mg, 1.36 mmol, 1.2 eq.), HATU (517 mg, 1.36 mmol, 1.2 eq.) and DIPEA (587 µL, 3.39 mmol, 3.0 eq.) were dissolved in anh. deg. DMF (5 mL). The mixture was stirred for 30 min at 0°C. A solution of the deprotected dipeptide in anh. deg. DMF (5 mL) was added at 0°C and the mixture was stirred overnight at RT. The reaction was quenched with H_2O (50 mL) and brine (20 mL), and was extracted with EtOAc (3x50 mL). The combined organic extracts were dried (Na_2SO_4) and conc. *in vacuo* to give a pale-yellow oil. The crude product was purified chromatographically (NP, EtOAc/Hex=75/25, R_F(EtOAc/Hex=75/25)=0.26): The desired product was obtained as a pale-yellow, amorphous solid (920 mg, 1.04 mmol, 92% yield and 73% yield over four steps). This procedure was adapted from the literature.[324]



^{1}H-NMR (600 MHz, CD_3OD): δ = 7.82-7.74 (m, 2H, *H*5), 7.67-7.59 (m, 2H, *H*4), 7.41-7.33 (m, 2H, *H*3), 7.33-7.23 (m, 7H, *H*2+*H*35-37), 5.09 (d, $^{2}J_{33ab}$=12.2 Hz, 1H, *H*33a), 5.04 (d, $^{2}J_{33ab}$=12.2 Hz, 1H, *H*33b), 4.49-4.30 (m, 3H, *H*22+24+8a), 4.27-4.13 (m, 3H, *H*7+10+8b), 3.88-3.78 (m, 1H, *H*19a), 3.60-3.45 (m, 1H, *H*19b), 3.12-2.78 (m, 4H, *H*14+28), 2.16-1.93 (m, 2H, *H*21), 1.93-1.81 (m, 2H, *H*20), 1.81-1.53 (m, 4H, *H*11+25), 153-1.14 (m, 26H, *H*12+13+17+26+27+31).

^{13}C-NMR (151 MHz, CD_3OD): δ = 173.8 (*C*23), 173.0, 172.9 (*C*18+32), 158.7, 158.5, 158.3 (*C*9+15+29), 145.1 (*C*1), 142.5 (*C*6), 137.1 (*C*34), 129.5, 29.4, 129.3 (*C*35-37), 128.8 (*C*3), 128.2 (*C*2), 126.3 (*C*4), 120.9 (*C*5), 79.79, 79.75 (*C*16+30), 68.0, 67.9 (*C*8+33), 62.1 (*C*22), 54.4 (*C*10), 53.5 (*C*24), 48.3, 48.2 (*C*7+19), 41.0, 40.9 (*C*14+28), 31.8, 31.5 (*C*11+25), 30.6 (*C*21), 28.8 (*C*17+31), 25.1 (*C*20), 24.1, 23.9, 23.7, 23.4 (*C*12+13+26+27).

IR: v = 3318 (-N-H), 1687 (C=O), 1516 (-N-H deform.), 1164 (C-O-C).

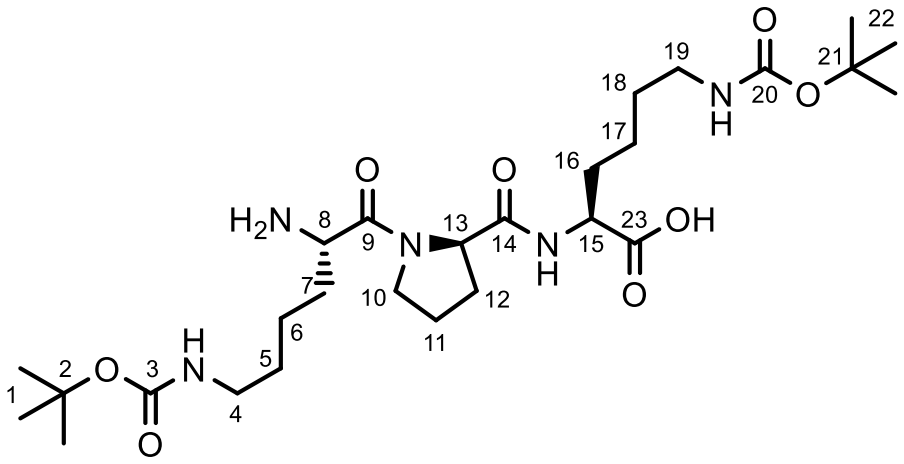
LC/MS (Method 1, 254 nm, ESI+): t_R = 11.7 min, m/z: 906.3 [M+Na^+] (24), 884.4 [M+H^+] (44), 784.3 [M-$C_5H_7O_2^-$] (100).

HRMS (ESI+): $C_{33}H_{38}N_2NaO_6^+$ [M+Na^+] calc.: 581.2622 found: 581.2621.

98 (BNH-099) *N*^{6}-(*tert*-Butoxycarbonyl)-*N*^{2}-*N*^{6}-(*tert*-butoxycarbonyl)-L-lysyl-D-prolyl-L-lysine

The protected tripeptide **100** (759 mg, 859 µmol) was dissolved in piperidine/DMF (1/4, 10 mL) and the solution was stirred for 3 h at RT. The solution was conc. in high vacuum overnight at RT. The residue was dissolved in MeOH (25 mL) and the solution

was N_2 sparged for 1 min. Pd/C (10% with 50% H_2O, 183 mg, 85.9 µmol, 10 mol%) was added. The mixture was H_2-sparged for 1 min and stirred for 48 h at RT. The crude product was purified chromatographically (RP, PF-15C18AQ-F0080, H_2O+FA/MeCN+FA=9/1 to 1/9 over 13 CV): The desired product was obtained as a white, amorphous solid (461 mg, 807 µmol, 94% yield).



^{1}H-NMR (600 MHz, CD_3OD): δ = 4.53-4.43 (m, 1H, *H*13), 4.38-4.29 (m, 1H, *H*15), 4.25-4.17 (m, 1H, *H*8), 3.86-3.76 (m, 1H, *H*10a), 3.65-3.55 (m, 1H, *H*10b), 3.10-2.97 (m, 4H, *H*4+19), 2.29-2.20 (m, 1H, *H*12a), 2.16-1.98 (m, 3H, *H*11+12b), 1.95-1.75 (m, 3H, *H*16+7a), 1.75-1.63 (m, 1H, *H*7b), 1.59-1.28 (m, 26H, *H*1+5+6+17+18+22).

^{13}C-NMR (151 MHz, CD_3OD): δ = 176.7 (*C*23), 173.6 (*C*14), 169.3 (*C*9), 158.6, 158.5 (*C*3+20), 80.0, 79.8 (C2+21), 62.0 (*C*13), 54.6 (*C*15), 53.2 (*C*8), 48.6 (*C*10), 41.2, 40.7 (*C*4+19), 33.1 (*C*7), 31.1, 31.0 (*C*12+16), 30.6, 30.5 (*C*5+18), 28.8 (*C*1+22), 25.5 (*C*11), 24.2 (*C*17), 23.0 (*C*6).

IR: v = 3316 (-N-H), 1682 (C=O), 1166 (C-O-C).

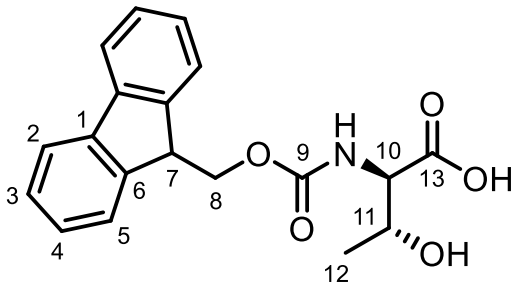
LC/MS (Method 2, 200 nm, ESI+): t_R = 5.6 min, m/z: 572.4 [M+H$^+$] (100).

HRMS (ESI+): $C_{33}H_{38}N_2NaO_6^+$ [M+Na$^+$] calc.: 581.2622 found: 581.2621.

104 (BNH-103) (((9*H*-Fluoren-9-yl)methoxy)carbonyl)-*D*-allothreonine

(*R*,*R*)-Thr-OH (5.00 g, 42.0 mmol) was dissolved in sat. aq. $NaHCO_3$ (100 mL). A solution of Fmoc-ONSU (14.2 g, 42.0 mmol, 1.0 eq.) in 1,4-dioxane (100 mL) was added and the mixture was stirred overnight at RT. Aq. HCl (1 M) was added until

pH=3. The mixture was extracted with EtOAc (4x100 mL). The combined organic extracts were dried (Na_2SO_4) and conc. *in vacuo*. The crude product was purified chromatographically (RP, PF-15C18AQ-F0080 H_2O+FA/MeCN+FA=9/1 to 1/9 over 13 CV): The desired product was obtained as a white, amorphous solid (13.9 g, 40.7 mmol, 97% yield). This procedure was adapted from the literature.[254] The spectroscopic data agrees with the literature.



^{1}H-NMR (600 MHz, CD_3OD): δ = 7.80-7.73 (m, 2H, *H*2), 7.69-7.57 (m, 2H, *H*2), 7.41-7.32 (m, 2H, *H*4), 7.32-7.25 (m, 2H, *H*3), 4.45-4.29 (m, 2H, *H*8), 4.27-4.22 (m, 1H, *H*10), 4.22-4.16 (m, 1H, *H*7), 4.15-4.00 (m, 1H, *H*11), 1.29-1.18 (m, 3H, *H*12).

^{13}C-NMR (151 MHz, CD_3OD): δ = 173.8 (*C*13), 158.6 (*C*9), 145.2 (*C*6), 142.5 (*C*1), 128.7 (*C*4), 128.1 (*C*3), 126.2 (*C*5), 120.9 (*C*2), 68.8 (*C*11), 68.1 (*C*8), 61.3 (*C*10), 48.3 (*C*7), 19.4 (*C*12).

IR: v = 3311 (-N-H), 1681 (C=O), 735 (=C-H deform.).

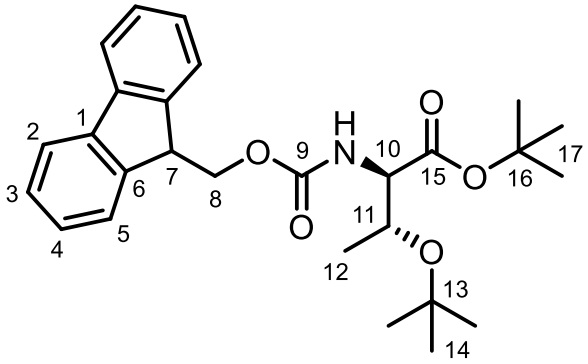
LC/MS (Method 3, 254 nm, ESI+): t_R = 5.8 min, m/z: 342.1 [$M+H^+$] (100).

HRMS (ESI+): $C_{19}H_{19}NNaO_5^+$ [$M+Na^+$] calc.: 364.1155 found: 364.1158.

105 (BNH-104) *tert*-Butyl *N*-(((9*H*-fluoren-9-yl)methoxy)carbonyl)-*O*-(*tert*-butyl)-D-allothreoninate

(*R*,*R*)-Fmoc-Thr-OH 104 (1.0 g, 2.93 mmol) was suspended in AcOtBu (25 mL). Aq. $HClO_4$ (60%, 10 µL, 59.7 µmol, 2 mol%) was added and the mixture was sonicated for 5 min, and stirred overnight at RT. A clear solution was obtained. The reaction was quenched with sat. $NaHCO_3$ and extracted with EtOAc (2x30 mL). The combined organic extracts were dried (Na_2SO_4) and conc. *in vacuo*. The crude product was purified chromatographically (NP, PF-15SIHP-F0080, EtOAc/Hex=1/9, R_F=0.34): The desired product was obtained alongside the starting material. The latter was reacted

and purified again under the same conditions. The desired product was obtained as a white, amorphous solid. (887 mg, 1.96 mmol, 66% yield overall). This procedure was adapted from the literature.[254] The spectroscopic data agrees with the literature.



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.81-7.72 (m, 2H, *H*2), 7.66-7.57 (m, 2H, *H*5), 7.43-7.36 (m, 2H, *H*4), 7.35-7.28 (m, 2H, *H*3), 5.63-5.51 (m, 1H, N*H*), 4.47-4.33 (m, 2H, *H*8) 4.33-4.28 (m, 1H, *H*10), 4.28-4.19 (m, 1H, *H*7), 4.06-3.95 (m, 1H, *H*11), 1.51 (s, 9H, *H*17), 1.26-1.22 (m, 3H, H12), 1.20 (s, 9H, *H*14).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 169.3 (*C*15), 156.1 (*C*9), 144.0 (*C*6), 141.4 (*C*1), 127.7 (*C*4), 127.1 (*C*3), 125.3 (*C*5), 120.0 (*C*2), 82.2 (*C*13), 74.1 (*C*16), 68.6 (*C*11), 67.1 (*C*8), 60.0 (*C*10), 47.3 (*C*7), 28.4 (*C*14), 28.2 (*C*17), 19.7 (*C*12).

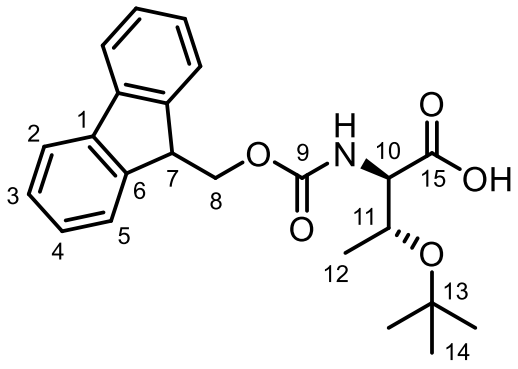
IR: v = 3377 (-N-H), 1699 (C=O), 1537 (-N-H deform.), 729 (=C-H deform.).

LC/MS (Method 3, 254 nm, ESI+): t_R = 14.8 min, m/z: 476.3 [M+Na$^+$] (64), 454.3 [M+H$^+$] (94), 398.2 [M-$C_4H_7^-$] (34), 342.2 [M-($C_4H_7^-$)x2] (62).

HRMS (ESI+): $C_{27}H_{35}NNaO_5^+$ [M+Na$^+$] calc.: 476.2410 found: 476.2410.

103 (BNH-105) *N*-(((9*H*-Fluoren-9-yl)methoxy)carbonyl)-*O*-(*tert*-butyl)-D-allothreonine

(*R*,*R*)-Fmoc-Thr(tBu)-OtBu **105** (546 mg, 1.20 mmol) was dissolved in toluene (25 mL) and SiO_2 (5.86 g) was added. The suspension was refluxed for 30 min and filtered. The filter cake was washed with MeOH/DCM (1/9, 50 mL) and the filtrate was conc. *in vacuo*. The crude product was purified chromatographically (RP, HPLC, H_2O+FA/MeCN+FA=95/5 to 5/95 over 13 CV): The desired product was obtained as a white, amorphous solid (94.9 mg, 239 µmol, 19% yield). This procedure was adapted from the literature.[254] The spectroscopic data agrees with the literature.



In $CDCl_3$ at 300 K rotamers were observed.

^{1}H-NMR (600 MHz, DMSO-d_6, 353 K): δ = 7.88-7.81 (m, 2H, *H*2), 7.74-7.66 (m, 2H, *H*5), 7.44-7.36 (m, 2H, *H*4), 7.36-7.27 (m, 2H, *H*3), 6.74 (bs, 1H, N*H*), 4.40-4.29 (m, 2H, *H*10+11), 3.28-3.21 (m, 1H, *H*7), 4.16-4.07 (m, 1H, *H*8a), 4.04-3.94 (m, 1H, *H*8b), 1.15 (s, 9H, *H*14), 1.14-1.09 (m, 3H, *H*14).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 147.2 (*C*15), 156.3 (*C*9), 143.8 (*C*6), 141.4 (*C*1), 127.2 (*C*4), 125.2 (*C*3), 125.2 (*C*5), 120.1 (*C*2), 75.0 (*C*13), 68.4 (*C*11), 67.3 (*C*8), 59.4 (*C*10), 47.2 (*C*7), 28.3 (*C*14), 19.4 (*C*12).

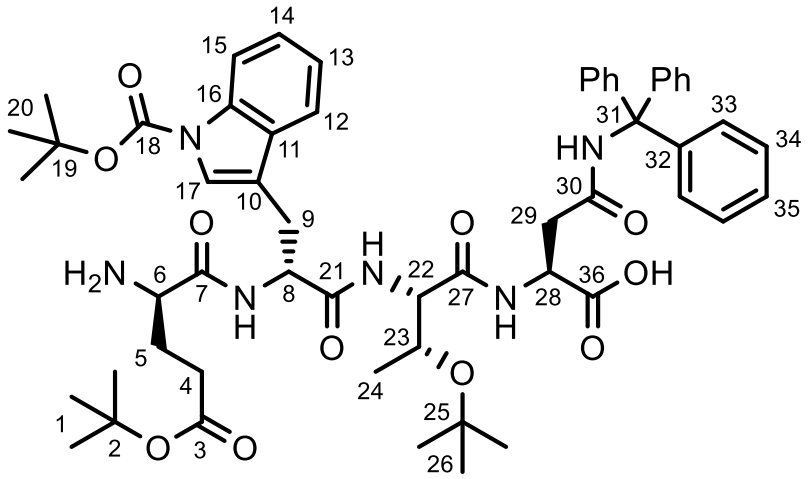
IR: v = 1713 (C=O), 1190 (C-O-C), 737 (=C-H deform.).

LC/MS (Method 1, 254 nm, ESI+): t_R = 9.0 min, m/z: 420.1 [M+Na$^+$] (100), 398.1 [M+H$^+$] (27), 342.1 [M-$C_4H_7^-$] (65).

HRMS (ESI+): $C_{23}H_{27}NNaO_5^+$ [M+Na$^+$] calc.: 420.1781 found: 420.1781.

107 (BNH-106) *N*^{2}-(*N*-(N$^\alpha$-((*R*)-2-Amino-5-(*tert*-butoxy)-5-oxopentanoyl)-1-(*tert*-butoxycarbonyl)-D-tryptophyl)-*O*-(*tert*-butyl)-L-threonyl)-*N*^{4}-trityl-L-asparagine

This tetrapeptide was synthesised according to general procedure A. CTC resin (1.29 g, 2.00 mmol) was used. The desired product was obtained as an off-white, amorphous solid (1.38 g, 1.38 mmol, 68% yield).



^{1}H-NMR (600 MHz, CD_3OD): δ = 8.11-8.02 (m, 1H, *H*12), 7.68-7.63 (m, 1H, *H*15), 7.55 (s, 1H, *H*17), 7.32-7.26 (m, 1H, *H*14), 7.26-7.14 (m, 16H, *H*13+*H*33-35), 4.83-4.76 (m, 1H, *H*28), 4.54-4.48 (m, 1H, *H*8), 4.26-4.21 (m, 1H, *H*22), 3.91-3.85 (m, 1H, *H*6), 3.80-3.72 (m, 1H, *H*23), 3.28-3.06 (m, 2H, *H*29), 2.92-2.70 (m, 2H, *H*9), 2.38-2.21 (m, 2H, *H*4), 2.11-1.95 (m, 2H, *H*5), 1.64 (s, 9H, *H*1), 1.42(s, 9H, *H*20), 1.05 (s, 9H, *H*26), 0.63-0.59 (M, 3H, H24).

^{13}C-NMR (151 MHz, CD_3OD): δ = 175.6 (*C*36), 173.5 (*C*3), 172.5 (*C*30), 172.0 (*C*21), 171.0 (*C*27), 170.4 (*C*7), 150.9 (*C*18), 146.0 (*C*32), 136.7 (*C*16), 131.6 (*C*10), 130.0, 128.7, 127.8 (*C*33-35), 125.7, 125.5 (*C*14+17), 123.8 (*C*13), 120.2 (*C*15), 116.8 (*C*11), 116.3 (*C*12), 85.0 (*C*2), 82.3 (*C*19), 75.8 (*C*25), 71.7 (*C*31), 67.8 (*C*23), 59.4 (*C*22), 55.7 (*C*28), 53.8 (*C*6), 52.3 (*C*8), 40.6 (*C*9), 31.67 (*C*4), 28.8 (*C*26), 28.5, 28.4 (*C*1+20), 27.9, 27.8 (*C*5+29), 18.6 (*C*24).

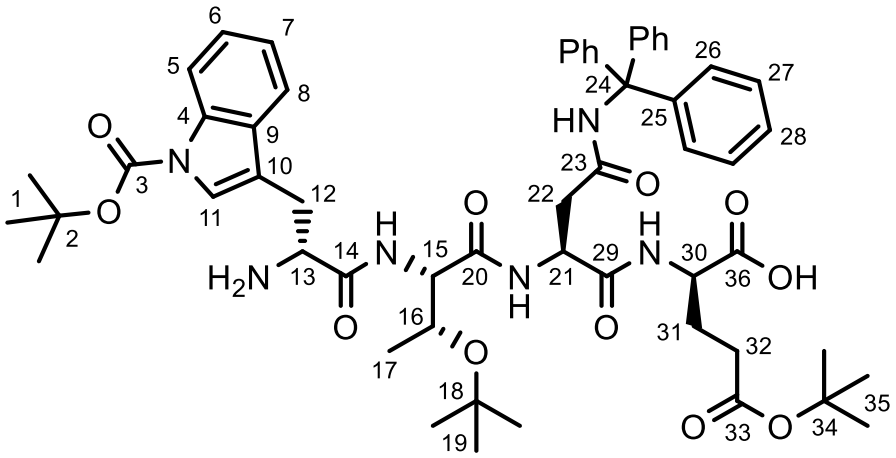
IR: v = 3323 (N-H), 1661 (C=O), 1152 (C-O-C), 699 (=C-H deform.).

LC/MS (Method 2, 254 nm, ESI+): t_R = 9.2 min, m/z: 1003.6 [M+H$^+$] (100).

HRMS (ESI+): $C_{56}H_{71}N_6O_{11}^+$ [M+H$^+$] calc.: 1003.5175 found: 1003.5173.

108 (BNH-107) (*R*)-2-((*S*)-2-((2*S*,3*R*)-2-((*R*)-2-amino-3-(1-(*tert*-butoxycarbonyl)-1*H*-indol-3-yl)propanamido)-3-(*tert*-butoxy)butanamido)-4-oxo-4-(tritylamino)butanamido)-5-(*tert*-butoxy)-5-oxopentanoic acid

This tetrapeptide was synthesised according to general procedure A. CTC resin (1.20 g, 1.86 mmol) was used. The desired product was obtained as an off-white, amorphous solid (1.64 g, 1.63 mmol, 87% yield).



^{1}H-NMR (600 MHz, CD_3OD): δ = 8.17-8.09 (m, 1H, *H*8), 7.69-7.63 (m, 1H, *H*5), 7.63-7.57 (m, 1H, *H*11), 7.37-7.30 (m, 1H, *H*6), 7.30-7.11 (m, 16H, *H*7+26-28), 4.76-4.68 (m, 1H, *H*30), 4.36-4.24 (m, 2H, *H*13+21), 4.18-4.11 (m, 1H, *H*15), 3.93-3.85 (m, 1H, *H*16), 3.31-3.18 (m, 2H, *H*32), 3.09-2.99 (m, 1H, *H*31a), 2.67-2.58 (m, 1H, *H*31b), 2.35-2.20 (m, 2H, *H*22), 2.20-2.08 (m, 1H, *H*12a), 2.02-1.88 (m, 1H, *H*12b), 1.66 (s, 9H, *H*35), 1.43 (s, 9H, *H*19), 1.01 (s, 9H, *H*1), 0.78-0.70 (m, 3H, *H*17).

^{13}C-NMR (151 MHz, CD_3OD): δ = 176.6 (*C*29), 174.2 (*C*23), 172.5 (*C*33), 172.2 (*C*20), 171.8 (*C*36), 171.1 (*C*14), 150.8 (*C*3), 145.7 (*C*25), 137.0 (*C*4), 131.1 (*C*10), 130.0, 128.8, 127.9 (*C*26-28), 126.3 (*C*11), 125.9 (*C*6), 124.0 (*C*7), 120.1 (*C*5), 116.4 (*C*8), 114.7 (*C*9), 85.2 (*C*34), 81.5 (*C*18), 75.5 (*C*2), 71.8 (*C*24), 67.5 (*C*16), 61.6 (*C*15), 54.7, 54.3 (*C*13+21), 51.3 (*C*30), 38.0 (*C*31), 32.7 (*C*22), 29.0 (*C*12), 28.9, 28.42, 28.38 (*C*1+19+35), 28.3 (*C*32), 20.4 (*C*17).

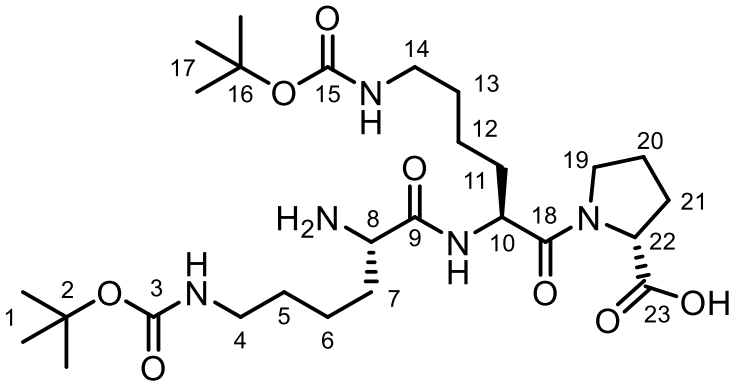
IR: v = 1667 (C=O), 1367 (-C-N), 1152 (C-O-C), 698 (=C-H deform.).

LC/MS (Method 2, 254 nm, ESI+): t_R = 9.8 min, m/z: 1003.7 [M+H$^+$] (100).

HRMS (ESI+): $C_{56}H_{71}N_6O_{11}^+$ [M+H$^+$] calc.: 1003.5176 found: 1003.5173.

101 (BNH-108) N^6-(*tert*-Butoxycarbonyl)-N^2-(N^6-(*tert*-butoxycarbonyl)-L-lysyl)-L-lysyl-D-proline

This tripeptide was synthesised according to general procedure A. CTC resin (1.00 g, 1.55 mmol) was used. The desired product was obtained as an off-white, amorphous solid (868 mg, 1.52 mmol, 98% yield).



Rotamers were observed in the NMR spectra.

^{1}H-NMR (600 MHz, DMSO-d_6): δ = 8.54-8.44 (m, 1H, N*H*$_2$ terminus), 6.79-6.67 (m, 2H, Lys-ε-N*H*), 4.65-4.57, 4.35-4.27 (m, 1H, *H*8), 4.57-4.51, 4.23-4.16 (m, 1H, *H*22), 3.68-3.49 (m, 2H, *H*10+19a), 3.39-3.27 (m, 1H, *H*19b), 2.93-2.79 (m, 4H, *H*4+14), 2.16-1.96 (m, 1H, *H*21a), 1.96-1.81 (m, 2H, *H*11a+21b), 1.80-1.71 (m, 1H, *H*20a), 1.71-1.55 (m, 3H, *H*7a+11b+20b), 1.55-1.43 (m, 1H, *H*7b), 1.43-1.05 (m, 26H, *H*1+5+6+12+13+17).

^{13}C-NMR (151 MHz, DMSO-d_6): δ = 174.5, 173.4 (*C*23), 170.0(*C*18), 169.1(*C*9), 155.56, 155.55 (*C*3+15), 77.4, 77.3 (*C*2+16), 60.2, 58.9 (*C*22), 52.6, 52.3 (*C*10), 50.4, 49.9 (*C*8), 46.4, 46.1 (*C*19), 40.1 (*C*4+14), 31.9, 31.6 (*C*7), 31.2, 30.9 (*C*11), 29.3, 29.0 (*C*21), 28.3 (*C*1+17), 22.7, 22.3 (*C*20), 24.1, 22.2, 21.76, 21.5 (*C*5+6+12+13).

IR (ATR): v = 3298 (N-H), 1681,1624 (C=O), 1519 (N-H deform.), 1164 (C-O-C).

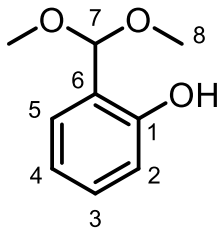
LC/MS (Method 2, 200 nm, ESI+): t_R = 5.3 min, m/z: 572.4 [M+H$^+$] (100).

HRMS (ESI+): $C_{27}H_{49}N_5NaO_8^+$ [M+Na$^+$] calc.: 594.3473 found: 594.3476.

120 (BNH-110) 2-(Dimethoxymethyl)phenol

Salicylaldehyde (15 mL, 143 mmol), trimethyl orthoformate (15.7 mL, 143 mmol, 1.0 eq.), and $LiBF_4$ (402 mg, 4.29 mmol, 3 mol%) were dissolved in anh. MeOH (50 mL). The solution was refluxed overnight under N_2. The mixture was left to cool to rt and sat. aq. $NaHCO_3$ (150 mL) was added. The mixture was extracted with EtOAc (3x50 mL) and the combined organic extracts were washed with sat. aq. $NaHCO_3$ (2x100 mL). The extracts were dried (Na_2SO_4) and conc. *in vacuo*. The obtained yellow oil was distilled at 5 mbar. The fraction with a boiling point of 88°C was collected. The desired product was obtained as a colourless oil (9.49 g, 56.4 mmol, 39% yield). This

procedure was adapted from the literature.[259] The spectroscopic data agrees with the literature.[325]



^{1}H-NMR (400 MHz, $CDCl_3$): δ = 8.06 (bs, 1H, O*H*), 7.28-7.21 (m, 1H, *H*3), 7.21-7.15 (m, 1H, *H*5), 6.92-6.85 (m, 2H, *H*2+4), 5.56 (s, 1H, *H*7), 3.40 (s, 6H, *H*8).

^{13}C-NMR (101 MHz, $CDCl_3$): δ = 155.8 (*C*1), 130.4 (*C*3), 128.5 (*C*5), 121.0 (*C*6), 119.8 (*C*4), 117.1 (*C*2), 104.7 (*C*7), 52.9 (*C*8).

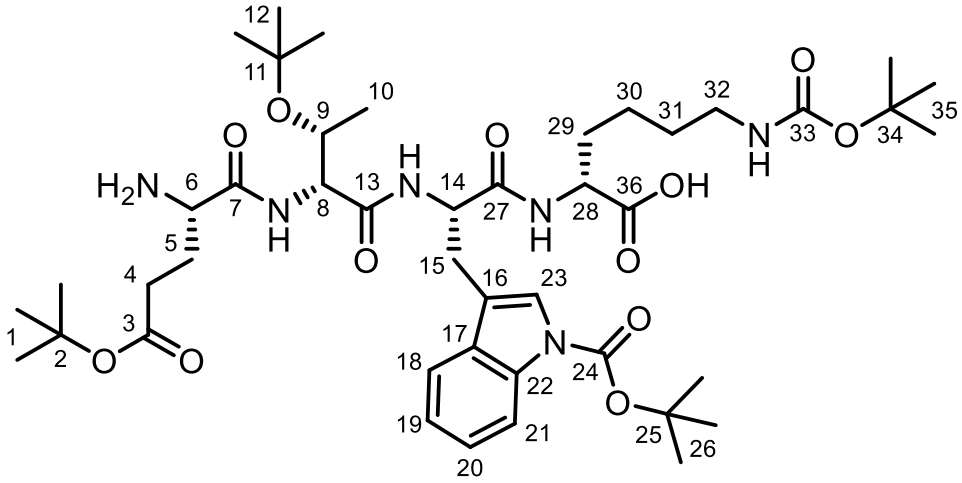
IR: v = 3361 (O-H), 2832 (O-CH_3), 1586 (ring vibr.), 1038 (C-O-C), 755 (=C-H deform.).

LC/MS (Method 1, 254 nm, ESI-): t_R = 4.2 min, m/z: 167.0 [M-H$^+$] (9), 153.0 [M-CH_3^+] (7), 137.0 [M-CH_3O^+] (7), 121.0 [Aldehyde-H$^+$] (100).

HRMS (ESI-): $C_9H_{11}O_3^-$ [M-H$^+$] calc.: 167.0714 found: 167.0714.

106 (BNH-112) *N*^{2}-(*N*$^\alpha$-(*N*-((*S*)-2-Amino-5-(*tert*-butoxy)-5-oxopentanoyl)-*O*-(*tert*-butyl)-D-allothreonyl)-1-(*tert*-butoxycarbonyl)-L-tryptophyl)-*N*^{6}-(*tert*-butoxycarbonyl)-D-lysine

This tetrapeptide was synthesised according to general procedure A. CTC resin (500 mg, 775 µmol) was used. The desired product was obtained as an off-white, amorphous solid (671 mg, 767 µmol, 99% yield).



^{1}H-NMR (600 MHz, CD_3OD): δ = 8.15-8.01 (m, 1H, *H*18), 7.70-7.63 (m, 1H, *H*21), 7.56 (s, 1H, *H*23), 7.33-7.27 (m, 1H, *H*20), 7.27-7.20 (m, 1H, *H*19), 4.88-4.76 (m, 1H, *H*28), 4.44 (d, $^{3}J_{8\text{-}9}$=7.0 Hz, 1H, *H*8), 4.17 (t, $^{3}J_{14\text{-}15}$=5.7 Hz, 1H, *H*14), 3.87 (t, $^{3}J_{5\text{-}6}$=6.3 Hz, 1H, *H*6), 3.81-3.70 (m, 1H, *H*9), 3.46-3.36 (m, 1H, *H*29a), 3.06-2.92 (m, 3H, *H*29b+32), 2.511-2.34 (m, 2H, *H*4), 2.13-2.01 (m, 2H, *H*5), 1.93-1.81 (m, 1H, *H*15a), 1.72-1.1.63 (m, 10H, *H*15b+26), 1.49-1.38 (m, 20H, *H*1+31+35), 1.32-1.15 (m, 2H, *H*30), 0.94-0.78 (m, 12H, *H*10+12).

^{13}C-NMR (151 MHz, CD_3OD): δ = 177.8 (*C*27), 173.2 (*C*3), 172.3 (*C*36), 172.0 (*C*13), 170.1 (*C*7), 158.5 (*C*33), 151.0 (*C*24), 137.0 (*C*23), 131.5 (*C*18), 125.5 (*C*17), 125.5 (*C*21), 123.8 (*C*20), 120.2 (*C*22), 117.8 (*C*16), 116.3 (*C*19), 84.9, 82.3, 79.8, 75.2 (*C*2+11+25+34), 67.9 (*C*9), 59.8 (*C*8), 55.9 (*C*14), 54.5 (*C*28), 53.9 (*C*6), 41.3 (*C*32), 32.8 (*C*15), 31.7 (*C*4), 30.8 (*C*31), 28.9, 28.5, 28.5, 28.3 (*C*1+12+26+35), 27.8 (*C*5), 27.6 (*C*29), 23.6 (*C*30), 19.3 (*C*10).

IR: v = 3290 (N-H), 1644 (C=O), 1366 (C-H deform.), 1155 (C-O-C), 743 (=C-H deform.).

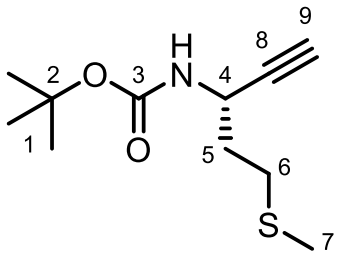
LC/MS (Method 2, 254 nm, ESI+): t_R = 8.8 min, m/z: 875.8 [M+H$^+$] (100).

HRMS (ESI+): $C_{44}H_{71}N_6O_{12}^+$ [M+H$^+$] calc.: 875.5124 found: 875.5122.

143 (BNH-123) *tert*-Butyl (*S*)-(5-(methylthio)pent-1-yn-3-yl)carbamate

This α-amino alkyne was synthesised according to general procedure B from (*S*)-Boc-Met-OH (2.00 g, 8.02 mmol). The reduction gave the desired aldehyde in more than 95% purity, according to ^{1}H-NMR analysis (1.70 g, 7.27 mmol, 90% yield,

R_F(EE/Hex=4/6)=0.69). The crude aldehyde was added in anh. MeOH (4.5 mL) to the homologation reaction. The crude product was purified chromatographically (NP, EE/Hex=1/9, R_F =0.35): The desired product was obtained as a colourless oil (835 mg, 3.64 mmol, 50% yield and 45% yield over three steps). The spectroscopic data agrees with the literature.[274]



[1]H-NMR (400 MHz, $CDCl_3$): δ = 5.00-4.71 (m, 1H, N*H*), 4.65-4.34 (m, 1H, *H*4), 2.68-2.50 (m, 12, *H*6), 2.32-2.26 (m, 1H, *H*9), 2.12-2.06 (m, 3H, *H*7), 2.03-1.83 (m, 2H, *H*5), 1.43 (s, 9H, *H*1).

[13]C-NMR (101 MHz, $CDCl_3$): δ = 154.8 (*C*3), 82.8 (*C*8), 80.2 (*C*2), 71.8 (*C*9), 42.2 (*C*4), 35.5 (*C*5), 30.1 (*C*6), 28.4 (*C*1), 15.6 (*C*7).

IR: v = 3297 (≡C-H), 1688 (C=O), 1504 (N-H deform.), 1159 (C-O-C), 644 (C-S).

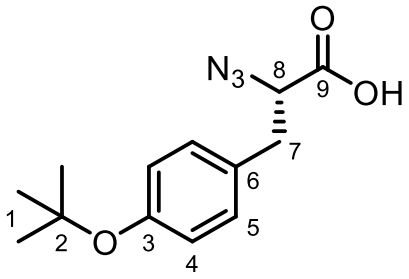
LC/MS (Method 1, 200 nm, ESI+): t_R = 7.6 min, m/z: 230.1 [$M+H^+$] (10), 174.1 [$M-C_4H_7^-$] (100), 130.1 [$M-C_5H_7O_2^-$] (11).

HRMS (ESI+): $C_{11}H_{19}NNaO_2S^+$ [$M+Na^+$] calc.: 252.1029 found: 252.1029.

Optical Rotation: $[\alpha]_D^{20}$-14.6 (c=1.0, DCM).

134 (BNH-124) (*S*)-2-Azido-3-(4-(*tert*-butoxy)phenyl)propanoic acid

This α-azido acid was synthesised according to general procedure C from (*S*)-H_2N-Tyr(tBu)-OH (4.31 g, 18.2 mmol). The crude product was purified chromatographically (NP, AcOH/EtOAc/Hex=1/30/69, R_F=0.26) and (RP, PF-15C18AQ-F0080 H_2O+FA/MeOH+FA=5/3 to 1/9 over 13 CV with 25 mL/min): The desired product was obtained as a yellow, viscous oil (2.95 g, 11.2 mmol, 61% yield). The spectroscopic data agrees with the literature.[266]



^{1}H-NMR (400 MHz, $CDCl_3$): δ = 8.88 (bs, 1H, COO*H*), 7.18-7.13 (m, 2H, *H*5), 6.99-6.93 (m, 2H, *H*4), 4.16-4.06 (m, 2H, *H*8), 3.24-2.94 (m, 2H, *H*7), 1.34 (s, 9H, *H*1).

^{3}C-NMR (101 MHz, $CDCl_3$): δ = 175.1 (*C*9), 154.5 (*C*3), 130.8 (*C*6), 129.8 (*C*5), 124.6 (*C*4), 79.1 (*C*2), 63.3 (*C*8), 37.1 (*C*7), 28.9 (*C*1).

IR: v = 2105 (-N_3), 1719 (C=O), 1505 (ring vibr.), 1154 (C-O-C).

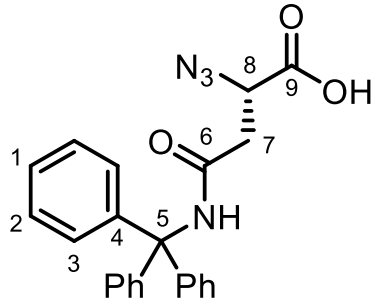
LC/MS (Method 3, 220 nm, ESI-): t_R = 5.2 min, m/z: 262.1 [M-H$^+$] (100).

HRMS (ESI-): $C_{13}H_{16}N_3O_3^-$ [M-H$^+$] calc.: 262.1197 found: 262.1194.

Optical Rotation: $[\alpha]_D^{20}$ -48.8 (c=1.0, DCM).

135 (BNH-127) (*S*)-2-Azido-4-oxo-4-(tritylamino)butanoic acid

This α-azido acid was synthesised according to general procedure C from (*S*)-H_2N-Asn(Trt)-OH (7.49 g, 20.0 mmol). The crude product was purified chromatographically (NP, AcOH/EtOAc/Hex=1/35/64): The desired product was obtained as a white, amorphous solid (7.10 g, 17.7 mmol, 88% yield).



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.31-7.22 (m, 9H, *H*1+2), 7.22-7.13 (m, 6H, *H*3), 7.13-6.83 (m, 2H, N*H*+COO*H*), 4.45-4.32 (m, 1H, *H*8), 2.86-2.52 (m, 2H, *H*7).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 172.6 (*C*9), 168.7 (*C*6), 144.0 (*C*4), 128.8 (*C*3), 128.2 (*C*2), 127.4 (*C*1), 71.4 (*C*5), 58.7 (*C*8), 39.0 (*C*7).

IR: v = 3260 (O-H), 2104 (-N_3), 1644 (C=O), 1531 (N-H deform.), 1175 (C-O).

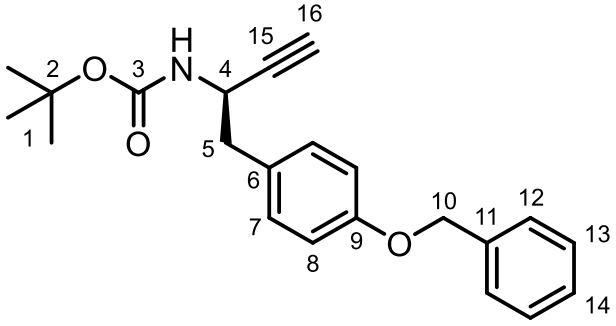
LC/MS (Method 3, 220 nm, ESI-): t_R = 6.9 min, m/z: 399.1 [M-H^+] (100).

HRMS (ESI-): $C_{23}H_{19}N_4O_3^-$ [M-H^+] calc.: 399.1463 found: 399.1463.

Optical Rotation: $[\alpha]_D^{20}$ -21.0 (c=1.0, DCM).

144 (BNH-129) *tert*-Butyl (*R*)-(1-(4-(benzyloxy)phenyl)but-3-yn-2-yl)carbamate

This α-amino alkyne was synthesised according to general procedure B from (*R*)-Boc-Tyr(Bn)-OH (2.97 g, 8.00 mmol). The reduction gave the desired aldehyde (2.82 g, 7.92 mmol, 99% yield, R_F(EE/Hex=4/6)=0.62). The crude aldehyde was added in anh. MeOH (6.5 mL) to the homologation reaction. The crude product was purified chromatographically (NP, EE/Hex=7.5/92.5, R_F=0.18): The desired product was obtained as a colourless oil (893 mg, 2.54 mmol, 32% yield and 31% yield over three steps).



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.47-7.42 (m, 2H, *H*13), 7.42-7.36 (m, 2H, *H*12), 7.36-7.30 (m, 1H, *H*14), 7.23-7.16 (m, 2H, *H*7), 6.97-6.90 (m, 2H, *H*8), 5.06 (s, 2H, *H*10), 4.80-4.53 (m, 2H, N*H*+*H*4), 2.99-2.85 (m, 1H, *H*5), 2.28 (s, 1H, *H*16), 1.44 (s, 9H, *H*1).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 158.1 (*C*9), 154.7 (*C*3), 137.2 (*C*11), 130.9 (*C*7), 128.8 (*C*6), 128.7 (*C*12), 128.0 (*C*14), 127.6 (*C*13), 114.8 (*C*8), 83.1 (*C*15), 80.1 (*C*2), 72.2 (*C*16), 70.1 (*C*10), 44.1 (*C*4), 40.9 (*C*5), 28.5 (*C*1).

IR: v = 3352 (-N-H), 3287 (≡C-H), 1683 (C=O), 1509 (-N-H deform.), 1239 (C-O-C), 1161 (C-O-C).

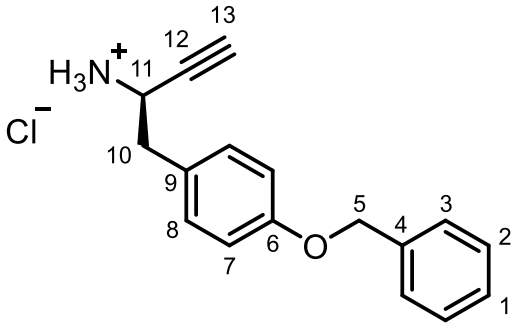
LC/MS (Method 3, 220 nm, ESI+): t_R = 12.3 min, m/z: 374.2 [M+Na^+] (10), 296.1 [M-$C_4H_7^-$] (100).

HRMS (ESI+): $C_{22}H_{25}NNaO_3^+$ [M+Na$^+$] calc.: 374.1727 found: 374.1728.

Optical Rotation: $[\alpha]_D^{20}$ +2.2 (c=1.0, DCM).

146 (BNH-133) (*R*)-1-(4-(Benzyloxy)phenyl)but-3-yn-2-amine hydrochloride

The Boc-α-amino alkyne **144** (648 mg, 1.84 mmol) was dissolved in DCM (12 mL). HCl in dioxane (4.0 M, 3.23 mL, 12.9 mmol, 7.0 eq.) was added and the mixture was stirred overnight at RT under N_2. LC/MS analysis indicated remaining starting material. DCM (1 mL) and HCl in dioxane (4.0 M, 3.68 mL, 14.8 mmol, 8 eq.) were added. The mixture was stirred for 72 h at RT under N_2. The suspension was conc. under a stream of compressed air and was subsequently conc. *in vacuo*. The desired product was obtained as a white, amorphous solid (525 mg, 1.83 mmol, 99% yield) The product was used without further purification.



^{1}H-NMR (600 MHz, CD_3OD): δ = 8.60 (bs, 2H, N*H*$_2$), 7.44-7.38 (m, 2H, *H*3), 7.38-7.32 (m, 2H, *H*2), 7.32-7.26 (m, 1H, *H*1), 7.26-7.20 (m, 2H, *H*7), 6.99-6.93 (m, 2H, *H*8), 5.04 (s, 2H, *H*5), 4.20 (bs, 1H, *H*13), 3.11-3.03 (m, 2H, *H*10a+11), 3.02-2.93 (m, 1H, *H*10b).

^{13}C-NMR (151 MHz, CD_3OD): δ = 169.7 (*C*9), 159.7 (*C*6), 138.7 (*C*4), 131.7 (*C*7), 129.5 (*C*2), 128.9 (*C*1), 128.5 (*C*3), 116.2 (*C*8), 80.4 (*C*12), 77.7 (*C*11), 71.0 (*C*5), 45.4 (*C*13), 40.3 (*C*10).

IR: v = 3233 (≡C-H), 1583 (N-H deform.), 1512 (ring vibr.), 1254 (C-O-C), 773, 731 (=C-H deform.).

LC/MS (Method 3, 220 nm, ESI+): t_R = 9.8 min, m/z: 252.1 [M] (97), 296.1 [M-NH_3] (100).

HRMS (ESI+): $C_{17}H_{18}NO^+$ [M+H$^+$] calc.: 252.1383 found: 252.1387.

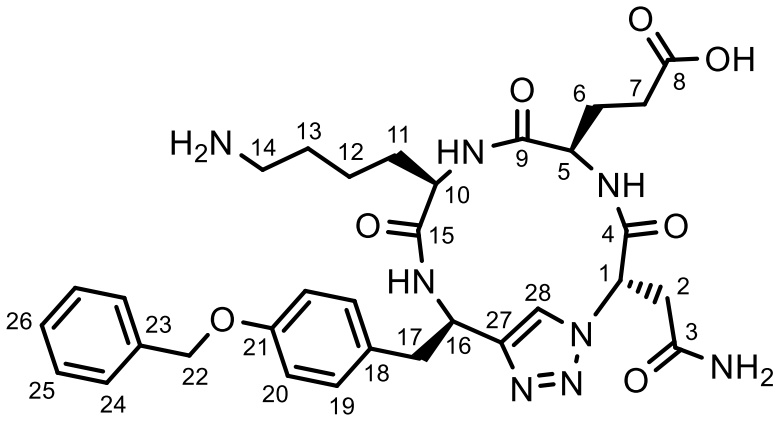
168 (BNH-137) 3-((2*R*,5*R*,8*R*,11*S*,*Z*)-11-(2-Amino-2-oxoethyl)-5-(4-aminobutyl)-2-(4-(benzyloxy)benzyl)-4,7,10-trioxo-1^1*H*-3,6,9-triaza-1(4,1)-triazolacycloundecaphane-8-yl)propanoic acid

Linear Pseudo Tetrapeptide:

This click cyclic tetrapeptide was synthesised according to general procedure D with CTC resin (1.00 g, 1.55 mmol/g). The crude product was purified chromatographic-ally (NP, TEA/MeOH/DCM=1/3/96, R_F=0.09): The desired azido tripeptide **163** was obtained as an off-white, amorphous solid (1.25 g, 1.53 mmol, 99% yield). It was coupled to the α-amino alkyne **146** in anh. MeCN. The desired pseudo tetrapeptide **165** (BNH-135) was obtained as an off-white, amorphous solid (1.22 g, 1.16 mmol, 76% yield). The crude product was used without further purification.

Click Cyclisation and Deprotection:

The pseudo tetrapeptide **165** (100 mg, 95.5 µmol) was cyclised for 48 h. The crude cyclisation product was deprotected. The crude deprotection product was purified chromatographically (RP, HPLC, H_2O+FA/MeCN+FA=95/5 to 7/3 over 21 CV with 18 mL/min): The desired click cyclic tetrapeptide **168** was obtained as a white, amorphous solid (33.5 mg, 51.6 µmol, 54% yield rel. to pseudo tetrapeptide **165**).



^{1}H-NMR (600 MHz, DMSO-d_6): δ = 8.00 (s, 1H, Glu-N*H*), 7.79-7.72 (m, 1H, *H*28), 7.46-7.40 (m, 2H, *H*24), 7.40-7.34 (m, 2H, *H*25), 7.34-7.27 (m, 1H, *H*26), 7.23 (d, $^3J_{19\text{-}20}$=8.5 Hz, 2H, *H*19), 6.92 (d, $^3J_{19\text{-}20}$=8.5 Hz, 2H, *H*20), 5.59-5.53(m, 1H, *H*16), 5.07 (s, 2H, *H*22), 5.03-4.96 (m, 1H, *H*1), 4.21-4.15 (m,1H, *H*5),4.04-3.98 (m, 1H, *H*10), 3.37 (dd, $^2J_{2ab}$=6.5 Hz, $^3J_{1\text{-}2a}$=14.2 Hz, 1H, *H*2a), 3.29 (dd, $^2J_{2ab}$ =9.1 Hz, $^3J_{1\text{-}2a}$ =14.1 Hz, 1H, *H*2b), 3.20 (dd, $^2J_{17ab}$=7.9 Hz, $^3J_{16\text{-}17a}$=15.9 Hz, 1H, *H*17a), 2.98 (dd, $^2J_{17ab}$=6.4 Hz,

$^{3}J_{16\text{-}17b}$=15.9 Hz, 1H, *H*17b), 2.31-2.09 (m, 2H, *C*7), 1.98-1.81 (m, 2H, *H*6), 1.69-1.54 (m, 2H, *H*11), 1.52-1.36 (m, 2H, *H*13), 1.35-1.13 (m, 2H, *H*12).

^{13}C-NMR (151 MHz, DMSO-d_6): δ = 171.1 (*C*4), 170.5 (*C*3), 170.4 (*C*8), 170.3 (*C*9), 165.9 (*C*15), 156.8 (*C*21), 150.7 (*C*27), 137.1 (*C*23), 129.9 (*C*18), 128.4 (*C*19), 128.2 (*C*25), 127.5 (*C*26), 127.3 (*C*24), 123.4 (*C*28), 114.4 (*C*20), 69.2 (*C*22), 59.8 (*C*16), 55.9 (*C*5), 54.6 (*C*10), 49.1 (*C*1), 37.7 (*C*2), 36.7 (*C*14), 33.6 (*C*17), 29.5 (*C*11), 28.0 (*C*13), 26.8 (*C*6), 22.1 (*C*12).

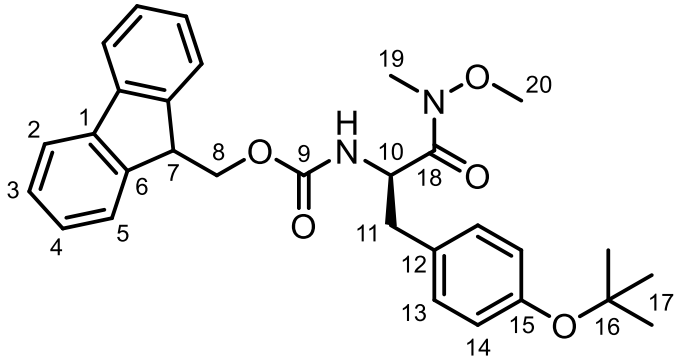
IR: v = 3432 (N-H), 3289 (-NH_3^+), 1671, 1635 (C=O), 1530, 1512 (N-H deform.), 1402 (CO in COO^-), 1227 (C-N).

LC/MS (Method 1, 200 nm, ESI+): t_R = 5.3 min, m/z: 649.3 [$M+H^+$] (100).

HRMS (ESI+): $C_{32}H_{41}N_8O_7^+$ [$M+H^+$] calc.: 649.3093 found: 649.3093.

147 (BNH-140) (9*H*-Fluoren-9-yl)methyl (*R*)-(3-(4-(*tert*-butoxy)phenyl)-1-(methoxy(methyl)amino)-1-oxopropan-2-yl)carbamate

(*R*)-Fmoc-Tyr(tBu)-OH (1.00 g, 2.18 mmol) was dissolved in DCM (218 mL). DIPEA (1.37 mL, 8.05 mmol, 3.7 eq.) and PyBOP (1.25 g, 2.39 mmol, 1.1 eq.) were added. The solution was stirred for 15 min at RT. *N,O*-Dimethylhydroxylamine HCl (255 mg, 2.61 mmol, 1.2 eq.) was added and the solution was stirred overnight at RT. The mixture was washed with aq. HCl (500 mM, 3x100 mL) and sat. $NaHCO_3$ (100 mL). The mixture was dried (Na_2SO_4) and conc. *in vacuo*. The crude product was purified chromatographically (NP, $CHCl_3$, R_F=0.21): The desired product was obtained as a white, amorphous solid (1.08 g, 2.14 mmol, 98% yield). This procedure was adapted from the literature.[271]



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.79-7.72 (m, 2H, *H*4), 7.63-7.53 (m, 2H, *H*5), 7.44-7.36 (m, 2H, *H*2), 7.36-7.28 (m, 2H, *H*3), 7.14-7.04 (m, 2H, *H*13), 6.95-6.84 (m, 2H, *H*14), 5.64-5.45 (m, 1H, N*H*), 5.09-4.92 (m, 1H, *H*10), 4.41-4.23 (m, 2H, *H*8), 4.23-4.12 (m, 1H, *H*7), 3.62 (s, 3H, *H*20), 3.16 (s, 3H, *H*19), 3.10-2.85 (m, 2H, *H*11), 1.30 (s, 9H, *H*17).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 172.1 (*C*18), 155.8 (*C*9), 154.4 (*C*15), 144.0 (*C*6), 141.4 (*C*1), 131.3 (*C*12), 130.0 (*C*13), 127.8 (*C*2), 127.2 (*C*3), 125.3 (*C*5), 124.2 (*C*4), 120.0 (*C*14), 78.4 (*C*16), 67.1 (*C*8), 61.6 (*C*20), 52.2 (*C*10), 47.3 (*C*7), 38.4 (*C*11), 32.2 (*C*19), 28.9 (*C*17).

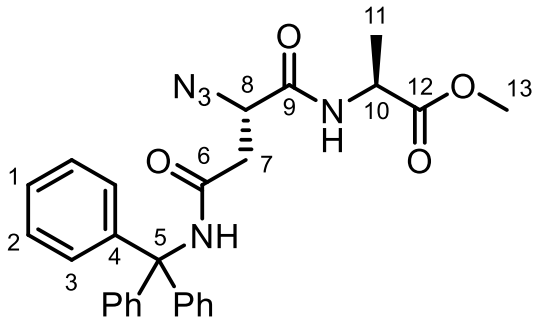
IR: v = 3304 (N-H), 1712, 1661 (C=O), 1236, 1160 (C-O-C), 739 (C=C-H deform.).

LC/MS (Method 3, 254 nm, ESI+): t_R = 12.8 min, m/z: 503.3 [$M+H^+$] (100).

HRMS (ESI+): $C_{30}H_{34}N_2NaO_5^+$ [$M+Na^+$] calc.: 525.2360 found: 525.2361.

137 (BNH-141) Methyl ((*S*)-2-azido-4-oxo-4-(tritylamino)butanoyl)-L-alaninate

This azido dipeptide was synthesised according to general procedure E from the α-azido acid **135** (100 mg, 250 µmol). The crude product was purified chromatographically (RP, HPLC, H_2O+FA/MeCN+FA=7/3 over 2 CV, 7/3 to 5/95 over 10 CV, 5/95 over 2 CV with 18 mL/min): The desired product was obtained as a white, amorphous solid (78 mg, 161 µmol, 64% yield).



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.39-7.15 (m, 15H, *H*1-3), 7.11-7.01 (m, 1H, Ala-α-N*H*), 6.93-6.83 (m, 1H, Asn-γ-N*H*), 4.57-4.43 (m, 2H, *H*8+10), 3.77 (s, 3H, *H*13), 3.16-3.03 (m, 1H, *H*7a), 2.74-2.63 (m, 1H, *H*7b), 1.43-1.30 (m, 3H, *H*11).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 172.7 (*C*12), 168.5 (*C*9), 167.9 (*C*6), 144.3 (*C*4), 128.6 (*C*3), 128.0 (*C*2), 127.2 (*C*1), 71.0 (*C*5), 59.8 (*C*8), 52.3 (*C*13), 48.3 (*C*10), 39.7 (*C*7), 18.0 (*C*11).

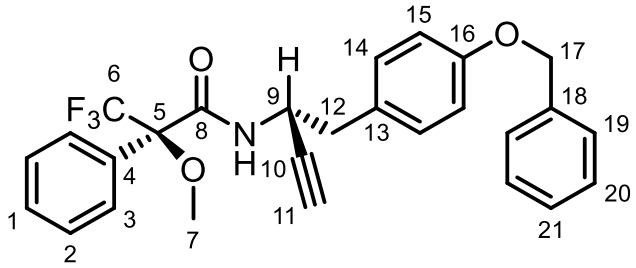
IR: v = 3270 (N-H), 2099 ($-N_3$), 1743 (C=O), 1651 (C=O), 1525 (N-H deform.), 1208 (C-O-C), 699 (=C-H deform.).

LC/MS (Method 3, 220 nm, ESI+): t_R = 10.9 min, m/z: 586.2 [M+H$^+$] (22), 243.1 [M-$C_{19}H_{14}^-$] (100).

HRMS (ESI+): $C_{27}H_{27}N_5NaO_4^+$ [M+Na$^+$] calc.: 508.1955 found: 508.1952.

153 (BNH-145) (*R*)-*N*-((*R*)-1-(4-(Benzyloxy)phenyl)but-3-yn-2-yl)-3,3,3-trifluoro-2-methoxy-2-phenylpropanamide

This Mosher amide was synthesised according to general procedure F from α-amino alkyne **146** (50.0 mg, 174 µmol). The crude product was purified chromatographically (RP, HPLC, H_2O+FA/MeCN+FA=7/3 over 2 CV, 7/3 to 5/95 over 10 CV, 5/95 over 2 CV with 18 mL/min): The desired product was isolated as a clear, sticky, gum-like solid (56 mg, 120 µmol, 68% yield).



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.49-7.30 (m, 10H, *H*1-3+19-21), 7.05-6.98 (m, 2H, *H*15), 6.89-6.83 (m, 2H, *H*14), 6.83-6.75 (m, 1H, N*H*), 5.13-4.99 (m, 2H, *H*9+17), 3.45-3.38 (m, 3H, *H*7), 2.96-2.88 (m, 2H, *H*12), 2.36-2.31 (m, 1H, *H*11).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 165.5 (*C*8), 158.1 (*C*16), 137.1 (*C*5), 132.6, 130.8, 129.6, 128.7, 128.6, 128.1, 128.1, 127.6 (*C*1-3+13+15+19-21), 124.7 (*C*4), 122.8 (*C*18), 114.8 (*C*14), 83.9 (*C*6), 82.0 (*C*10), 72.7 (*C*11), 70.1 (*C*17), 55.1 (*C*7), 42.3 (*C*9), 40.3 (*C*12).

IR: v = 3400 (N-H), 3294 (≡C-H), 1688 (C=O), 1510 (N-H deform.), 1155 (C-O-C), 712 (=C-H deform.).

LC/MS (Method 3, 220 nm, ESI+): t_R = 12.7 min, m/z: 485.2 [M+NH_4^+] (47), 468.2 [M+H^+] (100).

HRMS (ESI+): $C_{22}H_{24}F_3NNaO_3^+$ [M+Na^+] calc.: 490.1600 found: 490.1600.

149 (BNH-150) (*R*)-1-(4-(*tert*-Butoxy)phenyl)but-3-yn-2-amine

<u>$NaBH_4$ Reduction:</u>

(*R*)-Fmoc-Tyr(tBu)-OH (10.0 g, 21.8 mmol) was dissolved in anh. THF (80 mL). The solution was cooled to 0°C under N_2. *N*-Methylmorpholine (99%, 2.67 mL, 26.1 mmol, 1.2 eq.) and isobutyl chloroformate (98%, 3.34 mL, 23.9 mmol, 1.1 eq.) were added sequentially. The opaque mixture was stirred for 30 min at 0°C under N_2. $NaBH_4$ (2.47 g, 65.3 mmol, 3.0 eq.) was suspended in anh. THF/anh. MeOH (3/1, 40 mL) at -94°C under N_2. The opaque Tyr mixture was added to the $NaBH_4$ suspension. The resulting mixture was stirred for 45 min at -94°C under N_2. The reaction was quenched with AcOH/H_2O (1/9, 100 mL) and warmed to RT. The organic solvents were removed *in vacuo*. The mixture was extracted with EtOAc (3x50 mL) and the combined organic extracts were washed with sat. $NaHCO_3$ (2x60 mL) and H_2O (60 mL). The extracts were dried (Na_2SO_4) and conc. *in vacuo*. The desired alcohol was obtained as an off-white, amorphous solid (9.52 g, 21.4 mmol, 98% yield). The crude product was used without further purification. This procedure was adapted from the literature.[326]

<u>DMP Oxidation:</u>

The crude alcohol was dissolved in DCM (150 mL). H_2O (771 µL, 42.8 mmol, 2 eq.) and DMP (9.98 g, 23.5 mmol, 1.1 eq.) were added. The mixture was stirred for 1 h at RT and half of the DCM was removed under a stream of compressed air at RT. Meanwhile a solution of $Na_2S_2O_3$*$5H_2O$ (49 g, 200 mmol) and $NaHCO_3$ (38 g, 452 mmol) in H_2O (500 mL) was prepared. The reaction mixture was diluted with MTBE (100 mL) and the previously prepared solution (150 mL). The mixture was stirred for 30 min at RT. The phases were separated, and the aq. phase was extracted with MTBE (2x100 mL). The combined organic phases were washed with sat. $NaHCO_3$ (100 mL) and half-saturated brine (150 mL). The extracts were dried (Na_2SO_4) and conc. *in vacuo*. The desired aldehyde was obtained as a colourless oil (9.02 g, 20.3 mmol, 95% yield). The crude product was used immediately (!) without further purification. This procedure was adapted from the literature.[327]

Ohira-Bestmann Homologation:

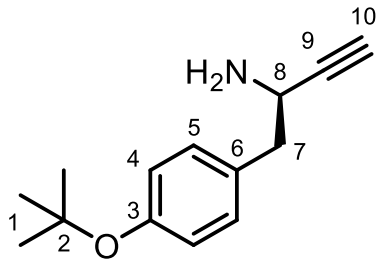
The Fmoc-protected α-amino alkyne **150** was synthesised according to general procedure B from the previously prepared crude aldehyde (9.02 mg, 20.3 mmol), which was added in anh. MeOH (50.0 mL) to the homologation reaction. The crude product was used without further purification.

Reprotection:

The obtained residue was suspended in DCM (150 mL). DIPEA (3.94 mL, 30.5 mmol, 1.5 eq.) and Fmoc-ONSU (4.52 g, 13.4 mmol, 0.66 eq.) were added. The mixture was stirred overnight at RT, was conc. *in vacuo,* and was suspended in EtOAc (150 mL). The organic phase was washed with aq. HCl (1 M, 3x100 mL) and sat. $NaHCO_3$ (100 mL). The combined organic extracts were dried (Na_2SO_4) and conc. *in vacuo*. The crude product was purified chromatographically (NP, EtOAc/Hex=5/95, R_F(EtOAc/Hex=1/9)=0.17). The desired Fmoc-protected α-amino alkyne **150** was obtained as an off-white, amorphous solid (2.23 g, 5.08 mmol, 25% yield rel. to the respective aldehyde). This procedure was adapted from the literature.[271]

Deprotection:

Fmoc-protected α-amino alkyne **150** (2.23 g, 5.08 mmol) and octadecyl mercaptan (14.6 g, 50.8 mmol, 10 eq.) were dissolved in THF (100 mL). DBU (76.0 µL, 508 µmol, 10 mol%) was added. The mixture was stirred for 1 h at RT. The mixture was conc. *in vacuo* and purified chromatographically (Washed apolar impurities off with DCM, then MeOH/DCM=2/98 R_F=0.12): The desired product was isolated as a pale-yellow oil (1.06 g, 4.88 mmol, 96% yield and 22% yield over five steps). This procedure was adapted from the literature.[286]



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.88 (bs, 2H, N*H*$_2$), 7.19-7.13 (m, 2H, *H*5), 6.94-6.88 (m, 2H, *H*4), 4.10-3.97 (m, 1H, *H*8), 3.12-2.94 (m, 2H, *H*7), 2.42 (s, 1H, *H*10), 1.32 (s, 9H, *H*1).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 167.8 (*C*3), 154.8 (*C*6), 130.3 (*C*5), 124.3 (*C*4), 81.5 (*C*9), 78.6 (*C*2), 75.0 (*C*10), 44.3 (*C*8), 40.5 (*C*7), 29.0 (*C*1).

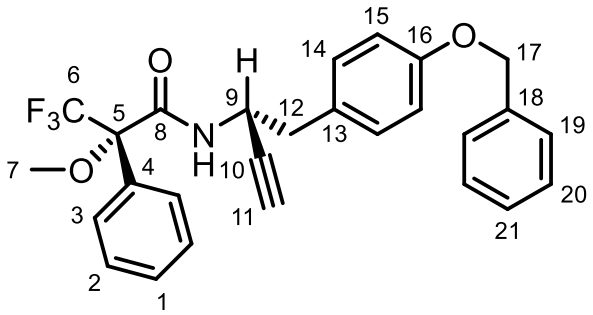
IR: v = 3285 (≡C-H), 1505 (N-H deform.), 1160 (C-O-C), 895 (=C-H deform.).

LC/MS (Method 3, 220 nm, ESI+): t_R = 7.4 min, m/z: 218.2 [M+H$^+$] (71), 145.1 [M-$C_4H_9O^-$] (100).

HRMS (ESI+): $C_{14}H_{20}NO^+$ [M+H$^+$] calc.: 218.1539 found: 218.1538.

152 (BNH-157) (*S*)-*N*-((*R*)-1-(4-(Benzyloxy)phenyl)but-3-yn-2-yl)-3,3,3-trifluoro-2-methoxy-2-phenylpropanamide

This Mosher amide was synthesised according to general procedure F from α-amino alkyne **146** (50.0 mg, 174 µmol). The crude product was purified chromatographically (RP, HPLC, H_2O+FA/MeCN+FA=7/3 over 2 CV, 7/3 to 5/95 over 10 CV, 5/95 over 2 CV with 18 mL/min): The desired product was isolated as a clear, sticky, gum-like solid (58 mg, 125 µmol, 72% yield).



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.56-7.30 (m, 10H, *H*1-3+19-21), 7.25-7.19 (m, 2H, *H*15), 7.04-6.98 (m, 1H, N*H*), 6.98-6.93 (m, 2H, *H*14), 5.10-4.99 (m, 3H, *H*9+17), 3.28 (s, 3H, *H*7), 3.05-2.94 (m, 2H, *H*12), 2.35-2.29 (m, 1H, *H*11).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 165.5 (*C*8), 158.2 (*C*16), 137.2 (*C*5), 132.2, 130.9, 129.7, 128.7, 128.7, 128.1, 127.9, 127.6 (*C*1-3+13+15+19-21), 124.8 (*C*4), 122.9 (*C*18), 114.9 (*C*14), 84.1 (*C*6), 82.0 (*C*10), 72.8 (*C*11), 70.1 (*C*17), 55.0 (*C*7), 42.8 (*C*9), 40.3 (*C*12).

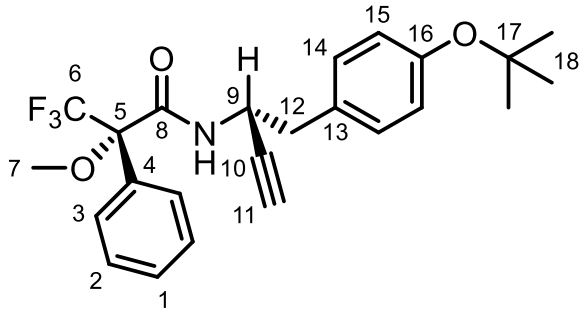
IR: v = 3408 (N-H), 3294 (≡C-H), 1687 (C=O), 1510 (N-H deform.), 1156 (C-O-C), 712 (=C-H deform.).

LC/MS (Method 3, 220 nm, ESI+): t_R = 12.7 min, m/z: 485.2 [M+NH_4^+] (39), 468.2 [M+H^+] (100).

HRMS (ESI+): $C_{27}H_{24}F_3NNaO_3^+$ [M+Na^+] calc.: 490.1600 found: 490.1601.

154 (BNH-158) (*S*)-*N*-((*R*)-1-(4-(*tert*-Butoxy)phenyl)but-3-yn-2-yl)-3,3,3-trifluoro-2-methoxy-2-phenylpropanamide

This Mosher amide was synthesised according to general procedure F from α-amino alkyne **149** (50.0 mg, 230 µmol). The crude product was purified chromatographically (RP, HPLC, H_2O+FA/MeCN+FA=8/2 over 2 CV, 8/2 to 5/95 over 10 CV, 5/95 over 2 CV with 18 mL/min): The desired product was isolated as a clear, sticky, gum-like solid (58 mg, 135 µmol, 58% yield).



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.54-7.34 (m, 5H, *H*1-3), 7.21-7.13 (m, 2H, *H*14), 7.02-6.91 (m, 3H, N*H*+*H*15), 5.09-4.97 (m, 2H, *H*9), 3.29-3.20 (m, 3H, *H*7), 3.08-2.94 (m, 2H, *H*12), 2.35-2.28 (m, 1H, *H*11), 1.33 (s, 9H, *H*18).

^{13}C-NMR (101 MHz, $CDCl_3$): δ = 165.5 (*C*8), 154.6 (*C*16), 132.2 (*C*4), 130.7 (*C*13), 130.2 (*C*14), 129.7 (*C*1), 128.7 (*C*2), 127.8 (*C*3), 125.2 (*C*5), 124.2 (*C*15), 84.0 (*C*6), 82.0 (*C*10), 78.5 (*C*17), 72.6 (*C*11), 55.0 (*C*7), 42.5 (*C*9), 40.4 (*C*12), 28.9 (*C*18).

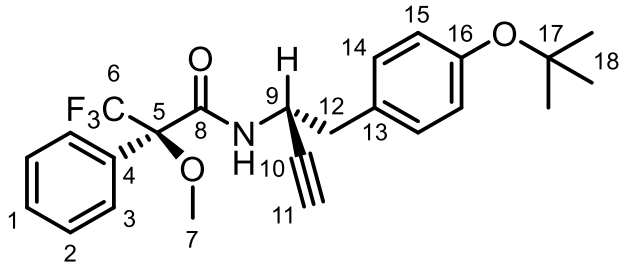
IR: v = 3307 (≡C-H), 2978 (CH_3), 1687 (C=O), 1505 (N-H deform.), 1155 (C-O-C), 717 (=C-H deform.).

LC/MS (Method 3, 220 nm, ESI+): t_R = 12.5 min, m/z: 451.2 [M+NH_4^+] (43), 434.2 [M+H^+] (100).

HRMS (ESI+): $C_{24}H_{26}F_3NNaO_3^+$ [M+Na^+] calc.: 456.1757 found: 456.1757.

155 (BNH-159) (*R*)-*N*-((*R*)-1-(4-(*tert*-Butoxy)phenyl)but-3-yn-2-yl)-3,3,3-trifluoro-2-methoxy-2-phenylpropanamide

This Mosher amide was synthesised according to general procedure F from α-amino alkyne **149** (50.0 mg, 230 µmol). The crude product was purified chromatographically (RP, HPLC, H_2O+FA/MeCN+FA=8/2 over 2 CV, 8/2 to 5/95 over 10 CV, 5/95 over 2 CV with 18 mL/min): The desired product was isolated as a clear, sticky, gum-like solid (67 mg, 154 µmol, 67% yield).



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.54-7.32 (m, 5H, *H*1-3), 7.02-6.94 (m, 2H, *H*14), 6.89-6.82 (m, 2H, *H*15), 6.82-6.72 (m, 1H, N*H*), 5.10-4.98 (m, 1H, *H*9), 3.43-3.34 (m, 3H, *H*7), 2.99-2.84 (m, 2H, *H*12), 2.36-2.27 (m, 1H, *H*11), 1.33 (s, 9H, *H*18).

^{13}C-NMR (101 MHz, $CDCl_3$): δ = 165.5 (*C*8), 154.6 (*C*16), 132.5 (*C*4), 130.5 (*C*13), 130.2 (*C*14), 129.6 (*C*1), 128.7 (*C*2), 127.6 (*C*3), 125.1 (*C*5), 124.0 (*C*15), 83.9 (*C*6), 81.9 (*C*10), 78.5 (*C*17), 72.7 (*C*11), 55.1 (*C*7), 42.3 (*C*9), 40.4 (*C*12), 29.0 (*C*18).

IR: v = 3291 (≡C-H), 2980 (CH_3), 1674 (C=O), 1505 (N-H deform.), 1154 (C-O-C), 696 (=C-H deform.).

LC/MS (Method 3, 220 nm, ESI+): t_R = 12.6 min, m/z: 451.2 [M+NH_4^+] (45), 434.2 [M+H^+] (100).

HRMS (ESI+): $C_{24}H_{26}F_3NNaO_3^+$ [M+Na^+] calc.: 456.1757 found: 456.1757.

95 (BNH-166) 3-((2*R*,5*R*,8*R*,11*S*,*Z*)-11-(2-Amino-2-oxoethyl)-5-(4-aminobutyl)-2-(4-hydroxybenzyl)-4,7,10-trioxo-1^1*H*-3,6,9-triaza-1(4,1)-triazolacycloundecaphane-8-yl)propanoic acid

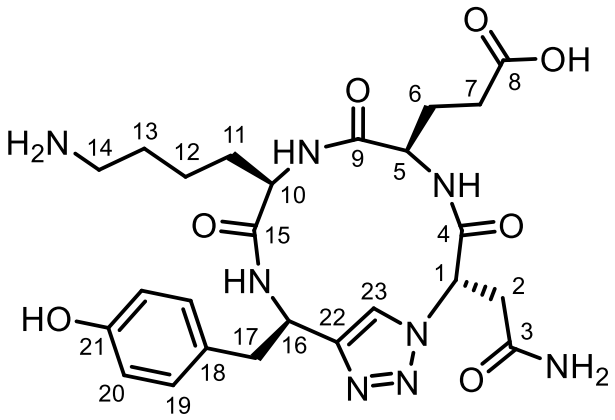
<u>Linear Pseudo Tetrapeptide:</u>

This click cyclic tetrapeptide was synthesised according to general procedure D with CTC resin (1.00 g, 1.55 mmol/g). The crude product was purified chromatographically

(NP, TEA/MeOH/DCM=1/3/96, R_F=0.09): The desired azido tripeptide **163** was obtained as an off-white, amorphous solid (1.25 g, 1.53 mmol, 99% yield). It was coupled to the α-amino alkyne **149** in anh. MeCN. The desired pseudo tetrapeptide **166** was obtained as an off-white, amorphous solid (1.13 g, 1.12 mmol, 73% yield). The crude product was used without further purification.

Click Cyclisation and Deprotection:

The pseudo tetrapeptide **166** (100 mg, 98.7 µmol) was cyclised for 48 h and the crude cyclisation product was deprotected. The crude deprotection product was purified chromatographically (NP, HPLC, aq. NH_4OAc (10 mM)/MeCN=5/95 to 95/5 over 14 CV with 18 mL/min): The desired click cyclic tetrapeptide **95** was obtained as a white, amorphous solid (27.9 mg, 50.0 µmol, 50% yield rel. to the pseudo tetrapeptide **166**).



^{1}H-NMR (600 MHz, DMSO-d_6): δ = 8.99-8.81 (m, 1H, Glu-N*H*), 8.81-8.64 (m, 1H, *H*23), 8.60-8.42 (m, 1H, Tyr-N*H*), 8.29-8.10 (m, 1H, Lys-N*H*), 7.57 (s, 1H, Asn-NH_2), 7.06 (d, $^3J_{19\text{-}20}$=7.9 Hz, 2H, *H*19), 6.92 (s, 1H, Lys-H_2), 6.65 (d, $^3J_{19\text{-}20}$=7.5 Hz, 2H, *H*20), 5.54-5.40 (m, 1H, *H*16), 4.96-4.83 (m, 1H, *H*1), 4.09-3.91 (m, 2H, *H*5+10), 3.30-3.18 (m, 1H, *H*17a), 3.13-2.92 (m, 5H, *H*2+14+17b), 2.30-1.80 (m, 4H, *H*6+7), 1.69-1.35 (m, 4H, *H*11+13*)*, 1.29-0.96 (m, 2H, *H*12).

^{13}C-NMR (151 MHz, DMSO-d_6): δ = 171.1 (*C*4), 170.7 (*C*3), 170.59 (*C*8), 170.56 (*C*9), 166.0 (*C*15), 155.7 (*C*21), 151.2 (*C*22), 130.0 (*C*19), 128.6 (*C*18), 124.0 (*C*23), 114.9 (*C*20), 59.9 (*C*16), 56.1 (*C*5), 54.6 (*C*10), 49.4 (*C*1), 38.3 (*C*2), 33.3 (*C*17), 29.6 (*C*13), 28.2 (*C*6), 26.6 (*C*11), 22.1 (*C*12).

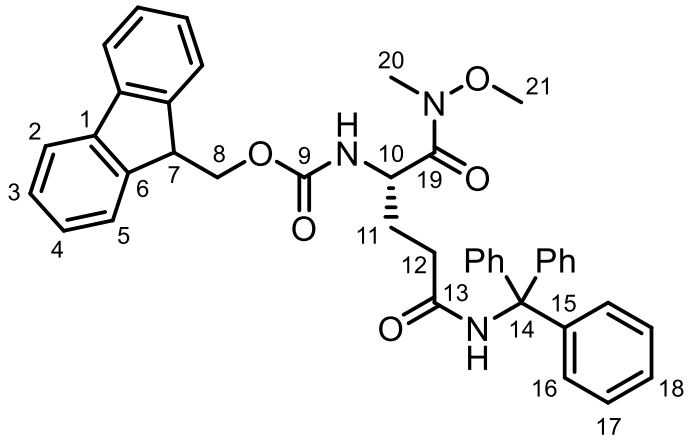
IR: v = 3426 (NH_3^+), 3287 (≡C-H), 1656 (C=O), 1512 (N-H deform.), 1399 (CO in COO^-), 730 (=C-H).

LC (Method 6, 220 nm): t_R = 16.0 min.

HRMS (ESI+): $C_{25}H_{35}N_8O_7^+$ [M+H$^+$] calc.: 559.2623 found: 559.2623.

148 (BNH-167) (9*H*-Fluoren-9-yl)methyl (*S*)-(1-(methoxy(methyl)amino)-1,5-dioxo-5-(tritylamino)pentan-2-yl)carbamate

(*S*)-Fmoc-Gln(Trt)-OH (1.0 g, 1.64 mmol) was dissolved in DCM (20 mL). DIPEA 613 µL, 3.60 mmol, 2.2 eq.) and PyBOP (937 mg, 1.80 mmol, 1.1 eq.) were added. The solution was stirred for 15 min at RT and *N,O*-Dimethylhydroxylamine HCl (192 mg, 1.96 mmol, 1.2 eq.) was added. The solution was stirred overnight at RT. The mixture was washed with aq. HCl (0.5 M, 3x100 mL) and sat. $NaHCO_3$ (100 mL) The mixture was dried (Na_2SO_4) and conc. *in vacuo*. The crude product was purified chromatographically (NP, $CHCl_3$): The desired product was obtained as a white, amorphous solid (974 mg, 1.49 mmol, 91% yield). This procedure was adapted from the literature.[271] The spectroscopic data agrees with the literature.[328]



^{1}H-NMR (400 MHz, DMSO-d$_6$): δ = 8.59 (s, 1H, δ-N*H*), 7.93-7.85 (m, 2H, *H*5), 7.78-7.71 (m, 2H, *H*2), 7.69-7.57 (m, 1H, α-N*H*), 7.45-7.38 (m, 2H, *H*4), 7.36-7.30 (m, 2H, *H*3), 7.30-7.22 (m, 6H, *H*17), 7.22-7.13 (m, 9H, *H*16+18), 4.49 (m, 1H, *H*10), 4.36-4.17 (m, 3H, *H*7+8), 3.66 (s, 3H, *H*21), 3.10 (s, 3H, *H*20), 2.44-2.26 (m, 2H, *H*12), 1.87-1.61 (m, 2H, *H*11).

^{13}C-NMR (101 MHz, DMSO-d$_6$): δ = 171.2 (*C*13), 156.1 (*C*9), 144.9 (*C*15), 143.9 (*C*19), 143.8 (*C*1), 140.7 (*C*6), 128.5 (*C*16), 127.7 (*C*4), 127.5 (*C*17), 127.1 (*C*3), 126.3 (*C*18), 125.4 (*C*2), 120.1 (*C*5), 69.2 (*C*14), 65.7 (*C*7), 61.1 (*C*21), 50.5 (*C*10), 46.7 (*C*8), 32.4 (*C*12), 31.8 (*C*20), 26.9 (*C*11).

IR: v = 3306 (N-H), 3051 (=C-H), 1715, 1653 (C=O), 1490 (ring vibr.), 1246 (C-O-C), 759, 740 (=C-H deform.).

LC/MS (Method 3, 220 nm, ESI+): t_R = 13.7 min, m/z: 671.3 [M+NH_4^+] (43), 654.3 [M+H^+] (100).

HRMS (ESI+): $C_{41}H_{39}N_3NaO_5^+$ (M+Na^+) calc.: 676.2782 found: 676.2782.

151 (BNH-169) (*S*)-4-Amino-*N*-tritylhex-5-ynamide

Reduction Homologation Sequence:

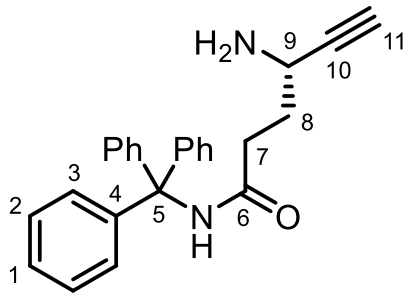
This α-amino alkyne was synthesised according to general procedure B from (*S*)-Fmoc-Gln(Trt)-OH (8.4 g, 13.8 mmol). The reduction yielded a mixture (8.81 g) of the desired aldehyde and starting material. The crude aldehyde was added in anh. MeOH (56.0 mL) to the homologation reaction. The crude product was used without further purification.

Reprotection:

The obtained residue was suspended in DCM (140 mL). DIPEA (2.40 mL, 13.8 mmol, 1.0 eq. rel. to acid) and Fmoc-ONSU (2.32 g, 6.88 mmol, 0.50 eq. rel. to acid) were added. The mixture was stirred overnight at RT. The mixture was conc. *in vacuo* and suspended in EtOAc (50 mL). The organic phase was washed with aq. HCl (1 M, 50 mL) and sat. $NaHCO_3$ (50 mL). The combined organic extracts were dried (Na_2SO_4) and conc. *in vacuo*. The crude product was purified chromatographically (NP, EtOAc/Hex=2/8, R_F(EtOAc/Hex=3/7)=0.32). The desired Fmoc-protected α-amino alkyne was obtained with impurities (2.55 g). This procedure was adapted from the literature.[271]

Deprotection:

Fmoc-protected α-amino alkyne (2.55 g, 4.32 mmol assuming 100% purity) and octadecylmercaptan (12.4 g, 43.2 mmol, 10 eq.) were dissolved in THF (100 mL). DBU (64.6 µL, 432 µmol, 10 mol%) was added. The mixture was stirred for 1 h at RT. The mixture was conc. *in vacuo* and purified chromatographically (Washed apolar impurities off with DCM, then MeOH/DCM=3/97 R_F(MeOH/DCM=5/95)=0.33): The desired α-amino alkyne 151 was isolated as a white, amorphous solid (541 mg, 1.47 mmol, 10% yield over four steps). This procedure was adapted from the literature.[286]



^{1}H-NMR (600 MHz, CD_3OD): δ = 8.51 (s, 1H, α-N*H*), 7.33-7.20 (m, 15H, *H*1-3), 4.03-3.90 (m, 1H, *H*9), 3.21-3.17 (m,1H, *H*11), 2.72-2.60 (m, 2H, *H*7), 2.12-1.83 (m, 2H, *H*8).

^{13}C-NMR (151 MHz, CD_3OD): δ = 173.5 (*C*6), 145.9 (*C*4), 130.0 (*C*3), 128.7 (*C*2), 127.8 (*C*1), 79.8 (*C*10), 77.9 (*C*11), 71.7 (*C*5), 43.4 (*C*9), 33.1 (*C*7), 30.4 (*C*8).

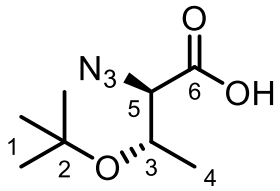
IR: v = 3274 (≡C-H), 1641 (C=O), 1531 (N-H deform.), 1490 (ring vibration), 764, 698 (=C-H deform.).

LC/MS (Method 2, 220 nm, ESI+): t_R = 5.7 min, m/z: 369.2 [M+H$^+$] (100).

HRMS (ESI-): $C_{25}H_{25}N_2O^+$ [M+H$^+$] calc.: 369.1961 found: 369.1961.

136 (BNH-170) (2*R*,3*S*)-2-Azido-3-(*tert*-butoxy)butanoic acid

This α-azido acid was synthesised according to general procedure C from (2*R*,3*S*)-H_2N-Thr(tBu)-OH (9.90 g, 56.5 mmol). The crude product was purified chromatographically (NP, AcOH/EtOAc/Hex=1/9/90, R_F(AcOH/EtOAc/Hex=1/19/80)= 0.15, stain: bromocresol green): The desired product was obtained as a white, amorphous solid (11.3 g, 55.9 mmol, 99% yield).



^{1}H-NMR (400 MHz, $CDCl_3$): δ = 9.92 (bs, 1H, COO*H*), 4.23-4.09 (m, 1H, *H*3), 3.75-3.59 (m, 1H, *H*5), 1.30-1.25 (m, 3H, *H*4), 1.21 (s, 9H, *H*1).

^{13}C-NMR (101 MHz, $CDCl_3$): δ = 173.9 (*C*6), 75.6 (*C*2), 68.8 (*C*3), 66.2 (*C*5), 28.3 (*C*1), 20.1 (*C*4).

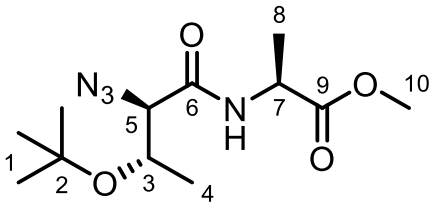
IR: v = 2977 (C-H), 2117 (-N_3), 1718 (C=O), 1230 (C-O-C).

LC/MS (Method 2, 220 nm, ESI-): t_R = 6.7 min, m/z: 401.2 [M+M-H$^+$] (100), 199.9 [M-H$^+$] (72).

HRMS (ESI-): $C_8H_{14}N_3O_3^-$ [M-H$^+$] calc.: 200.1041 found: 200.1040.

138 (BNH-171) Methyl ((2*R*,3*S*)-2-azido-3-(*tert*-butoxy)butanoyl)-L-alaninate

This azido dipeptide was synthesised according to general procedure E from α-azido acid **136** (100 mg, 497 µmol). The crude product was purified chromatographically (RP, HPLC, H_2O+FA/MeCN+FA=95/5 over 2 CV, 95/5 to 5/95 over 10 CV, 5/95 over 2 CV with 18 mL/min): The desired product was obtained as a white, amorphous solid (118 mg, 412 µmol, 83% yield).



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.33-7.19 (m, 1H, N*H*), 4.57-4.49 (m, 1H, *H*7), 3.93-3.87 (m, 1H, *H*3), 3.87-3.82 (m, 1H, *H*5), 3.70 (s, 3H, *H*10), 1.42-1.34 (m, 3H, *H*8), 1.23-1.09 (m, 12H, *H*1+4).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 172.9 (*C*9), 167.7 (*C*6), 75.3 (*C*2), 68.2 (*C*3), 67.7 (*C*5), 52.4 (*C*10), 48.1 (*C*7), 28.2 (*C*1), 19.1 (*C*4), 18.5 (*C*8).

IR: v = 3302 (N-H), 2978 (CH_3), 2104 (-N_3), 1741, 1654 (C=O), 1210 (C-O-C).

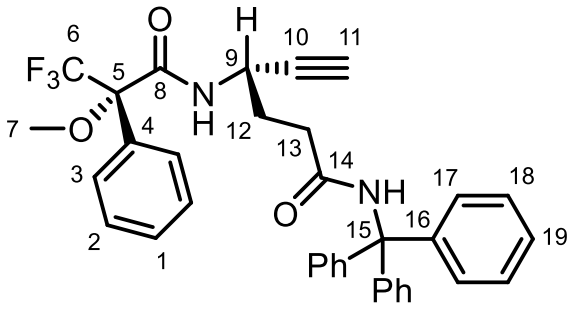
LC/MS (Method 2, 220 nm, ESI+): t_R = 7.6 min, m/z: 309.1 [M+Na$^+$] (13), 287.1 [M+H$^+$] (6), 231.0 [M-$C_4H_7^-$] (100).

HRMS (ESI+): $C_{12}H_{22}N_4NaO_4^+$ [M+Na$^+$] calc.: 309.1533 found: 309.1533.

156 (BNH-175) (*S*)-4-((*S*)-3,3,3-Trifluoro-2-methoxy-2-phenylpropanamido)-*N*-tritylhex-5-ynamide

This Mosher amide was synthesised according to general procedure F from α-amino alkyne **151** (50 mg, 136 µmol). The crude product was purified chromatographically

(RP, HPLC, H_2O+FA/MeCN+FA=6/4 over 2 CV, 6/4 to 5/95 over 10 CV, 5/95 over 2 CV with 18 mL/min): The desired product was isolated as a clear, sticky, gum-like solid (53 mg, 90.7 µmol, 66% yield).



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.55-7.48 (m, 2H, *H*2), 7.45-7.40 (m, 1H, *H*1), 7.40-7.33 (m, 2H, *H*3), 7.31-7.27 (m, 6H, *H*18), 7.27-7.24 (m, 3H, *H*19), 7.22-7.16 (m, 6H, *H*17), 6.67 (s, 1H, δ-N*H*), 4.82-4.72 (m, 1H, *H*9), 3.40 (s, 3H, *H*7), 2.47-2.36 (m, 1H, *H*13a), 2.36-2.27 (m, 2H, *H*13b+11), 2.08-1.92 (m, 2H, *H*12).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 170.6 (*C*14), 166.0 (*C*8), 144.7 (*C*16), 132.7 (*C*5), 129.6 (*C*4), 128.8 (*C*3), 128.71 (*C*1), 128.65 (*C*17), 128.1 (*C*18), 127.5 (*C*2), 127.1 (*C*19), 83.8 (*C*6), 81.6 (*C*10), 72.5 (*C*1), 70.8 (*C*15), 55.1 (*C*7), 41.0 (*C*9), 33.5 (*C*13), 30.8 (*C*12).

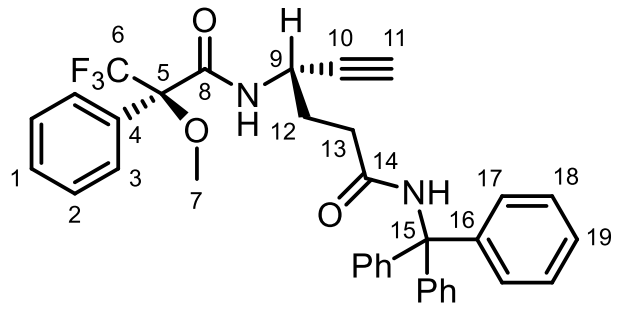
IR: v = 3415 (N-H), 3298 (≡C-H), 3057 (=C-H), 1668 (C=O), 1490 (ring virb.), 1155 (C-O-C), 766 (=C-H).

LC/MS (Method 2, 254 nm, ESI+): t_R = 11.8 min, m/z: 607.3 [M+Na$^+$] (65), 585.4 [M+H$^+$] (33), 243.0 [Trt$^+$] (100).

HRMS (ESI+): $C_{35}H_{31}F_3N_2NaO_3^+$ [M+Na$^+$] calc.: 607.2179 found: 607.2176.

157 (BNH-176) (*S*)-4-((*R*)-3,3,3-Trifluoro-2-methoxy-2-phenylpropanamido)-*N*-tritylhex-5-ynamide

This Mosher amide was synthesised according to general procedure F from α-amino alkyne **151** (50 mg, 136 µmol). The crude product was purified chromatographically (RP, HPLC, H_2O+FA/MeCN+FA=6/4 over 2 CV, 6/4 to 5/95 over 10 CV, 5/95 over 2 CV with 18 mL/min): The desired product was isolated as a white, amorphous solid (54 mg, 92.4 µmol, 68% yield).



^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.55-7.50 (m, 2H, *H*2), 7.45-7.39 (m, 3H, *H*1+3), 7.32-7.27 (m, 6H, *H*18), 7.27-7.23 (m, 3H, *H*19), 7.23-7.18 (m, 6H, *H*17), 6.79 (s, 1H, δ-N*H*), 4.82-4.73 (m, 1H, *H*9), 3.30 (s, 3H, *H*7), 2.54-2.37 (m, 2H, *H*13), 2.34-2.29 (m, 1H, *H*11), 2.15-1.98 (m, 2H, *H*12).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 170.7 (*C*14), 165.9 (*C*8), 144.7 (*C*16), 132.0 (*C*5), 129.7 (*C*4), 128.8 (*C*1), 128.71 (*C*3), 128.65 (*C*17), 128.1 (*C*2), 128.0 (*C*18), 127.1 (*C*19), 84.1 (*C*6), 81.8 (*C*10), 72.3 (*C*11), 70.8 (*C*15), 55.0 (*C*7), 41.1 (*C*9), 33.5 (*C*13), 31.0 (*C*12).

IR: v = 3298 (≡C-H), 1668 (C=O), 1490 (ring vibr.), 1154 (C-O-C), 765 (=C-H).

LC/MS (Method 2, 254 nm, ESI+): t_R = 11.8 min, m/z: 607.3 [M+Na$^+$] (47), 585.4 [M+H$^+$] (27), 243.0 [Trt$^+$] (100).

HRMS (ESI+): $C_{35}H_{31}F_3N_2NaO_3^+$ [M+Na$^+$] calc.: 607.2179 found: 607.2176.

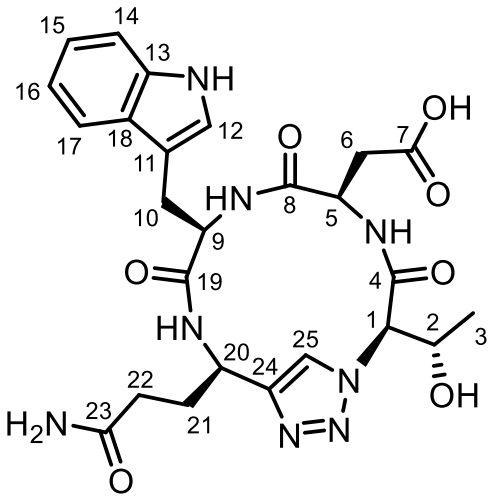
96 (BNH-177) 2-((11*R*,2*R*,5*R*,8*R*,11*R*,*Z*)-5-((1*H*-Indol-3-yl)methyl)-2-(3-amino-3-oxopropyl)-11-((*S*)-1-hydroxyethyl)-4,7,10-trioxo-1^1*H*-3,6,9-triaza-1(4,1)-triazolacycloundecaphane-8-yl)acetic acid

Linear Pseudo Tetrapeptide:

This click cyclic tetrapeptide was synthesised according to general procedure D with CTC resin (1.00 g, 1.55 mmol/g). The crude product was purified chromatographic-ally (NP, TEA/MeOH/DCM=1/1/98 to 1/3/96, R_F(TEA/MeOH/DCM= 1/3/96)=0.25): The desired azido tripeptide **164** was obtained as an off-white, amorphous solid (921 mg, 1.40 mmol, 90% yield). It was coupled to the α-amino alkyne **151** in anh. MeCN, yielding the desired pseudo tetrapeptide **167** was obtained as a white, amorphous solid (1.13 g, 1.16 mmol, 82% yield). The crude product was used without further purification.

Click Cyclisation and Deprotection:

The pseudo tetrapeptide **167** (100 mg, 99.1 µmol) was cyclised for 72 h. The crude cyclisation product was deprotected. The crude deprotection product was purified chromatographically (NP, HPLC, aq. NH_4OAc (10 mM)/MeCN=5/95 over 5 CV, 5/95 to 6/4 over 10 CV, 6/4 to 95/5 over 2 CV with 18 mL/min): The desired click cyclic tetrapeptide **96** was obtained as a white, amorphous solid (11.5 mg, 20.7 µmol, 20% yield rel. to pseudo tetrapeptide **167**).



^{1}H-NMR (600 MHz, DMSO-d_6): δ = 12.44 (bs, 1H, COO*H*), 10.77 (s, 1H, Trp-δ-N*H*), 8.40-8.26 (m, 1H, Gln-α-N*H*), 8.18-7.98 (m, 1H, *H*25), 7.72 (bs, 1H, Asp-N*H*), 7.58-7.44 (m, 1H, *H*17), 7.38-7.26 (m, 1H, *H*14), 7.19 (s, 1H, Gln-δ-N*H*$_2$), 7.12-7.00 (m, 2H, *H*12+15), 7.00-6.91 (m, 1H, *H*16), 6.71 (s, 1H, Trp-α-N*H*), 5.65 (bs, 1H, Thr-O*H*), 5.14-5.02 (m, 1H, *H*20), 4.93-4.78 (m, 1H, *H*1), 4.52-4.41 (m, 1H, *H*2), 4.41-4.29 (m, 1H, *H*5), 3.21-3.07 (m, 1H, *H*6a), 2.83-2.67 (m, 1H, *H*6b), 2.36-2.25 (m, 1H, *H*21a), 2.25-2.13 (m, 2H, *H*22), 2.00-1.87 (m, 1H, *H*21b), 1.21 (bs, 3H, *H*3).

^{13}C-NMR (151 MHz, DMSO-d_6): δ = 173.6 (*C*23), 171.1 (*C*19), 165.2 (*C*4), 150.1 (*C*24), 135.9 (*C*13), 127.2 (*C*18), 122.7 (*C*12), 121.4 (*C*25), 120.9 (*C*15), 118.21 (*C*16), 118.18 (*C*17), 111.2 (*C*14), 109.8 (*C*11), 72.6 (*C*1), 63.6 (*C*2), 53.5 (*C*5), 45.8 (*C*20), 31.8 (*C*22), 26.5 (*C*6), 25.2 (*C*21), 20.1 (*C*3).

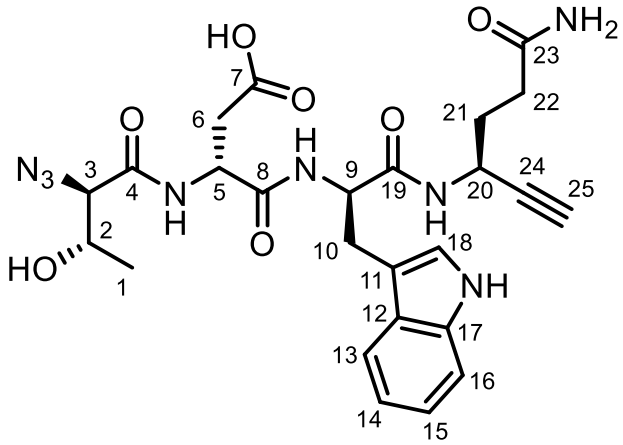
IR: v = 3426 (O-H), 3291 (N-H), 1652 (C=O), 1527 (N-H deform.), 730 (=C-H).

LC/MS (Method 2, 220 nm, ESI+): t_R = 4.3 min, m/z: 555.3 [M+H$^+$] (100).

HRMS (ESI+): $C_{25}H_{30}N_8NaO_7^+$ [M+Na$^+$] calc.: 577.2130 found: 577.2129.

171 (BNH-178) (*R*)-4-(((*R*)-1-(((*S*)-6-Amino-6-oxohex-1-yn-3-yl)amino)-3-(1*H*-indol-3-yl)-1-oxopropan-2-yl)amino)-3-((2*R*,3*S*)-2-azido-3-hydroxybutanamido)-4-oxobutanoic acid

This unprotected, linear pseudo tetrapeptide was synthesised from the protected pseudo tetrapeptide **167** (100 mg, 99.1 µmol) according to the deprotection protocol in general procedure D. The crude product was purified chromatographically (NP, HPLC, aq. NH_4OAc (10 mM)/MeCN=5/95 over 5 CV, 5/95 to 6/4 over 10 CV, 6/4 to 95/5 over 3 CV, 95/5 over 2 CV with 18 mL/min). The desired product was obtained as white, amorphous solid (26.0 mg, 46.9 µmol, 47% yield).



^{1}H-NMR (600 MHz, DMSO-d_6): δ = 10.79 (s, 1H, Trp-δ-N*H*), 8.48-8.39 (m, 1H, Asp-N*H*), 8.39-8.32 (m, 1H, Gln-N*H*), 8.10-8.01 (m, 1H, Trp-α-N*H*), 7.63-7.55 (m, 1H, *H*13), 7.35-7.29 (m, 1H, *H*16), 7.25 (bs, 1H, Gln-α-N*H*), 7.15-7.10 (m, 1H, *H*18), 7.09-7.02 (m, 1H, *H*15), 7.00-6.94 (m, 1H, *H*14), 6.73 (bs, 1H, Gln-δ-NH_2), 4.65-4.56 (m, 1H, *H*5), 4.56-4.46 (m, 1H, *H*20), 4.46-4.39 (m, 1H, *H*9), 3.94-3.85 (m, 1H, *H*2), 3.52-3.46 (m, 1H, *H*3), 3.20-3.16 (m,1H, *H*25), 3.14-3.06 (m, 1H, *H*10a), 2.99-2.90 (m, 1H, *H*10b), 2.72-2.63 (m, 1H, *H*6a), 2.51-2.43 (m, 1H, *H*6b), 2.11-2.05 (m, 2H, *H*22), 1.79-1.67 (m, 2H, *H*21), 1.04-0.99 (m, 3H, *H*1).

^{13}C-NMR (151 MHz, DMSO-d_6): δ = 173.4 (*C*23), 171.9 (*C*7), 170.5 (*C*19), 170.2 (*C*8), 168.1 (*C*4), 136.0 (*C*17), 127.2 (*C*12), 123.5 (*C*18), 120.8 (*C*15), 118.4 (*C*13), 118.2 (*C*14), 111.2 (*C*16), 109.7 (*C*11), 83.7 (*C*24), 73.2 (*C*25), 67.3 (*C*2), 67.2 (*C*3), 53.8 (*C*9), 49.5 (*C*5), 40.0 (*C*20), 36.5 (*C*6), 31.2 (*C*22), 30.6 (*C*21), 27.7 (*C*10), 19.9 (*C*1).

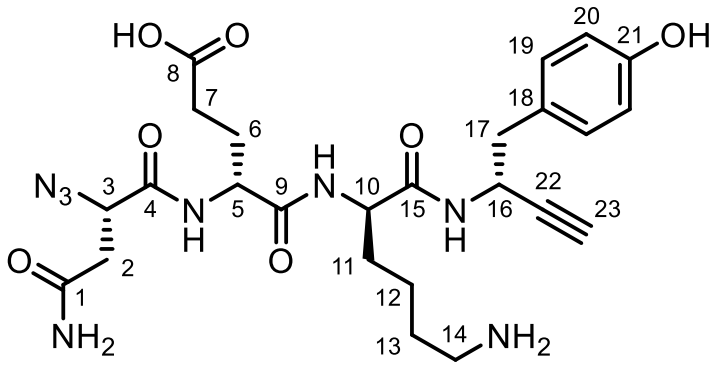
IR: v = 3260 (≡C-H), 2111 (-N_3), 1651 (C=O), 1515 (N-H), 1393 (CO in COO^-), 1231 (C-N), 1096 (C-O), 744 (=C-H).

LC (Method 6, 220 nm): t_R = 8.6 min.

HRMS (ESI+): $C_{25}H_{30}N_8NaO_7^+$ [M+Na$^+$] calc.: 577.2130 found: 577.2129.

170 (BNH-179) (*R*)-5-(((*R*)-6-Amino-1-(((*R*)-1-(4-hydroxyphenyl)but-3-yn-2-yl)amino)-1-oxohexan-2-yl)amino)-4-((*S*)-4-amino-2-azido-4-oxobutanamido)-5-oxopentanoic acid

This unprotected, linear pseudo tetrapeptide was synthesised from the protected pseudo tetrapeptide **166** (100 mg, 98.7 µmol) according to the deprotection protocol in general procedure D. The crude product was purified chromatographically (NP, HPLC, aq. NH_4OAc (10 mM)/MeCN=5/95 over 2 CV, 5/95 to 95/5 over 10 CV, 95/5 over 2 CV with 18 mL/min). The desired product was obtained as a white, amorphous solid (28.0 mg, 50.1 µmol, 50% yield).



^{1}H-NMR (600 MHz, DMSO-d_6): δ = 9.44-9.30 (m, 1H, Lys-α-N*H*), 8.28-8.07 (m, 2H, Tyr-N*H*+Glu-N*H*), 7.69-7.54 (m, 1H, Asn-γ-NH_2), 7.13-6.95 (m, 3H, Lys-ε-NH_2+*H*19), 6.72-6.59 (m, 2H, *H*20), 4.64-4.50 (m, 1H, *H*3), 4.35-4.25 (m, 1H, *H*16), 4.23-4.07 (m, 2H, *H*5+10), 3.10-3.07 (m, 1H, *H*23), 2.77-2.56 (m, 6H, *H*2+14+17), 2.22-2.04 (m, 2H, *H*7), 1.95-1.75 (m, 2H, *H*6), 1.65-1.38 (m, 4H, *H*11+13), 1.38-1.19 (m, 2H, *H*12).

^{13}C-NMR (151 MHz, DMSO-d_6): δ = 176.3 (*C*8), 170.8 (*C*4), 170.7 (*C*9), 170.5 (*C*1), 169.2 (*C*15), 156.1 (*C*21), 130.2 (*C*19), 127.0 (*C*18), 114.9 (*C*20), 83.5 (*C*22), 73.5 (*C*23), 59.1 (*C*16), 53.6 (*C*5), 52.5 (*C*10), 42.1 (*C*3), 40.2 (*C*2), 38.5 (*C*14), 36.7 (*C*17), 33.4 (*C*7), 30.7 (*C*11), 27.7 (*C*6), 27.2 (*C*13), 22.3 (*C*12).

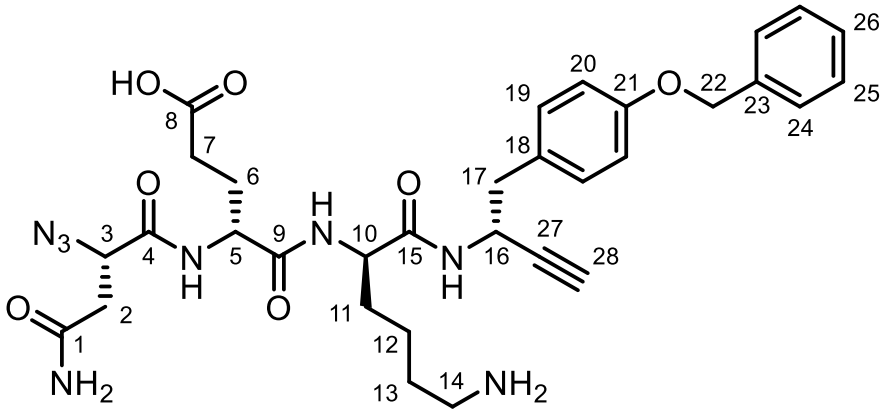
IR: v = 3271 (≡C-H), 2412 (O-H), 2113 (-N_3), 1629 (C=O), 1537, 1513 (N-H deform.), 1396 (CO in COO^-), 1231 (C-N).

LC (Method 6, 220 nm): t_R = 13.3 min.

HRMS (ESI+): $C_{25}H_{34}N_8O_7^+$ [M+H$^+$] calc.: 559.2623 found: 559.2620.

169 (BNH-180) (*R*)-5-(((*R*)-6-Amino-1-(((*R*)-1-(4-(benzyloxy)phenyl)but-3-yn-2-yl)amino)-1-oxohexan-2-yl)amino)-4-((*S*)-4-amino-2-azido-4-oxobutanamido)-5-oxopentanoic acid

This unprotected, linear pseudo tetrapeptide was synthesised from the protected pseudo tetrapeptide **165** (100 mg, 95.5 µmol) according to the deprotection protocol in general procedure D. The crude product was purified chromatographically (NP, HPLC, aq. NH_4OAc (10 mM)/MeCN=5/95 over 2 CV, 5/95 to 95/5 over 10 CV, 95/5 over 2 CV with 18 mL/min). The desired product was obtained as white, amorphous solid (24.0 mg, 37.0 µmol, 38% yield).



^{1}H-NMR (600 MHz, DMSO-d$_6$): δ = 9.44-9.27 (m, 1H, Lys-α-N*H*), 8.31-8.19 (m, 1H, Tyr-N*H*), 8.19-8.09 (m, 1H, Glu-N*H*), 7.70-7.56 (m, 1H, Asn-γ-N*H*$_2$), 7.47-7.41 (m, 2H, *H*24), 7.41-7.35 (m, 2H, *H*25), 7.35-7.29 (m, 1H, *H*26), 7.19-7.12 (m, 2H, *H*19), 7.11-7.03 (m, 1H, Lys-ε-N*H*$_2$), 6.95-6.87 (m, 2H, *H*20), 5.06 (s, 2H, *H*22), 4.66-4.55 (m, 1H, *H*3), 4.35-4.27 (m, 1H, *H*16), 4.32-4.10 (m, 2H, *H*5+10), 3.13-3.07 (m, 1H, *H*28), 2.87-2.57 (m, 6H, *H*2+14+17), 2.26-2.04 (m, 2H, *H*7), 1.96-1.77 (m, 2H, *H*6), 1.69-1.41 (m, 4H, *H*11+13), 1.41-1.13 (m, 2H, *H*12).

^{13}C-NMR (151 MHz, DMSO-d$_6$): δ = 176.1 (*C*8), 170.9 (*C*4), 170.7 (*C*9), 170.6 (*C*1), 169.3 (*C*15), 157.1 (*C*21), 137.2 (*C*23), 130.4 (*C*19), 129.3 (*C*18), 128.4 (*C*25), 127.7 (*C*26), 127.6 (*C*24), 114.4 (*C*20), 83.3 (*C*27), 73.8 (*C*28), 69.1 (*C*22), 59.1 (*C*16), 53.6 (*C*5) 52.5 (*C*10), 42.1 (*C*3), 40.1 (*C*2), 38.5 (*C*14), 36.8 (*C*17), 33.1 (*C*7), 30.6 (*C*11), 27.5 (*C*6), 27.0 (*C*13), 22.3 (*C*12).

IR: v = 3276 (≡C-H), 3063 (=C-H), 2115 (-N$_3$), 1632 (C=O), 1513 (N-H deform.), 1393

(C-N), 1243 (C-O-C).

LC/MS (Method 2, 254 nm, ESI+): t_R = 5.4 min, m/z: 649.4 [M+H$^+$] (100).

HRMS (ESI+): $C_{32}H_{40}N_8NaO_7^+$ [M+Na$^+$] calc.: 671.2912 found: 671.2912.

3.4 Biochemical Assays

All biological tests were performed by Lead Discovery Center in Dortmund. A detailed description of the assay procedures is included here for completeness. Upon arrival each compound was dissolved in DMSO and stored at RT.

The nucleotide exchange assay protocol was copied, with slight modifications, from BENARY *et al*. in 2025.[49]

Nucleotide Exchange Assay

Inactive, GDP-bound K-RAS is incubated with SOS1 and GTP. K-RAS is transferred to its active GTP-bound state which leads to a release of GDP. The K-RAS bound GTP is hydrolysed to GDP even in the absence of the corresponding GAP protein. In the GDP GloBioluminescent GDP detection assay for glycosyltransferases (Promega, Madison, WI, USA) used, GDP is converted to ATP which can be quantified using a luciferase/luciferin reaction. The resulting luminescence signal is then measured with a suitable microplate reader. This assay was developed for K-RAS wild type protein as well as for the G12D and G12V mutant. For every sample, 5 µL K-RAS working solution in assay buffer (50 mM HEPES pH 7.5, 4 mM MgCl2, 2 mM EGTA, 0.01% Brij35, 1 mM TCEP) were transferred into a suitable assay plate (e. g. Greiner #784075). The test compound was added with an echo acoustic dispenser (Beckman Coulter, Brea, CA, USA) in a concentration range from 3000 µM to 3 µM (8-point dilution). After addition of the test compound 5 µL of SOS1-GTP mix in assay buffer were added. The reaction mix was incubated overnight at room temperature followed by the addition of 10 µL GPD detection reagent. After a second incubation period of 1 h at room temperature the luminescence signal was measured with an Envision spectrophotometer (Perkin Elmer, Waltham, MA, USA). IC_{50} values were determined from the sigmoidal dose response curves with the software Quattro Workflow (Quattro GmbH, Munich, Germany). To filter out compounds which interfere with the nucleotide exchange assay independent from K-RAS and SOS1, a control assay was developed. For this, GDP was titrated and detected with the GDP Glo bioluminescent GDP detection assay for

glycosyltransferases (Promega, Madison, WI, USA) to generate a luminescence signal comparable to the positive control of the nucleotide exchange assay. Compounds were checked for assay interference by performing dose response curves in the same concentration range as for the nucleotide exchange assay. For every sample,10 µL GDP (125 nM final assay concentration) in assay buffer (50 mM HEPES pH 7.5, 4 mM $MgCl_2$, 2 mM EGTA, 0.01% Brij35, 1 mM TCEP) was transferred into a suitable assay plate (e. g. Greiner #784075). The test compound was added with an echo acoustic dispenser (Beckman Coulter, Brea, CA, USA) in a concentration range from 3000 µM to 3 µM (8-point dilution). The reaction mix was incubated over night at room temperature followed by the addition of 10 µL GPD detection reagent. After a second incubation period of 1 h at room temperature the luminescence signal was measured with an Envision spectrophotometer (Perkin Elmer, Waltham, MA, USA). IC_{50} values were determined from the sigmoidal dose response curves with the software Quattro Work-flow (Quattro GmbH, Munich, Germany).

The CellTiter-Glo assay protocol was copied, with slight modifications, from JEUKEN *et al*. in 2022.[96]

CellTiter-Glo Assay

The CellTiter-Glo Luminescent Cell Viability Assay (Promega) is a homogeneous method of determining the number of viable cells in culture. It is based on quantification of ATP, indicating the presence of metabolically active cells. On day 1 25 µL of the cell suspension are seeded at a cell number that assure assay linearity and optimal signal intensity. After incubation for 24 h at 37°C/5% CO_2-compounds dissolved in DMSO are added at different concentrations by Echo Liquid Handling Technology. Cells are further incubated in humidified chambers for 72 h at 37°C and 5% CO_2. Cells treated with the compound vehicle DMSO are used as positive controls and cells treated with 10 µM staurosporine serve as negative controls. At day 5–72 h after compound addition - the CellTiter Glo Reagent is prepared according to the instructions of the kit (Promega Inc.). Thereon, mixture and assay plates are equilibrated at room temperature for 20 min. Equal volumes of the reagent-medium-mixture is added to the volume of culture medium present in each well. The plates are mixed at ~300 rpm for 2 minutes on an orbital shaker. The microplates are then incubated at room temperature for 10 minutes for stabilization of the luminescent signal. Following

incubation, the luminescence is recorded on a Victor microplate reader (Perkin Elmer) using a 200 ms integration time. The data is then analysed with Excel using the XLFIT Plugin (dose response Fit 205) for IC_{50}-determination. As quality control the Z'-factor is calculated from 16 positive and negative control values. Only assay results showing a Z'-factor ≥0.5 are used for further analysis.

4 Abbreviations

AA	Amino acid
AcOH	Acetic acid
AI	Artificial intelligence
ASL	Atom selection language
CADD	Computer-Aided Drug Design
CataCXium® PtB	2-(Di-tert-butyl-phosphino)-1-phenyl-1H-pyrrole
CCF	Cyclopentyl chloroformate
CDI	1,1'-Carbonyldiimidazole
CPC	Cyclopentyl carbamate
CRD	Cysteine-rich domain
CTC	2-Chlorotrityl chloride
CuAAC	Cu-catalysed azide-alkyne cycloaddition
CV	Column volume *or* cross validation
DARPin	Designed ankyrin repeat protein
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DCM	Dichloromethane
ΔG_{bind}	Binding free energy
Cyctetpep	Cyclic tetrapeptides
Cyctripep	Cyclic tripeptides
DARPins	Designed ankyrin repeat proteins
DIBAL-H	Diisobutylaluminium hydride
DIC	*N*,*N*′-Diisopropylcarbodiimide
DIPEA	*N,N*-Diisopropylethylamine
DMA	*N*,*N*-Dimethylacetamide
DMAP	4-Dimethylaminopyridine
DMF	*N*,*N*-Dimethylformamide
DMP	Dess-Martin periodinane
DMTMM BF_4	4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium tetrafluoroborate
dppf	1,1'-bis(diphenylphosphino)ferrocene
ECFP	Extended connectivity fingerprint or Morgan fingerprint
EDCI	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	Ethylenediaminetetraacetate
EGFR	Epithelial growth factor receptor
ESI	Electrospray ionisation
eq.	Equivalents
FA	Formic acid
FDA	Food and drug administration
Fmoc-ONSU	*N*-(9-Fluorenylmethoxycarbonyloxy)succinimide
GDP	Guanosine 5′-diphosphate
GEF	Guanine nucleotide exchange factor
GTP	Guanosine 5′-triphosphate
GUI	Graphical user interface
h	Hours
HATU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate
HBA	Hydrogen bond acceptors
HBD	Hydrogen bond donors

Hex	Cyclohexane
HILIC	Hydrophilic Interaction Liquid Chromatography
HOBt	Hydroxybenzotriazole
HPLC	High performance liquid chromatography
HVR	Hypervariable region
Hz	Hertz
IC_{50}	Half-maximal inhibitory concentration
iPrOH	Isopropanol
J	Coupling constant
K_D	Dissociation constant
KRAS	Kirsten rat sarcoma virus GTPase
KNN	K-nearest neighbours
LDC	Lead Discovery Center
MAPK	Mitogen-activated protein kinase
MBMB	Methyl 4-(bromomethyl)-3-methoxybenzoate
MC	Methyl carbamate
MCF	Methyl chloroformate
MDS	Molecular dynamics simulation
MeCN	Acetonitrile
µM	Micromolar
MMGBSA	Molecular mechanics generalized born surface area
µL	Microliter
ms	Milliseconds
MSA	Methanesulfonamide
MTBE	Methyl tert-butyl ether
mTORC2	Mammalian target of rapamycin complex 2
MW	Molecular weight
nM	Nanomolar
NRB	Number of rotatable bonds
t_R	Retention time
LC/MS	Liquid chromatography mass spectrometry
MeOH	Methanol
min	Minutes
NBS	*N*-Bromosuccinimide
NF1	Neurofibromin
NIS	*N*-Iodosuccinimide
NMM	*N*-Methylmorpholine
PEG200	Polyethylene glycol with mean molar mass of 200 g/mol
PIP	Piperidine
PIP_2	Phosphatidylinositol-4,5-bisphosphate
PIP_3	Phosphatidylinositol (3,4,5)-trisphosphate
PI3K	Phosphoinositide 3-kinase
PMI	Principal moment of inertia
PPI	Protein-protein interaction
PPII	Protein-protein interaction inhibitor
ppm	parts per million
PyBOP	Benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate
pAKT	Phosphorylated AKT
pERK	Phosphorylated ERK

QSAR	Quantitative structure activity relationship
RAF	Rapidly accelerated fibrosarcoma kinase
RAL	RAS-like GTPase
RALGDS	RAS-like guanine nucleotide dissociation stimulator protein
RBD	RAS binding domain
RF	Random forest
R_F	Retardation factor
RMSE	Root mean square error
RO5	Lipinski's rule of five
RO4	Rule of four
SAR	Structure activity relationship
SOS	Son of Sevenless
SPPS	Solid-phase peptide synthesis
SVR	Support vector
tAKT	Total AKT
TBAC	Tetrabutylammonium chloride
TBAI	Tetrabutylammonium iodide
TBHP	*t*-Butyl hydroperoxide
TBTA	Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine
TEA	Triethylamine
tERK	Total ERK
Tf	Triflyl, trifluoromethyl
TFA	Trifluoroacetic acid
THF	Tetrahydrofurane
TIPS	Triisopropyl silane
TLC	Thin layer chromatography
TMP	2,2,6,6-Tetramethylpiperidine
TMS	Trimethylsilyl
TSA	*o*-Toluenesulfonamide
t-SNE	t-Distributed stochastic neighbour embedding
v	Wavenumber

5 References

(1) Kwan, A. K.; Piazza, G. A.; Keeton, A. B.; Leite, C. A. The Path to the Clinic: A Comprehensive Review on Direct KRAS^{G12C} Inhibitors. *J. Exp. Clin. Cancer Res.* **2022**, *41* (1), 27. DOI: 10.1186/s13046-021-02225-w.

(2) Punekar, S. R.; Velcheti, V.; Neel, B. G.; Wong, K. K. The Current State of the Art and Future Trends in RAS-Targeted Cancer Therapies. *Nat. Rev. Clin. Oncol.* **2022**, *19* (10), 637–655. DOI: 10.1038/s41571-022-00671-9.

(3) Smith, M. D.; Quarles, L. D.; Demerdash, O.; Smith, J. C. Drugging the Entire Human Proteome: Are We There Yet? *Drug Discov. Today* **2024**, *29* (3), 103891. DOI: 10.1016/j.drudis.2024.103891.

(4) Xiong, D.; Qiu, Y.; Zhao, J.; Zhou, Y.; Lee, D.; Gupta, S.; Torres, M.; Lu, W.; Liang, S.; Kang, J. J.; Eng, C.; Loscalzo, J.; Cheng, F.; Yu, H. A Structurally Informed Human Protein-Protein Interactome Reveals Proteome-Wide Perturbations Caused by Disease Mutations. *Nat. Biotechnol.* **2024**. DOI: 10.1038/s41587-024-02428-4.

(5) Nolan, A.; Raso, C.; Kolch, W.; von Kriegsheim, A.; Wynne, K.; Matallanas, D. Proteomic Mapping of the Interactome of KRAS Mutants Identifies New Features of RAS Signalling Networks and the Mechanism of Action of Sotorasib. *Cancers* **2023**, *15* (16), 4141. DOI: 10.3390/cancers15164141.

(6) Moine-Franel, A.; Mareuil, F.; Nilges, M.; Ciambur, C. B.; Sperandio, O. A Comprehensive Dataset of Protein-Protein Interactions and Ligand Binding Pockets for Advancing Drug Discovery. *Sci. Data* **2024**, *11* (1), 402. DOI: 10.1038/s41597-024-03233-z.

(7) Wu, S.; Gao, X.; Wu, D.; Liu, L.; Yao, H.; Meng, X.; Zhang, X.; Bai, F. Motif-Guided Identification of KRAS-Interacting Proteins. *BMC Biol.* **2024**, *22* (1), 264. DOI: 10.1186/s12915-024-02067-w.

(8) Weng, C.; Faure, A. J.; Escobedo, A.; Lehner, B. The Energetic and Allosteric Landscape for KRAS Inhibition. *Nature* **2024**, *626* (7999), 643–652. DOI: 10.1038/s41586-023-06954-0.

(9) Keskin, O.; Gursoy, A.; Ma, B.; Nussinov, R. Principles of Protein–Protein Interactions: What Are the Preferred Ways for Proteins to Interact? *Chem. Rev.* **2008**, *108* (4), 1225–1244. DOI: 10.1021/cr040409x.

(10) Rennella, E.; Henry, C.; Dickson, C. J.; Georgescauld, F.; Wales, T. E.; Erdmann, D.; Cotesta, S.; Maira, M.; Sedrani, R.; Brachmann, S. M.; Ostermann, N.; Engen, J. R.; Kay, L. E.; Beyer, K. S.; Wilcken, R.; Jahnke, W. Dynamic Conformational Equilibria in the Active States of KRAS and NRAS. *RSC Chem. Biol.* **2025**, *6* (1), 106–118. DOI: 10.1039/d4cb00233d.

(11) Vithani, N.; Zhang, S.; Thompson, J. P.; Patel, L. A.; Demidov, A.; Xia, J.; Balaeff, A.; Mentes, A.; Arnautova, Y. A.; Kohlmann, A.; Lawson, J. D.; Nicholls, A.; Skillman, A. G.; LeBard, D. N. Exploration of Cryptic Pockets Using Enhanced Sampling Along Normal Modes: A Case Study of KRAS^{G12D}. *J. Chem. Inf. Model.* **2024**, *64* (21), 8258–8273. DOI: 10.1021/acs.jcim.4c01435.

(12) Feng, H.; Zhang, Y.; Bos, P. H.; Chambers, J. M.; Dupont, M. M.; Stockwell, B. R. K-Ras^{G12D} Has a Potential Allosteric Small Molecule Binding Site. *Biochemistry* **2019**, *58* (21), 2542–2554. DOI: 10.1021/acs.biochem.8b01300.

(13) Wells, J. A.; McClendon, C. L. Reaching for High-Hanging Fruit in Drug Discovery at Protein-Protein Interfaces. *Nature* **2007**, *450* (7172), 1001–1009. DOI: 10.1038/nature06526.

(14) Sperandio, O.; Reynès, C. H.; Camproux, A. C.; Villoutreix, B. O. Rationalizing the Chemical Space of Protein-Protein Interaction Inhibitors. *Drug Discov. Today* **2010**, *15* (5–6), 220–229. DOI: 10.1016/j.drudis.2009.11.007.
(15) Morelli, X.; Bourgeas, R.; Roche, P. Chemical and Structural Lessons from Recent Successes in Protein-Protein Interaction Inhibition (2P2I). *Curr. Opin. Chem. Biol.* **2011**, *15* (4), 475–481. DOI: 10.1016/j.cbpa.2011.05.024.
(16) Ikeda, K.; Maezawa, Y.; Yonezawa, T.; Shimizu, Y.; Tashiro, T.; Kanai, S.; Sugaya, N.; Masuda, Y.; Inoue, N.; Niimi, T.; Masuya, K.; Mizuguchi, K.; Furuya, T.; Osawa, M. DLiP-PPI Library: An Integrated Chemical Database of Small-to-Medium-Sized Molecules Targeting Protein-Protein Interactions. *Front. Chem.* **2023**, *10*, 1090643. DOI: 10.3389/fchem.2022.1090643.
(17) Bosc, N.; Muller, C.; Hoffer, L.; Lagorce, D.; Bourg, S.; Derviaux, C.; Gourdel, M. E.; Rain, J. C.; Miller, T. W.; Villoutreix, B. O.; Miteva, M. A.; Bonnet, P.; Morelli, X.; Sperandio, O.; Roche, P. Fr-PPIChem: An Academic Compound Library Dedicated to Protein-Protein Interactions. *ACS Chem. Biol.* **2020**, *15* (6), 1566–1574. DOI: 10.1021/acschembio.0c00179.
(18) Higueruelo, A. P.; Jubb, H.; Blundell, T. L. TIMBAL v2: Update of a Database Holding Small Molecules Modulating Protein-Protein Interactions. *Database* **2013**, *2013*, bat039. DOI: 10.1093/database/bat039.
(19) Torchet, R.; Druart, K.; Ruano, L. C.; Moine-Franel, A.; Borges, H.; Doppelt-Azeroual, O.; Brancotte, B.; Mareuil, F.; Nilges, M.; Ménager, H.; Sperandio, O. The IPPI-DB Initiative: A Community-Centered Database of Protein-Protein Interaction Modulators. *Bioinformatics* **2021**, *37* (1), 89–96. DOI: 10.1093/bioinformatics/btaa1091.
(20) Basse, M. J.; Betzi, S.; Morelli, X.; Roche, P. 2P2Idb v2: Update of a Structural Database Dedicated to Orthosteric Modulation of Protein-Protein Interactions. *Database* **2016**, *2016*. DOI: 10.1093/database/baw007.
(21) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* **2012**, *64*, 4–17. DOI: 10.1016/j.addr.2012.09.019.
(22) Dougherty, P. G.; Qian, Z.; Pei, D. Macrocycles as Protein-Protein Interaction Inhibitors. *Biochem. J.* **2017**, *474* (7), 1109–1125. DOI: 10.1042/bcj20160619.
(23) Tanada, M.; Tamiya, M.; Matsuo, A.; Chiyoda, A.; Takano, K.; Ito, T.; Irie, M.; Kotake, T.; Takeyama, R.; Kawada, H.; Hayashi, R.; Ishikawa, S.; Nomura, K.; Furuichi, N.; Morita, Y.; Kage, M.; Hashimoto, S.; Nii, K.; Sase, H.; Ohara, K.; Ohta, A.; Kuramoto, S.; Nishimura, Y.; Iikura, H.; Shiraishi, T. Development of Orally Bioavailable Peptides Targeting an Intracellular Protein: From a Hit to a Clinical KRAS Inhibitor. *J. Am. Chem. Soc.* **2023**, *145* (30), 16610–16620. DOI: 10.1021/jacs.3c03886.
(24) Sarojini, V.; Cameron, A. J.; Varnava, K. G.; Denny, W. A.; Sanjayan, G. Cyclic Tetrapeptides from Nature and Design: A Review of Synthetic Methodologies, Structure, and Function. *Chem. Rev.* **2019**, *119*, 10318–10359. DOI: 10.1021/acs.chemrev.8b00737.
(25) Stouten, P. F. W.; Sander, C.; Wittinghofer, A.; Valencia, A. How Does the Switch II Region of G-Domains Work? *FEBS Lett.* **1993**, *320* (1), 1–6. DOI: 10.1016/0014-5793(93)81644-f.
(26) Wennerberg, K.; Rossman, K. L.; Der, C. J. The Ras Superfamily at a Glance. *J. Cell Sci.* **2005**, *118* (5), 843–846. DOI: 10.1242/jcs.01660.

(27) Gray, J. L.; von Delft, F.; Brennan, P. E. Targeting the Small GTPase Superfamily through Their Regulatory Proteins. *Angew. Chem. Int. Ed.* **2020**, *59* (16), 6342–6366. DOI: 10.1002/anie.201900585.
(28) Pantsar, T. The Current Understanding of KRAS Protein Structure and Dynamics. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 189–198. DOI: 10.1016/j.csbj.2019.12.004.
(29) Whitley, M. J.; Tran, T. H.; Rigby, M.; Yi, M.; Dharmaiah, S.; Waybright, T. J.; Ramakrishnan, N.; Perkins, S.; Taylor, T.; Messing, S.; Esposito, D.; Nissley, D. V.; McCormick, F.; Stephen, A. G.; Turbyville, T.; Cornilescu, G.; Simanshu, D. K. Comparative Analysis of KRAS4a and KRAS4b Splice Variants Reveals Distinctive Structural and Functional Properties. *Sci. Adv.* **2024**, *10* (7), eadj4137. DOI: 10.1126/sciadv.adj4137.
(30) Nussinov, R.; Tsai, C. J.; Jang, H. Oncogenic Ras Isoforms Signaling Specificity at the Membrane. *Cancer Res.* **2018**, *78* (3), 593–602. DOI: 10.1158/0008-5472.CAN-17-2727.
(31) Pálfy, G.; Menyhárd, D. K.; Ákontz-Kiss, H.; Vida, I.; Batta, G.; Tőke, O.; Perczel, A. The Importance of Mg^{2+}-Free State in Nucleotide Exchange of Oncogenic K-Ras Mutants. *Chem. Eur. J.* **2022**, *28* (59), e202201449. DOI: 10.1002/chem.202201449.
(32) Hall, A.; Self, A. J. The Effect of Mg^{2+} on the Guanine Nucleotide Exchange Rate of $p21^{N\text{-}Ras}$. *J. Biol. Chem.* **1986**, *261* (24), 10963–10965. DOI: 10.1016/S0021-9258(18)67333-8.
(33) Scheffzek, K.; Ahmadian, M. R.; Kabsch, W.; Wiesmüller, L.; Lautwein, A.; Schmitz, F.; Wittinghofer, A. The Ras-RasGAP Complex: Structural Basis for GTPase Activation and Its Loss in Oncogenic Ras Mutants. *Science* **1997**, *277* (5324), 333–338. DOI: 10.1126/science.277.5324.333.
(34) Huang, L.; Guo, Z.; Wang, F.; Fu, L. KRAS Mutation: From Undruggable to Druggable in Cancer. *Signal Transduct. Target Ther.* **2021**, *6*, 386. DOI: 10.1038/s41392-021-00780-4.
(35) Rubinson, D. A.; Tanaka, N.; de la Cruz, F. F.; Kapner, K. S.; Rosenthal, M. H.; Norden, B. L.; Barnes, H.; Ehnstrom, S.; Morales-Giron, A. A.; Brais, L. K.; Lemke, C. T.; Aguirre, A. J.; Corcoran, R. B. Sotorasib Is a Pan-RAS^{G12C} Inhibitor Capable of Driving Clinical Response in $NRAS^{G12C}$ Cancers. *Cancer Discov.* **2024**, *14* (5), 727–736. DOI: 10.1158/2159-8290.CD-23-1138.
(36) Windrich, J.; Ney, G. M.; Rosenberg, P. S.; Kim, J.; Zenker, M.; Stewart, D. R.; Kratz, C. P. Cancer in Multilineage Mosaic RASopathies Due to Pathogenic Variants in HRAS or KRAS: A Systematic Review and Meta-Analysis. *Clin. Cancer Res.* **2024**, *30* (22), 5116–5121. DOI: 10.1158/1078-0432.CCR-24-1928.
(37) Prior, I. A.; Hood, F. E.; Hartley, J. L. The Frequency of Ras Mutations in Cancer. *Cancer Res.* **2020**, 80 (14), 2969–2974. DOI: 10.1158/0008-5472.CAN-19-3682.
(38) Prior, I. A.; Lewis, P. D.; Mattos, C. A Comprehensive Survey of Ras Mutations in Cancer. *Cancer Res.* **2012**, *72* (10), 2457–2467. DOI: 10.1158/0008-5472.CAN-11-2612.
(39) Cook, J. H.; Melloni, G. E. M.; Gulhan, D. C.; Park, P. J.; Haigis, K. M. The Origins and Genetic Interactions of KRAS Mutations Are Allele- and Tissue-Specific. *Nat. Commun.* **2021**, *12* (1), 1808. DOI: 10.1038/s41467-021-22125-z.

(40) Johnson, C.; Burkhart, D. L.; Haigis, K. M. Classification of KRAS Activating Mutations and the Implications for Therapeutic Intervention. *Cancer Discov.* **2022**, *12* (4), 913–923. DOI: 10.1158/2159-8290.CD-22-0035.
(41) Nussinov, R.; Jang, H. Direct K-Ras Inhibitors to Treat Cancers: Progress, New Insights, and Approaches to Treat Resistance. *Annu. Rev. Pharmacol. Toxicol.* **2024**, *64*, 231–253. DOI: 10.1146/annurev-pharmtox-022823-113946.
(42) Bahar, M. E.; Kim, H. J.; Kim, D. R. Targeting the RAS/RAF/MAPK Pathway for Cancer Therapy: From Mechanism to Clinical Studies. *Signal Transduct. Target. Ther.* **2023**, *8* (1), 455. DOI: 10.1038/s41392-023-01705-z.
(43) Parikh, K.; Banna, G.; Liu, S. V.; Friedlaender, A.; Desai, A.; Subbiah, V.; Addeo, A. Drugging KRAS: Current Perspectives and State-of-Art Review. *J. Hematol. Oncol.* **2022**, *15* (1), 152. DOI: 10.1186/s13045-022-01375-4.
(44) Zhou, Y.; Prakash, P.; Gorfe, A. A.; Hancock, J. F. Ras and the Plasma Membrane: A Complicated Relationship. *Cold Spring Harb. Perspect. Med.* **2017**, 8 (10), a031831. DOI: 10.1101/cshperspect.a031831.
(45) Baltanás, F. C.; Zarich, N.; Rojas-Cabañeros, J. M.; Santos, E. SOS GEFs in Health and Disease. *Biochim. Biophys. Acta, Rev. Cancer* **2020**, *1874* (2), 188445. DOI: 10.1016/j.bbcan.2020.188445.
(46) Schlessinger, J. How Receptor Tyrosine Kinases Activate Ras. *Trends Biochem. Sci.* **1993**, *18* (8), 273–275. DOI: 10.1016/0968-0004(93)90031-H.
(47) Kholodenko, B. N.; Hoek, J. B.; Westerhoff, H. V. Why Cytoplasmic Signalling Proteins Should Be Recruited to Cell Membranes. *Trends Cell Biol.* **2000**, *10* (5), 173–178. DOI: 10.1016/S0962-8924(00)01741-4.
(48) Boriack-Sjodin, P. A.; Margarit, S. M.; Bar-Sagi, D.; Kuriyan, J. The Structural Basis of the Activation of Ras by Sos. *Nature* **1998**, *394* (6691), 337–343. DOI: 10.1038/28548.
(49) Benary, G. E.; Kilgenstein, F.; Koller, S.; Scherkenbeck, J. Monophthalates of Betulinic Acid and Related Pentacyclic Triterpenes Inhibit Efficiently the SOS-Mediated Nucleotide Exchange and Impact PI3K/AKT Signaling in Oncogenic K-RAS4B Proteins. *RSC Adv.* **2025**, *15* (2), 883–895. DOI: 10.1039/D4RA08503E.
(50) Vetter, I. R.; Wittinghofer, A. The Guanine Nucleotide-Binding Switch in Three Dimensions. *Science* **2001**, *294* (5545), 1299–1304. DOI: 10.1126/science.1062023.
(51) Goody, R. S.; Frech, M.; Wittinghofer, A. Affinity of Guanine Nucleotide Binding Proteins for Their Ligands: Facts and Artefacts. *Trends Biochem. Sci.* **1991**, *16* (9), 327–328. DOI: 10.1016/0968-0004(91)90134-H.
(52) Green, M. Avidin and Streptavidin. *Methods in Enzymology,* Vol. 184; Academic Press, **1990**; pp 51–67. DOI: 10.1016/0076-6879(90)84259-J.
(53) Rudack, T.; Xia, F.; Schlitter, J.; Kötting, C.; Gerwert, K. The Role of Magnesium for Geometry and Charge in GTP Hydrolysis, Revealed by Quantum Mechanics/Molecular Mechanics Simulations. *Biophys. J.* **2012**, *103* (2), 293–203. DOI: 10.1016/j.bpj.2012.06.015.
(54) Mondal, S.; Hsiao, K.; Goueli, S. A. A Homogenous Bioluminescent System for Measuring GTPase, GTPase Activating Protein, and Guanine Nucleotide Exchange Factor Activities. *Assay Drug Dev. Technol.* **2015**, *13* (8), 444–455. DOI: 10.1089/adt.2015.643.
(55) Sun, Q.; Burke, J. P.; Phan, J.; Burns, M. C.; Olejniczak, E. T.; Waterson, A. G.; Lee, T.; Rossanese, O. W.; Fesik, S. W. Discovery of Small Molecules That

Bind to K-Ras and Inhibit Sos-Mediated Activation. *Angew. Chem. Int. Ed.* **2012**, *51* (25), 6140–6143. DOI: 10.1002/anie.201201358.

(56) Abbott, J. R.; Patel, P. A.; Howes, J. E.; Akan, D. T.; Kennedy, J. P.; Burns, M. C.; Browning, C. F.; Sun, Q.; Rossanese, O. W.; Phan, J.; Waterson, A. G.; Fesik, S. W. Discovery of Quinazolines That Activate SOS1-Mediated Nucleotide Exchange on RAS. *ACS Med. Chem. Lett.* **2018**, *9* (9), 941–946. DOI: 10.1021/acsmedchemlett.8b00296.

(57) Winter, J. J. G.; Anderson, M.; Blades, K.; Brassington, C.; Breeze, A. L.; Chresta, C.; Embrey, K.; Fairley, G.; Faulder, P.; Finlay, M. R. V.; Kettle, J. G.; Nowak, T.; Overman, R.; Patel, S. J.; Perkins, P.; Spadola, L.; Tart, J.; Tucker, J. A.; Wrigley, G. Small Molecule Binding Sites on the Ras:SOS Complex Can Be Exploited for Inhibition of Ras Activation. *J. Med. Chem.* **2015**, *58* (5), 2265–2274. DOI: 10.1021/jm501660t.

(58) Burns, M. C.; Sun, Q.; Daniels, R. N.; Camper, D.; Kennedy, J. P.; Phan, J.; Olejniczak, E. T.; Lee, T.; Waterson, A. G.; Rossanese, O. W.; Fesik, S. W. Approach for Targeting Ras with Small Molecules That Activate SOS-Mediated Nucleotide Exchange. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111* (9), 3401–3406. DOI: 10.1073/pnas.1315798111.

(59) Hillig, R. C.; Sautier, B.; Schroeder, J.; Moosmayer, D.; Hilpmann, A.; Stegmann, C. M.; Werbeck, N. D.; Briem, H.; Boemer, U.; Weiske, J.; Badock, V.; Mastouri, J.; Petersen, K.; Siemeister, G.; Kahmann, J. D.; Wegener, D.; Böhnke, N.; Eis, K.; Graham, K.; Wortmann, L.; von Nussbaum, F.; Bader, B. Discovery of Potent SOS1 Inhibitors That Block RAS Activation *via* Disruption of the RAS-SOS1 Interaction. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116* (7), 2551–2560. DOI: 10.1073/pnas.1812963116.

(60) Maurer, T.; Garrenton, L. S.; Oh, A.; Pitts, K.; Anderson, D. J.; Skelton, N. J.; Fauber, B. P.; Pan, B.; Malek, S.; Stokoe, D.; Ludlam, M. J. C.; Bowman, K. K.; Wu, J.; Giannetti, A. M.; Starovasnik, M. A.; Mellman, I.; Jackson, P. K.; Rudolph, J.; Wang, W.; Fang, G. Small-Molecule Ligands Bind to a Distinct Pocket in Ras and Inhibit SOS-Mediated Nucleotide Exchange Activity. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (14), 5299–5304. DOI: 10.1073/pnas.1116510109.

(61) Bos, J. L.; Rehmann, H.; Wittinghofer, A. GEFs and GAPs: Critical Elements in the Control of Small G Proteins. *Cell* **2007**, *129* (5), 865–877. DOI: 10.1016/j.cell.2007.05.018.

(62) John, J.; Schlichting, I.; Schiltz, E.; Rösch, P.; Wittinghofer, A. C-Terminal Truncation of $P21^{H}$ Preserves Crucial Kinetic and Structural Properties. *J. Biol. Chem.* **1989**, *264* (22), 13086–13092. DOI: 10.1016/S0021-9258(18)51599-4.

(63) Gideon, P.; John, J.; Frech, M.; Lautwein, A.; Clark, R.; Scheffler, J. E.; Wittinghofer, A. Mutational and Kinetic Analyses of the GTPase-Activating Protein (GAP)-p21 Interaction: The C-Terminal Domain of GAP Is Not Sufficient for Full Activity. *Mol. Cell. Biol.* **1992**, *12* (5), 2050–2056. DOI: 10.1128/mcb.12.5.2050-2056.1992.

(64) Pai, E. F.; Krengel, U.; Petsko, G. A.; Goody, R. S.; Kabsch, W.; Wittinghofer, A. Refined Crystal Structure of the Triphosphate Conformation of H-Ras p21 at 1.35 Å Resolution: Implications for the Mechanism of GTP Hydrolysis. *EMBO J.* **1990**, *9* (8), 2351–2359. DOI: 10.1002/j.1460-2075.1990.tb07409.x.

(65) Lu, S.; Jang, H.; Nussinov, R.; Zhang, J. The Structural Basis of Oncogenic Mutations G12, G13 and Q61 in Small GTPase K-Ras4B. *Sci. Rep.* **2016**, *6* (1), 21949. DOI: 10.1038/srep21949.

(66) Mozzarelli, A. M.; Simanshu, D. K.; Castel, P. Functional and Structural Insights into RAS Effector Proteins. *Mol. Cell* **2024**, *84* (15), 2807–2821. DOI: 10.1016/j.molcel.2024.06.027.
(67) Junk, P.; Kiel, C. Structure-Based Prediction of Ras-Effector Binding Affinities and Design of “Branchegetic” Interface Mutations. *Structure* **2023**, *31* (7), 870–883.e5. DOI: 10.1016/j.str.2023.04.007.
(68) Nussinov, R.; Tsai, C. J.; Jang, H. Is Nanoclustering Essential for All Oncogenic KRas Pathways? Can It Explain Why Wild-Type KRas Can Inhibit Its Oncogenic Variant? *Semin. Cancer Biol.* **2019**, *54*, 114–120. DOI: 10.1016/j.semcancer.2018.01.002.
(69) Lee, K. Y. Membrane-Driven Dimerization of the Peripheral Membrane Protein KRAS: Implications for Downstream Signaling. *Int. J. Mol. Sci.* **2024**, *25* (5), 2530. DOI: 10.3390/ijms25052530.
(70) Shaul, Y. D.; Seger, R. The MEK/ERK Cascade: From Signaling Specificity to Diverse Functions. *Biochim. Biophys. Acta, Mol. Cell Res.* **2007**, *1773* (8), 1213–1226. DOI: 10.1016/j.bbamcr.2006.10.005.
(71) Cruz-Migoni, A.; Canning, P.; Quevedo, C. E.; Bataille, C. J. R.; Bery, N.; Miller, A.; Russell, A. J.; Phillips, S. E. V.; Carr, S. B.; Rabbitts, T. H. Structure-Based Development of New RAS-Effector Inhibitors from a Combination of Active and Inactive RAS-Binding Compounds. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116* (7), 2545–2550. DOI: 10.1073/pnas.1811360116.
(72) Wang, X.; Allen, S.; Blake, J. F.; Bowcut, V.; Briere, D. M.; Calinisan, A.; Dahlke, J. R.; Fell, J. B.; Fischer, J. P.; Gunn, R. J.; Hallin, J.; Laguer, J.; Lawson, J. D.; Medwid, J.; Newhouse, B.; Nguyen, P.; O’Leary, J. M.; Olson, P.; Pajk, S.; Rahbaek, L.; Rodriguez, M.; Smith, C. R.; Tang, T. P.; Thomas, N. C.; Vanderpool, D.; Vigers, G. P.; Christensen, J. G.; Marx, M. A. Identification of MRTX1133, a Noncovalent, Potent, and Selective $KRAS^{G12D}$ Inhibitor. *J. Med. Chem.* **2022**, *65* (4), 3123–3133. DOI: 10.1021/acs.jmedchem.1c01688.
(73) Kessler, D.; Gmachl, M.; Mantoulidis, A.; Martin, L. J.; Zoephel, A.; Mayer, M.; Gollner, A.; Covini, D.; Fischer, S.; Gerstberger, T.; Gmaschitz, T.; Goodwin, C.; Greb, P.; Häring, D.; Hela, W.; Hoffmann, J.; Karolyi-Oezguer, J.; Knesl, P.; Kornigg, S.; Koegl, M.; Kousek, R.; Lamarre, L.; Moser, F.; Munico-Martinez, S.; Peinsipp, C.; Phan, J.; Rinnenthal, J.; Sai, J.; Salamon, C.; Scherbantin, Y.; Schipany, K.; Schnitzer, R.; Schrenk, A.; Sharps, B.; Siszler, G.; Sun, Q.; Waterson, A.; Wolkerstorfer, B.; Zeeb, M.; Pearson, M.; Fesik, S. W.; McConnell, D. B. Drugging an Undruggable Pocket on KRAS. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116* (32), 15823–15829. DOI: 10.1073/pnas.1904529116.
(74) Takács, T.; Kudlik, G.; Kurilla, A.; Szeder, B.; Buday, L.; Vas, V. The Effects of Mutant Ras Proteins on the Cell Signalome. *Cancer Metastasis Rev.* **2020**, *39* (4), 1051–1065. DOI: 10.1007/s10555-020-09912-8.
(75) Wortzel, I.; Seger, R. The ERK Cascade: Distinct Functions within Various Subcellular Organelles. *Genes Cancer* **2011**, *2* (3), 195–209. DOI: 10.1177/1947601911407328.
(76) acold, M. E.; Suire, S.; Perisic, O.; Lara-Gonzalez, S.; Davis, C. T.; Walker, E. H.; Hawkins, P. T.; Stephens, L.; Eccleston, J. F.; Williams, R. L. Crystal Structure and Functional Analysis of Ras Binding to Its Effector Phosphoinositide 3-Kinase γ. *Cell* **2000**, *103* (6), 931–944. DOI: 10.1016/S0092-8674(00)00196-3.
(77) Cuesta, C.; Arévalo-Alameda, C.; Castellano, E. The Importance of Being PI3K in the RAS Signaling Network. *Genes* **2021**, *12* (7), 1094.

DOI: 10.3390/genes12071094.
(78) Fruman, D. A.; Chiu, H.; Hopkins, B. D.; Bagrodia, S.; Cantley, L. C.; Abraham, R. T. The PI3K Pathway in Human Disease. *Cell* **2017**, *170* (4), 605–635. DOI: 10.1016/j.cell.2017.07.029.
(79) Lien, E. C.; Dibble, C. C.; Toker, A. PI3K Signaling in Cancer: Beyond AKT. *Curr. Opin. Cell Biol.* **2017**, *45*, 62–71. DOI: 10.1016/j.ceb.2017.02.007.
(80) Quevedo, C. E.; Cruz-Migoni, A.; Bery, N.; Miller, A.; Tanaka, T.; Petch, D.; Bataille, C. J. R.; Lee, L. Y. W.; Fallon, P. S.; Tulmin, H.; Ehebauer, M. T.; Fernandez-Fuentes, N.; Russell, A. J.; Carr, S. B.; Phillips, S. E. V.; Rabbitts, T. H. Small Molecule Inhibitors of RAS-Effector Protein Interactions Derived Using an Intracellular Antibody Fragment. *Nat. Commun.* **2018**, *9* (1), 3169. DOI: 10.1038/s41467-018-05707-2.
(81) Huang, L.; Hofer, F.; Martin, G. S.; Kim, S. H. Structural Basis for the Interaction of Ras with RalGDS. *Nat. Struct. Mol. Biol.* **1998**, *5* (6), 422–426. DOI: 10.1038/nsb0698-422.
(82) Vetter, I. R.; Linnemann, T.; Wohlgemuth, S.; Geyer, M.; Kalbitzer, H. R.; Herrmann, C.; Wittinghofer, A. Structural and Biochemical Analysis of Ras–Effector Signaling *via* RalGDS. *FEBS Lett.* **1999**, *451* (2), 175–180. DOI: 10.1016/S0014-5793(99)00555-4.
(83) Ferro, E.; Trabalzini, L. RalGDS Family Members Couple Ras to Ral Signalling and That's Not All. *Cell. Signal.* **2010**, *22* (12), 1804–1810. DOI: 10.1016/j.cellsig.2010.05.010.
(84) Yan, C.; Liu, D.; Li, L.; Wempe, M. F.; Guin, S.; Khanna, M.; Meier, J.; Hoffman, B.; Owens, C.; Wysoczynski, C. L.; Nitz, M. D.; Knabe, W. E.; Ahmed, M.; Brautigan, D. L.; Paschal, B. M.; Schwartz, M. A.; Jones, D. N. M.; Ross, D.; Meroueh, S. O.; Theodorescu, D. Discovery and Characterization of Small Molecules That Target the GTPase Ral. *Nature* **2014**, *515* (7527), 443–447. DOI: 10.1038/nature13713.
(85) Wu, Y.; Reiner, D. J. A Signalling Cascade for Ral. *Small GTPases* **2022**, *13* (1), 128–135. DOI: 10.1080/21541248.2021.1917953.
(86) van Dam, E. M.; Robinson, P. J. Ral: Mediator of Membrane Trafficking. *Int. J. Biochem. Cell Biol.* **2006**, *38* (11), 1841–1847. DOI: 10.1016/j.biocel.2006.04.006.
(87) Traut, T. W. Physiological Concentrations of Purines and Pyrimidines. *Mol. Cell. Biochem.* **1994**, *140* (1), 1–22. DOI: 10.1007/BF00928361.
(88) Hunter, J. C.; Gurbani, D.; Ficarro, S. B.; Carrasco, M. A.; Lim, S. M.; Choi, H. G.; Xie, T.; Marto, J. A.; Chen, Z.; Gray, N. S.; Westover, K. D. In Situ Selectivity Profiling and Crystal Structure of SML-8-73-1, an Active Site Inhibitor of Oncogenic K-Ras G12C. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111* (24), 8895–8900. DOI: 10.1073/pnas.1404639111.
(89) Xiong, Y.; Lu, J.; Hunter, J.; Li, L.; Scott, D.; Choi, H. G.; Lim, S. M.; Manandhar, A.; Gondi, S.; Sim, T.; Westover, K. D.; Gray, N. S. Covalent Guanosine Mimetic Inhibitors of G12C KRAS. *ACS Med. Chem. Lett.* **2017**, *8* (1), 61–66. DOI: 10.1021/acsmedchemlett.6b00373.
(90) Müller, M. P.; Jeganathan, S.; Heidrich, A.; Campos, J.; Goody, R. S. Nucleotide-Based Covalent Inhibitors of KRas Can Only Be Efficient *In Vivo* If They Bind Reversibly with GTP-like Affinity. *Sci. Rep.* **2017**, *7* (1), 3687. DOI: 10.1038/s41598-017-03973-6.

(91) Grant, B. J.; Lukman, S.; Hocker, H. J.; Sayyah, J.; Brown, J. H.; McCammon, J. A.; Gorfe, A. A. Novel Allosteric Sites on Ras for Lead Generation. *PLoS One* **2011**, *6* (10), e25711. DOI: 10.1371/journal.pone.0025711.
(92) Lanman, B. A.; Allen, J. R.; Allen, J. G.; Amegadzie, A. K.; Ashton, K. S.; Booker, S. K.; Chen, J. J.; Chen, N.; Frohn, M. J.; Goodman, G.; Kopecky, D. J.; Liu, L.; Lopez, P.; Low, J. D.; Ma, V.; Minatti, A. E.; Nguyen, T. T.; Nishimura, N.; Pickrell, A. J.; Reed, A. B.; Shin, Y.; Siegmund, A. C.; Tamayo, N. A.; Tegley, C. M.; Walton, M. C.; Wang, H.-L.; Wurz, R. P.; Xue, M.; Yang, K. C.; Achanta, P.; Bartberger, M. D.; Canon, J.; Hollis, L. S.; McCarter, J. D.; Mohr, C.; Rex, K.; Saiki, A. Y.; San Miguel, T.; Volak, L. P.; Wang, K. H.; Whittington, D. A.; Zech, S. G.; Lipford, J. R.; Cee, V. J. Discovery of a Covalent Inhibitor of $KRAS^{G12C}$ (AMG 510) for the Treatment of Solid Tumors. *J. Med. Chem.* **2020**, *63* (1), 52–65. DOI: 10.1021/acs.jmedchem.9b01180.
(93) Study of MRTX1133 in Patients With Advanced Solid Tumors Harboring a KRAS G12D Mutation. *ClinicalTrials.gov*. https://clinicaltrials.gov/study/NCT05737706 (accessed 2025-01-02).
(94) Vasta, J. D.; Peacock, D. M.; Zheng, Q.; Walker, J. A.; Zhang, Z.; Zimprich, C. A.; Thomas, M. R.; Beck, M. T.; Binkowski, B. F.; Corona, C. R.; Robers, M. B.; Shokat, K. M. KRAS Is Vulnerable to Reversible Switch-II Pocket Engagement in Cells. *Nat. Chem. Biol.* **2022**, *18* (6), 596–604. DOI: 10.1038/s41589-022-00985-w.
(95) Kessler, D.; Bergner, A.; Böttcher, J.; Fischer, G.; Döbel, S.; Hinkel, M.; Müllauer, B.; Weiss-Puxbaum, A.; McConnell, D. B. Drugging All RAS Isoforms With One Pocket. *Future Med. Chem.* **2020**, *12* (21), 1911–1923. DOI: 10.4155/fmc-2020-0221.
(96) Jeuken, S.; Shkura, O.; Röger, M.; Brickau, V.; Choidas, A.; Degenhart, C.; Gülden, D.; Klebl, B.; Koch, U.; Stoll, R.; Scherkenbeck, J. Synthesis, Biological Evaluation, and Binding Mode of a New Class of Oncogenic K-Ras4b Inhibitors. *ChemMedChem* **2022**, *17* (22). DOI: 10.1002/cmdc.202200392.
(97) Schöpel, M.; Jockers, K. F. G.; Düppe, P. M.; Autzen, J.; Potheraveedu, V. N.; Ince, S.; Yip, K. T.; Heumann, R.; Herrmann, C.; Scherkenbeck, J.; Stoll, R. Bisphenol A Binds to Ras Proteins and Competes with Guanine Nucleotide Exchange: Implications for GTPase-Selective Antagonists. *J. Med. Chem.* **2013**, *56* (23), 9664–9672. DOI: 10.1021/jm401291q.
(98) Zhou, X.; Ji, Y.; Zhou, J. Multiple Strategies to Develop Small Molecular KRAS Directly Bound Inhibitors. *Molecules* **2023**, *28* (8), 3615. DOI: 10.3390/molecules28083615.
(99) Rosnizeck, I. C.; Graf, T.; Spoerner, M.; Tränkle, J.; Filchtinski, D.; Herrmann, C.; Gremer, L.; Vetter, I. R.; Wittinghofer, A.; König, B.; Kalbitzer, H. R. Stabilizing a Weak Binding State for Effectors in the Human Ras Protein by Cyclen Complexes. *Angew. Chem. Int. Ed.* **2010**, *49* (22), 3830–3833. DOI: 10.1002/anie.200907002.
(100) Holderfield, M.; Lee, B. J.; Jiang, J.; Tomlinson, A.; Seamon, K. J.; Mira, A.; Patrucco, E.; Goodhart, G.; Dilly, J.; Gindin, Y.; Dinglasan, N.; Wang, Y.; Lai, L. P.; Cai, S.; Jiang, L.; Nasholm, N.; Shifrin, N.; Blaj, C.; Shah, H.; Evans, J. W.; Montazer, N.; Lai, O.; Shi, J.; Ahler, E.; Quintana, E.; Chang, S.; Salvador, A.; Marquez, A.; Cregg, J.; Liu, Y.; Milin, A.; Chen, A.; Bar Ziv, T.; Parsons, D.; Knox, J. E.; Klomp, J. E.; Roth, J.; Rees, M.; Ronan, M.; Cuevas-Navarro, A.; Hu, F.; Lito, P.; Santamaria, D.; Aguirre, A. J.; Waters, A. M.; Der, C. J.; Ambrogio, C.; Wang, Z.; Gill, A. L.; Koltun, E. S.; Smith, J. A. M.; Wildes, D.;

Singh, M. Concurrent Inhibition of Oncogenic and Wild-Type RAS-GTP for Cancer Therapy. *Nature* **2024**, *629* (8013), 919–926. DOI: 10.1038/s41586-024-07205-6.
(101) Cuevas-Navarro, A.; Pourfarjam, Y.; Hu, F.; Rodriguez, D. J.; Vides, A.; Sang, B.; Fan, S.; Goldgur, Y.; de Stanchina, E.; Lito, P. Pharmacological Restoration of GTP Hydrolysis by Mutant RAS. *Nature* **2025**, *637* (8044), 224–229. DOI: 10.1038/s41586-024-08283-2.
(102) Schöpel, M.; Herrmann, C.; Scherkenbeck, J.; Stoll, R. The Bisphenol A Analogue Bisphenol S Binds to K-Ras4B – Implications for 'BPA-Free' Plastics. *FEBS Lett.* **2016**, *590* (3), 369–375. DOI: 10.1002/1873-3468.12056.
(103) Franz, M.; Mörchen, B.; Degenhart, C.; Gülden, D.; Shkura, O.; Wolters, D.; Koch, U.; Klebl, B.; Stoll, R.; Helfrich, I.; Scherkenbeck, J. Sequence-Selective Covalent CaaX-Box Receptors Prevent Farnesylation of Oncogenic Ras Proteins and Impact MAPK/PI3K Signaling. *ChemMedChem* **2021**, *16* (16), 2504–2514. DOI: 10.1002/cmdc.202100167.
(104) Schöpel, M.; Shkura, O.; Seidel, J.; Kock, K.; Zhong, X.; Löffek, S.; Helfrich, I.; Bachmann, H. S.; Scherkenbeck, J.; Herrmann, C.; Stoll, R. Allosteric Activation of GDP-Bound Ras Isoforms by Bisphenol Derivative Plasticisers. *Int. J. Mol. Sci.* **2018**, *19* (4), 1133. DOI: 10.3390/ijms19041133.
(105) Jeuken, S. *Fragmentevolution von Bisphenol-Liganden Für Die Suche Nach Inhibitoren von K-Ras4B*. Doctoral Dissertation, Bergische Universität Wuppertal, **2021**. https://elekpub.bib.uni-wuppertal.de/urn/urn:nbn:de:hbz:468-20210713-120811-7.
(106) Peña, S.; Scarone, L.; Manta, E.; Serra, G. An Efficient Synthesis of 2,4′-Bi-1,3-Oxa(thia)zoles as Scaffolds for Bioactive Products. *Chem. Heterocycl. Compd.* **2011**, *47* (6), 703–709. DOI: 10.1007/s10593-011-0823-z.
(107) Hämmerle, J.; Schnürch, M.; Iqbal, N.; Mihovilovic, M. D.; Stanetty, P. A Guideline for the Arylation of Positions 4 and 5 of Thiazole *via* Pd-Catalyzed Cross-Coupling Reactions. *Tetrahedron* **2010**, *66* (40), 8051–8059. DOI: 10.1016/j.tet.2010.07.081.
(108) Specklin, S.; Decuypere, E.; Plougastel, L.; Aliani, S.; Taran, F. One-Pot Synthesis of 1,4-Disubstituted Pyrazoles from Arylglycines *via* Copper-Catalyzed Sydnone-Alkyne Cycloaddition Reaction. *J. Org. Chem.* **2014**, *79* (16), 7772–7777. DOI: 10.1021/jo501420r.
(109) Huang, X.; Cheng, C. C.; Fischmann, T. O.; Duca, J. S.; Yang, X.; Richards, M.; Shipps, G. W. Discovery of a Novel Series of Chk1 Kinase Inhibitors with a Distinctive Hinge Binding Mode. *ACS Med. Chem. Lett.* **2012**, *3* (2), 123–128. DOI: 10.1021/ml200249h.
(110) Blakemore, D. Suzuki–Miyaura Coupling. *Synthetic Methods in Drug Discovery;* Vol. 1; The Royal Society of Chemistry, **2016**; pp 1–69. DOI: 10.1039/9781782622086-00001.
(111) Amatore, C.; Le Duc, G.; Jutand, A. Mechanism of Palladium-Catalyzed Suzuki-Miyaura Reactions: Multiple and Antagonistic Roles of Anionic "Bases" and Their Countercations. *Chem. Eur. J.* **2013**, *19* (31), 10082–10093. DOI: 10.1002/chem.201300177.
(112) Jones, R. A.; Bean, G. P. The Chemistry of Pyrroles. *Organic Chemistry*; Vol. 34; Academic Press Inc. (London) Ltd., **1977**; p 209.
(113) Li, H. J.; Wu, Y. C.; Dai, J. H.; Song, Y.; Cheng, R.; Qiao, Y. Regioselective Electrophilic Aromatic Bromination: Theoretical Analysis and Experimental Verification. *Molecules* **2014**, *19* (3), 3401–3416.

DOI: 10.3390/molecules19033401.
(114) Dötz, K. H.; Glänzer, J. 3-Pyrrolylcarbene Complexes of Chromium: Synthesis and Carbene Annulation to the Isoindole Skeleton. *Z. Naturforsch. B,* **1993**, *48* (11), 1595–1602. DOI: 10.1515/znb-1993-1119.
(115) Nicolaou, K. C.; King, N. P.; Finlay, M. R. V.; He, Y.; Roschangar, F.; Vourloumis, D.; Vallberg, H.; Sarabia, F.; Ninkovic, S.; Hepworth, D. Total Synthesis of Epothilone E and Related Side-Chain Modified Analogues *via* a Stille Coupling Based Strategy. *Bioorg. Med. Chem.* **1999**, *7* (5), 665–697. DOI: 10.1016/S0968-0896(98)00153-9.
(116) Baines, K.; Dicke, R.; Neumann, W. P.; Vorspohl, K. New Ways of Generating Stannyl Radicals for Organic Syntheses. *Free Radicals in Synthesis and Biology*; Springer, **1989**; pp 107–114. DOI: 10.1007/978-94-009-0897-0_9.
(117) Stephan, M.; Panther, J.; Wilbert, F.; Ozog, P.; Müller, T. J. J. Heck Reactions of Acrolein or Enones and Aryl Bromides – Synthesis of 3-Aryl Propenals or Propenones and Consecutive Application in Multicomponent Pyrazole Syntheses. *Eur. J. Org. Chem.* **2020**, *2020* (14), 2086–2092. DOI: 10.1002/ejoc.202000066.
(118) Mattson, A. E.; Bharadwaj, A. R.; Scheidt, K. A. The Thiazolium-Catalyzed Sila-Stetter Reaction: Conjugate Addition of Acylsilanes to Unsaturated Esters and Ketones. *J. Am. Chem. Soc.* **2004**, *126* (8), 2314–2315. DOI: 10.1021/ja0318380.
(119) Ohno, T.; Sakai, M.; Ishino, Y.; Shibata, T.; Maekawa, H.; Nishiguchi, I. Mg-Promoted Regio- and Stereoselective C-Acylation of Aromatic α,β-Unsaturated Carbonyl Compounds. *Org. Lett.* **2001**, *3* (22), 3439–3442. DOI: 10.1021/ol016376e
(120) Clark, J. H.; Cork, D. G.; Gibbs, H. W. The Synthesis of 1,4-Diketones. *J. Chem.* Soc., Chem. Commun. **1982**, 635–636.
(121) Clark, J. H.; Miller, J. M.; So, K. H. Hydrogen Bonding in Organic Synthesis. Part 8. Reactions of 2-Nitropropane in the Presence of Tetraethylammonium Fluoride. The Interaction of the Nitro-Group with Fluoride. *J. Chem. Soc., Perkin Trans. 1* **1978**, 941–946. DOI: 10.1039/P19780000941.
(122) Salman, H.; Abraham, Y.; Tal, S.; Meltzman, S.; Kapon, M.; Tessler, N.; Speiser, S.; Eichen, Y. 1,3-Di(2-Pyrrolyl)Azulene: An Efficient Luminescent Probe for Fluoride. *Eur. J. Org. Chem.* **2005**, *2005* (11), 2207–2212. DOI: 10.1002/ejoc.200500012.
(123) Donohoe, T. J.; Headley, C. E.; Cousins, R. P. C.; Cowley, A. Flexibility in the Partial Reduction of 2,5-Disubstituted Pyrroles: Application to the Synthesis of DMDP. *Org. Lett.* **2003**, *5* (7), 999–1002. DOI: 10.1021/ol027504h.
(124) Zhang, Z.; Ray, S.; Imlay, L.; Callaghan, L. T.; Niederstrasser, H.; Mallipeddi, P. L.; Posner, B. A.; Wetzel, D. M.; Phillips, M. A.; Smith, M. W. Total Synthesis of (+)-Spiroindimicin A and Congeners Unveils Their Antiparasitic Activity. *Chem. Sci.* **2021**, *12* (30), 10388–10394. DOI: 10.1039/D1SC02838C.
(125) Hasse, K.; Willis, A. C.; Banwell, M. G. Modular Total Syntheses of Lamellarin G Trimethyl Ether and Lamellarin S. *Eur. J. Org. Chem.* **2011**, *2011* (1), 88–99. DOI: 10.1002/ejoc.201001133.
(126) Hager, K.; Franz, A.; Hirsch, A. Self-Assembly of Chiral Depsipeptide Dendrimers. *Chem. Eur. J.* **2006**, *12* (10), 2663–2679. DOI: 10.1002/chem.200501300.

(127) Mee, S. P. H.; Lee, V.; Baldwin, J. E. Stille Coupling Made Easier – The Synergic Effect of Copper(I) Salts and the Fluoride Ion. *Angew. Chem. Int. Ed.* **2004**, *43* (9), 1132–1136. DOI: 10.1002/anie.200352979.
(128) Bajwa, J. S. Chemoselective Deprotection of Benzyl Esters in the Presence of Benzyl Ethers, Benzyloxymethyl Ethers and N-Benzyl Groups by Catalytic Transfer Hydrogenation. *Tetrahedron Lett.* **1992**, *33* (17), 2299–2302. DOI: 10.1016/S0040-4039(00)74195-5.
(129) Khurana, J. M.; Arora, R. Rapid Chemoselective Deprotection of Benzyl Esters by Nickel Boride. *Synthesis* **2009**, *2009* (7), 1127–1130. DOI: 10.1055/s-0028-1087982.
(130) Sajiki, H.; Hattori, K.; Hirota, K. The Formation of a Novel Pd/C-Ethylenediamine Complex Catalyst: Chemoselective Hydrogenation without Deprotection of the *O*-Benzyl and *N*-Cbz Groups. *J. Org. Chem.* **1998**, *63* (22), 7990–7992. DOI: 10.1021/jo9814694.
(131) Goverdhan, G.; Raghupathi Reddy, A.; Himabindu, V.; Mahesh Reddy, G. Synthesis and Characterization of Critical Process Related Impurities of an Asthma Drug – Zafirlukast. *J. Saudi Chem. Soc.* **2014**, *18* (2), 129–138. DOI: 10.1016/j.jscs.2011.06.002.
(132) Goverdhan, G.; Reddy, A. R.; Himabindu, V.; Reddy, G. M. Concise and Alternative Synthesis of Zafirlukast, an Anti-Asthma Drug. *Synth. Commun.* **2013**, *43* (4), 498–504. DOI: 10.1080/00397911.2011.603875.
(133) Goverdhan, G.; Reddy, A. R.; Srinivas, K.; Himabindu, V.; Reddy, G. M. Identification, Characterization and Synthesis of Impurities of Zafirlukast. *J. Pharm. Biomed. Anal.* **2009**, *49* (4), 895–900. DOI: 10.1016/j.jpba.2009.01.023.
(134) Goverdhan, G.; Reddy, A. R.; Reddy, N. R.; Srinivas, K.; Himabindu, V.; Reddy, G. M. Synthesis and Characterization of New Analogs of Zafirlukast. *Org. Commun.* **2012**, *5* (1), 27–41.
(135) Schierle, S.; Flauaus, C.; Heitel, P.; Willems, S.; Schmidt, J.; Kaiser, A.; Weizel, L.; Goebel, T.; Kahnt, A. S.; Geisslinger, G.; Steinhilber, D.; Wurglics, M.; Rovati, G. E.; Schmidtko, A.; Proschak, E.; Merk, D. Boosting Anti-Inflammatory Potency of Zafirlukast by Designed Polypharmacology. *J. Med. Chem.* **2018**, *61* (13), 5758–5764. DOI: 10.1021/acs.jmedchem.8b00458.
(136) Paladugu, S.; Mainkar, P. S.; Chandrasekhar, S. Synthesis of Asthma Drug Zafirlukast (Accolate) Using Intramolecular Oxidative Coupling *via* sp^3 C–H Bond Activation. *ACS Omega* **2018**, *3* (4), 4289–4294. DOI: 10.1021/acsomega.8b00476.
(137) Matassa, V. G.; Maduskuie, T. P.; Shapiro, H. S.; Hesp, B.; Aharony, D.; Krell, R. D.; Keith, R. A.; Snyder, D. W. Evolution of a Series of Peptidoleukotriene Antagonists: Synthesis and Structure/Activity Relationships of 1,3,5-Substituted Indoles and Indazoles. *J. Med. Chem.* **1990**, *33* (6), 1781–1790. DOI: 10.1021/jm00168a037.
(138) Bredael, K.; Geurs, S.; Clarisse, D.; De Bosscher, K.; D'hooghe, M. Carboxylic Acid Bioisosteres in Medicinal Chemistry: Synthesis and Properties. *J. Chem.* **2022**, *2022* (1), 2164558. DOI: 10.1155/2022/2164558.
(139) Jaradat, D. M. M. Thirteen Decades of Peptide Synthesis: Key Developments in Solid Phase Peptide Synthesis and Amide Bond Formation Utilized in Peptide Ligation. *Amino Acids* **2018**, *50* (1), 39–68. DOI: 10.1007/s00726-017-2516-0.
(140) Weissman, S. A.; Zewge, D. Recent Advances in Ether Dealkylation. *Tetrahedron* **2005**, *61* (33), 7833–7863. DOI: 10.1016/j.tet.2005.05.041.

(141) Nagy, A.; Novák, Z.; Kotschy, A. Sequential and Domino Sonogashira Coupling: Efficient Tools for the Synthesis of Diarylalkynes. *J. Organomet. Chem.* **2005**, *690* (20), 4453–4461. DOI: 10.1016/j.jorganchem.2004.12.036.
(142) Capani, J. S.; Cochran, J. E.; Liang, J. CsF-Mediated in Situ Desilylation of TMS-Alkynes for Sonogashira Reaction. *J. Org. Chem.* **2019**, *84* (14), 9378–9384. DOI: 10.1021/acs.joc.9b01307.
(143) Gogoi, A.; Modi, A.; Guin, S.; Rout, S. K.; Das, D.; Patel, B. K. A Metal-Free Domino Synthesis of 3-Aroylindoles *via* Two sp^3 C–H Activation. *Chem. Commun.* **2014**, *50* (72), 10445–10447. DOI: 10.1039/c4cc04407j.
(144) Crisalli, P.; Kool, E. T. Water-Soluble Organocatalysts for Hydrazone and Oxime Formation. *J. Org. Chem.* **2013**, *78* (3), 1184–1189. DOI: 10.1021/jo302746p.
(145) Gemal, A. L.; Luche, J. L. Lanthanoids in Organic Synthesis. 6. The Reduction of α-Enones by Sodium Borohydride in the Presence of Lanthanoid Chlorides: Synthetic and Mechanistic Aspects. *J. Am. Chem. Soc.* **1981**, *103* (18), 5454–5459. DOI: 10.1021/ja00408a029.
(146) Zhang, X.; Li, S.; Yu, W.; Xie, Y.; Tung, C. H.; Xu, Z. Asymmetric Azide-Alkyne Cycloaddition with Ir(I)/Squaramide Cooperative Catalysis: Atroposelective Synthesis of Axially Chiral Aryltriazoles. *J. Am. Chem. Soc.* **2022**, *144* (14), 6200–6207. DOI: 10.1021/jacs.2c02563.
(147) Boren, B. C.; Narayan, S.; Rasmussen, L. K.; Zhang, L.; Zhao, H.; Lin, Z.; Jia, G.; Fokin, V. V. Ruthenium-Catalyzed Azide-Alkyne Cycloaddition: Scope and Mechanism. *J. Am. Chem. Soc.* **2008**, *130* (28), 8923–8930. DOI: 10.1021/ja0749993.
(148) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739–1749. DOI: 10.1021/jm0306430.
(149) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49* (21), 6177–6196. DOI: 10.1021/jm051256o.
(150) Sadybekov, A. V.; Katritch, V. Computational Approaches Streamlining Drug Discovery. *Nature* **2023**, *616* (7958), 673–685. DOI: 10.1038/s41586-023-05905-z.
(151) Davies, J. W.; Glick, M.; Jenkins, J. L. Streamlining Lead Discovery by Aligning *in Silico* and High-Throughput Screening. *Curr. Opin. Chem. Biol.* **2006**, *10* (4), 343–351. DOI: 10.1016/j.cbpa.2006.06.022.
(152) Zabolotna, Y.; Lin, A.; Horvath, D.; Marcou, G.; Volochnyuk, D. M.; Varnek, A. Chemography: Searching for Hidden Treasures. *J. Chem. Inf. Model.* **2021**, *61* (1), 179–188. DOI: 10.1021/acs.jcim.0c00936.
(153) Sampath Kumar, H. M.; Herrmann, L.; Tsogoeva, S. B. Structural Hybridization as a Facile Approach to New Drug Candidates. *Bioorg. Med. Chem. Lett.* **2020**, *30* (23), 127514. DOI: 10.1016/j.bmcl.2020.127514.
(154) Willett, P. The Literature of Chemoinformatics: 1978–2018. *Int. J. Mol. Sci.* **2020**, *21* (15), 5576. DOI: 10.3390/ijms21155576.
(155) Engel, T.; Gasteiger, J. Applied Chemoinformatics. *Applied Chemoinformatics*; Wiley, **2018**; pp 15–17. DOI: 10.1002/9783527806539.

(156) Todeschini, R.; Consonni, V. *Molecular Descriptors for Chemoinformatics*; Vol. 1; Wiley Blackwell, **2009**. DOI: 10.1002/9783527628766.
(157) Bajusz, D.; Rácz, A.; Héberger, K. Why Is Tanimoto Index an Appropriate Choice for Fingerprint-Based Similarity Calculations? *J. Cheminf.* **2015**, *7* (1), 20. DOI: 10.1186/s13321-015-0069-3.
(158) Mellor, C. L.; Marchese Robinson, R. L.; Benigni, R.; Ebbrell, D.; Enoch, S. J.; Firman, J. W.; Madden, J. C.; Pawar, G.; Yang, C.; Cronin, M. T. D. Molecular Fingerprint-Derived Similarity Measures for Toxicological Read-Across: Recommendations for Optimal Use. *Regul. Toxicol. Pharmacol.* **2019**, *101*, 121–134. DOI: 10.1016/j.yrtph.2018.11.002.
(159) Riniker, S.; Landrum, G. A. Open-Source Platform to Benchmark Fingerprints for Ligand-Based Virtual Screening. *J. Cheminf.* **2013**, *5* (5), 26. DOI: 10.1186/1758-2946-5-26.
(160) Awale, M.; Reymond, J. L. Polypharmacology Browser PPB2: Target Prediction Combining Nearest Neighbors with Machine Learning. *J. Chem. Inf. Model.* **2019**, *59* (1), 10–17. DOI: 10.1021/acs.jcim.8b00524.
(161) Dunn, T. B.; Seabra, G. M.; Kim, T. D.; Juárez-Mercado, K. E.; Li, C.; Medina-Franco, J. L.; Miranda-Quintana, R. A. Diversity and Chemical Library Networks of Large Data Sets. *J. Chem. Inf. Model.* **2021**, *62* (9), 2186–2201. DOI: 10.1021/acs.jcim.1c01013.
(162) Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures - A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5* (2), 107–113. DOI: 10.1021/c160017a018.
(163) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50* (5), 742–754. DOI: 10.1021/ci100050t.
(164) Daylight Chemical Information Systems, Inc. *Daylight Theory: Fingerprints*. https://www.daylight.com/dayhtml/doc/theory/theory.finger.html (accessed 2024-12-16).
(165) Tahıl, G.; Delorme, F.; Le Berre, D.; Monflier, É.; Sayede, A.; Tilloy, S. Stereoisomers Are Not Machine Learning's Best Friends. *J. Chem. Inf. Model.* **2024**, *64* (14), 5451–5469. DOI: 10.1021/acs.jcim.4c00318.
(166) Orsi, M.; Reymond, J. L. One Chiral Fingerprint to Find Them All. *J. Cheminf.* **2024**, *16* (1), 53. DOI: 10.1186/s13321-024-00849-6.
(167) Sippl, W.; Robaa, D. QSAR/QSPR. *Applied Chemoinformatics*; John Wiley & Sons, Ltd: **2018**; pp 9–52. DOI: 10.1002/9783527806539.ch2.
(168) James, G.; Witten, D.; Hastie, T.; Tibshirani, R.; Taylor, J. Tree-Based Methods. *An Introduction to Statistical Learning*; Springer Texts in Statistics; Springer International Publishing: Cham, **2023**; pp 331–366. DOI: 10.1007/978-3-031-38747-0.
(169) Medina-Franco, J. L.; Chávez-Hernández, A. L.; López-López, E.; Saldívar-González, F. I. Chemical Multiverse: An Expanded View of Chemical Space. *Mol. Inform.* **2022**, *41* (11), 2200116. DOI: 10.1002/minf.202200116.
(170) Saldívar-González, F. I.; Medina-Franco, J. L. Chemoinformatics Approaches to Assess Chemical Diversity and Complexity of Small Molecules. *Small Molecule Drug Discovery: Methods, Molecules and Applications*; Elsevier, **2020**; pp 83–102. DOI: 10.1016/B978-0-12-818349-6.00003-0.
(171) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods. *J. Phys. Chem. A* **1998**, *102* (21), 3762–3772. DOI: 10.1021/jp980230o.

(172) Ohue, M.; Kojima, Y.; Kosugi, T. Generating Potential Protein-Protein Interaction Inhibitor Molecules Based on Physicochemical Properties. *Molecules* **2023**, *28* (15), 5652. DOI: 10.3390/molecules28155652.
(173) Aguilar Troyano, F. J.; Merkens, K.; Anwar, K.; Gómez-Suárez, A. Radical-Based Synthesis and Modification of Amino Acids. *Angew. Chem. Int. Ed.* **2021**, *60* (3), 1098–1115. DOI: 10.1002/anie.202010157.
(174) Aimetti, A. A.; Shoemaker, R. K.; Lin, C.-C.; Anseth, K. S. On-Resin Peptide Macrocyclization Using Thiol-Ene Click Chemistry. *Chem. Commun.* **2010**, *46* (23), 4061–4063. DOI: 10.1039/c001375g.
(175) Mochnáčová, E.; Petroušková, P.; Danišová, O.; Hudecová, P.; Bhide, K.; Kulkarni, A.; Bhide, M. Simple and Rapid Pipeline for the Production of Cyclic and Linear Small-Sized Peptides in *E. coli*. *Protein Expr. Purif.* **2022**, *191*, 106026. DOI: 10.1016/j.pep.2021.106026.
(176) Bruce, A.; Adebomi, V.; Czabala, P.; Palmer, J.; McFadden, W. M.; Lorson, Z. C.; Slack, R. L.; Bhardwaj, G.; Sarafianos, S. G.; Raj, M. A Tag-Free Platform for Synthesis and Screening of Cyclic Peptide Libraries. *Angew. Chem. Int. Ed.* **2024**, *63* (21), e202320045. DOI: 10.1002/anie.202320045.
(177) Mathiesen, I. R.; Calder, E. D. D.; Kunzelmann, S.; Walport, L. J. Discovering Covalent Cyclic Peptide Inhibitors of Peptidyl Arginine Deiminase 4 (PADI4) Using mRNA-Display with a Genetically Encoded Electrophilic Warhead. *Commun. Chem.* **2024**, *7* (1), 304. DOI: 10.1038/s42004-024-01388-9.
(178) Xiang, H.; Bai, L.; Zhang, X.; Dan, T.; Cheng, P.; Yang, X.; Ai, H.; Li, K.; Lei, X. A Facile Strategy for the Construction of a Phage Display Cyclic Peptide Library for the Selection of Functional Macrocycles. *Chem. Sci.* **2024**, *15* (30), 11847–11855. DOI: 10.1039/d4sc03207a.
(179) Tavassoli, A. SICLOPPS Cyclic Peptide Libraries in Drug Discovery. *Curr. Opin. Chem. Biol.* **2017**, *38*, 30–35. DOI: 10.1016/j.cbpa.2017.02.016.
(180) Millward, S. W.; Fiacco, S.; Austin, R. J.; Roberts, R. W. Design of Cyclic Peptides That Bind Protein Surfaces with Antibody-like Affinity. *ACS Chem. Biol.* **2007**, *2* (9), 625–634. DOI: 10.1021/cb7001126.
(181) Hill, T. A.; Shepherd, N. E.; Diness, F.; Fairlie, D. P. Constraining Cyclic Peptides to Mimic Protein Structure Motifs. *Angew. Chem. Int. Ed.* **2014**, *53* (48), 13020–13041. DOI: 10.1002/anie.201401058.
(182) Ji, X.; Nielsen, A. L.; Heinis, C. Cyclic Peptides for Drug Development. *Angew. Chem. Int. Ed.* **2024**, *63* (3), e202308251. DOI: 10.1002/anie.202308251.
(183) Lucana, M. C.; Arruga, Y.; Petrachi, E.; Roig, A.; Lucchi, R.; Oller-Salvia, B. Protease-Resistant Peptides for Targeting and Intracellular Delivery of Therapeutics. *Pharmaceutics* **2021**, *13* (12), 2065. DOI: 10.3390/pharmaceutics13122065.
(184) Qian, Z.; Rhodes, C. A.; McCroskey, L. C.; Wen, J.; Appiah-Kubi, G.; Wang, D. J.; Guttridge, D. C.; Pei, D. Enhancing the Cell Permeability and Metabolic Stability of Peptidyl Drugs by Reversible Bicyclization. *Angew. Chem. Int. Ed.* **2017**, *56* (6), 1525–1529. DOI: 10.1002/anie.201610888.
(185) Price, D. A.; Eng, H.; Farley, K. A.; Goetz, G. H.; Huang, Y.; Jiao, Z.; Kalgutkar, A. S.; Kablaoui, N. M.; Khunte, B.; Liras, S.; Limberakis, C.; Mathiowetz, A. M.; Ruggeri, R. B.; Quan, J.-M.; Yang, Z. Comparative Pharmacokinetic Profile of Cyclosporine (CsA) with a Decapeptide and a Linear Analogue. *Org. Biomol. Chem.* **2017**, *15* (12), 2501–2506. DOI: 10.1039/c7ob00096k.

(186) Kwon, Y. U.; Kodadek, T. Quantitative Comparison of the Relative Cell Permeability of Cyclic and Linear Peptides. *Chem. Biol.* **2007**, *14* (6), 671–677. DOI: 10.1016/j.chembiol.2007.05.006.
(187) Linker, S. M.; Schellhaas, C.; Kamenik, A. S.; Veldhuizen, M. M.; Waibl, F.; Roth, H. J.; Fouché, M.; Rodde, S.; Riniker, S. Lessons for Oral Bioavailability: How Conformationally Flexible Cyclic Peptides Enter and Cross Lipid Membranes. *J. Med. Chem.* **2023**, *66* (4), 2773–2788. DOI: 10.1021/acs.jmedchem.2c01837.
(188) Li, J.; Yanagisawa, K.; Sugita, M.; Fujie, T.; Ohue, M.; Akiyama, Y. CycPeptMPDB: A Comprehensive Database of Membrane Permeability of Cyclic Peptides. *J. Chem. Inf. Model.* **2023**, *63* (7), 2240–2250. DOI: 10.1021/acs.jcim.2c01573.
(189) Cao, L.; Xu, Z.; Shang, T.; Zhang, C.; Wu, X.; Wu, Y.; Zhai, S.; Zhan, Z.; Duan, H. Multi_CycGT: A Deep Learning-Based Multimodal Model for Predicting the Membrane Permeability of Cyclic Peptides. *J. Med. Chem.* **2024**, *67* (3), 1888–1899. DOI: 10.1021/acs.jmedchem.3c01611.
(190) Zhang, H.; Chen, S. Cyclic Peptide Drugs Approved in the Last Two Decades (2001–2021). *RSC Chem. Biol.* **2022**, *3* (1), 18–31. DOI: 10.1039/D1CB00154J.
(191) Sakamoto, K.; Kamada, Y.; Sameshima, T.; Yaguchi, M.; Niida, A.; Sasaki, S.; Miwa, M.; Ohkubo, S.; Sakamoto, J.; Kamaura, M.; Cho, N.; Tani, A. K-Ras(G12D)-Selective Inhibitory Peptides Generated by Random Peptide T7 Phage Display Technology. *Biochem. Biophys. Res. Commun.* **2017**, *484* (3), 605–611. DOI: 10.1016/j.bbrc.2017.01.147.
(192) Niida, A.; Sasaki, S.; Yonemori, K.; Sameshima, T.; Yaguchi, M.; Asami, T.; Sakamoto, K.; Kamaura, M. Investigation of the Structural Requirements of K-Ras(G12D) Selective Inhibitory Peptide KRpep-2d Using Alanine Scans and Cysteine Bridging. *Bioorg. Med. Chem. Lett.* **2017**, *27* (12), 2757–2761. DOI: 10.1016/j.bmcl.2017.04.063.
(193) Lim, S.; Boyer, N.; Boo, N.; Huang, C.; Venkatachalam, G.; Juang, Y.-C. A.; Garrigou, M.; Kaan, H. Y. K.; Duggal, R.; Peh, K. M.; Sadruddin, A.; Gopal, P.; Yuen, T. Y.; Ng, S.; Kannan, S.; Brown, C. J.; Verma, C. S.; Orth, P.; Peier, A.; Ge, L.; Yu, X.; Bhatt, B.; Chen, F.; Wang, E.; Li, N. J.; Gonzales, R. J.; Stoeck, A.; Henry, B.; Sawyer, T. K.; Lane, D. P.; Johannes, C. W.; Biswas, K.; Partridge, A. W. Discovery of Cell Active Macrocyclic Peptides with On-Target Inhibition of KRAS Signaling. *Chem. Sci.* **2021**, *12* (48), 15975–15987. DOI: 10.1039/D1SC05187C.
(194) Sakamoto, K.; Masutani, T.; Hirokawa, T. Generation of KS-58 as the First K-Ras(G12D)-Inhibitory Peptide Presenting Anti-Cancer Activity *in Vivo*. *Sci. Rep.* **2020**, *10* (1), 21671. DOI: 10.1038/s41598-020-78712-5.
(195) Sogabe, S.; Kamada, Y.; Miwa, M.; Niida, A.; Sameshima, T.; Kamaura, M.; Yonemori, K.; Sasaki, S.; Sakamoto, J.; Sakamoto, K. Crystal Structure of a Human K-Ras G12D Mutant in Complex with GDP and the Cyclic Inhibitory Peptide KRpep-2d. *ACS Med. Chem. Lett.* **2017**, *8* (7), 732–736. DOI: 10.1021/acsmedchemlett.7b00128.
(196) Sakamoto, K.; Qi, Y.; Miyako, E. Nanoformulation of the K-Ras(G12D)-Inhibitory Peptide KS-58 Suppresses Colorectal and Pancreatic Cancer-Derived Tumors. *Sci. Rep.* **2023**, *13* (1), 518. DOI: 10.1038/s41598-023-27825-8.
(197) Upadhyaya, P.; Qian, Z.; Selner, N. G.; Clippinger, S. R.; Wu, Z.; Briesewitz, R.; Pei, D. Inhibition of Ras Signaling by Blocking Ras–Effector Interactions with Cyclic Peptides. *Angew. Chem. Int. Ed.* **2015**, *54* (26), 7602–7606.

DOI: 10.1002/anie.201502763.
(198) Takeuchi, K.; Misaki, I.; Tokunaga, Y.; Fujisaki, M.; Kamoshida, H.; Takizawa, T.; Hanzawa, H.; Shimada, I. Conformational Plasticity of Cyclic Ras-Inhibitor Peptides Defines Cell Permeabilization Activity. *Angew. Chem. Int. Ed.* **2021**, *60* (12), 6567–6572. DOI: 10.1002/anie.202016647.
(199) Ng, S.; Juang, Y.-C.; Chandramohan, A.; Kaan, H. Y. K.; Sadruddin, A.; Yuen, T. Y.; Ferrer-Gago, F. J.; Lee, X. C.; Liew, X.; Johannes, C. W.; Brown, C. J.; Kannan, S.; Aronica, P. G.; Berglund, N. A.; Verma, C. S.; Liu, L.; Stoeck, A.; Sawyer, T. K.; Partridge, A. W.; Lane, D. P. De-Risking Drug Discovery of Intracellular Targeting Peptides: Screening Strategies to Eliminate False-Positive Hits. *ACS Med. Chem. Lett.* **2020**, *11* (10), 1993–2001. DOI: 10.1021/acsmedchemlett.0c00022.
(200) Li, A.; Li, X.; Zou, J.; Zhuo, X.; Chen, S.; Chai, X.; Gai, C.; Xu, W.; Zhao, Q.; Zou, Y. SOS1-Inspired Hydrocarbon-Stapled Peptide as a Pan-Ras Inhibitor. *Bioorg. Chem.* **2023**, *135*, 106500. DOI: 10.1016/j.bioorg.2023.106500.
(201) Fumagalli, G.; Carbajo, R. J.; Nissink, J. W. M.; Tart, J.; Dou, R.; Thomas, A. P.; Spring, D. R. Targeting a Novel KRAS Binding Site: Application of One-Component Stapling of Small (5–6-Mer) Peptides. *J. Med. Chem.* **2021**, *64* (23), 17287–17303. DOI: 10.1021/acs.jmedchem.1c01334.
(202) Elek, G. Z.; Koppel, K.; Zubrytski, D. M.; Konrad, N.; Järving, I.; Lopp, M.; Kananovich, D. G. Divergent Access to Histone Deacetylase Inhibitory Cyclopeptides *via* a Late-Stage Cyclopropane Ring Cleavage Strategy. Short Synthesis of Chlamydocin. *Org. Lett.* **2019**, *21* (20), 8473–8478. DOI: 10.1021/acs.orglett.9b03305.
(203) Bock, V. D.; Perciaccante, R.; Jansen, T. P.; Hiemstra, H.; Van Maarseveen, J. H. Click Chemistry as a Route to Cyclic Tetrapeptide Analogues: Synthesis of Cyclo-[Pro-Val-ψ(Triazole)-Pro-Tyr]. *Org. Lett.* **2006**, *8* (5), 919–922. DOI: 10.1021/ol053095o.
(204) Wong, C. T. T.; Lam, H. Y.; Song, T.; Chen, G.; Li, X. Synthesis of Constrained Head-to-Tail Cyclic Tetrapeptides by an Imine-Induced Ring-Closing/Contraction Strategy. *Angew. Chem. Int. Ed.* **2013**, *52* (39), 10212–10215. DOI: 10.1002/anie.201304773.
(205) Kawagishi, H.; Somoto, A.; Kuranari, J.; Kimura, A.; Chiba, S. A Novel Cyclotetrapeptide Produced by *Lactobacillus helveticus* as a Tyrosinase Inhibitor. *Tetrahedron Lett.* **1993**, *34* (21), 3439–3440. DOI: 10.1016/S0040-4039(00)79177-5.
(206) Rodionov, V. O.; Fokin, V. V.; Finn, M. G. Mechanism of the Ligand-Free Cu^{I}-Catalyzed Azide-Alkyne Cycloaddition Reaction. *Angew. Chem. Int. Ed.* **2005**, *44* (15), 2210–2215. DOI: 10.1002/anie.200461496.
(207) He, J.; Ghosh, P.; Nitsche, C. Biocompatible Strategies for Peptide Macrocyclisation. *Chem. Sci.* **2024**, *15* (7), 2300–2322. DOI: 10.1039/D3SC05738K.
(208) White, C. J.; Yudin, A. K. Contemporary Strategies for Peptide Macrocyclization. *Nat. Chem.* **2011**, *3* (7), 509–524. DOI: 10.1038/nchem.1062.
(209) Bechtler, C.; Lamers, C. Macrocyclization Strategies for Cyclic Peptides and Peptidomimetics. *RSC Med. Chem.* **2021**, *12* (8), 1325–1351. DOI: 10.1039/D1MD00083G.
(210) Naveja, J. J.; Saldívar-González, F. I.; Sánchez-Cruz, N.; Medina-Franco, J. L. Cheminformatics Approaches to Study Drug Polypharmacology. *Methods Pharmacol. Toxicol.* **2019**, pp 3–25. DOI: 10.1007/7653_2018_6.

(211) Liu, J. X.; Zhang, X.; Huang, Y. Q.; Hao, G. F.; Yang, G. F. Multi-Level Bioinformatics Resources Support Drug Target Discovery of Protein-Protein Interactions. *Drug Discov. Today* **2024**, *29* (5), 103979. DOI: 10.1016/J.DRUDIS.2024.103979.
(212) Labbé, C. M.; Laconde, G.; Kuenemann, M. A.; Villoutreix, B. O.; Sperandio, O. IPPI-DB: A Manually Curated and Interactive Database of Small Non-Peptide Inhibitors of Protein-Protein Interactions. *Drug Discov. Today* **2013**, *18* (19–20), 958–968. DOI: 10.1016/J.DRUDIS.2013.05.003.
(213) Enamine. *Protein-Protein Interaction Library.* https://enamine.net/compound-libraries/targeted-libraries/ppi-library (accessed 2024-12-09).
(214) Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; Bonino da Silva Santos, L.; Bourne, P. E.; Bouwman, J.; Brookes, A. J.; Clark, T.; Crosas, M.; Dillo, I.; Dumon, O.; Edmunds, S.; Evelo, C. T.; Finkers, R.; Gonzalez-Beltran, A.; Gray, A. J. G.; Groth, P.; Goble, C.; Grethe, J. S.; Heringa, J.; 't Hoen, P. A. C.; Hooft, R.; Kuhn, T.; Kok, R.; Kok, J.; Lusher, S. J.; Martone, M. E.; Mons, A.; Packer, A. L.; Persson, B.; Rocca-Serra, P.; Roos, M.; van Schaik, R.; Sansone, S.-A.; Schultes, E.; Sengstag, T.; Slater, T.; Strawn, G.; Swertz, M. A.; Thompson, M.; van der Lei, J.; van Mulligen, E.; Velterop, J.; Waagmeester, A.; Wittenburg, P.; Wolstencroft, K.; Zhao, J.; & Mons, B. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3* (1), 160018. DOI: 10.1038/sdata.2016.18.
(215) Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Sieb, C.; Thiel, K.; Wiswedel, B. KNIME: The Konstanz Information Miner. In *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*; Springer, **2007**.
(216) Mansouri, K.; Moreira-Filho, J. T.; Lowe, C. N.; Charest, N.; Martin, T.; Tkachenko, V.; Judson, R.; Conway, M.; Kleinstreuer, N. C.; Williams, A. J. Free and Open-Source QSAR-Ready Workflow for Automated Standardization of Chemical Structures in Support of QSAR Modeling. *J. Cheminform.* **2024**, *16* (1), 19. DOI: 10.1186/S13321-024-00814-3.
(217) Falcón-Cano, G.; Molina, C.; Cabrera-Pérez, M. Á. ADME Prediction with KNIME: Development and Validation of a Publicly Available Workflow for the Prediction of Human Oral Bioavailability. *J. Chem. Inf. Model.* **2020**, *60* (6), 2660–2667. DOI: 10.1021/acs.jcim.0c00019.
(218) Nava Höer, B. https://github.com/btnh/kras-ppi-inhibitor-qsar (accessed 2025-06-30).
(219) Wildman, S. A.; Crippen, G. M. Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39* (5), 868–873. DOI: 10.1021/CI990307L.
(220) mordred.SLogP module. *mordred 1.2.1a1 documentation*. https://mordred-descriptor.github.io/documentation/master/api/mordred.SLogP.html (accessed 2025-05-24).
(221) Sauer, W. H. B.; Schwarz, M. K. Molecular Shape Diversity of Combinatorial Libraries: A Prerequisite for Broad Bioactivity. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (3), 987–1003. DOI: 10.1021/CI025599W.
(222) Tipler, P. A.; Mosca, G. Berechnung von Trägheitsmomenten. *Physik*, 7th ed.; Springer, 2015; 287–296. DOI: 10.1007/978-3-642-54166-7.

(223) Duo, L.; Chen, Y.; Liu, Q.; Ma, Z.; Farjudian, A.; Ho, W. Y.; Low, S. S.; Ren, J.; Hirst, J. D.; Xie, H.; Tang, B. Discovery of Novel SOS1 Inhibitors Using Machine Learning. *RSC Med. Chem.* **2024**, *15* (4), 1392–1403. DOI: 10.1039/D4MD00063C.
(224) ChEMBL. *Bioactivity data for target CHEMBL2189121 (GTPase KRas) - IC_{50}*. https://www.ebi.ac.uk/chembl/explore/activities/STATE_ID:O9-j-vM_akhJyGF7piHdYw%3D%3D (accessed 2025-02-21).
(225) Van Der Maaten, L.; Hinton, G. Visualizing Data Using t-SNE. *J. Mach. Learn. Res.* **2008**, *9* (11), 2579–2605.
(226) Noroozi, Z.; Orooji, A.; Erfannia, L. Analyzing the Impact of Feature Selection Methods on Machine Learning Algorithms for Heart Disease Prediction. *Sci. Rep.* **2023**, *13* (1), 22588. DOI: 10.1038/s41598-023-49962-w.
(227) Moriwaki, H.; Tian, Y. S.; Kawashita, N.; Takagi, T. Mordred: A Molecular Descriptor Calculator. *J. Cheminform.* **2018**, *10* (1), 4. DOI: 10.1186/S13321-018-0258-Y.
(228) Descriptor List. *mordred 1.2.1a1 Documentation*. https://mordred-descriptor.github.io/documentation/master/descriptors.html (accessed 2024-12-17).
(229) Ross, B. C. Mutual Information between Discrete and Continuous Data Sets. *PLoS One* **2014**, *9* (2), e87357. DOI: 10.1371/JOURNAL.PONE.0087357.
(230) Hall, L. H.; Kier, L. B. Electrotopological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information. *J. Chem. Inf. Comput. Sci.* **1995**, *35* (6), 1039–1045. DOI: 10.1021/CI00028A014.
(231) Todeschini, R.; Consonni, V. Autocorrelation Descriptors. *Molecular Descriptors for Chemoinformatics*, Vol. 1; Wiley Blackwell, **2009**; p 29. DOI: 10.1002/9783527628766.
(232) Todeschini, R.; Consonni, V. Spectral Indices. *Molecular Descriptors for Chemoinformatics*, Vol. 1; Wiley Blackwell, **2009**; pp 719f. DOI: 10.1002/9783527628766.
(233) Todeschini, R.; Consonni, V. Connectivity Indices. *Molecular Descriptors for Chemoinformatics*, Vol. 1; Wiley Blackwell, **2009**; p 164. DOI: 10.1002/9783527628766.
(234) Todeschini, R.; Consonni, V. Electrotopological State Indices. *Molecular Descriptors for Chemoinformatics*, Vol. 1; Wiley Blackwell, 2009; pp 283ff. DOI: 10.1002/9783527628766.
(235) Roy, K.; Ghosh, G. QSTR with Extended Topochemical Atom Indices. 2. Fish Toxicity of Substituted Benzenes. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (2), 559–567. DOI: 10.1021/CI0342066.
(236) Todeschini, R.; Consonni, V. Indices of Neighborhood Symmetry. *Molecular Descriptors for Chemoinformatics*, Vol. 1; Wiley Blackwell, **2009**; p 410. DOI: 10.1002/9783527628766.
(237) Todeschini, R.; Consonni, V. P_VSA Descriptors. *Molecular Descriptors for Chemoinformatics*, Vol. 1; Wiley Blackwell, **2009**; pp 609ff. DOI: 10.1002/9783527628766.
(238) Todeschini, R.; Consonni, V. ID Numbers. *Molecular Descriptors for Chemoinformatics*, Vol. 1; Wiley Blackwell, **2009**; p 395. DOI: 10.1002/9783527628766.
(239) Comesana, A. E.; Huntington, T. T.; Scown, C. D.; Niemeyer, K. E.; Rapp, V. H. A Systematic Method for Selecting Molecular Descriptors as Features When

Training Models for Predicting Physiochemical Properties. *Fuel* **2022**, *321*, 123836. DOI: 10.1016/J.FUEL.2022.123836.
(240) Todeschini, R.; Consonni, V. Walk Counts. *Molecular Descriptors for Chemoinformatics*, Vol. 1; Wiley Blackwell, **2009**; pp 879f. DOI: 10.1002/9783527628766.
(241) Garrigou, M.; Sauvagnat, B.; Duggal, R.; Boo, N.; Gopal, P.; Johnston, J. M.; Partridge, A.; Sawyer, T.; Biswas, K.; Boyer, N. Accelerated Identification of Cell Active KRAS Inhibitory Macrocyclic Peptides Using Mixture Libraries and Automated Ligand Identification System (ALIS) Technology. *J. Med. Chem.* **2022**, *65* (13), 8961–8974. DOI: 10.1021/ACS.JMEDCHEM.2C00154.
(242) Shichiri, M.; Tanaka, Y. Inhibition of Cancer Progression by Rifampicin: Involvement of Antiangiogenic and Anti-Tumor Effects. *Cell Cycle* **2010**, *9* (1), 64–68. DOI: 10.4161/CC.9.1.10354.
(243) Henry, J. R.; Kaufman, M. D.; Peng, S.-B.; Ahn, Y. M.; Caldwell, T. M.; Vogeti, L.; Telikepalli, H.; Lu, W.-P.; Hood, M. M.; Rutkoski, T. J.; Smith, B. D.; Vogeti, S.; Miller, D.; Wise, S. C.; Chun, L.; Zhang, X.; Zhang, Y.; Kays, L.; Hipskind, P. A.; Wrobleski, A. D.; Lobb, K. L.; Clay, J. M.; Cohen, J. D.; Walgren, J. L.; McCann, D.; Patel, P.; Clawson, D. K.; Guo, S.; Manglicmot, D.; Groshong, C.; Logan, C.; Starling, J. J.; Flynn, D. L. Discovery of 1-(3,3-Dimethylbutyl)-3-(2-Fluoro-4-Methyl-5-(7-Methyl-2-(Methylamino)Pyrido[2,3-*d*]Pyrimidin-6-Yl)Phenyl)Urea (LY3009120) as a Pan-RAF Inhibitor with Minimal Paradoxical Activation and Activity against *BRAF* or *RAS* Mutant Tumor Cells. *J. Med. Chem.* **2015**, *58* (10), 4165–4179. DOI: 10.1021/ACS.JMEDCHEM.5B00067.
(244) Sabnis, R. W. Novel Quinoline Compounds as KRAS Inhibitors for Treating Cancer. *ACS Med. Chem. Lett.* **2023**, *14* (6), 707–708. DOI: 10.1021/acsmedchemlett.3c00195.
(245) Wei, J.; Kitada, S.; Rega, M. F.; Stebbins, J. L.; Zhai, D.; Cellitti, J.; Yuan, H.; Emdadi, A.; Dahl, R.; Zhang, Z.; Yang, L.; Reed, J. C.; Pellecchia, M. Apogossypol Derivatives as Pan-Active Inhibitors of Antiapoptotic B-Cell Lymphoma/Leukemia-2 (Bcl-2) Family Proteins. *J. Med. Chem.* **2009**, *52* (14), 4511–4523. DOI: 10.1021/JM900472S.
(246) Estrela, J. M.; Mena, S.; Obrador, E.; Benlloch, M.; Castellano, G.; Salvador, R.; Dellinger, R. W. Polyphenolic Phytochemicals in Cancer Prevention and Therapy: Bioavailability versus Bioefficacy. *J. Med. Chem.* **2017**, *60* (23), 9413–9436. DOI: 10.1021/ACS.JMEDCHEM.6B01026.
(247) Chakraborty, S.; Lin, S. H.; Shiuan, D.; Tai, D. F. Syntheses of Some α-Cyclic Tripeptides as Potential Inhibitors for HMG-CoA Reductase. *Amino Acids* **2015**, *47* (8), 1495–1505. DOI: 10.1007/S00726-015-1977-2.
(248) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99* (6), 1129–1143. DOI: 10.1016/J.NEURON.2018.08.011.
(249) García-Martín, F.; Bayó-Puxan, N.; Cruz, L. J.; Bohling, J. C.; Albericio, F. Chlorotrityl Chloride (CTC) Resin as a Reusable Carboxyl Protecting Group. *QSAR Comb. Sci.* **2007**, *26* (10), 1027–1035. DOI: 10.1002/QSAR.200720015.
(250) Colomer, I.; Chamberlain, A. E. R.; Haughey, M. B.; Donohoe, T. J. Hexafluoroisopropanol as a Highly Versatile Solvent. *Nat. Rev. Chem.* **2017**, *1* (11), 0088. DOI: 10.1038/s41570-017-0088.
(251) Goldberg, R. N.; Kishore, N.; Lennen, R. M. Thermodynamic Quantities for the Ionization Reactions of Buffers. *J. Phys. Chem. Ref. Data* **2002**, *31* (2), 231–370. DOI: 10.1063/1.1416902.

(252) Boutonnet, J.-C.; Bingham, P.; Calamari, D.; de Rooij, C.; Franklin, J.; Kawano, T.; Libre, J.-M.; McCulloch, A.; Malinverno, G.; Odom, J. M.; Rusch, G. M.; Smythe, K.; Sobolev, I.; Thompson, R.; Tiedje, J. M. Environmental Risk Assessment of Trifluoroacetic Acid. *Hum. Ecol. Risk Assess.* **1999**, *5* (1), 59–124. DOI: 10.1080/10807039991289644.
(253) Chung, B. K. W.; White, C. J.; Yudin, A. K. Solid-Phase Synthesis, Cyclization, and Site-Specific Functionalization of Aziridine-Containing Tetrapeptides. *Nat. Protoc.* **2017**, *12* (6), 1277–1287. DOI: 10.1038/nprot.2017.035.
(254) Kikuchi, M.; Konno, H. Improved Synthesis of D-Allothreonine Derivatives from L-Threonine. *Tetrahedron* **2013**, *69* (34), 7098–7101. DOI: 10.1016/j.tet.2013.06.027.
(255) Aldrich, J. V.; Kulkarni, S. S.; Senadheera, S. N.; Ross, N. C.; Reilley, K. J.; Eans, S. O.; Ganno, M. L.; Murray, T. F.; McLaughlin, J. P. Unexpected Opioid Activity Profiles of Analogues of the Novel Peptide Kappa Opioid Receptor Ligand CJ-15,208. *ChemMedChem* **2011**, *6* (9), 1739–1745. DOI: 10.1002/cmdc.201100113.
(256) Horton, D. A.; Bourne, G. T.; Coughlan, J.; Kaiser, S. M.; Jacobs, C. M.; Jones, A.; Rühmann, A.; Turner, J. Y.; Smythe, M. L. Cyclic Tetrapeptides *via* the Ring Contraction Strategy: Chemical Techniques Useful for Their Identification. *Org. Biomol. Chem.* **2008**, *6* (8), 1386–1395. DOI: 10.1039/b800464a.
(257) Gilmore, K.; Alabugin, I. V. Cyclizations of Alkynes: Revisiting Baldwin's Rules for Ring Closure. *Chem. Rev.* **2011**, *111* (11), 6513–6556. DOI: 10.1021/cr200164y.
(258) Zhang, Y.; Xu, C.; Lam, H. Y.; Lee, C. L.; Li, X. Protein Chemical Synthesis by Serine and Threonine Ligation. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110* (17), 6657–6662. DOI: 10.1073/pnas.1221012110.
(259) Wu, H.; Zhang, Y.; Li, Y.; Xu, J.; Wang, Y.; Li, X. Chemical Synthesis and Biological Evaluations of Adiponectin Collagenous Domain Glycoforms. *J. Am. Chem. Soc.* **2021**, *143* (20), 7808–7818. DOI: 10.1021/jacs.1c02382.
(260) Blanco-Canosa, J. B.; Dawson, P. E. An Efficient Fmoc-SPPS Approach for the Generation of Thioester Peptide Precursors for Use in Native Chemical Ligation. *Angew. Chem., Int. Ed.* **2008**, *47* (36), 6851–6855. DOI: 10.1002/anie.200705471.
(261) Zhao, J. F.; Zhang, X. H.; Ding, Y. J.; Yang, Y. S.; Bi, X. B.; Liu, C. F. Facile Synthesis of Peptidyl Salicylaldehyde Esters and Its Use in Cyclic Peptide Synthesis. *Org. Lett.* **2013**, *15* (20), 5182–5185. DOI: 10.1021/ol402279h.
(262) Anselme, J. P.; Fischer, W. The Reaction of Anions of Primary Amines and Hydrazones with *p*-Toluenesulfonyl Azide. *Tetrahedron* **1969**, *25* (4), 855–859. DOI: 10.1016/0040-4020(69)85018-0.
(263) Fischer, W.; Anselme, J. P. The Reaction of Amine Anions with *p*-Toluenesulfonyl Azide. Novel Azide Synthesis. *J. Am. Chem. Soc.* **1967**, *89* (20), 5284–5285. DOI: 10.1021/ja00996a036.
(264) Nyffeler, P. T.; Liang, C. H.; Koeller, K. M.; Wong, C. H. The Chemistry of Amine- Interconversion: Catalytic Diazotransfer and Regioselective Azide Reduction. *J. Am. Chem. Soc.* **2002**, *124* (36), 10773–10778. DOI: 10.1021/ja0264605.
(265) Pandiakumar, A. K.; Sarma, S. P.; Samuelson, A. G. Mechanistic Studies on the Diazo Transfer Reaction. *Tetrahedron Lett.* **2014**, *55* (18), 2917–2920. DOI: 10.1016/j.tetlet.2014.03.057.

(266) Lundquist, J. T., IV; Pelletier, J. C. Improved Solid-Phase Peptide Synthesis Method Utilizing α-Azide-Protected Amino Acids. *Org. Lett.* **2001**, *3* (5), 781–783. DOI: 10.1021/ol0155485.
(267) Treitler, D. S.; Leung, S. How Dangerous Is Too Dangerous? A Perspective on Azide Chemistry. *J. Org. Chem.* **2022**, *87* (17), 11293–11295. DOI: 10.1021/acs.joc.2c01402.
(268) Titz, A.; Radic, Z.; Schwardt, O.; Ernst, B. A Safe and Convenient Method for the Preparation of Triflyl Azide, and Its Use in Diazo Transfer Reactions to Primary Amines. *Tetrahedron Lett.* **2006**, *47* (14), 2383–2385. DOI: 10.1016/j.tetlet.2006.01.157.
(269) Stanley, N. J.; Pedersen, D. S.; Nielsen, B.; Kvist, T.; Mathiesen, J. M.; Bräuner-Osborne, H.; Taylor, D. K.; Abell, A. D. 1,2,3-Triazolyl Amino Acids as AMPA Receptor Ligands. *Bioorg. Med. Chem. Lett.* **2010**, *20* (24), 7512–7515. DOI: 10.1016/j.bmcl.2010.09.139.
(270) Cavender, C. J.; Shiner, V. Jr. Trifluoromethanesulfonyl Azide. Its Reaction with Alkyl Amines to Form Alkyl Azides. *J. Org. Chem.* **1972**, *37* (22), 3567–3569. DOI: 10.1021/jo00795a052.
(271) Mascarin, A.; Valverde, I. E.; Vomstein, S.; Mindt, T. L. 1,2,3-Triazole Stabilized Neurotensin-Based Radiopeptidomimetics for Improved Tumor Targeting. *Bioconjugate Chem.* **2015**, *26* (10), 2143–2152. DOI: 10.1021/acs.bioconjchem.5b00444.
(272) Johansson, H.; Pedersen, D. S. Azide- and Alkyne-Derivatised α-Amino Acids. *Eur. J. Org. Chem.* **2012**, *2012* (23), 4267–4281. DOI: 10.1002/ejoc.201200496.
(273) Ivkovic, J.; Lembacher-Fadum, C.; Breinbauer, R. A Rapid and Efficient One-Pot Method for the Reduction of *N*-Protected α-Amino Acids to Chiral α-Amino Aldehydes Using CDI/DIBAL-H. *Org. Biomol. Chem.* **2015**, *13* (42), 10456–10460. DOI: 10.1039/c5ob01838b.
(274) Engel-Andreasen, J.; Wellhöfer, I.; Wich, K.; Olsen, C. A. Backbone-Fluorinated 1,2,3-Triazole-Containing Dipeptide Surrogates. *J. Org. Chem.* **2017**, *82* (21), 11613–11619. DOI: 10.1021/acs.joc.7b01744.
(275) Gilbert, J. C.; Weerasooriya, U. Diazoethenes: Their Attempted Synthesis from Aldehydes and Aromatic Ketones by Way of the Horner-Emmons Modification of the Wittig Reaction. A Facile Synthesis of Alkynes. *J. Org. Chem.* **1982**, *47* (10), 1837–1845. DOI: 10.1021/jo00349a007.
(276) Ohira, S. Methanolysis of Dimethyl (1-Diazo-2-Oxopropyl) Phosphonate: Generation of Dimethyl (Diazomethyl) Phosphonate and Reaction with Carbonyl Compounds. *Synth. Commun.* **1989**, *19* (3–4), 561–564. DOI: 10.1080/00397918908050700.
(277) Kürti, L.; Czakó, B. Seyferth-Gilbert Homologation. *Strategic Applications of Named Reactions in Organic Synthesis: Background and Detailed Mechanism*; Elsevier Academic Press, 2005; pp 402f.
(278) Brewitz, L.; Dumjahn, L.; Zhao, Y.; Owen, C. D.; Laidlaw, S. M.; Malla, T. R.; Nguyen, D.; Lukacik, P.; Salah, E.; Crawshaw, A. D.; Warren, A. J.; Trincao, J.; Strain-Damerell, C.; Carroll, M. W.; Walsh, M. A.; Schofield, C. J. Alkyne Derivatives of SARS-CoV-2 Main Protease Inhibitors Including Nirmatrelvir Inhibit by Reacting Covalently with the Nucleophilic Cysteine. *J. Med. Chem.* **2023**, *66* (4), 2663–2680. DOI: 10.1021/acs.jmedchem.2c01627.
(279) Wen, J. J.; Crews, C. M. Synthesis of 9-Fluorenylmethoxycarbonyl-Protected Amino Aldehydes. *Tetrahedron: Asymmetry* **1998**, *9* (11), 1855–1858. DOI: 10.1016/S0957-4166(98)00183-9.

(280) Wang, G.; Mahesh, U.; Chen, G. Y. J.; Yao, S. Q. Solid-Phase Synthesis of Peptide Vinyl Sulfones as Potential Inhibitors and Activity-Based Probes of Cysteine Proteases. *Org. Lett.* **2003**, *5* (5), 737–740. DOI: 10.1021/ol0275567.
(281) Ngo, C.; Fried, W.; Aliyari, S.; Feng, J.; Qin, C.; Zhang, S.; Yang, H.; Shanaa, J.; Feng, P.; Cheng, G.; Chen, X. S.; Zhang, C. Alkyne as a Latent Warhead to Covalently Target SARS-CoV-2 Main Protease. *J. Med. Chem.* **2023**, *66* (17), 12237–12248. DOI: 10.1021/acs.jmedchem.3c00810.
(282) Falorni, M.; Porcheddu, A.; Taddei, M. Mild Reduction of Carboxylic Acids to Alcohols Using Cyanuric Chloride and Sodium Borohydride. *Tetrahedron Lett.* **1999**, *40* (23), 4395–4396. DOI: 10.1016/S0040-4039(99)00734-0.
(283) Meyer, S. D.; Schreiber, S. L. Acceleration of the Dess-Martin Oxidation by Water. *J. Org. Chem.* **1994**, *59* (24), 7549–7552. DOI: 10.1021/jo00103a067.
(284) Myers, A. G.; Zhong, B.; Movassaghi, M.; Kung, D. W.; Lanman, B. A.; Kwon, S. Synthesis of Highly Epimerizable *N*-Protected α-Amino Aldehydes of High Enantiomeric Excess. *Tetrahedron Lett.* **2000**, *41* (9), 1359–1362. DOI: 10.1016/S0040-4039(99)02293-5.
(285) Fields, G. B. Methods for Removing the Fmoc Group. *Peptide Synthesis Protocols*, Vol. 35; Humana Press: Totowa, NJ, **1994**; pp 17–27. DOI: 10.1385/0-89603-273-6:17.
(286) Sheppeck, J. E.; Kar, H.; Hong, H. A Convenient and Scalable Procedure for Removing the Fmoc Group in Solution. *Tetrahedron Lett.* **2000**, *41* (28), 5329–5333. DOI: 10.1016/S0040-4039(00)00853-4.
(287) Ralhan, K.; Krishnakumar, V. G.; Gupta, S. Piperazine and DBU: A Safer Alternative for Rapid and Efficient Fmoc Deprotection in Solid Phase Peptide Synthesis. *RSC Adv.* **2015**, *5* (126), 104417–104425. DOI: https://doi.org/10.1039/c5ra23441g.
(288) Wade, J. D.; Mathieu, M. N.; Macris, M.; Tregear, G. W. Base-Induced Side Reactions in Fmoc-Solid Phase Peptide Synthesis: Minimization by Use of Piperazine as N(α)-Deprotectionreagent. *Lett. Pept. Sci.* **2000**, *7* (2), 107–112. DOI: https://doi.org/10.1023/A:1008966207751.
(289) Hoye, T. R.; Jeffrey, C. S.; Shao, F. Mosher Ester Analysis for the Determination of Absolute Configuration of Stereogenic (Chiral) Carbinol Carbons. *Nat. Protoc.* **2007**, *2* (10), 2451–2458. DOI: https://doi.org/10.1038/nprot.2007.354.
(290) Springer, J.; De Cuba, K. R.; Calvet-Vitale, S.; Geenevasen, J. A. J.; Hermkens, P. H. H.; Hiemstra, H.; van Maarseveen, J. H. Backbone Amide Linker Strategy for the Synthesis of 1,4-Triazole-Containing Cyclic Tetra- and Pentapeptides. *Eur. J. Org. Chem.* **2008**, *2008* (15), 2592–2600. DOI: https://doi.org/10.1002/ejoc.200800143.
(291) Turner, R. A.; Oliver, A. G.; Lokey, R. S. Click Chemistry as a Macrocyclization Tool in the Solid-Phase Synthesis of Small Cyclic Peptides. *Org. Lett.* **2007**, *9* (24), 5011–5014. DOI: https://doi.org/10.1021/ol702228u.
(292) Horne, W. S.; Olsen, C. A.; Beierle, J. M.; Montero, A.; Ghadiri, M. R. Probing the Bioactive Conformation of an Archetypal Natural Product HDAC Inhibitor with Conformationally Homogeneous Triazole-Modified Cyclic Tetrapeptides. *Angew. Chem. Int. Ed.* **2009**, *48* (26), 4718–4724. DOI: https://doi.org/10.1002/anie.200805900.
(293) Beierle, J. M.; Horne, W. S.; van Maarseveen, J. H.; Waser, B.; Reubi, J. C.; Ghadiri, M. R. Conformationally Homogeneous Heterocyclic Pseudotetrapeptides as Three-Dimensional Scaffolds for Rational Drug Design:

Receptor-Selective Somatostatin Analogues. *Angew. Chem. Int. Ed.* **2009**, *48* (26), 4725–4729. DOI: https://doi.org/10.1002/anie.200805901.
(294) Worrell, B. T.; Malik, J. A.; Fokin, V. V. Direct Evidence of a Dinuclear Copper Intermediate in Cu(I)-Catalyzed Azide-Alkyne Cycloadditions. *Science* **2013**, *340* (6131), 457–460. DOI: https://doi.org/10.1126/science.1229506.
(295) Jin, L.; Tolentino, D. R.; Melaimi, M.; Bertrand, G. Isolation of Bis(Copper) Key Intermediates in Cu-Catalyzed Azide-Alkyne "Click Reaction". *Sci. Adv.* **2015**, *1* (5), e1500304. DOI: https://doi.org/10.1126/sciadv.1500304.
(296) Chan, T. R.; Hilgraf, R.; Sharpless, K. B.; Fokin, V. V. Polytriazoles as Copper(I)-Stabilizing Ligands in Catalysis. *Org. Lett.* **2004**, *6* (17), 2853–2855. DOI: https://doi.org/10.1021/ol0493094.
(297) Amaro, R. E.; Baron, R.; McCammon, J. A. An Improved Relaxed Complex Scheme for Receptor Flexibility in Computer-Aided Drug Design. *J. Comp. Aided Mol. Des.* **2008**, *22* (9), 693–705. DOI: https://doi.org/10.1007/s10822-007-9159-2.
(298) Armarego, W. L. F.; Chai, C. L. L. *Purification of Laboratory Chemicals*, 7th ed.; Butterworth-Heinemann: Oxford, **2013**; p 486. DOI: https://doi.org/10.1016/C2009-0-64000-9.
(299) Cai, L. Thin Layer Chromatography. *Curr. Prot. Essent. Lab. Tech.* **2014**, *8* (1), 6.3.1–6.3.18. DOI: https://doi.org/10.1002/9780470089941.et0603s08.
(300) Becker, H. G. O.; Berger, W.; Domschke, G. Infrarotspektroskopie. *Organikum: Organisch-Chemisches Grundpraktikum*, 21st ed.; Wiley-VCH: Weinheim, **2001**; pp 92–94.
(301) Babij, N. R.; McCusker, E. O.; Whiteker, G. T.; Canturk, B.; Choy, N.; Creemer, L. C.; De Amicis, C. V.; Hewlett, N. M.; Johnson, P. L.; Knobelsdorf, J. A.; Li, F.; Lorsbach, B. A.; Nugent, B. M.; Ryan, S. J.; Smith, M. R.; Yang, Q. NMR Chemical Shifts of Trace Impurities: Industrially Preferred Solvents Used in Process and Green Chemistry. *Org. Proc. Res. & Dev.* **2016**, *20* (3), 661–667. DOI: https://doi.org/10.1021/acs.oprd.5b00417.
(302) Akbaşlar, D.; Demirkol, O.; Giray, S. Paal-Knorr Pyrrole Synthesis in Water. Synth Commun **2014**, *44* (9), 1323–1332. DOI: 10.1080/00397911.2013.857691.
(303) Pati, M. L.; Fanizza, E.; Hager, S.; Groza, D.; Heffeter, P.; Laurenza, A. G.; Laquintana, V.; Curri, M. L.; Depalo, N.; Abate, C.; Denora, N. Quantum Dot Based Luminescent Nanoprobes for Sigma-2 Receptor Imaging. Mol. Pharm. **2018**, *15* (2), 458–471. DOI: 10.1021/ACS.MOLPHARMACEUT.7B00825.
(304) Nicolaou, K. C.; Pratt, B.; Arseniyadis, S. Chemical Synthesis of a Highly Potent Epothilone. US8003801B2, **2007**.
(305) Qiu, W.-W.; Surendra, K.; Yin, L.; Corey, E. J. Selective Formation of Six-Membered Oxa- and Carbocycles by the In(III)-Activated Ring Closure of Acetylenic Substrates. *Org. Lett.* **2011**, *13* (21), 5893–5895. DOI: 10.1021/ol202621g.
(306) Yu, M.; Teo, T.; Yang, Y.; Li, M.; Long, Y.; Philip, S.; Noll, B.; Heinemann, G. K.; Diab, S.; Eldi, P.; Mekonnen, L.; Anshabo, A. T.; Rahaman, M. H.; Milne, R.; Hayball, J. D.; Wang, S. Potent and Orally Bioavailable CDK8 Inhibitors: Design, Synthesis, Structure-Activity Relationship Analysis and Biological Evaluation. *Eur. J. Med. Chem*. **2021**, *214* (15), 113248. DOI: 10.1016/J.EJMECH.2021.113248.
(307) Khusnutdinov, R. I.; Baiguzina, A. R.; Mukminov, R. R.; Akhmetov, I. V; Gubaidullin, I. M.; Spivak, S. I.; Dzhemilev, U. M. New Synthesis of Pyrrole-2-

Carboxylic and Pyrrole-2,5-Dicarboxylic Acid Esters in the Presence of Iron-Containing Catalysts. Russian *J. Org. Chem.* **2010**, *46* (7), 1053–1059. DOI: 10.1134/S1070428010070158.
(308) Brown, F. J.; Bernstein, P. R.; Yee, Y. K.; Matassa, V. G. Heterocyclic Amide Derivatives, EUR1986/0199543A3, **1986**.
(309) Doyle, M. P.; Van Lente, M. A.; Mowat, R.; Fobare, W. F. Alkyl Nitrite-Metal Halide Deamination Reactions. 7. Synthetic Coupling of Electrophilic Bromination with Substitutive Deamination for Selective Synthesis of Multiply Brominated Aromatic Compounds from Arylamines. J. Org. Chem. **1980**, *45* (13), 2570–2575. DOI: 10.1021/JO01301A004.
(310) KNIME. *Getting Started Guide*. https://www.knime.com/getting-started-guide (accessed 2024-12-05).
(311) Roughley, S. D. Five Years of the KNIME Vernalis Cheminformatics Community Contribution. *Curr. Med. Chem.* **2020**, *27* (38), 6495–6522. DOI: 10.2174/0929867325666180904113616.
(312) Sydow, D.; Wichmann, M.; Rodríguez-Guerra, J.; Goldmann, D.; Landrum, G.; Volkamer, A. TeachOpenCADD-KNIME: A Teaching Platform for Computer-Aided Drug Design Using KNIME Workflows. *J. Chem. Inf. Model.* **2019**, *59* (10), 4083–4086. DOI: 10.1021/ACS.JCIM.9B00662.
(313) Moreira-Filho, J. T.; Ranganath, D.; Conway, M.; Schmitt, C.; Kleinstreuer, N.; Mansouri, K. Democratizing Cheminformatics: Interpretable Chemical Grouping Using an Automated KNIME Workflow. *J. Cheminf.* **2024**, *16* (1), 101. DOI: 10.1186/S13321-024-00894-1.
(314) Fillbrunn, A.; Dietz, C.; Pfeuffer, J.; Rahn, R.; Landrum, G. A.; Berthold, M. R. KNIME for Reproducible Cross-Domain Analysis of Life Science Data. *J. Biotechnol.* **2017**, *261*, 149–156. DOI: 10.1016/J.JBIOTEC.2017.07.028.
(315) Project Jupyter Documentation. https://docs.jupyter.org/en/latest/ (accessed 2024-12-05).
(316) Volkamer Lab. *TeachOpenCADD*. https://volkamerlab.org/projects/teachopencadd/ (accessed 2024-12-05).
(317) Sydow, D.; Rodríguez-Guerra, J.; Kimber, T. B.; Schaller, D.; Taylor, C. J.; Chen, Y.; Leja, M.; Misra, S.; Wichmann, M.; Ariamajd, A.; Volkamer, A. TeachOpenCADD 2022: Open Source and FAIR Python Pipelines to Assist in Structural Bioinformatics and Cheminformatics Research. *Nucleic Acids Res.* **2022**, *50* (W1), W753–W760. DOI: 10.1093/NAR/GKAC267.
(318) RDKit: Open-Source Cheminformatics. https://www.rdkit.org/ (accessed 2024-12-05).
(319) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12* (85), 2825–2830.
(320) Schwarze, M. KNIME. *RDKit Diversity Picker Node*. https://hub.knime.com/manuelschwarze/extensions/org.rdkit.knime.feature/latest/org.rdkit.knime.nodes.diversitypicker.RDKitDiversityPickerNodeFactory (accessed 2024-12-7).
(321) Ruskey, F.; Savage, C.; Min Yih Wang, T. Generating Necklaces. J.Algorithms **1992**, *13* (3), 414–430. DOI: 10.1016/0196-6774(92)90047-G.
(322) GDP-Glo™ Glycosyltransferase Assay Technical Manual. Promega Corporation, **2017**.

(323) Chan, W. C.; White, P. D. Analytical Procedures. *Fmoc Solid Phase Peptide Synthesis - A Practical Approach*; Oxford University Press, **2000**; p 61.
(324) Du, X.; Qian, J.; Wang, Y.; Zhang, M.; Chu, Y.; Li, Y. Identification and Immunological Evaluation of Novel TLR2 Agonists through Structure Optimization of Pam_3CSK_4. *Bioorg. Med. Chem.* **2019**, *27* (13), 2784–2800. DOI: 10.1016/J.BMC.2019.05.005.
(325) Clerici, A.; Pastori, N.; Porta, O. Efficient Acetalisation of Aldehydes Catalyzed by Titanium Tetrachloride in a Basic Medium. Tetrahedron **1998**, *54* (51), 15679–15690. DOI: 10.1016/S0040-4020(98)00982-X.
(326) Mindt, T.; Michel, U.; Dick, F. Synthesis and Evaluation of Enantiomeric Purity of Protected α-Amino and Peptide Aldehydes. *Helv. Chim. Acta* **1999**, *82* (11), 1960–1968. DOI: 10.1002/(SICI)1522-2675(19991110)82:11<1960::AID-HLCA1960>3.0.CO;2-2.
(327) Al-Gharabli, S. I.; Ali Shah, S. T.; Weik, S.; Schmidt, M. F.; Mesters, J. R.; Kuhn, D.; Klebe, G.; Hilgenfeld, R.; Rademann, J. An Efficient Method for the Synthesis of Peptide Aldehyde Libraries Employed in the Discovery of Reversible SARS Coronavirus Main Protease (SARS-CoV M^{pro}) Inhibitors. *ChemBioChem* **2006**, *7* (7), 1048–1055. DOI: 10.1002/CBIC.200500533.
(328) Brewer, M.; Oost, T.; Sukonpan, C.; Pereckas, M.; Rich, D. H. Sequencing Hydroxyethylamine-Containing Peptides *via* Edman Degradation. *Org Lett.* **2002**, *4* (20), 3469–3472. DOI: 10.1021/ol026590i.

6 Appendix

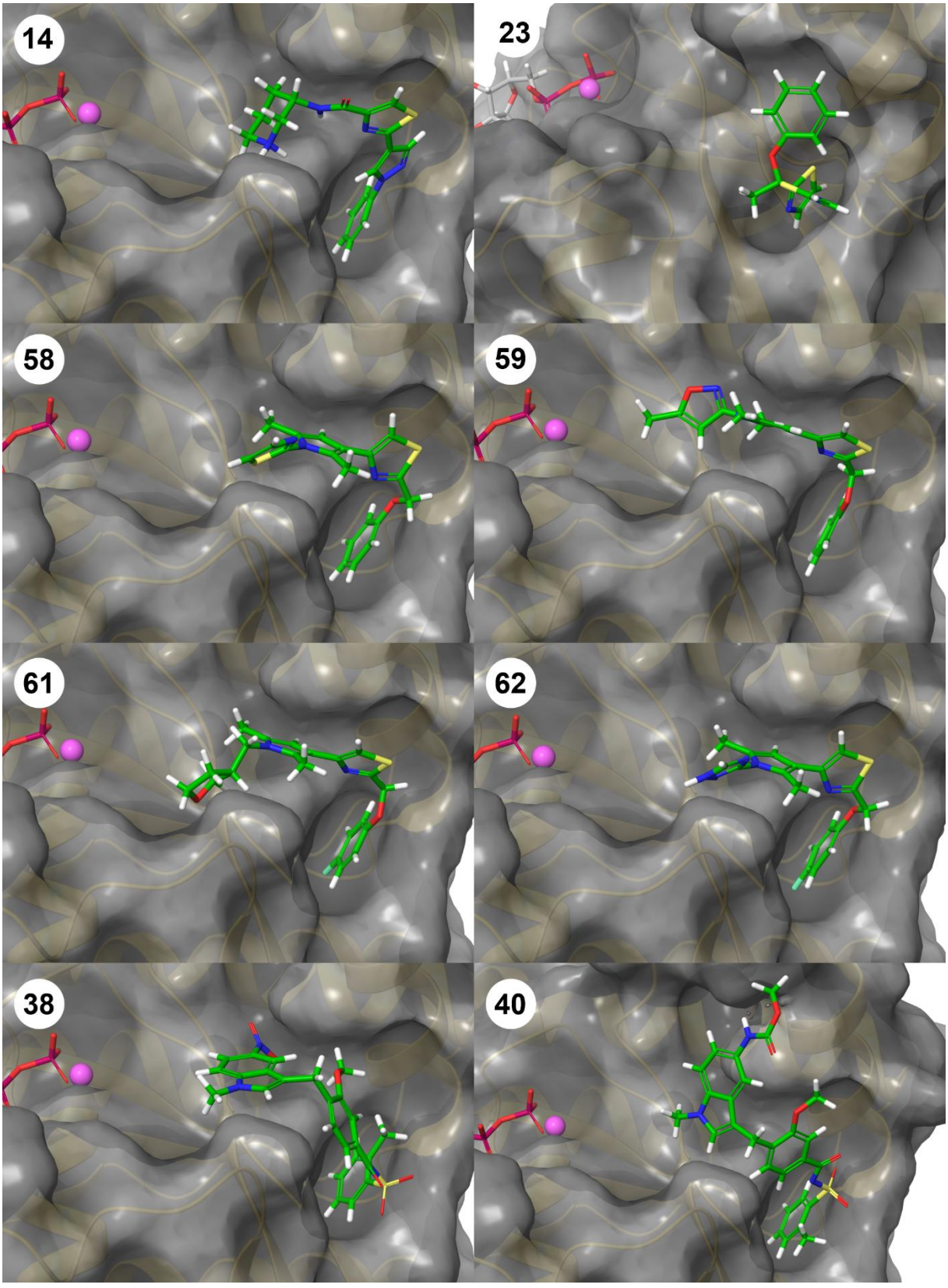


Figure 65: Predicted binding poses of biazoles **14**, **23**, **58**, **59**, **61**, **62**, as well as of zafirlukast derivatives **38**, and **40** at P1 in 6ZL5.

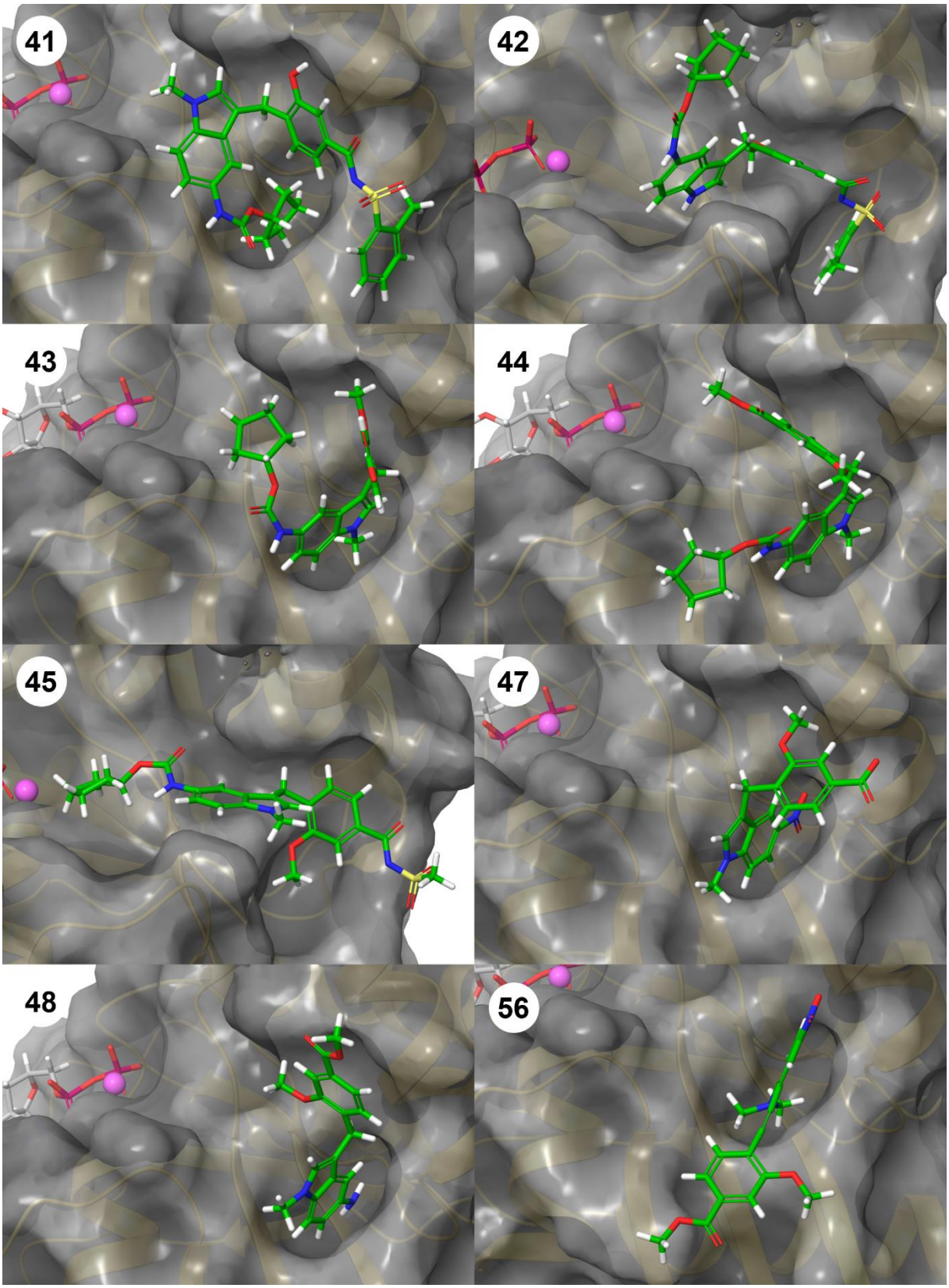


Figure 66: Predicted binding poses of zafirlukast derivatives **41**–**45**, **47**, **48**, and **56** at P1 in 6ZL5.

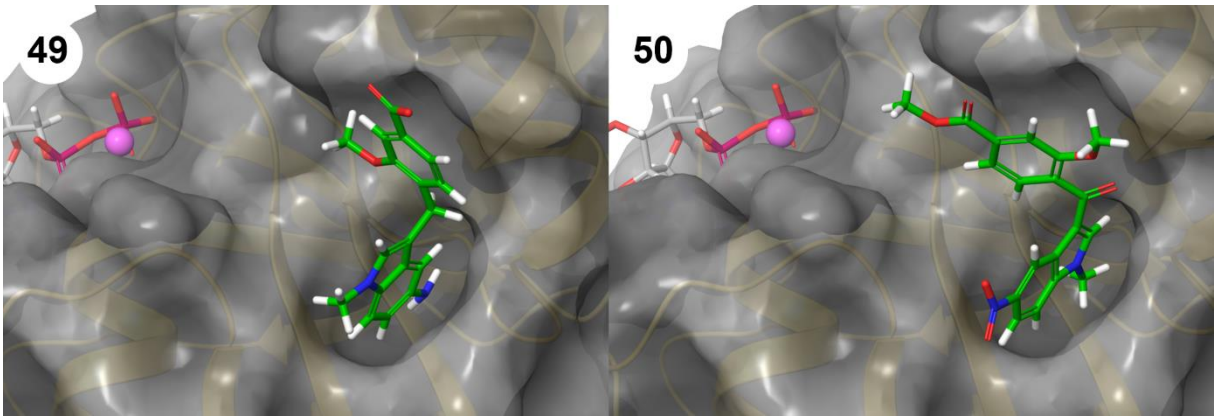


Figure 67: Predicted binding poses of zafirlukast derivatives **49** and **50** at P1 in 6ZL5.