Freie Universität Berlin

Fachbereich Mathematik und Informatik

Institut für Bioinformatik

Development of a Virtual Screening Pipeline for the Identification of Small Molecule PPIP5K2 Inhibitors

---

Master Thesis Bioinformatics (M.Sc.)

Submitted by:

Mareike Royar (geb. Leja)

Supervised by:

JProf. Dr. Andrea Volkamer

Dr. Clara Christ

28. June – 09. December 2021

Abstract

Phosphate homeostasis is essential for survival in eucaryotes. It is mediated by inositol phosphates, which are synthesized by a sequence of kinases. One such kinase in humans is diphosphoinositol pentakisphosphate kinase 2 (PPIP5K2). However, the exact mechanism of phosphate homeostasis in general and PPIP5K2 specifically are still subject to research. To date knockouts and mutations are used for research, but an inhibitor for PPIP5K2 is heavily sought after. Bartsch and Fiedler, who carry out research concerning PPIP5K2, performed a high-throughput screening of 60 000 compounds revealing a single ATP-competitive inhibitor. With an IC_{50} of 15.5 μM the affinity is comparably poor. The presented thesis aims at utilizing computer-aided drug design (CADD) methods to identify additional and ideally more potent PPIP5K2 ligands.

We employed the binding mode of the lead compound to perform a pharmacophore-based virtual screening. A second approach deriving pharmacophore features from the PPIP5K2 apo structure was included to broaden the chemical space of potential active compounds further.

We wrapped the pharmacophore screenings in a pipeline with additional CADD methods to evaluate the screening hits in more detail. Hits not fitting the binding pocket were discarded first. Then we used molecular docking to place the molecules in the binding pocket and filtered them based on matching the respective pharmacophore hypothesis and on the binding affinity score. Selected hits were simulated using molecular dynamics (MD) and the stability was evaluated by calculating the root-mean-square deviation (RMSD) and fluctuation (RMSF). We identified nine candidates from the lead-based screening and eight from the apo structure-based screening. All candidates conform to their respective pharmacophore, score sufficiently high and are stable in MD simulations.

Proposed candidates were handed over to our collaboration partners at FMP and will be tested *in-vitro*. If confirmed active, these compounds will aid building a library of PPIP5K2 ligands to be used in research and for further virtual screening and lead optimization campaigns.

Contents

1 Introduction

Growth and general survival of eucaryotes rely heavily on phosphates. Phosphate is used as building material and makes up around 1% of the body mass, above all as a main element in teeth an bones. [1] Besides, phosphate is used for many other purposes; it is best known for generation and storage of energy in the highly energetic bonds of Adenosine triphosphate (ATP). [2] However, phosphate excess can be toxic. The body needs to finely regulate the phosphate balance and serum phosphate concentrations to maintain the healthy body homeostasis. This calls for a complicated regulatory network. In fact, it spans the entire body and involves intestines, bones and kidneys. The exact mechanisms of para- and transcellular phosphate transport are, to this date, not fully understood. [3] A crucial part of regulatory pathways are enzymes and the up- or down-regulation thereof. Effects of specific enzymes can be measured using inhibitors for sentient regulation *in-vitro*. Concerning the phosphate homeostasis, the problem is that not for all enzymes contributing to phosphate regulation inhibitors are known. One of these enzymes is diphosphoinositol pentakisphosphate kinases 2 (PPIP5K2). A drug discovery campaign has been started and computer-aided drug design (CADD) may contribute in form of a pipeline constructed to discover potential inhibitors for PPIP5K2.

1.1 Regulation of Phosphate Homeostasis

One regulatory component for the phosphate homeostasis are inositol phosphates and inositol pyrophosphates. [4] Inositol, also called cyclohexanehexol, is a carbocyclic sugar and can be altered into finely differentiated neurotransmitters discerned by the amount of phosphate groups attached to the six-membered ring structure. Inositol phosphates (InsPs or IPs) have up to six single phosphate groups; inositol pyrophosphates (PP-InsPs or PP-IPs) include at least one phosphoanhydride bond linking two phosphate groups together and thus are able to pack up to eight phosphates around the six-membered ring. The occurring and relevant pyrophosphates are named IP_7 and IP_8, with

seven or eight phosphate atoms respectively. Possible combinations of positions and quantity of phosphates create a huge signaling network around IPs and PP-IPs, which we are only now beginning to understand. [5]

A variety of enzymes, kinases and phosphatases, synthesize and modify inositol phosphates. Kinases transfer phosphate groups from ATP to their target molecule, in this case the inositol phosphates, thus phosphorylating the molecule. [6] The reverse reaction is catalyzed by phosphatases, which cleave phosphate groups. Interestingly, some of the enzymes involved in inositol phosphate transformation posses a kinase and a phosphatase domain and are theoretically able to perform the reverse reactions as well. [7, 8]

The aforementioned pyrophosphates IP_7 and IP_8 are generated by further phosphorylation of inositol phosphate IP_6. In mammals the pyrophosphate production is based on two enzymes. Inositol hexakisphosphate kinases IP6K synthesize the pyrophosphate $5\text{-}IP_7$ ($5\text{-}PP\text{-}IP_5$), which diphosphoinositol pentakisphosphate kinases (PPIP5K) then further phosphorylate to IP_8 ($1, 5\text{-}[PP]_2\text{-}IP_4$). Another possibility for this pathway is that first PPIP5K synthesizes $1\text{-}IP_7$ ($1\text{-}PP\text{-}IP_5$) from IP_6 and then IP6K1 produces IP_8 (Figure 1.1). PPIP5Ks additionally possess a phosphatase-like domain with potential significance, but are better known for kinase activity as demonstrated in the introduced pathways. [9]

This does not lessen the impact PPIP5K has on the metabolism. The enzyme is conserved over species, which further hints towards a crucial role in intracellular phosphate homeostasis. [10] Humans have two paralogues of enzyme PPIP5K — PPIP5K1 and PPIP5K2; homologues called Asp1 or Vip1 appear in yeast species, VIH is found in plants. Mutations or knockouts of PPIP5K2 are associated with hearing loss and familial keratoconus in humans and mice. [11–13] The deletion of VIH1/VIH2 in plants leads to impaired plant growth and phosphate starvation responses [8]. Thus, PPIP5K2 and its homologues prove essential for phosphate homeostasis. [14, 15]

However, the mechanisms leading to the mentioned symptoms are poorly understood. Therefore, Simon Bartsch and Dorothea Fiedler from the Leibniz-Forschungsinstitut für Molekulare Pharmakologie (FMP) study the function of PPIP5K2. Research is conducted with mutants, knockouts of PPIP5K2 and comparisons to PPIP5K1. Wilson et.al. used IP6K1 and IP6K2 knockouts to prove the relevance of PP-IPs to phosphate homeostasis. [17] Nevertheless, an inhibitor allowing for the elaborate functional control of PPIP5K2 would elevate their and other researchers efforts.

Bartsch and Fiedler initiated a high-throughput screening (HTS), an automated experiment where
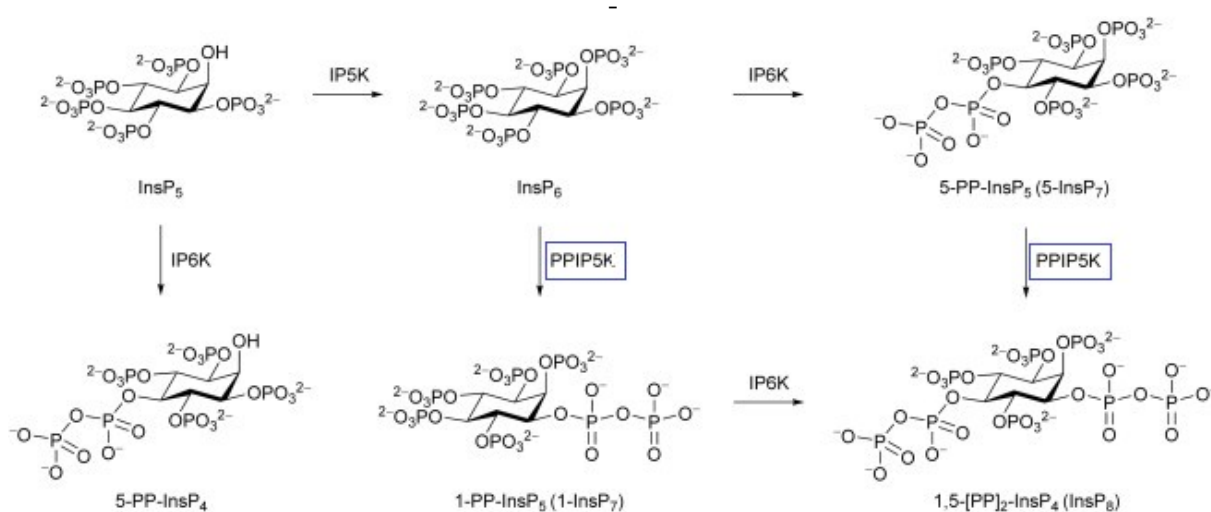
Figure 1.1: **Biosynthesis of diphosphoinositol phosphates.** Depicted are two possible synthesis pathways for IP_8 involving PPIP5K2 and IP6K. Figure taken from Wang et.al., 2014 [16].

more than 60000 compounds were physically tested for activity. With a hit rate of only 0.000017% (one hit in 60000 tested molecules) they identified the first and to date only known ATP-competitive inhibitor 105229 shown in Figure 1.2 (unpublished). Although promising, the lead compound requires micromolar concentrations to show significant activity with a half maximal inhibitory concentration (IC_{50}) of $15.5\,\mu$M. Optimally, an inhibitor is effective in nanomolar concentrations. As a consequence, several attempts were conducted to optimize the compound. First, they tried commercially available analogues, Autogramin-1 (Appendix A, Figure A.1) and Methly quinanzolinone (Appendix A, Figure A.2). Both performed no better than the original compound. Second, they added modifications to compound 105229. They found that small ethyl- and methyl substitutions at the 4-carbon-position of the thiazole moiety improve the inhibitory activity slightly. However, the synthesis of analogues based on trial and error could not improve this further.

1.2 Computer-Aided Drug Design and the ATP-Grasp Fold

To tackle the challenges of drug design and optimization with a more systematic as well as ideally cheaper and faster approach, computer-aided drug design (CADD) methods have emerged. [18, 19] Over the past decades the wider availability of computing power enabled the use of the entire algorithmic space of CADD, including computationally intensive methods. This kick-started the development and use of CADD methods further. Developers from genomic and proteomic initiatives
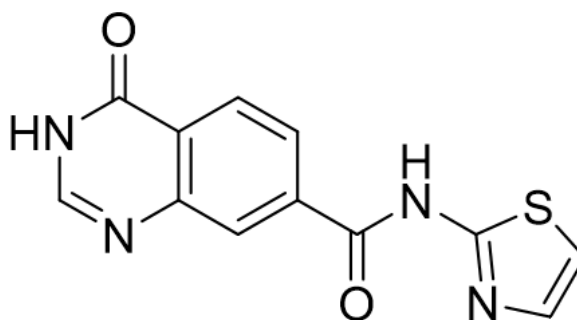
Figure 1.2: **Molecule 105229** comprised of a quinazolinone and a thiazole moiety with an IC_{50} of 15.5 μM against PPIP5K2, as measured by Bartsch and Fiedler (unpublished)

created more accurate and reliable methods to computationally identify new drug targets purely from the knowledge of protein structures and the possibility to simulate and evaluate ligand-target interactions. At the same time, the amount of generated data has increased exponentially. Databases for molecule structures and chemical compounds emerged. Scientific libraries like ChEMBL [20] with over 2 Million unique chemical structures supplemented by activity and genomic data exist alongside commercial variants like MolPort [21], which instead holds exclusive information to availability and pricing of their more than 7 Million offered compounds. The emerging ocean of data can only be maneuvered with new tools able to handle big data.

Many CADD methods are centered around searching these openly available libraries for structures that could be potentially active. In this process, called virtual screening, the millions of compounds in a library are filtered by chosen properties, which improves the hit rate when compared to high-throughput screening. [22] This process can be accompanied by methods to predict the activity of compounds, mainly molecular docking and molecular dynamics (MD) simulations. Note that the individual concepts used in CADD will be introduced in more detail in Chapter 2 with focus on the methods used within this thesis.

One example for the accomplishments of CADD in the research area covered in this work, is the discovery of novel inhibitors for some enzymes of the ATP-grasp superfamily, to which PPIP5K2 belongs. The ATP-grasp superfamily is a family of enzymes that share an atypical ATP-binding site, the ATP-grasp fold. [23] The main challenge in targeting the ATP-grasp fold is designing a selective and competitive inhibitor. Nevertheless, different CADD methods have been applied in the past to find inhibitors for the ATP-grasp fold. Wu et.al. used *in-silico* modeling to identify favourable electrostatic interactions with the binding site of pyruvate phosphate dikinase (PPDK). Compounds synthesized

according to the findings turned out to be potent and selective inhibitors for PPDK [24] Škedelj et.al. used a set of ligands with low activity to perform a ligand-based pharmacophore screening *in-silico* to find an inhibitor for bacterial d-aspartate ligase [25] while Mochalkin *et.al.* used a combination of fragment screening and virtual screening for their antibacterial biotin carboxylase inhibitors. [26]

1.3 Aims

Generally it seems hard to identify inhibitors for PPIP5K2, exemplified by only one hit in a HTS of 60000 compounds (hit rate of 0.0017%), while the protein is involved in several diseases and very important for phosphate homeostasis. This thesis aims at creating a virtual screening pipeline to support the discovery of novel inhibitors of PPIP5K2 incorporating CADD methods.

Computational approaches evolved to be an integral part of drug design for more than 15 years. Several CADD methods can be combined in more or less standardized pipelines to identify novel drug candidates, joining methods like molecular docking — the prediction of preferred binding positions of a ligand in a target — and molecular dynamic simulations — the analysis of molecular movement and interactions in a simulations — with virtual screenings. [27]

In our pipeline, we perform a pharmacophore-based virtual screening to assist the discovery and optimization of novel PPIP5K2 inhibitors. Two structure-based approaches are applied to derive pharmacophores, i.e. energetically-favorable hot spots in the binding site that are important for ligand binding. In the first approach, pharmacophore features are derived from a binding hypothesis of the lead compound, identified by Bartsch and Fiedler, to PPIP5K2. In the second approach, we work entirely with the apo protein structure of PPIP5K2 without bound ligand. We analyze the apo protein structure with the help of PyRod, a software able to predict pharmacophore features based solely on MD simulations of the target. [28]

With the created pharmacophores we perform virtual screenings. The most promising compounds are selected using the two complementary CADD methods, namely molecular docking and molecular dynamics simulation as well as manual visual inspections, which remain a crucial part of drug design despite all technological advances. [29]

We wish to present between 15 and 20 promising molecules to our collaboration partners at FMP for further analysis.

2 Concepts of Computational Methods

In our pipeline, we combine several CADD methods. Each of these is a highly specialised and elaborate tool, usually developed by either open-source developers or for commercial purposes. Here, only a short introduction to the most prevalent methods for this thesis is given.

2.1 Molecular Docking

Molecular docking predicts the preferred binding mode of a ligand to a macromolecule, often a protein. During docking the conformational space of the ligand is explored and sampled; hereby the protein is often kept rigid. The sampled poses are scored to estimate their spatial and physicochemical fit. Scoring functions are generally used to discern viable from unlikely poses and are able to rule out a lot of results based on protein clashes and unfavourable interactions. Most tools use their own implementation of a scoring algorithm based on free energy equations and only return the best scoring poses. Docking is an optimization problem, whose efficiency is defined by the used algorithms and the extent to which conformations are searched. It is still a relatively fast procedure and suitable to be performed on large libraries. [30, 31]

Several competing docking methods exist. They differ by search algorithms, scoring functions and treatment of ligand and protein flexibility.

During our work, we use the HYBRID algorithm included in the OpenEye OEDocking toolkit. The HYBRID method "uses bound ligand information to improve virtual screening performance". [32] This increases the docking performance of HYBRID compared to OpenEye's other docking algorithm FRED (Fast Exhaustive Docking). [33, 34] HYBRID accounts for full flexibility of the ligand, but treats the protein receptor as rigid as the algorithm compromises between optimizing the docking pose and minimizing computational cost. Emulating receptor flexibility multiplies the sampled space due to large variation leading to a steep increase in resource requirements.

HYBRID works in two steps. In the first step, it performs an exhaustive search. During the search, every possible rotation of each conformer of the ligand is placed in the protein active site in several variations. The arisen poses are pre-filtered and scored using OpenEye's Chemical Gaussian Overlay function, a primarily ligand-based scoring function that requires a bound ligand pose. It calculates similarities between docking pose and bound ligand based on shape, hydrogen bonds and metal chelating groups. The top scoring poses are passed to the second step: Optimization. Optimized poses are scored again, this time with the scoring function Chemgauss4, which calculates interaction potentials based on interactions of atom pairs and introduces a smoothing function to the potentials. The smoothed potentials are less sensitive to small changes in position and to an extent emulate protein flexibility. HYBRID returns the optimized poses and their corresponding Chemgauss4 score. [35]

2.2 Molecular Dynamics Simulation

Molecular dynamics (MD) simulations are computational simulations modeling movement and interactions of molecules like proteins and ligands in a solvent. Motions of atoms are approximated using Newtons law of motion. All forces affecting the atoms and bonds are encapsulated in force fields, which are parametric equations with terms for the distinct forces like bond length, bond angle and torsion. Several families of force fields exists; four well used families are AMBER, CHARMM, GRO-MOS and OPLS, who share the same functional form. [36] Often families consist of a set of force fields, each optimized for specific types of molecules — proteins, nucleic acids, carbohydrates or lipids or small molecules. [37] Other developers focus on perfecting specific types of force field, like the Open Force Field Initiative, who develop the small molecule force fields we use during this thesis. [38]

The development of the system is based on solving these equations and is expressed in a series of frames characterizing the position, velocity and potential energy of all particles at a given time point. The collection of frames and the resulting vectors are called a molecular trajectory. The collection of atoms forming the system and the bonds connecting them are called a topology. Common steps for performing a MD simulation include the solvation of the system, an equilibration phase and then the actual computation of trajectories. Length and time steps between frames depend on the aim of the simulation.

MD simulations account for full flexibility of protein and ligand. Ligands may induce small changes in protein conformation, that are not observed in rigid methods. [39] The simulations are especially suited to estimate the drug-target recognition and binding process, but can also be used to estimate stability of computationally derived ligand poses. This results in MD simulations being a valuable tool in the drug discovery process. [40]

However, these advantages come at a price: Their computational intensiveness is the biggest drawback, especially for larger systems or longer time scales. Since computational power has become cheaper and more widely accessible, MD simulations have seen increased use in biophysics, microbiology and bioinformatics.

2.3 Molecular Interaction Fields

A Molecular Interaction Field (MIF) describes the interaction energies between a molecular target and a chosen probe. The target is usually a protein or complex, while the probe is a sphere representing a specific atom, molecule or ion type. The interaction energies are calculated based on a structured placement of the probe along the target, most often in a grid pattern. Usually fields are calculated for specific interaction types, e.g. ionic interactions or hydrogen bonds. [41]

It is also possible to track the interaction over the course of an MD simulation. The field then becomes a dynamic Molecular Interaction Field (dMIF). The tool PyRod uses this techniques and traces water molecules and their protein environment in MD simulations to create dMIFs. [28]

Classic applications for MIFs and dMIFs include guiding the process of structure-based ligand design, docking of ligands to macromolecules, assessment of quantitative structure–activity relationships (QSARs) and many more. MIFs can also be applied to the prediction of pharmacokinetic properties. [42]

2.4 Virtual Screening

Screenings generally describe processes in which large libraries of molecules are searched for active compounds. Experimental high-throughput screening (HTS) is classically in use during drug design and is an automated process to quickly conduct tests on many compounds. Virtual Screening (VS) follows the same principle of quickly filtering through many compounds, with the difference that it operates entirely *in-silico*. [43] Virtual screenings are usually faster and can process larger libraries than experimental HTS. [44] Usually, the best hits identified in a VS, i.e. compounds predicted to bind well, are then tested experimentally. Compared to empirical screenings, the hit rates (ligands discovered per molecules tested) in libraries produced from virtual screenings are drastically improved. [45]

Virtual screenings differ by the method used to predict activity. Docking-based virtual screenings are a common approach and use molecular docking and its inherent scoring function to predict active compounds. [31, 46] An alternative approach - pharmacophore screening - caters for even larger libraries with millions of compounds and queries the database based on predefined chemical properties, which are thought to determine activity. [47]

2.5 Pharmacophores

A pharmacophore is "an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response". [48] Functional groups are classified into few pharmacophore feature types. The types include hydrogen bonds (acceptors or donors), ionic and hydrophobic areas, aromatic ring systems and exclusion volumes — spheres representing the boundaries of the binding pocket and limiting the size of the found molecule. A 3D pharmacophore describes a set of such features and their respective positions in three dimensional space.

There are several strategies to derive pharmacophores, which are classified in two big categories: ligand-based and structure-based. [22] Ligand-based methods utilize a set of active compounds to look for shared features. Structure-based methods are performed directly on the macromolecule targets, mostly proteins, or on macromolecule-ligand complexes.

In this work, we will focus on the latter of the two strategies. A straightforward approach utilizes mentioned macromolecule-ligand complexes, where the ligand is co-crystallized or docked into the

target site. The most probable interactions between the ligand and the target are calculated using chemical and geometric rules and then pose as blueprints for the pharmacophore features. Selected features are summarized in a pharmacophore. [49]

Other methods focus on macromolecule models without a ligand called apo structures. They place pharmacophore features inside the cavity *de novo*, which is especially useful when no ligands are known. Nonetheless, apo structures are also viable for the search of novel active regions unbiased by existing ligands. The aforementioned MIFs can be used to predict protein-ligand interactions and even elude pharmacophore features based on hotspots, points where the molecular interaction field reaches a minimum. [50]

3 Material, Methods and Tools

The data needed to perform our virtual screening is comprised of the target structure and, if available, a potential lead. Methods and tools are introduced in the scope of the pipeline.

3.1 PPIP5k2 and Compound 105229

Crystallographic structures of PPIP5K2 are publicly available at the Protein Data Bank (PDB). [51, 52] We chose the structure named 4NZN (Figure 3.1) due to its good resolution at 1.75Å. [16, 53] It is crystallized in complex with ATP analogue Adenylyl-imidodiphosphate (AMP-PNP) bound in the catalytic pocket and an inositol phosphate analogue bound to the external substrate capture site. [16] AMP-PNP is a non-selective ATP competitive inhibitor for most ATP binding sites and we assume it follows the same or a similar binding mode as ATP.

The discovered active lead compound 105229, which inhibits PPIP5K2 with an $IC_{50} = 15.5$ μM, is described by the Simplified Molecular Input Line Entry Specification (SMILES) O=C(NC=N1)C2=C1C=C(C(NC3=NC=CS3)=O)C=C2 and consists of two parts: a quinanzolinone moiety and a thiazole functional group (Figure 1.2).

Bartsch and Fiedler tested 18 *in-vitro* synthesised analogues and determined their IC_{50}. They provided the analogues in the SMILES format (Appendix B).

3.2 The Pipeline

Our pipeline compiles every computational process in chronological order as demonstrated in Figure 3.2. It showcases the combination of the different CADD methods and how they can complement each other with their unique strengths and weaknesses. Each step is described in detail in the following section.
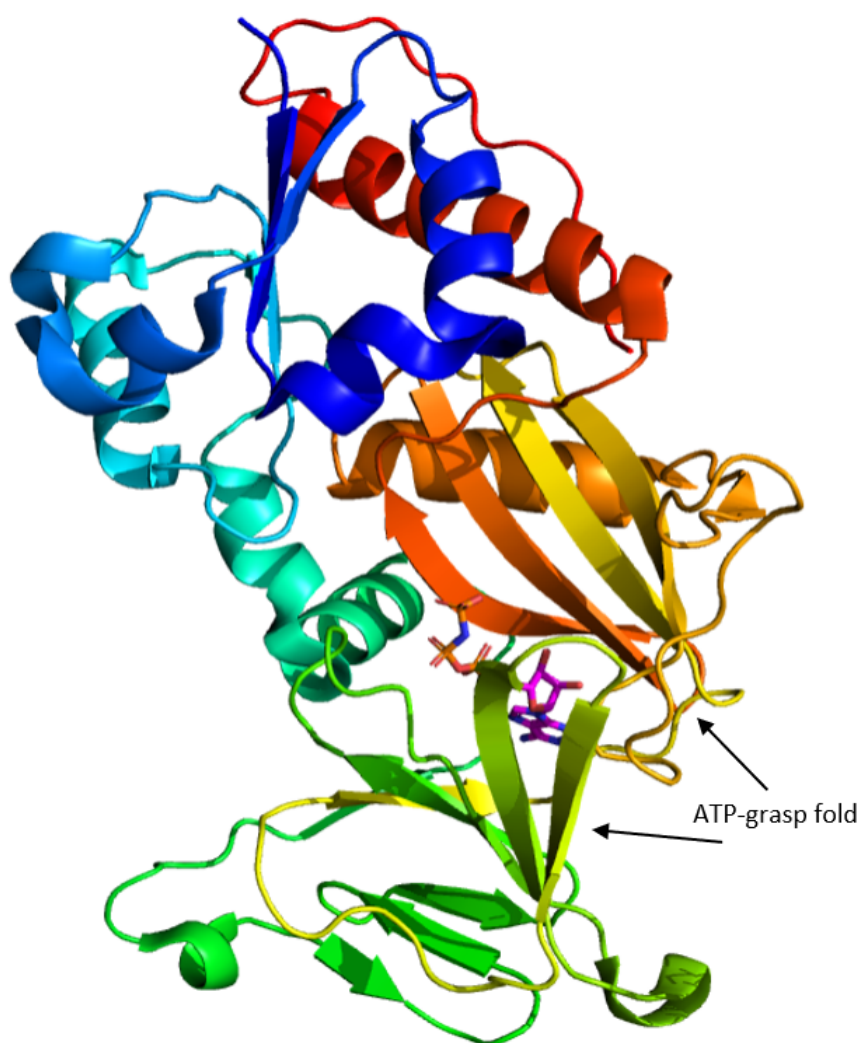
Figure 3.1: **Crystal structure of the catalytic domain of PPIP5K2**. It is in complex with AMP-PNP (magenta) bound inside the ATP-grasp fold. Structure with PDB id 4NZN. [16, 53]

Binding site identification is not required, since the ATP-grasp fold is a well known motif and in our model it is occupied by the bound AMP-PNP molecule. Instead, the pipeline can either be started with an identified binding site and a lead inhibitor for the ligand-target interaction-based method (green), or with just the apo protein structure for the second approach (turquoise). The pipeline then continues with the focal point virtual screening and subsequent hit filtering. The last node represents our goal: The identification of a small set of candidate structures for synthesis and *in-vitro* testing in the lab (orange).
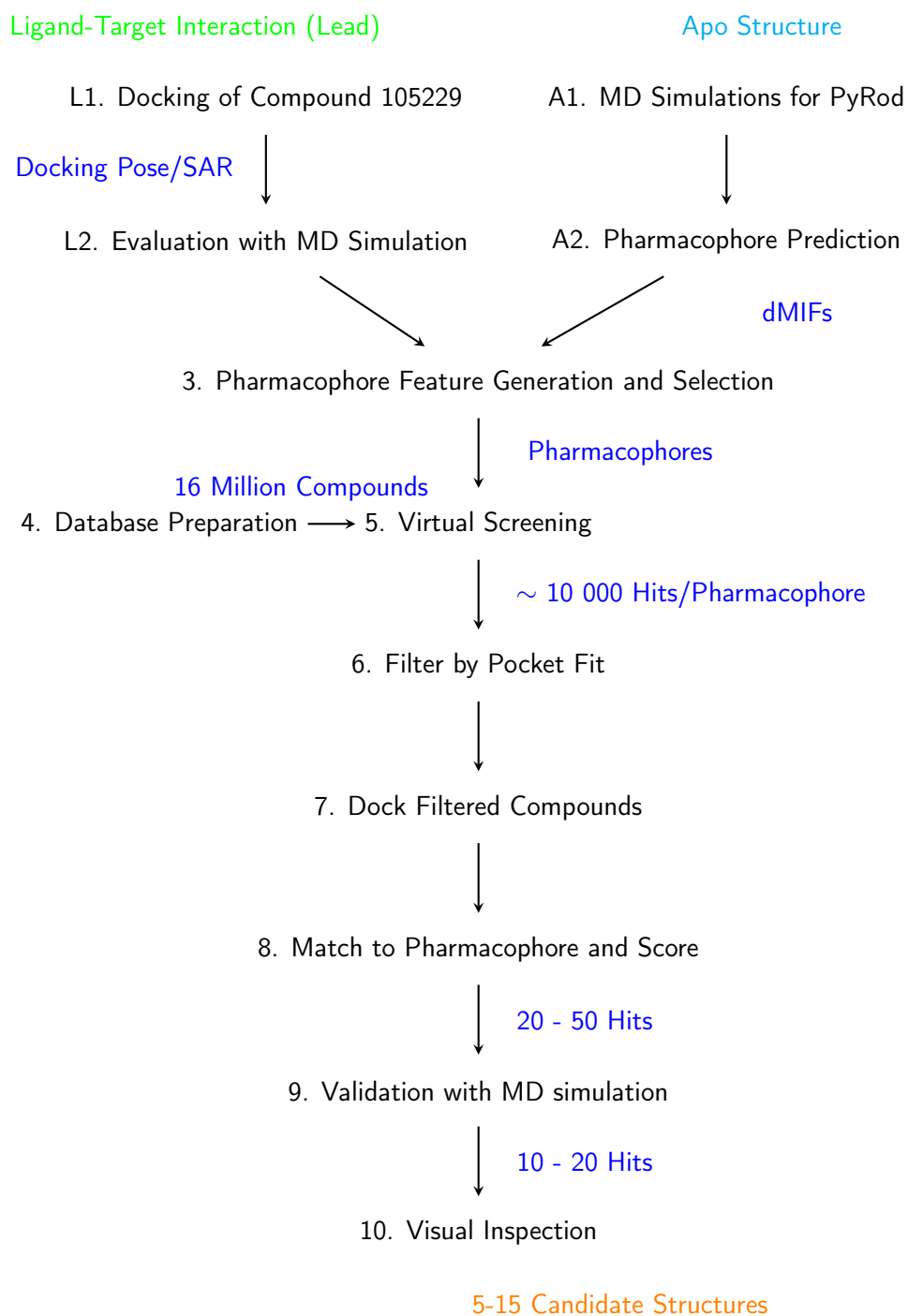
Figure 3.2: **Pharmacophore-based virtual screening pipeline for active compound discovery for PPIP5K2.** Included are active steps (black) over two approaches (green and turqoise). It is supplemented with expected intermittent outputs (blue) and the desired final result (orange). Detailed descriptions for Steps L1/A1 - 10 are given in Section 3.2.1.

3.2.1 Detailed Descriptions of Pipeline Steps

Here the Steps 1 - 10 of the pipeline (Figure 3.2) are described in more detail. The beginning of the pipeline differs for both methods. The ligand-target interaction (lead)-based approach includes Steps L1 and L2, the apo structure-based path likewise A1 and A2. Commencing from Step 3: Pharmacophore Feature Generation and Selection, all steps are carried out for both approaches, albeit sometimes with differences in the exact procedure.

L1. Docking of Compound 105229

Docking results highly depend on the setup of the docking procedure.

Structure preparation: We prepared the structure 4NZN for docking and proceeded to dock the lead and selected analogues. Before preparation, we removed water molecules, ions and other ligands from the structure 4NZN. For the final docking of 105229, we additionally stripped off all atoms from the AMP-PNP molecule not belonging to the adenine core. This procedure should decrease the bias by the co-crystallized AMP-PNP molecule in the HYBRID docking protocol.

Docking procedure: OESpruce from OpenEye's toolkit version 2020.2.2, which is imported in the KinoML framework, was used to prepare the cleaned structure including modeling missing side chains, capping of termini with NME and ACE as well as protonation at pH 7.4. [32, 54, 55] Ligands for docking were generated from SMILES and docked into the prepared structure using the HYBRID method. [35] In total, 20 docking poses were generated for the lead and its analogues to sufficiently sample possible binding modes.

Result visualization: We used LigandScout 4.4 to visualize the results. Docking poses were optimized by minimizing their energy in the Merck molecular force field 94 (MMFF94) employed by LigandScout. The MMFF94 is optimized for a wide range of organic chemistry calculations including proteins and small molecules. [56] We calculated the binding affinity score for each optimized pose and returned the best scoring pose as our result. [57]

L2. Evaluation with MD Simulation

We used the open-source tools OpenMM 7.5.0 [58] and OpenFF Toolkit 0.9.1 [38] to conduct the simulation. For simulation of docking poses the topologies of the ligand docking pose, as predicted

in Section 3.2.1, and the protein structure 4NZN were merged into a single system following the procedure as implemented in Talktorial T019: Molecular dynamics simulation in TeachOpenCADD. [59] This system was then solvated and simulated. The simulations were executed at 300 K using 2 fs time steps for a total of 100 ns of simulation time preceded by a 5 ns equilibration phase. As force fields, we used Amber14s ff14SB [60] and TIP3P [61] force fields for protein and solvent as well as openff-1.3.0 [62] for small molecules.

A1. MD Simulations for PyRod

Starting from an apo structure, PyRod analyses the protein environment of water molecules during MD simulations to place pharmacophore features at positions occupied by stable water molecules, whose replacement with ligand moieties is favourable. [28] It is recommended to perform more than one MD simulation as input for PyRod. The PyRod repository [63] provides a script running MD simulations in a serial fashion and additionally aligns and centers the protein in the water box. The script also performs simulations using OpenMM [58] with default parameters set similarly to what we use in Step L2 and 9. It uses the Amber14 as protein field and TIP4P-Ew [64] as water model. We performed ten 10 ns simulations on the solvated protein with no bound ligands.

A2. Pharmacophore Prediction

The last 5 ns from each of the ten replicates from the previous Step A1 were used as input for the PyRod trajectory analysis. The analysis was thereby focused on a 20 Å x 20 Å x 20 Å cubic box placed to hold the entire pocket without incorporating the external substrate capture site of PPIP5K2. Placement and size of the box were verified before any analysis was conducted using the "test_grid" functionality PyRod offers, where the defined box is visualized in the structure. Next, we executed PyRod to place 20 features per feature type in the defined box encapsulating the binding site.

3. Pharmacophore Feature Generation and Selection

Note that from here one, the steps were executed for both strategies introduced above.

Lead-based: Based on the docked pose of 105229 in the structure 4NZN, the lead-based pharmacophore features were generated by LigandScout from the most probable interactions in the energetically most favourable docking position. To do this, structure 4NZN was loaded into LigandScout.

Then we deleted water molecules, two magnesium ions and the co-crystallized ligand in the external capture site before inserting the 20 docking poses of 105229. We chose the pose with the best binding affinity score for pharmacophore featrue generation. LigandScout placed eight features, six hydrogen bonds and two hydrophobic interactions from which we selected five hydrogen bonds.

Apo structure-based: In the apo structure-based approach, features were generated as described in the Step A2. All $6 * 20$ features are collected in a so called "super pharmacophore" containing a vast amount of possible interactions. We inserted the pharmacophore into 4NZN and roughly deleted all features placed outside the pocket as well as ionizable and hydrophobic regions. From there, we chose sets with three to five features each for individual screenings.

4. Database Preparation

We used the commercially available compound database from MolPort, as of the 28th June 2021. [21] The database contains more than seven million compounds in SMILES format. For all individual SMILES we generated missing enantiomers, reasonable tautomers and the most likely protonation state with OpenEye's OEQuacPac [65] functionality, then used LigandScout's tools *idbgen* to create a screenable database from our enumerated smiles and generate 25 conformations per compound. The collection included some overly big structures and long lipids, which were not eligible for our case. Thus, we set a weight limit of 1000 Da. The completed database had more than 16 Million compounds.

5. Pharmacophore-Based Virtual Screening

LigandScout offers the command line tool *iscreen* to screen the generated libraries with selected pharmacophores. For screening, we generally opted for full feature overlap and no optional features. Only the first matching conformation was selected. When two pharmacophores are screened at once, only the first matching pharmacophore is noted as match. LigandScout accumulates all matched compounds, called hits, in a single file, which we subsequently sorted into different files by their "Pharmacophore Match" property.

6. Filter by Pocket Fit

In the chosen protein structure, the bound ATP defined the location of the binding pocket, but it did not show the volume and form of the binding pocket. To predict more precise extents of the binding pocket, we used the open-source binding site detection tool fpocket. [66] Fpocket detects cavities in a protein based and is suited for *de novo* prediction in large protein structures. It returns a set of placeholder atom coordinates dispersed along the surface of each pocket. Fpocket additionally scores the size and druggability of the pockets. We used fpockets original *fpocket* function on the apo protein of 4NZN and the centroid structures. It detected several pockets and the correct set of placeholder atoms was determined visually in the molecular visualization system PyMol version 2.4.1. [67]

Lead-based: The pocket area was defined as a 5 Å perimeter around the pocket descriptor points generated from fpocket. The 5 Å perimeter is needed to cover the entire pocket space, as the defining points are placed only on the pocket surface.

The pocket needed to be extended to fit the position of the lead compound pose, which is slightly solvent exposed in a favourable way. We added a 3 Å perimeter around the lead docking pose to the pocket volume (Figure 3.3), which is large enough to accommodate even the better scoring, bigger analogues as tests with dockings from the described analogues confirmed.

To check if atoms from compound poses (hits) returned by the pharmacophore screening are inside of the defined scope, we calculated pairwise distances between hit atoms and the placeholders atoms describing our predicted pocket and additionally between hit atoms and the docking pose coordinates of the lead. The total distance of each hit atom to the pocket is described by its minimum distance to any placeholder atom defining the pocket. A pharmacophore screening hit was discarded if at least one atom was outside of the defined pocket volume.

Apo structure-based: Since the apo structure pharmacophores were derived from a MD simulation, the position and conformation of the protein changed when compared to the original crystallographic structure, even if minimally only. Although simulation frames are centered and aligned, this does not negate the effect. PyRod gives the option to search the simulation for frames that are able to fulfill the features of a given pharmacophore. The frames are clustered to select a single structure called centroid. We selected the centroid structure for all screened pharmacophores respectively. For the identification of the centroid structures, the pharmacophore feature tolerances and weights were adjusted to the recommended defaults.
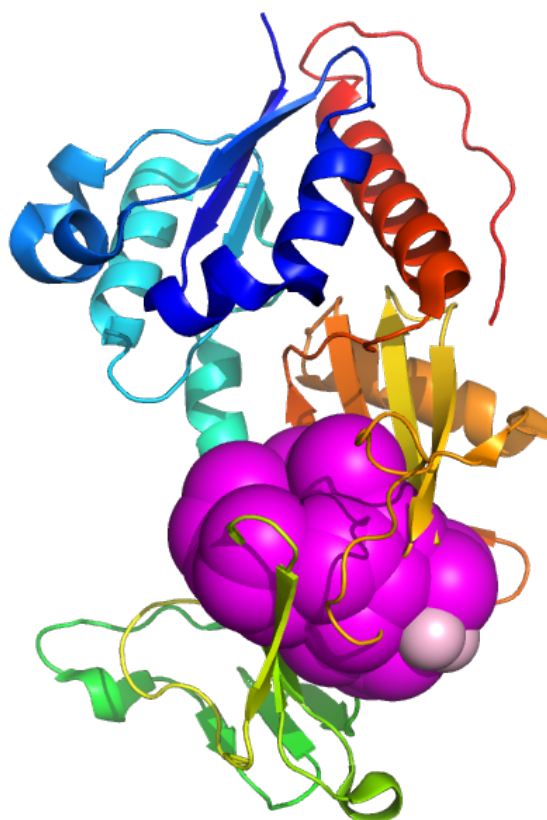
Figure 3.3: **Viable area of PPIP5K2s binding pocket.** Pocket volume is defined as 5Å around the fpocket coordinates (magenta) and 3Å around the lead docking pose (light pink).

Pockets volumes for the centroids were again determined with *fpocket*, as described earlier, without extension by the lead docking pose.

Hits from apo-pharmacophores were compared with the pockets of their respective centroid; those matching the lead pharmacophore with 4NZN and the docking pose of the lead.

7. Dock Filtered Compounds

After filtering the pharmacophore hits, re-docking was performed with OpenEye's HYBRID docking algorithm as described in Step L1: Docking procedure with the difference, that only three docking poses per hit were generated. Hits were always docked into the same structure as which was used for pharmacophore creation.

Lead-based: Hits found with the pharmacophore from the lead are docked into the receptor structure

4NZN, which we prepared in Step L1: Structure preparation.

Apo structure based: For re-docking of the filtered hits found with pharmacophores generated from the apo structure, we used the PyRod generated centroid structures. Since the centroids result from MD simulations, they do not contain AMP-PNP. HYBRID requires a ligand for its guided docking procedure. To reproduce the guided docking, we inserted AMP-PNP into the binding cavity. We used PyMol's [67] command *super 4NZN, centroid* to superimpose 4NZN over the centroid structure and then copied the ligand into the latter.

8. Pharmacophore Matching and Scoring

The next examination and filtering was performed using LigandScout's graphical user interface.

It was important, that the correct target protein structure (4NZN or centroid) and pharmacophore were loaded for the library of re-docked molecules. If applicable, we deleted water and co-crystallized molecules or ions (except AMP-PNP as it is replaced by the inserted molecules) to insure that optimization and scoring were not disturbed by molecules placed in or near the cavity. Energy minimization of the docking poses with the MMFF94 force field [56] was performed to relax any artificial clashes or unfavorable conformations introduced by the docking algorithm.

We used two independent scoring functions from LigandScout to evaluate compounds: Pharamcophore fit score and binding affinity score.

The pharmacophore fit score is a measurement for comparing the pharmacophoric properties of screening hits with a template pharmacophore, usually the one used for screening. We retained only hits which achieve a fit score of at least ten-times the number of features, which indicates full feature overlap as each pharmacophore feature, that can be matched by the hit scores ten points. A maximum of ten extra points can be given for extra good fits.

The binding affinity score estimated the affinity of a ligand to the target. We retained hits with a binding affinity score lower than -20, which roughly corresponds to estimated micromolar activity.

Pharmacophore match scoring and filtering of docked hits was conducted with the following steps all performed with inherent LigandScout library functions:

1. Load corresponding protein structure and pharmacophore into LigandScout

2. Insert library of docking poses

3. Delete water, two magnesium ions and other ligands, if present

4. Minimize MMFF94 energy of library of docking poses

5. Calculate unaligned pharmacophore score without exclusion volumes

6. Calculate the binding affinity score

7. Filter for an unaligned pharmacophore score $> 10*$ number of features)

8. Filter for a binding affinity score < -20

Duplicates, which stem from the sampled docking poses or simply duplicates in the database, were removed retaining the generally better scoring alternative. From the remaining hits, candidates for MD simulation were hand-picket in a visual inspection. The selection was based on:

- Interactions with the target (Additional interactions beyond the pharmacophore features)

- Polarity matching with pocket surface (Pocket surface visualized in LigandScout)

- Similarities to other compounds (Not to many compounds with minimal deviations)

.

9. Validation with MD Simulation

The last step of the selection of candidate molecules was MD simulation. Here, we assessed the stability of the binding pose in 100 ns simulations with the same settings for the simulation as described earlier. Again, it is important to use the prepared structure fitting to the respective query compound. We calculated the per-atom root mean-square fluctuation (RMSF) of the candidate ligand atoms and the root-mean-square deviation (RMSD) of the candidate ligand throughout the simulation using the Python package MDAnalysis. [68, 69] Trajectories were aligned along the C-alpha atoms before calculating these scores. For the RMSD calculation the first frame of the simulation was used as reference. We allowed a maximum RMSF/RMSD of 2 Å for the lead-based and 2.1 Å for the apo structre-based approach.

10. Visual Inspection

From the remaining compounds the final set was again selected manually. The reduced library was loaded into LigandScout as described in Step 8. Energy minimization was performed per molecule and this time included the side chains. The optimized compound poses were judged based on quality of interactions and uniqueness and the final set of candidates was selected.

Visualisations of candidates in the results (Section 4.3) are based on this final, minimized pose.

3.3 Code and Availability

Scripts used during this thesis are publicly available at `https://github.com/volkamerlab/ppip5k2`. All resource intensive calculations (virtual screening, molecular docking, molecular dynamcis simulation) were performed on the high-performance cluster (HPC) Curta run by Freie-Universität Berlin. [70]

4 Results

PPIP5K2 is important for phosphate homeostasis and its deregulation leads to several diseases, however its function and regulation are not well understood yet. No potent inhibitor usable for research is known and an HTS screening only revealed a single hit compound out of 60000 screened molecules, with insufficient inhibitory activity. Thus arises need for a computer-aided pipeline with access to more compounds and faster estimation of affinity to support the optimization and identification of novel potential active compounds.

The pipeline we designed was already introduced in its entirety in Section 3.2. However, while we did have a concept in the beginning, the details of the pipeline were developed over the course of this work. They were fine-tuned to fit the purpose of the study in hand, but were implemented in a way that they could easily be adapted to study other proteins and libraries. Thus, the pipeline itself should be understood as part of the results.

For the identification of novel PPIP5K inhibitors, we ran the complete pipeline twice; once based on interactions derived from the lead compound 105229 to PPIP5K2 and once based on the PPIP5K2 apo structure. Any intermittent outputs produced by the processes in the pipeline are more than just the input for the next step. They are essential for quality control and often have intrinsic informational value.

The final output is a set of compounds, our candidates, for both approaches.

4.1 Ligand-Target Interaction-Based Screening

The first run was based on the known active molecular structure 105229, the lead compound discovered by Bartsch and Fiedler in a HTS screen. The goal was to derive pharmacophore features from the structure-activity relationship (SAR) of the lead with PPIP5K2 and thus find compounds with similar properties.

4.1.1 Binding Hypothesis

To derive pharmacophore features from the interactions between the lead and PPIP5K2, we first needed a binding hypothesis. Since no co-crystallized structures of PPIP5K2 with the lead are available, we explored it with hybrid docking. The first hybrid docking (using the OpenEye software) of 105229 to the PPIP5K2 structure with PDB id 4NZN was guided by the complete co-crystallized AMP-PNP; the best pose including the generated pharmacophore using LigandScout is shown in Figure 4.1. The orientation of 105229 in the binding pocket closely resembled that of ATP. Moreover, the quinanzolinone moiety and the protein backbone of MET240 and GLU238 formed hydrogen bonds also visible in the binding mode of ATP.
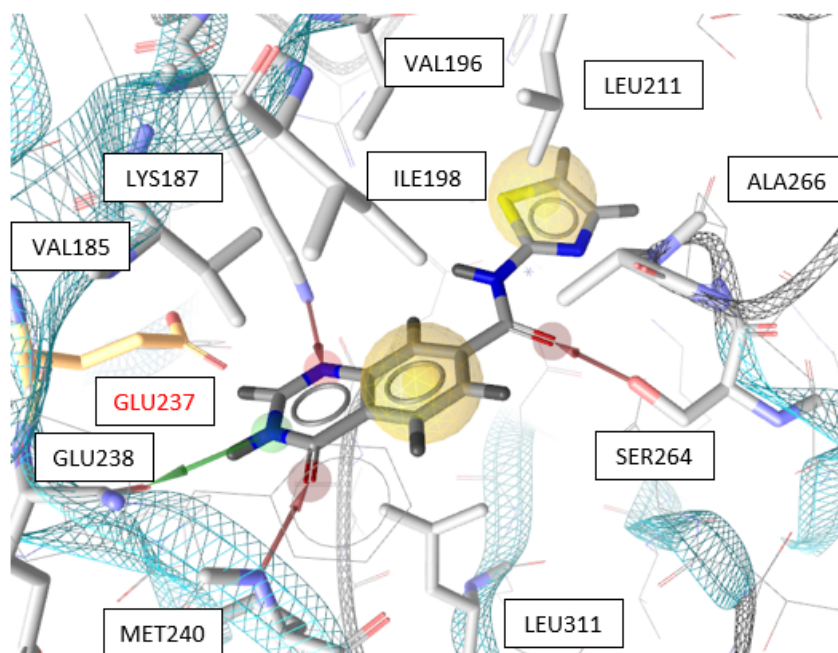
Figure 4.1: **Docking pose of lead compound 105229 with full AMP-PNP as guide for the hybrid docking algorithm.** Interacting protein residues are visualized as white "sticks". GLU137 shown in orange showcases an unfavourable proximity to the lead. The binding site of the PPIP5K2 structure with PDB id 4NZN is shown as "snake" and "lines" representation. Arrows depict hydrogen bonds. Green arrows mark ligand hydrogen bond donors, red arrows acceptors. Yellow spheres indicate hydrophobic interactions. Interacting protein residues are visualized as white "sticks". Figure created using LigandScout.

However, we questioned the results. The hydrogen bonds with MET240 and LYS187 are promising but sparse. Furthermore, there was a hydrogen pointing towards the side chain GLU237, which is not favourable. The thiazole moiety of the lead reached into a voluminous, flexible and charged part of the binding pocket — perfect for the bulky and charged phosphate groups of ATP but not suitable

for a small, hydrophobic moiety like the thiazole of 105229. The binding affinity score of -22.71 calculated by LigandScout was not outstanding enough to disperse our doubts. A 100 ns molecular dynamics (MD) simulation confirmed our apprehension: 105229 did not bind stable, it drifted and the thiazole moiety moved freely.

We hypothesised that the hybrid docking is biased too much by the position of AMP-PNP. We performed the docking again with only the adenosine ring of AMP-PNP as a template for hybrid docking, trying to retain key interactions while allowing for more diverse poses.

The second docking attempt resulted in a flipped docking pose, which was more stable in a subsequent 100 ns MD simulation and avoided the highly flexible, charged area targeted by the phosphate groups of the ATP analogue. The quinanzolinone double ring structure of 105229 still overlapped with the AMP-PNP adenosine ring, retaining interactions around the core, but it reached through the pocket orthogonal to AMP-PNP (Figure 4.2).
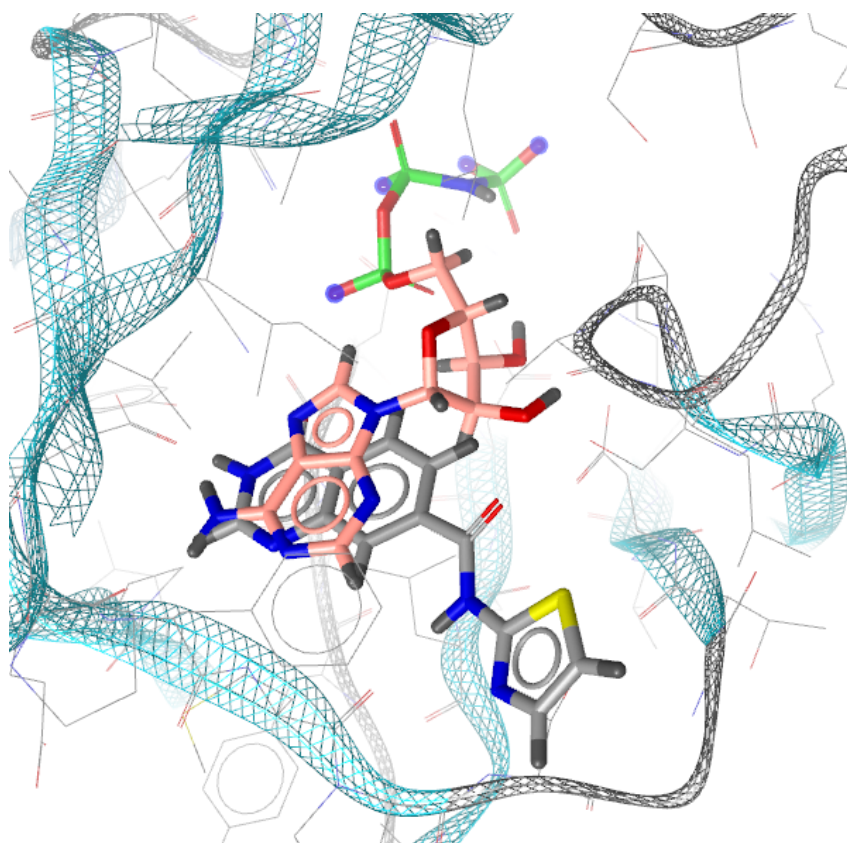
Figure 4.2: **Orientation comparison of co-crystallized AMP-PNP (gray) and docking pose of 105229 (orange)** in the ATP-grasp fold of PPIP5K2 [PDB id 4NZN]. Binding site shown as "snake" and "lines" representation.

The revisited docking pose (Figure 4.3) was characterized by three hydrogen bonds between the core double ring structure and the side chains of LYS187, GLU237 as well as the backbone of MET240. Three more hydrogen bonds interact with the thiazole moiety and the connecting amide. Pharmacophore feature generation from LigandScout further identified hydrophobic contacts in both ring structures, namely in the core's quinazolinone and the thiazole ring. Generally, the calculated pose of 105229 expressed a good fit for the pocket with the thiazole ring pointing towards the solvent. This binding pose had a binding affinity score of -34.83, which is better when compared to the earlier docking results with full AMP-PNP as guide.
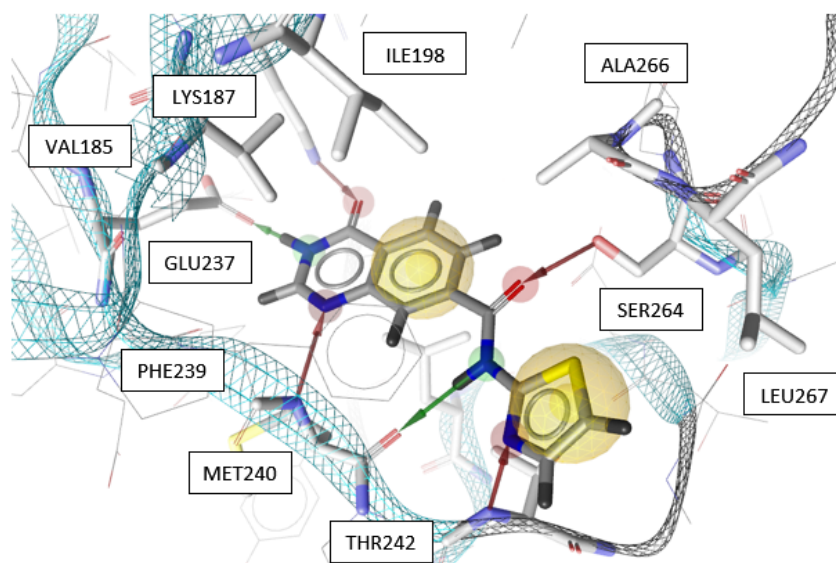
Figure 4.3: **Binding hypothesis of 105229 with pharmacophore features.** Binding site of PPIP5K2 structure [PDB id 4NZN] shown as "snake" and "lines" representation. Arrows depict hydrogen bonds. Green arrows mark ligand hydrogen bond donors, red arrows acceptors. Yellow spheres indicate hydrophobic interactions. Interacting protein residues are visualized as sticks. Figure created using LigandScout.

In order to verify our binding hypothesis, we tried to reproduce the SAR experiments from Bartsch and Fiedler (unpublished). To improve the binding affinity of 105229, which is only active in micromolar concentrations of 15.5 μM, they synthesized analogues in a first lead optimization campaign. This revealed that small alkyl substitutions with methyl- or ethyl-groups at the 4-position of the thiazole moiety slightly enhanced the binding affinity. However, bigger substitutes like cyclopropyl or aromatic rings worsen the binding affinity. Any modifications on the quinazolinone moiety were not favourable (Table 4.1).

We docked the favourable thiazole analogues into the PPIP5K2 binding site using OpenEye's hybrid docking functionality and scored the results with LigandScout's binding affinity score. Our calculated binding affinity scores matched the ranking based on the IC_{50} values from the SAR study as seen in Table 4.1, which gave us confidence in our binding hypothesis. Our hypothesis explains why some hydrophobic additions to the thiazole moiety are unfavourable: The thiazole moiety is already solvent exposed and while hydrophobic moieties at this place are generally good, too much solvent exposure with bigger additions is adversary. Furthermore, it is clearly visible why modifications to the quinanzolinone moiety, like N-Methyl-Quinanzolinone, were unfavourable: The double ring structure forms the core hydrogen bonds with the protein backbone. The tested additions must have disturbed these interactions, weakening the binding affinity.
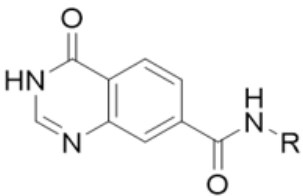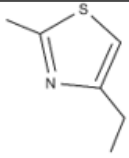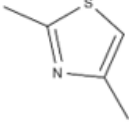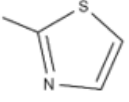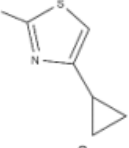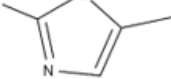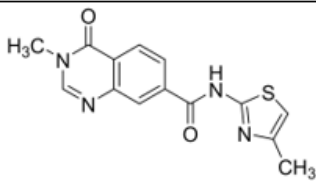
Base Structure	Compound	Modification	Score	IC50
	4-Ethyl-105229		- 35.96	8.1 µM
	4-Methyl-105229		- 35.48	9.5 µM
	105229		- 34.83	15.5 µM
	4-Cyclopropyl-105229		—	40.6 µM
	5-Methyl-105229		- 34.74	160.7 µM
	N-Methyl-Quinanzolinone-105229		—	259.7 µM

Table 4.1: **Binding affinity scores of docked analogues of 105229 and measured IC_{50} values.** The lead structure is modified at the position of the thiazole moiety (R). Binding affinity scores were calculated in LigandScout from the best scored docking pose, IC_{50} values were taken from Bartsch and Fiedler's SAR study (unpublished).

4.1.2 Pharmacophore Derived from the Docking Hypothesis

For our pharmacophore model, we picked a subset of features from the interactions between the lead and the target proposed by our binding hypothesis. A total of eight features are detected in the binding hypothesis pictured in Figure 4.3: Four hydrogen bond donors, two acceptors and two hydrophobic regions. We focused on the core interactions deep inside the pocket and forwent the hydrophobic interactions entirely, since they are not present in the AMP-PNP complex and we therefore deemed them the least important. The pharmacophore was still rather complex with six hydrogen bonds. Not wanting to disturb key interactions inside the pocket, we removed the acceptor involving the thiazole moiety at the end of the molecule. The refined pharmacophore with five features (leadPharm) used for screening still provides suitable anchor points that span around the core structure and could position the compound tightly in the binding site (Figure 4.4).

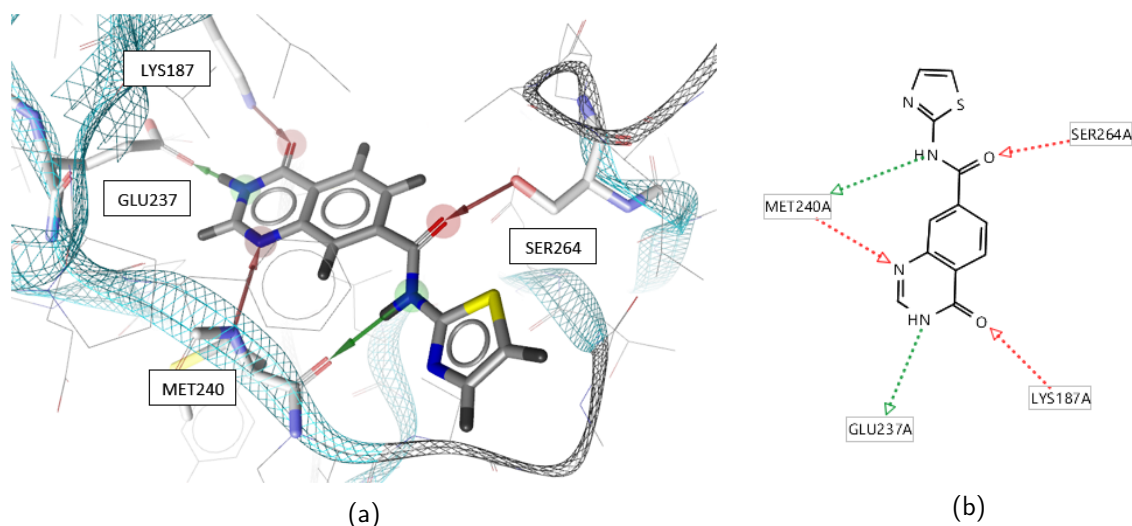(a)   (b)

Figure 4.4: **Refined pharmacophore leadPharm.** Best docking pose of the lead compound with PPIP5K2 [PDB id 4NZN] and selected pharmacophoric features in 3D (a) and 2D (b) representation. Included are five selected hydrogen bonds and no hydrophobic interactions. Red arrows depict hydrogen bond acceptors, while green arrows represent donors. Binding site in (a) shown as "snake" and "lines" representation. Interaction partners are visualized as "sticks".

4.1.3 Screening Results and Compound Selection

Pharmacophore screening of the prepared MolPort library containing more than 16 Million compounds with the leadPharm pharmacophore using LigandScout lead to around 13432 hits. 10708 of those fit into the binding pocket, the rest was removed by our pocket-fit filtering (Step 6 of the pipeline

in Section 3.2.1) For the next filtering step all those 10708 compounds were docked with OpenEye's HYBRID algorithm. The docking poses were loaded into LigandScout with structure 4NZN and the leadPharm pharmacophore and minimized without protein side chains, Then the pharmacophore match scores and binding affinity scores were calculated. The docking poses do not necessarily align with the pharmacophore features as the procedure aims to minimize energy without prior knowledge of ligand orientation or features. Thus, although all pharmacophore screening hits were selected in the previous step to fulfill all five features, not all compounds still matched those features after docking. The binding affinity scores painted a similar picture, although generally satisfactory. This left us with 175 compounds fulfilling both imposed thresholds: A binding affinity score better than -20 and a pharmacophore fit score larger than 50 after docking. Hit numbers evolution per computational step is summarized in Table 4.2, the first row describes the above experiment.

	Screening	Pocket Fit	Scoring	Visual Insp.	MD Sim.	**Candidates**
leadPharm	13 432	10 708	176	53	13	9
apoPharm-1A	156 902	—	—	—	—	—
apoPharm-1B	38 620	—	—	—	—	—
apoPharm-2	1 094	570	0	—	—	—
apoPharm-3A	3 000	2 863	2 (12)	—	—	—
apoPharm-3B	4 540	4 269	59 (376)	28	10	8

Table 4.2: **Number of hits retained after each filtering step** for all six pharmacophores. Striked cells denote that calculations were not performed. Scoring thresholds were binding affinity score < -20 and full pharmacophore match (One feature missed).

We manually selected 53 structures based on their interactions of the pocket apart from those defined by the pharmacophore features, similarities to other structures and polarity matching with the pocket surface (Section 3.2.1). If there were duplicates — either different docking poses of the same compound or duplicates in the database — the one with the better binding affinity score was selected.

For this final set of 53 identified hits, MD simulations of 200 ns were performed to check the stability of the compounds. For each resulting MD trajectory, the RMSD and RMSF values were calculated. Most compounds performed well, with very few outliers (Figure 4.5).

The required thresholds of 2 Å for both measurements, RMSF and RMSD were fulfilled by 13 compounds. From these results, we picked nine diverse compounds with seven unique cores. Quality of interactions and uniqueness were the main attributes for the selection. The picked molecules are introduced in Section 4.3.1.
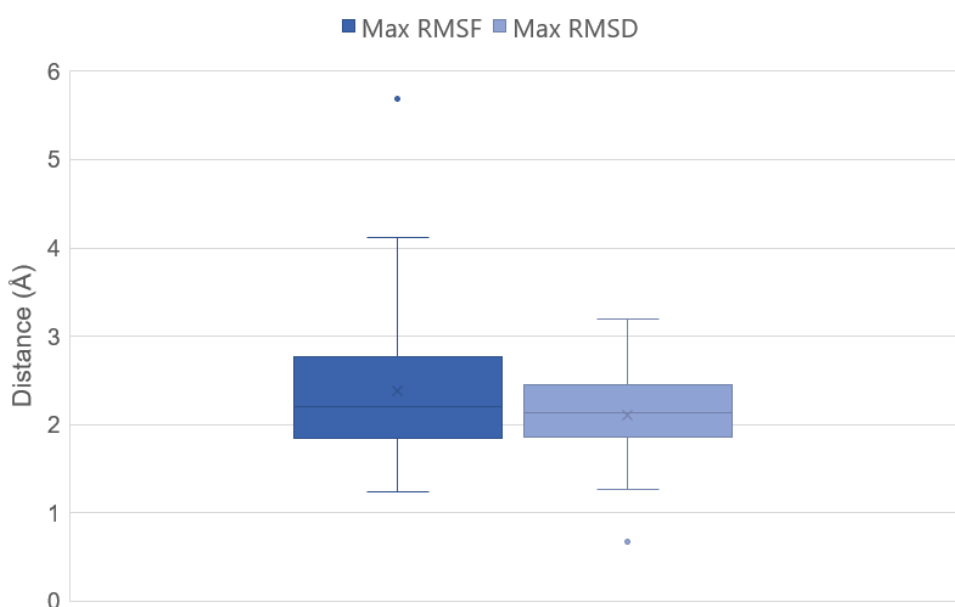
Figure 4.5: **Boxplot of the maximum RMSF and RMSD values for the 53 simulated structures from screening the lead derived pharmacophore leadPharm.** The medians are 2.26 Å (RMSF) and 2.20 Å (RMSD).

4.2 Apo Protein Structure-Based Screening

We only had one lead compound and were thus limited to a small interaction set for pharmacophore creation and little possibilities for variation. Thus we introduced a second approach: apo structure-based screening. Starting from an apo structure had the advantage of being unbiased and it presented the opportunity to discover potential interaction sites not covered by the lead ligand.

4.2.1 Initial Feature Set

PyRod is a tool that allows to identify pharmacophoric hot spots in apo structures by sampling the protein environment of water molecules during MD simulations. Ten 10 ns MD simulations were performed on PPIP5K2 structure with PDB id 4NZN without the co-crystallized ligands. PyRod identified a total of $6 * 20$ features. The features detected by PyRod were distributed over the entire analyzed box, but we observed an obvious grouping inside the ATP-grasp fold and the general pocket area (Figure 4.6). Hydrophobic interactions were mainly placed at the solvent exposed parts of the pocket. Negative ionic interactions were buried inside of the pocket, where the charged part of ATP resides when bound. Contrary, positive ionic interactions appear randomly placed and mainly outside
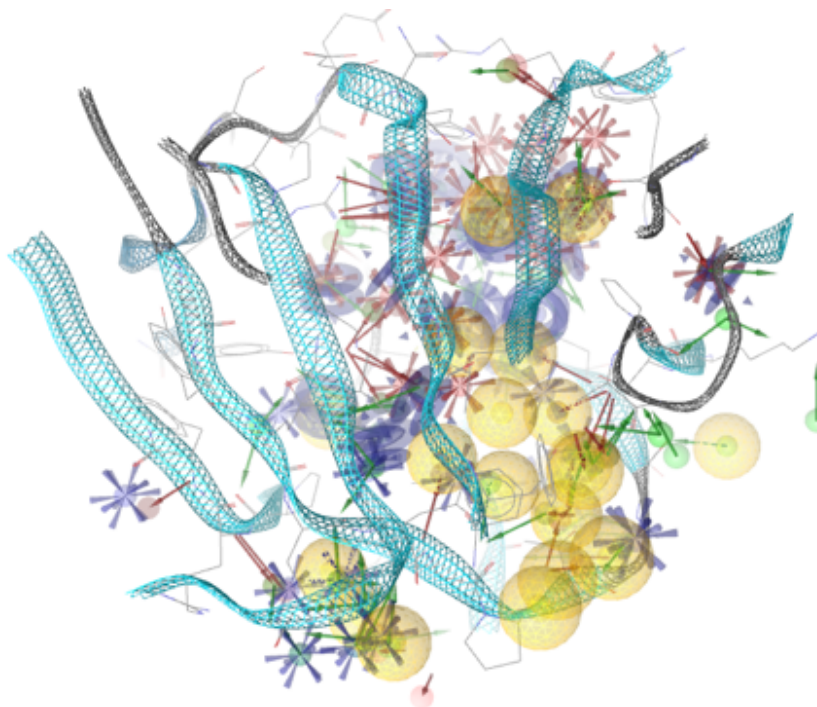
the pocket.

Figure 4.6: **Super pharmacophore generated by PyRod** placed over the binding pocket of structure 4NZN of PPIP5K2 drawn as "snake" and "lines". Hydrophobic interactions (yellow spheres) are grouped towards the front of the pocket. Hydrogen bonds (red and green arrows) are found along the protein backbone. Negative ionic interactions (red stars) and aromatic interactions (blue rings) are located deeper inside the pocket. Positive ionic interactions (blue stars) appear scattered outside of the fold.

The aromatic interactions grouped in the center of the pocket were intriguing, as they did not appear in any analysis before. One of the aromatic interactions caught our eye, because it was placed right above the core region filling the pocket space nicely. None of the synthesized FMP analogues had additions at this position and we did not know yet how favourable it his. To test the viability of aromatic features at the position suggested by the PyRod pharmacophore, we created SMILES for lead analogues with aromatic rings and other hydrophobic additions at the corresponding position. These analogues were docked into 4NZN using hybrid docking and the binding affinity score was calculated in LigandScout. As seen in Table 4.3, any hydrophobic addition improved the estimated binding affinity, with bigger and bulkier structures scoring even better.

Following the promising docking results, we selected the aromatic feature and declared it one of the key features. Apart from the aromatic interaction, we again focused on hydrogen bonds in the core pocket region and selected a handful of hydrogen bonds along the backbone of the protein reducing
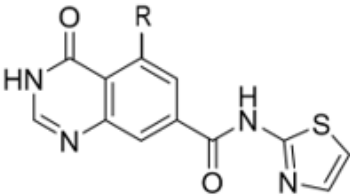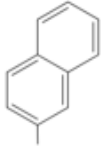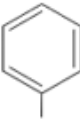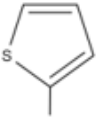
Base Structure	Compound	Modification	Score
	Naphtyl-105229		- 45.86
	Phenyl-105229		- 42.52
	Thiophen-105229		- 41.04
	Bromine-105229	Br	- 37.74
	105229	—	- 34.83

Table 4.3: **Scores of docked analogues of 105229.** The lead structure is modified at the quina-zolinone structure (R), i.e. larger substituents are placed there. Binding affinity scores are calculated from the best scoring docking pose.

the pharmacophore to a less overwhelming collection of features. All exclusion volumes were retained.

4.2.2 Iterations of Feature Selection and Screenings

As described before, too many features were exposed by PyRod for our apo structure of interest, so additional manual selection was needed. Note that in the following section three iterative pharmacophore feature selection attempts and their respective screening results are described.

Pharmacophore feature sets apoPharm-1A and 1B

At first, we selected two small subsets with three features each, both containing the aromatic interaction, described in detail above, and two hydrogen bond features (Figure 4.7). The first pharmacophore model was inspired by ATP, the second by the lead compound. The ATP-inspired pharmacophore has a double hydrogen donor with GLU237 and GLU238, thus matching the same feature present in ATP,

and an acceptor forming a hydrogen bond with the backbone of MET240 (Figure 4.7b). The pharmacophore inspired by the lead shared the same hydrogen bond acceptor with MET240, but has only a single hydrogen bond donor (Figure 4.7a). As declared earlier, we kept the aromatic feature. Thus, both pharmacophore models contained three features, counting the double-donor as single feature. We deliberately chose less restrictive pharmacophores in the beginning to test the general viability of the apo structure derived pharmacophores.
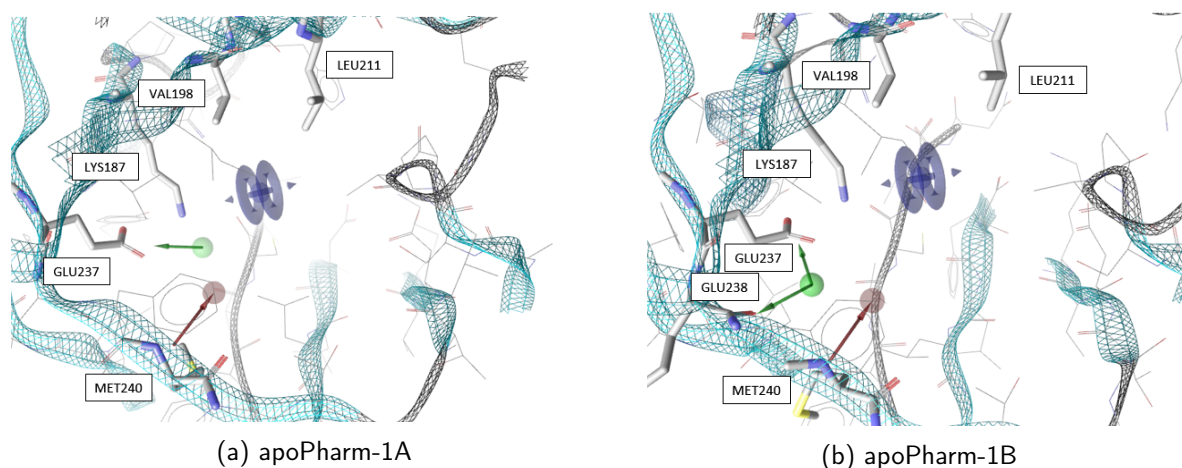

(a) apoPharm-1A   (b) apoPharm-1B

Figure 4.7: **Pharmacophore feature sets apoPharm-1A and apoPharm-1B selected from the PyRod generated super pharmacophore.** Both pharmacophores share the aromatic feature (blue ring) and a hydrogen bond acceptor (red arrow). (a) apoPharm-1A: The double donor is replaced with a single donor (green arrow), which is exhibited in the docking pose of the lead. (b) apoPharm-1B: A double donor (green arrows) similar to the interaction of ATP is added. The pharmacophores are visualized in their respective centroid in "snake" and "lines" representation. Interaction partners are visualized as "sticks".

A pre-screening with these two pharmacophores resulted in a total of 195522 hits. The huge number was not suitable for further analysis.

Pharmacophore feature set apoPharm-2

We decided not to continue working with these hits and instead elaborate on the feature selection. The large number of hits allowed us to be more restrictive with the pharmacophores and to add features, reducing the number of expected hits. We continued adding features to the ATP-inspired pharmacophore (Figure 4.7b) to ensure diversity compared to the lead interaction-based screening and because it had less hits than the lead inspired pharmacophore.

We added an acceptor at SER264 next to the aromatic feature to further span the binding pocket

(apoPharm-2, Figure 4.8). So far, pharmacophores probed mainly one side of the fold (GLU237 and MET240). Here, we included the opposite site into the apo structure derived pharmacophore. We then performed another screening with this four-feature pharmacophore. This lead to a immense decrease in hit numbers, resulting in merely 1094 hits. The same filtering steps as described in more detail above were imposed, namely filtering by pocket fit, and docking with subsequent scoring in LigandScout. No candidates were retained after scoring with LigandScout and we stopped analysis based on this pharmacophore.
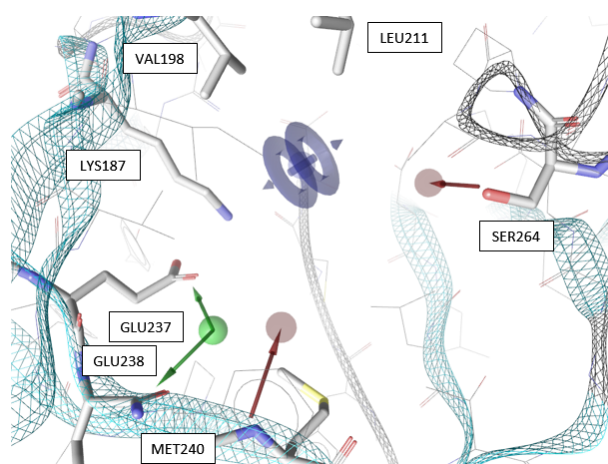

Figure 4.8: **Improved ATP-inspired pharmacophore apoPharm-2 with an acceptor added across the fold.** Arrows depict hydrogen bonds. Green arrows mark ligand donors, red arrows acceptors. Yellow spheres indicate hydrophobic interactions. The pharmacophores is visualized in its respective centroid in "snake" and "lines" representation. Interaction partners are visualized as "sticks".

Pharmacophore feature sets apoPharm-3A and 3B

Although we only used one additional feature, we underestimated the hit reduction and tried another selection to relax the search again. We decided to try adding other fourth features and opted once for a donor next to the other core interactions on the backbone (apoPharm-3A, Figure 4.9a) and once for an acceptor close to the aromatic ring (apoPharm-3B, Figure 4.9b). We expected more hits from the added donor since it is again more similar to the binding mode of the lead and thus the lead-inspired pharmacophore, which generated plenty of hits in the respective screening.

Pharmacophore screening with both pharmacophores, apoPharm-3A and 3B, in parallel generated 7540 hits; an improved result, both promising and workable. Surprisingly, we received more hits for apoPharm-3B — 4540 compared to 3000 with apoPharm-3A — although it was matched second.

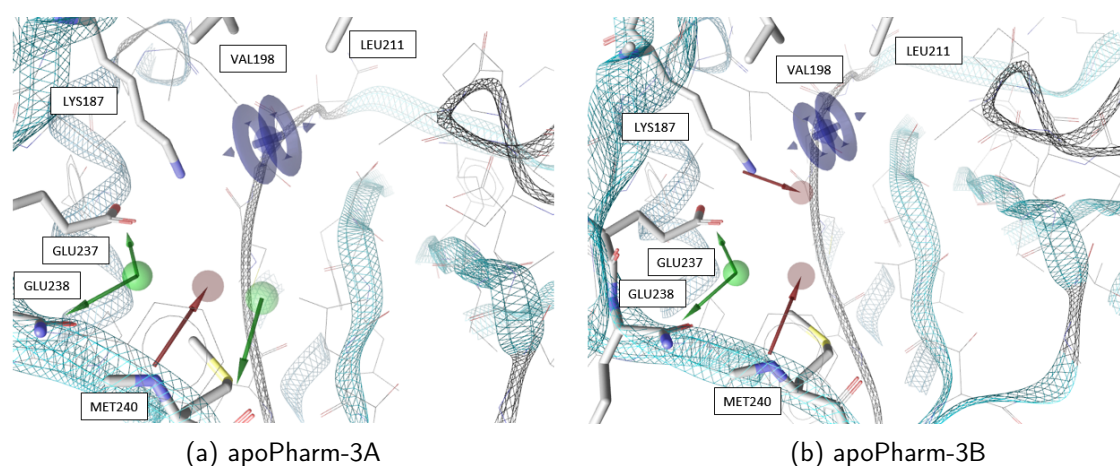(a) apoPharm-3A   (b) apoPharm-3B

Figure 4.9: **Improved ATP-inspired pharmacophore set apoPharm-3A and B.** The ATP-inspired pharmacophore from Figure 4.7b is extended with a donor (green arrow) next to the acceptor, apoPharm-3A (a) or with another acceptor (red arrow) below the aromatic interaction, apoPharm-3B (b). The pharmacophores are visualized in their respective centroid in "snake" and "lines" representation. Interaction partners are visualized as "sticks".

Filtering structures with to much solvent exposure removed only a few hits (271 for apoPharm-3A and 137 for apoPharm-3B). The final sets of $2863/4269$ compounds respectively were re-docked with OpenEye's hybrid docking functionality and scored in LigandScout.

The hits from the added donor pharmacophore apoPharm-3A scored generally poor for pharmacophore match and binding affinity. Thus, we reduced the pharmacophore match threshold from 50 to 40, which means four out of five features were hit. This was done due to the aromatic interaction being hard to match, as even small rotations of the compounds aromatic moiety during docking negated identification. There were still only a few pharmacophore feature matches, with overall low binding affinity scores. Merely a handful passed both thresholds and the selection was not appealing, so no compound was manually selected from this step.

The results from docking the hits from apoPharm-3B, with added acceptor close to the aromatic interaction, were more promising. 376 hits were retained after calculation of binding affinity and pharmacophore match, with a pharmacophore match threshold of 40. From these hits, 28 were selected for MD simulation during visual inspection. MD simulations were run for 200 ns and the RMSF and RMSD values calculated over the returned trajectory. The median RMSF of the apoPharm-3B screening candidates (Figure 4.10) was 2.26 Å and thus similar to the leadPharm simulations with 2.20 Å (Figure 4.5), but the deviation was bigger. The median RMSD is only 1.59 Å and much lower

compared to the median of 2.14 Å for the leadPharm simulations. We relaxed the RMSF and RMSD threshold to 2.1 Å to include two hits slightly above 2 Å. As a result, ten structures stayed below the adapted thresholds. The set included a variety of core structures. We selected eight candidates after visual inspection based on uniqueness.
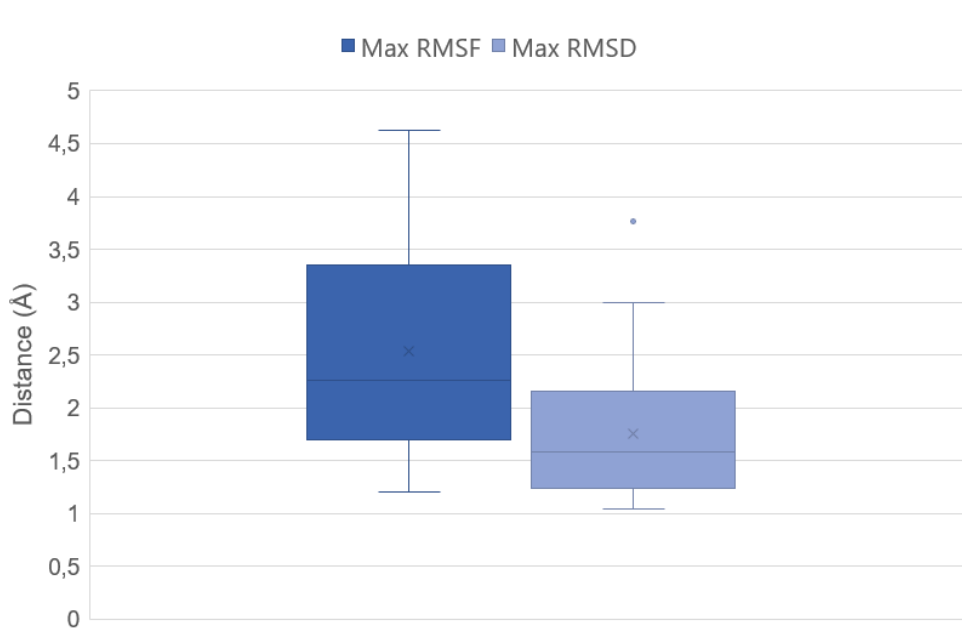

Figure 4.10: **Boxplot of the maximum RMSF and RMSD values for the 28 simulated structures from screening the apo derived pharmacophore apoPharm-3B.** The median for RMSF is 2.26 Å and slightly above the threshold of 2 Å, while the median RMSD is only 1.59 Å.

4.3 The Candidates

Our final selection of candidates included only such molecules that conform to all our criteria: Match all pharmacophore features — except of the aromatic interaction in the apo structure-based screening (apoPharm-3B) — after docking, have a decent binding affinity score, conform to our manual selection criteria and be stable in an MD simulation.

4.3.1 Ligand-Target Interaction-Based Candidates

We selected nine candidates from the lead-based screening (Figure 4.11) as described in Section 4.1. They had diverse cores and structures and covered a wide range of interaction patterns. Their names

and scores are summarized in Table 4.4. We present the selected candidates in more detail in the following section. SMILES are listed in Appendix C.1.
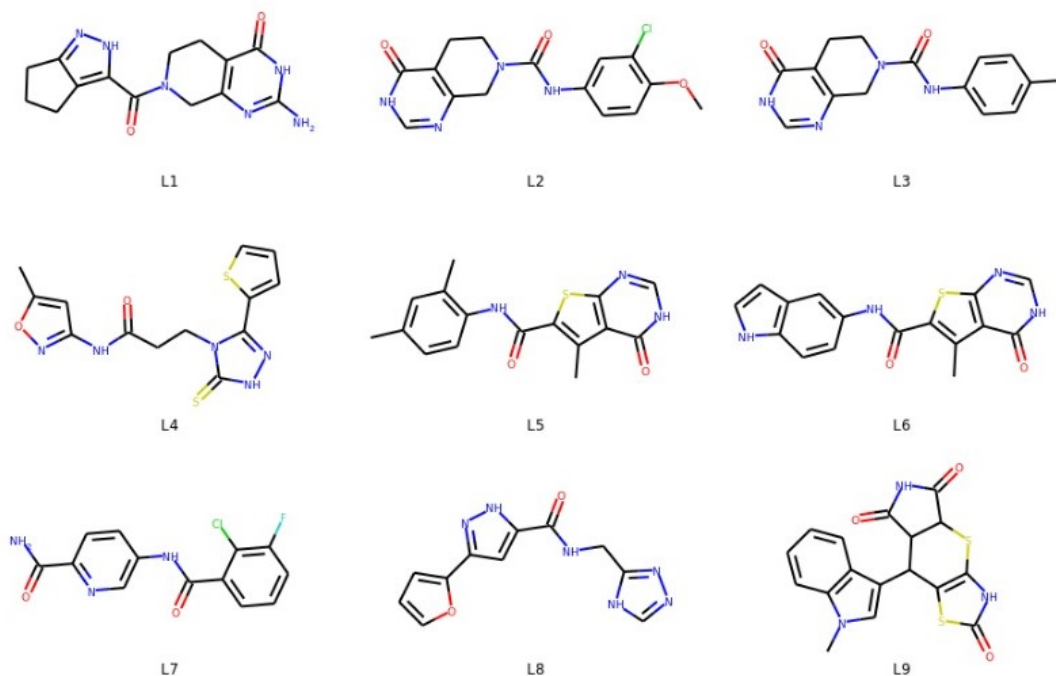

Figure 4.11: **Overview of the nine candidates L1 - L9 selected from the lead-based screening.** 2D images of candidate molecules were calculated and displayed in Jupyter Notebooks with python library RDKit version 2019.09.3 based on their SMILES representation.

ID	MolPort ID	Pharmacophore Match	Binding Affinity	RMSF (Å)	RMSD (Å)
L1	MolPort-020-232-854	56.25	-27.20	1.24	0.68
L2	MolPort-010-942-179	57.29	-31.65	1.69	1.91
L3	MolPort-044-520-775	56.75	-26.56	1.84	1.49
L4	MolPort-039-265-725	56.89	-24.15	1.63	1.50
L5	MolPort-000-675-772	56.73	-24.04	1.83	1.41
L6	MolPort-009-765-314	56.97	-23.69	1.59	1.74
L7	MolPort-027-923-237	56.46	-22.50	1.33	1.43
L8	MolPort-046-438-314	56.19	-21.89	1.57	1.46
L9	MolPort-000-437-543	55.98	-20.96	1.41	1.27

Table 4.4: **Scores of the nine selected candidates from the lead-based screening.** Given are the pharmacophore match score and the binding affinity score of the docking pose, minimized and calculated by LigandScout. Other measurements are the RMSF and RMSD from the MD simulations.

Candidate L1

The first candidate has the lowest RMSD at 0.68 Å of all candidates and also an extremely low RMSF of 1.24 Å based on the final MD simulation. In total, seven hydrogen bonds have been identified by LigandScout involving the backbone of GLU238, MET240 and THR242, as well as side chains of LYS187, GLU237 and SER264 (Figure 4.12). L1 has a similar aromatic double ring as the lead compound's quinazolinone with an additional amine, which acts as an hydrogen bond donor interacting with GLU238. It has the second highest binding affinity of the candidates (-27.20) as predicted by LigandScout, albeit lower than the lead (-34.80).
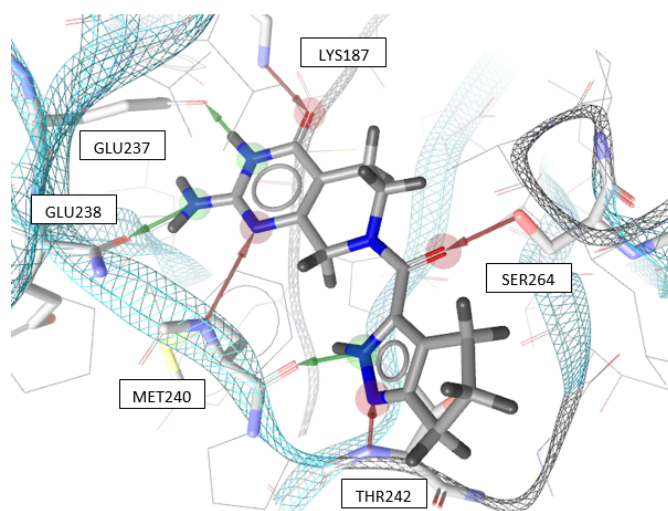

Figure 4.12: **Candidate L1**. Shown is the best OpenEye docking pose (MMFF94 energy minimized in LigandScout) of candidate L1 with identified interactions to the binding site of PPIP5K2 structure 4NZN without waters, magnesium ions and other co-crystallized ligands. Arrows depict hydrogen bonds. Green arrows mark ligand donors, red arrows acceptors. Binding site of PPIP5K2 structure [PDB id 4NZN] shown as "snake" and "lines" representation. Interaction partners are visualized as "sticks".

Candidates L2 and L3

The molecules L2 and L3 share a similar core and are thus discussed together (Figure 4.13). The core consists of a double ring, which forms the crucial hydrogen bonds with the backbone of MET240 and side chains of residues GLU237, LYS187 as well as SER264. The structures differ at the solvent exposed part. L2 has an hydrophobic chlorine interacting with PHE239 and a methoxy substituent (Figure 4.13a), whereas L3 has an hydrophobic methyl group additionally interacting with LEU267 (Figure 4.13b). Due to their similarities, their scores are also comparable. However, L2 has a better

binding affinity score than L3 (-31.65 to -26.56), and simultaneously a worse (higher) RMSD (1.91 Å to 1.49 Å).
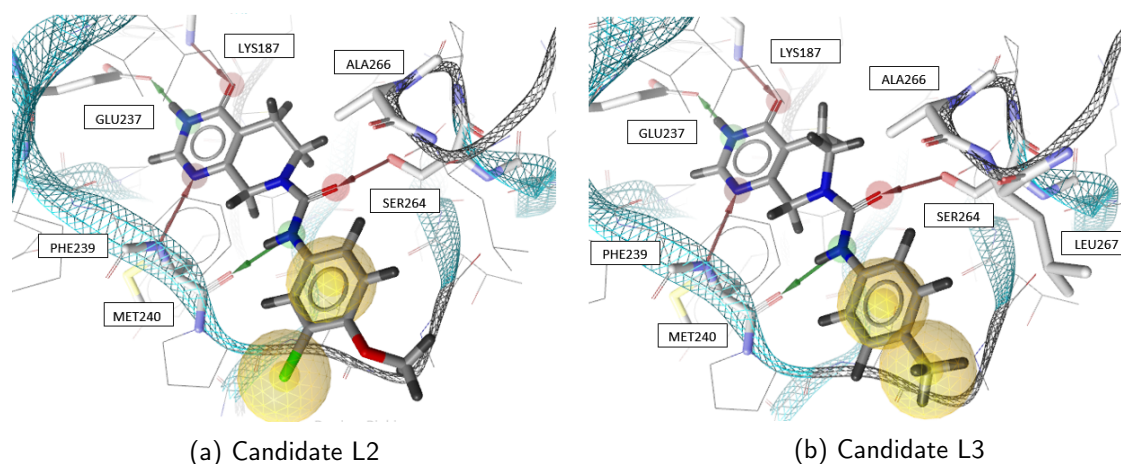

(a) Candidate L2

(b) Candidate L3

Figure 4.13: **Candidate L2 and L3**. Shown are the best OpenEye docking poses (MMFF94 energy minimized in LigandScout) of candidate L2 (a) and L3 (b) with identified interactions to the binding site of PPIP5K2 structure 4NZN without waters, magnesium ions and other co-crystallized ligands. Arrows depict hydrogen bonds. Green arrows mark ligand donors, red arrows acceptors. Yellow spheres indicate hydrophobic interactions. Binding site of PPIP5K2 structure [PDB id 4NZN] shown as "snake" and "lines" representation. Interaction partners are visualized as "sticks". For clarity, PHE239 remains drawn as lines.

Candidate L4

Candidate L4 does not contain a bi-cyclic core ring structure but a smaller 5-membered ring instead (Figure 4.14). It is elongated and L-shaped following the curve of the pocket and retains strong interactions with the backbone in form of hydrogen bonds with LYS187, GLU327, MET240 and THR242. An aromatic thiazole ring provides depth inside the pocket forming hydrophobic interactions with LEU211, VAL196 and ALA266. Contrary to the lead with its thiazole moiety, a isoxazole ring is positioned towards the solvent interacting with ALA266, PHE239 and LEU267.

Candidates L5 and L6

L5 and L6 again share a double ring core structure containing a sulfur atom. They form hydrogen bonds inside the pocket with LYS187, GLU237, MET240 and SER264. Both molecules have an ethyl group interacting with ALA266, ILE198 and LEU311 (Figure 4.15).
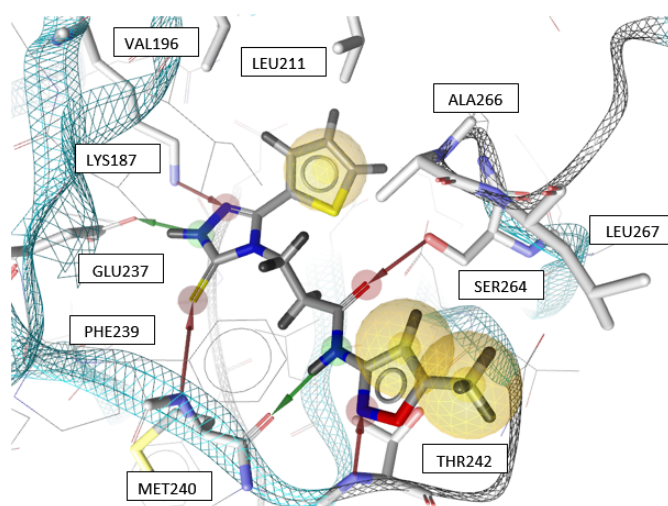
Figure 4.14: **Candidate L4.** Shown is the best OpenEye docking pose (MMFF94 energy minimized in LigandScout) of candidate L4 with identified interactions to the binding site of PPIP5K2 structure 4NZN without waters, magnesium ions and other co-crystallized ligands. Arrows depict hydrogen bonds. Green arrows mark ligand donors, red arrows acceptors. Yellow spheres indicate hydrophobic interactions. Binding site of PPIP5K2 structure [PDB id 4NZN] shown as "snake" and "lines" representation. Interaction partners are visualized as "sticks".

Similarly to L2 and L3, they have differing groups interacting with the solvent exposed part of the pocket. L5 has an extremely bulky hydrophobic ring with two methyl groups (Figure 4.15a); the other candidate L6 instead has an indoline with slight hydrophobic properties (Figure 4.15b).

Both hydrophobic substituents interact with THR242, PHE239, LEU267 and ALA266.

Candidate L7

This candidate is unique as it contains a carboxylic amide group which form hydrogen bonds with LYS187 and GLU237 inside the pocket (Figure 4.16). The connected pyridine ring bonds with MET240 and forms hydrophobic interaction with VAL185, THR242, ILE198, LEU311, PHE239 and ALA266. At the solvent exposed part there is a hydrophobic ring decorated with a chlorine and a fluorine interacting again with PHE239 and ILE198 as well as LEU267. Although less bulky than some of the other candidates, its binding affinity is on the lower end (-22.50 Å).

(a) Candidate L5

(b) Candidate L6
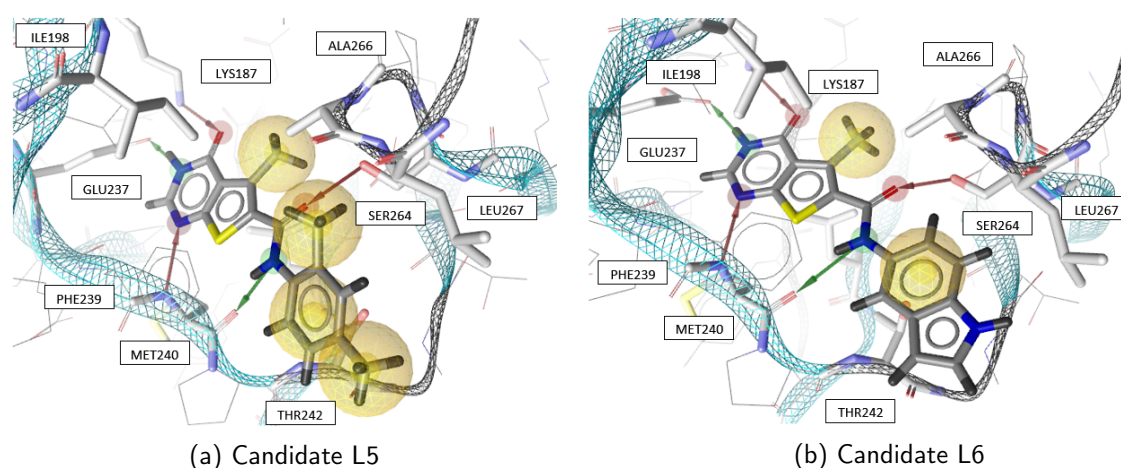
Figure 4.15: **Candidates L5 and L6**.Shown are the best OpenEye docking poses (MMFF94 energy minimized in LigandScout) of candidate L5 (a) and L6 (b) with identified interactions to the binding site of PPIP5K2 structure 4NZN without waters, magnesium ions and other co-crystallized ligands. Arrows depict hydrogen bonds. Green arrows mark ligand donors, red arrows acceptors. Yellow spheres indicate hydrophobic interactions. Binding site of PPIP5K2 structure [PDB id 4NZN] shown as "snake" and "lines" representation. Interaction partners are visualized as "sticks". For clarity, PHE239 remains drawn as lines. LEU311 is behind the molecule.
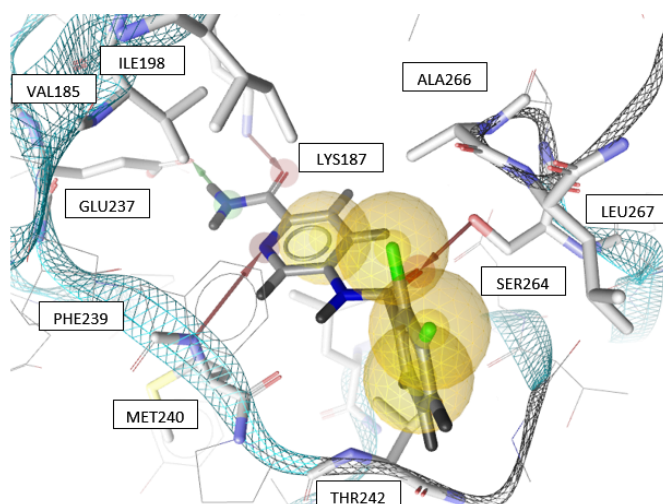

Figure 4.16: **Candidate L7.** Shown is the best OpenEye docking pose (MMFF94 energy minimized in LigandScout) of candidate L7 with identified interactions to the binding site of PPIP5K2 structure 4NZN without waters, magnesium ions and other co-crystallized ligands. Arrows depict hydrogen bonds. Green arrows mark ligand donors, red arrows acceptors. Yellow spheres indicate hydrophobic interactions. Binding site of PPIP5K2 structure with PDB id 4NZN shown as "snake" and "lines" representation. Interaction partners are visualized as "sticks". For clarity, PHE239 remains drawn as lines.

Candidate L8

Candidate L8 is curved and spans from a location deep inside the pocket all the way to the solvent exposed part (Figure 4.17). Due to the curve, it does not point into the key interaction area as prominently as other candidates, yet interactions with LYS187, GLU237, MET240 and SER264 are met. An additional bond forms with THR242. A tetrahydrofuran inside the pocket is able to form hydrophobic interactions with VAL196.
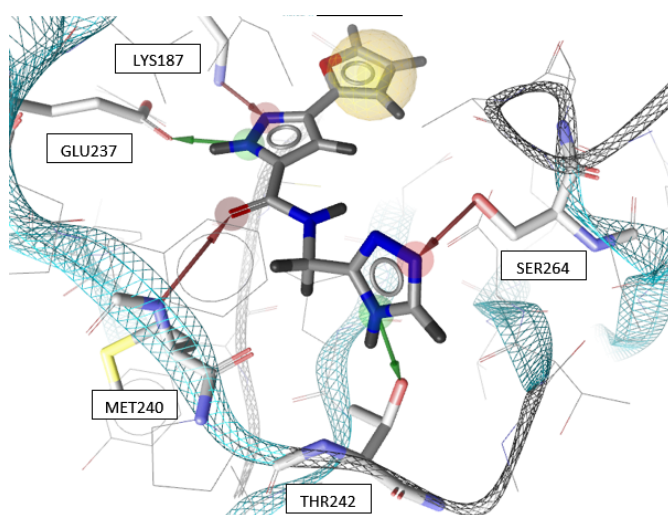

Figure 4.17: **Candidate L8**. Shown is the best OpenEye docking pose (MMFF94 energy minimized in LigandScout) of candidate L8 with identified interactions to the binding site of PPIP5K2 structure 4NZN without waters, magnesium ions and other co-crystallized ligands. Arrows depict hydrogen bonds. Green arrows mark ligand donors, red arrows acceptors. Yellow spheres indicate hydrophobic interactions. Binding site of PPIP5K2 structure [PDB id 4NZN] is shown as "snake" and "lines" representation. Interaction partners are visualized as "sticks". For clarity, PHE239 remains drawn as lines.

Candidate L9

Our last candidate L9 is quite unique with a combination of five rings, a double ring and three connected rings (Figure 4.18). A succinimide is located in the core area undergoing hydrogen bonds with LYS187, GLU237 and MET240. Additionally, the molecule forms one hydrogen bond with SER261 and another with MET240. Hydrophobic contacts inside the pocket occur with VAL196 and LEU211. Again there are favorable backbone interactions and a deeper reach into the pocket simultaneously. The compact shape keeps the interactions focused in the core region. With a value of -20.96 it has the worst binding affinity score of the candidates despite its size. At the same time

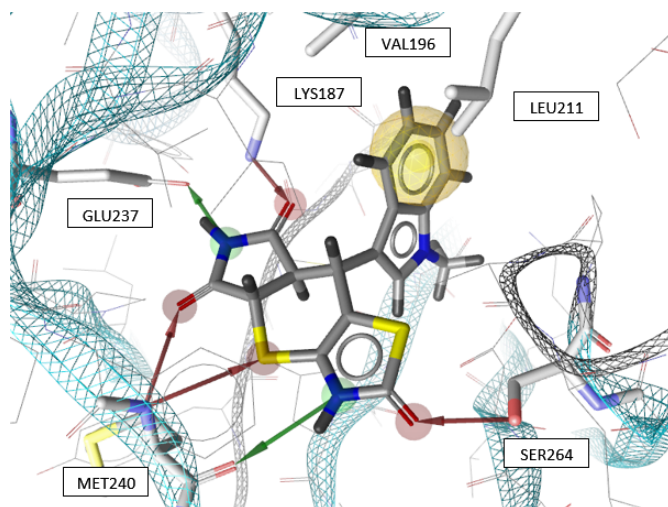the RMSD with 1.27 Å and the RMSF with 1.41 Å perform well.

Figure 4.18: **Candidate L9**. Shown is the best OpenEye docking pose (MMFF94 energy minimized in LigandScout) of candidate L9 with identified interactions to the binding site of PPIP5K2 structure 4NZN without waters, magnesium ions and other co-crystallized ligands. Arrows depict hydrogen bonds. Green arrows mark ligand donors, red arrows acceptors. Yellow spheres indicate hydrophobic interactions. Binding site of PPIP5K2 structure WITH PDB id 4NZN shown as "snake" and "lines" representation. Interaction partners are visualized as "sticks". For clarity, PHE239 remains drawn as lines.

4.3.2 Apo Protein Structure-Based Candidates

Candidates derived from the apo structure-based screening, apoPharm-3B, stem from the ATP-inspired four-feature pharmacophore including an additional acceptor (Figure 4.9b). We identified eight candidates A1-A8 that fulfilled our requirements (Figure 4.19), which represent a diverse set of molecules and clearly differ from the lead derived candidate set presented in the previous Section 4.3.1). On first glance, we noticed that these candidates favour nitrile groups, which we did not see in the other candidate set.

The scores of this set are in Table 4.5. SMILES are listed in Appendix C.2.
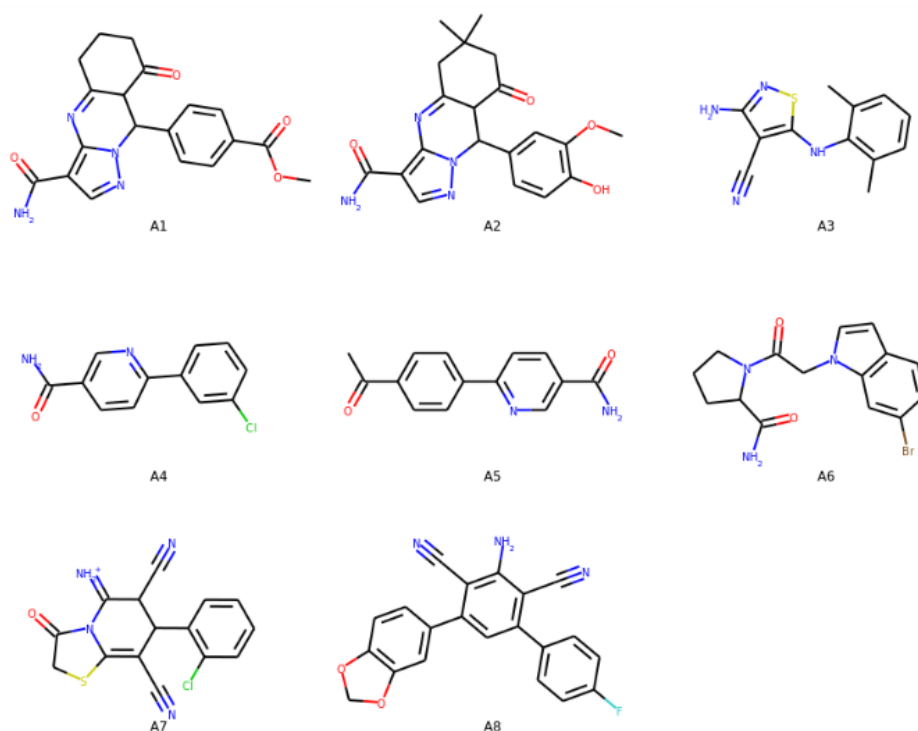
Figure 4.19: **Overview of the eight candidates A1 - A8 selected from the apo structure-based screening.** 2D images of candidate molecules were calculated and displayed in Jupyter Notebooks with the python library RDKit version 2019.09.3 based on their SMILES representation.

ID	MolPort ID	Pharmacophore Match	Binding Affinity	RMSF (Å)	RMSD (Å)
A1	MolPort-008-334-582	56.41	-20.69	1.90	1.12
A2	MolPort-008-344-232	56.45	-21.91	1.71	2.04
A3	MolPort-019-913-286	56.44	-23.29	1.59	1.17
A4	MolPort-005-066-645	47.59	-28.00	1.70	1.13
A5	MolPort-005-111-359	47.00	-26.53	2.07	1.41
A6	MolPort-020-182-708	56.15	-26.88	1.20	1.23
A7	MolPort-002-605-763	45.28	-27.29	1.57	1.43
A8	MolPort-009-753-846	54.84	-25.65	1.96	1.63

Table 4.5: **Scores of the eight selected candidates from the apo-based screening.** Given are the pharmacophore match score and the binding affinity score of the docking pose, minimized and calculated by LigandScout. Other measurements are the RMSF and RMSD from the MD simulations.

Candidates A1 and A2

Candidates A1 and A2 share a similar, bulky core structure constructed out of several rings (Figure 4.20). GLU237, GLU238, LYS187 and MET240 interact with the ring system composed of three joined rings with a carboxylic amide pointing towards the core region.

An aromatic ring above with a carboalcoxy (A1, Figure 4.20a) or an aldehyd and carbonyl group (A2, Figure 4.20b) provides hydrophobic interactions with just VAL196 (A1) or with VAl196, ILE198 and ALA266 (A2) in the more spacious part of the pocket. Candidate A2 additionally forms hydrophobic interactions with LEU227, LEU311 and THR242 at the solvent exposed part. Surprisingly and albeit their similarities, A1 has the best maximum RMSD (1.12 Å) while A2 has the worst value of this candidates (2.04 Å).

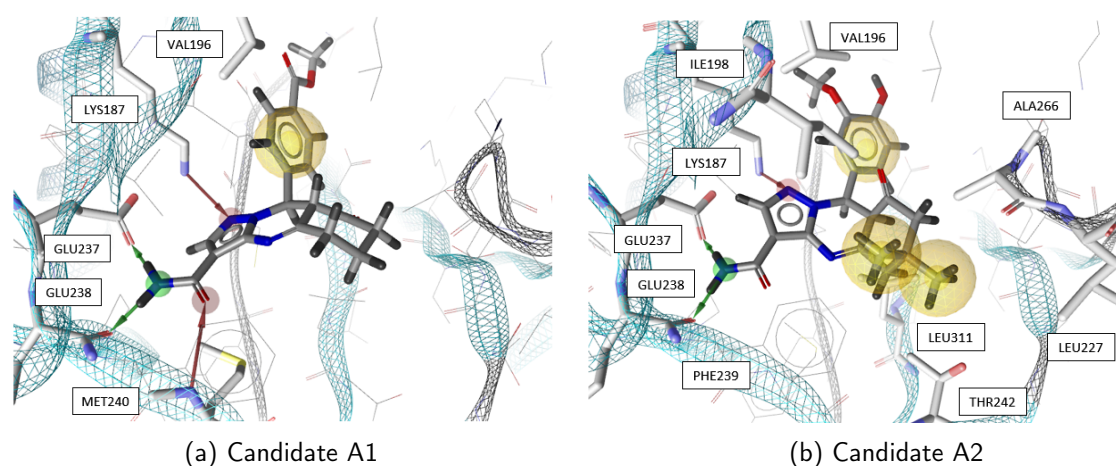(a) Candidate A1   (b) Candidate A2

Figure 4.20: **Candidates A1 and A2**. Shown are the best OpenEye docking poses (MMFF94 energy minimized in LigandScout) of candidates A1 (a) and A2 (b) with identified interactions to the binding site of the centroid structure, taken from a MD simulation of 4NZN. Arrows depict hydrogen bonds. Green arrows mark ligand donors, red arrows acceptors. Yellow spheres indicate hydrophobic interactions. The binding site of the centroid is shown as "snake" and "lines" representation. Interaction partners are visualized as "sticks".

Candidate A3

Candidate A3 consists of two rings connected by a secondary amine (Figure 4.21). The isothiazole provides the hydrogen bond with MET240, the connected amine is a double donor and binds GLU237 and GLU238 while a nitrite forms a hydrogen bond with LYS187. The six-membered ring decorated with two ethyl groups is a bulky hydrophobic moiety filling the pocket interacting with many residues, namely VAL185, ILE198, VAL196, LEU211, LEU311 and THR242.
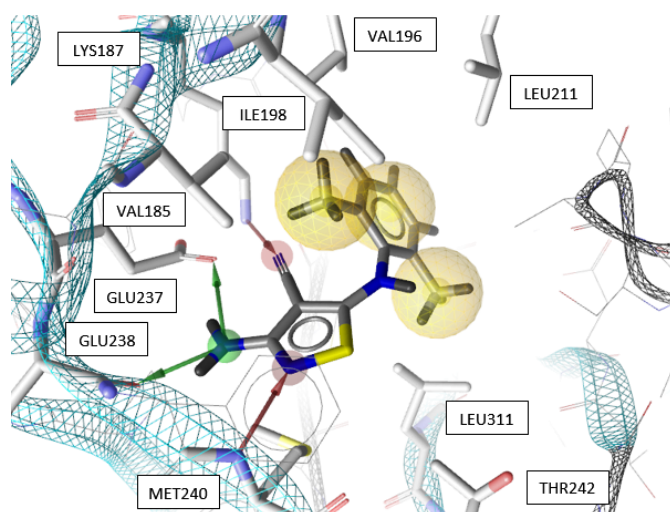
Figure 4.21: **Candidate A3**. Shown is the best OpenEye docking pose (MMFF94 energy minimized in LigandScout) of candidate A3 with identified interactions to the binding site of the centroid structure, taken from a MD simulation of 4NZN. Arrows depict hydrogen bonds. Green arrows mark ligand donors, red arrows acceptors. Yellow spheres indicate hydrophobic interactions. The binding site of the centroid is shown as "snake" and "lines" representation. Interaction partners are visualized as "sticks".

Candidates A4 and A5

Candidates A4 and A5 are two similar, simple and small structures (Figure 4.22). They both have a benzene ring and a pyridil ring decorated with a carboxamide crucial for forming the hydrogen bonds with LYS187, GLU237, GLU238 and MET240. The hydrophobic center is in contact with ILE198, VAL185, LEU311 and MET240.

Candidate A4 has a single chlorine atom at the benzene ring close to VAL196 and LEU211 (Figure 4.22a), while A5 features an aldehyde contacting LEU311 (Figure 4.22b).

They both have the lowest pharmacophore match scores (47.59 and 47.00), since they do not match the aromatic feature. Both compensate the deficit with high docking binding affinity scores (-28.00 and -26.53) and low RMSD scores (1.41 Å and 1.2 Å) from the MD simulation.

Candidate A6

With a hydrophobic double ring interacting with LEU311 and VAL196 not in the key interaction area but deeper in the pocket, this candidate A6 is different from the other candidates (Figure 4.23). Key hydrogen bond interactions with GLU237, GLU238 and MET240 are formed by a carboxylic amide
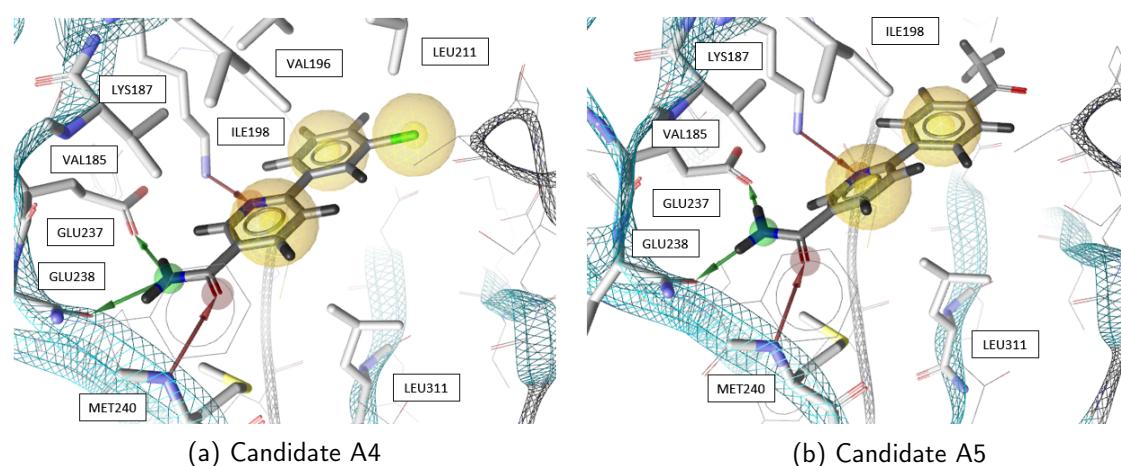
(a) Candidate A4   (b) Candidate A5

Figure 4.22: **Candidates A4 and A5**. Shown are the best OpenEye docking poses (MMFF94 energy minimized in LigandScout) of candidates A4 (a) and A5 (b) with identified interactions to the binding site of the centroid structure, taken from a MD simulation of 4NZN. Arrows depict hydrogen bonds. Green arrows mark ligand donors, red arrows acceptors. Yellow spheres indicate hydrophobic interactions. The binding site of the centroid is shown as "snake" and "lines" representation. Interaction partners are visualized as "sticks".

similar to the candidates A1 and A2. Another hydrogen bond is formed with LYS198. RMSF (1.20 Å) and RMSD (1.23 Å) values are both among the lowest of all eight candidates.

Candidate A7

Among all candidates presented in this thesis, A7 is our only candidate with an ionic interaction that acts as donor for the hydrogen bond with GLU237 (Figure 4.24). More hydrogen bonds are formed with MET240 and two nitrites function as hydrogen bond acceptors for bonds with LYS187 and SER264. The latter was a unique bond across the pocket and was not observed in other candidates. Hydrophobic interactions inside the pocket appear with VAL185, ILE198, VAL196, LEU211 and ALA266.
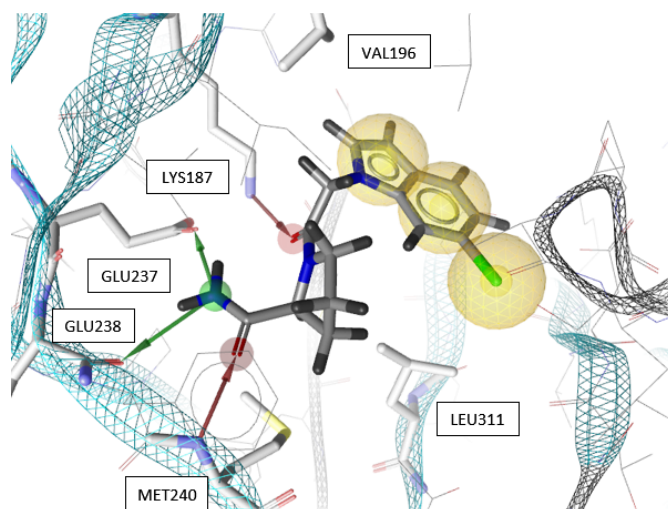
Figure 4.23: **Candidate A6**. Shown is the best OpenEye docking pose (MMFF94 energy minimized in LigandScout) of candidate A6 with identified interactions to the binding site of the centroid structure, taken from a MD simulation of 4NZN. Arrows depict hydrogen bonds. Green arrows mark ligand donors, red arrows acceptors. Yellow spheres indicate hydrophobic interactions. The binding site of the centroid is shown as "snake" and "lines" representation. Interaction partners are visualized as "sticks".
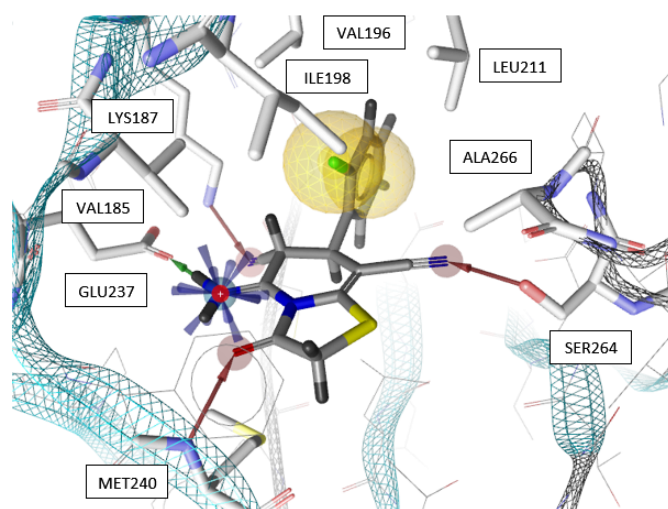

Figure 4.24: **Candidate A7**.Shown is the best OpenEye docking pose (MMFF94 energy minimized in LigandScout) of candidate A7 with identified interactions to the binding site of the centroid structure, taken from a MD simulation of 4NZN. Arrows depict hydrogen bonds. Green arrows mark ligand donors, red arrows acceptors. Yellow spheres indicate hydrophobic interactions. The binding site of the centroid is shown as "snake" and "lines" representation. Interaction partners are visualized as "sticks".

Candidate A8

Candidate A8 has three aromatic rings; the ring in the center is decorated with two nitrites and one amine (Figure 4.25). These three nitrogen atoms enable all hydrogen bonds with the nitrites acting as acceptors for LYS187 and MET240 and an amine as donor for GLU237. The other rings provide hydrophobic interactions towards the solvent and inside the pocket with ILE198, LEU211, THR242, ALA266, LEU311 and PHE239.
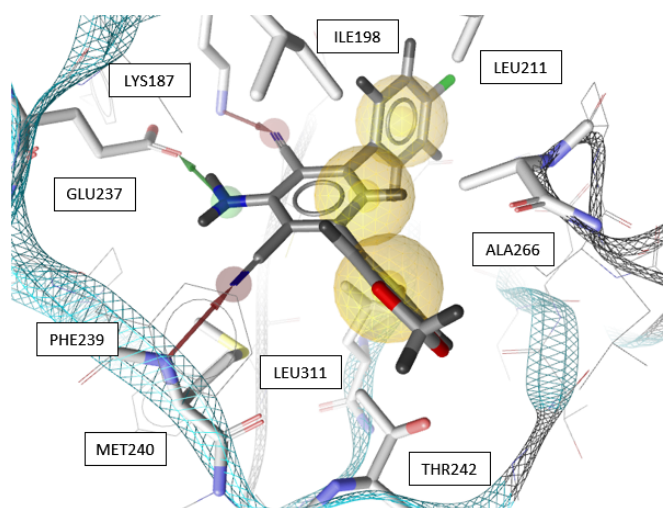

Figure 4.25: **Candidate A8**. Shown is the best OpenEye docking pose (MMFF94 energy minimized in LigandScout) of candidate A8 with identified interactions to the binding site of the centroid structure, taken from a MD simulation of 4NZN. Arrows depict hydrogen bonds. Green arrows mark ligand donors, red arrows acceptors. Yellow spheres indicate hydrophobic interactions. The binding site of the centroid is shown as "snake" and "lines" representation. Interaction partners are visualized as "sticks". For clarity, PHE239 remains drawn as lines.

5 Discussion

The kinase PPIP5K2 synthesizes pyrophosphates, a known regulatory component of phosphate home-ostasis. [4, 5] While pathological symptoms of knockouts are known, the mechanisms behind them are unclear. [11–13] Potent and selective inhibitors for PPIP5K2 could help unravel the enigma surrounding the regulation of phosphate in the cell. The aim of this thesis was to create a pipeline centered around a pharmacophore-based virtual screening to identify novel, potentially active inhibitors of the human kinase PPIP5K2.

5.1 The Pipeline

The pipeline setup allowed us to screen and filter pharmacophores efficiently, combining computational scoring methods with visual inspections. We ran six pharmacophore screenings of the prepared MolPort library containing over 16 Million compounds. From each screen, the returned hits were pre-filtered using the implemented script removing solvent exposed molecules. Next, the between 500 and 10000 selected hits per screen were re-docked, using the implemented OpenEye docking script, and further processed with LigandScout for minimization and subsequent calculation of pharmacophore fit score and binding affinity score. Around 99% of hits were filtered here. Finally, a total of 81 MD simulations of length 200 ns each were performed using the implemented scripts running OpenMM.

There is room for improvement in the automation of individual steps.

Since LigandScout offers binding affinity calculation and pharmacophore alignment as command line tools (*ialign* and *iaffinity*), this scoring could potentially be more automated. However, we did not simulate all molecules that passed the scoring but selected a subset manually during visual inspection and thus used the graphical user interface of LigandScout.

5.1.1 Computing Power and Times

Steps performed in the graphical user interface of LigandScout and the filtering by pocket fit required little computational power. Filtering of hits by pocket fit, which remove molecules exposed to the solvent to an extent assumed unfavourable, may be the most unique part of the pipeline and it is highly efficient. We calculated pairwise distances between the pocket descriptor points and the ligand atoms. Since the number of points of pocket descriptors and molecule atoms were sufficiently small, the calculations were fast. The implemented script processed several tens of thousands of molecules in seconds on a mediocre laptop (4 cores, 8GB RAM). It reduced the number of hits early and saved time for more computationally expensive filtering steps.

The computationally expensive operations, performed on the HPC, were library generation, virtual screening, molecular docking and MD simulations. These methods either require handling extensive amounts of data or they include running a simulation. The most resource and time intensive method per compound is MD simulation and it was thus performed as a last step on a handful of promising candidates.

Generating the library took the longest and it is therefore recommended to do this only once. The rest of the pipeline is meant to be iterated to improve results as desired.

5.1.2 Adjusting the Output

Adjustments and iterations were mainly carried out if the number of hits after any filtering step did not match the expectations. This was either the case if there were too many hits to handle or too few promising candidates.

Number and Type of Pharmacophore Features

Changing the pharmacophore had the biggest effect on hit numbers. A rule of thumb is, that the more features a pharmacophore contains, the more restrictive it is. Less hits will be returned from the database, but as a trade-off the screening will run faster. As described above, pharmacophore models were iteratively adapted throughout this thesis to find a well defined pharmacophore, with sufficient hit numbers.

The first apo derived pharmacophores were not restrictive enough (Figure 4.7) . They only consisted

of three features and returned an overabundance of hits (Table 4.2). Upon adding a single acceptor across the pocket, hit rates dropped below a feasible amount. This demonstrated, that hit rates also depend on the specific combination pharmacophores features, not just the number. Other fourth features returned a decent amount of hits (Figure 4.9).

Compared to this, the five feature pharmacophore derived from the lead returned pleasantly many hits, albeit having one feature more.

Scoring

Smaller adjustments can be made to the filtering thresholds to either include or exclude more hits. Scoring algorithms are handy scales; they give us a rough idea if a compound or docking pose can be viable, but they should not be the only mean of measure as scoring functions are known to not necessarily correlate with experimental binding affinities. [71, 72] The binding affinity score depends largely on the docking pose and the conformation of the protein, which is optimized for the co-crystallized ligand. We did minimize the MMFF94 energy of docking poses before pharmacophore scoring to optimize the docking pose further, however the side chains of the protein were not affected. In reality, the protein is flexible so that ligand and binding site accommodate each other. [39] This means the score can only ever be an approximation.

We introduced the MD simulations into our pipeline to counteract docking inaccuracies. The RMSD was used to determine the total ligand movement compared to the simulations first frame after equilibration. If the RMSD was low, we assumed the ligand remains in the binding pocket. The RMSF on the other side calculates the atom fluctuation per ligand and discloses if certain parts of the ligand moved a lot, even if the overall pose is stable. As described in literature, movements of $1-3$ Å are expected. [73] We chose a relatively strict cutoff at 2Å for the lead-based methods and 2.1 Å for the apo structure-based approach to reduce the candidate sets to our preferred size.

5.2 Screening Results

We ran Steps 3 - 10 of the pipeline several times without much alteration and produced reliable results.

The key hydrogen bond interactions involved GLU237, MET240 and LYS187. Those three features were present in both pharmacophores leadPharm and apoPharm-3B and the bonding partners were occupied by the co-crystallized AMP-PNP as well. Hence, we referred to the interaction area as core area.

Lead-Based Pharmacophore Screening

Originating from the core area, the lead derived pharmacophore added interactions along the backbone towards the solvent orthogonal to the binding mode of ATP. This forced the hits to follow the binding mode of the lead. Most hits have a similar binding mode, although cores are diverse. We kept two pairs with similar cores (L2 and L3, L5 an L6) and differing terminal groups to determine the impact of the solvent exposed groups. Candidates L4 and L8 differ the most from the lead binding mode. They are curved and have an additional hydrophobic contact deeper in the pocket. The most unique structure is candidate L9. It is a compact molecule that hits the interactions expressed by the lead and has a big, hydrophobic structure inside the pocket. However, the better scoring candidates are similar to the lead and rather dissimilar to ATP. The lead-like compounds target an area, which is less buried in the binding pocket, smaller and with fewer charges.

The feature combination of the lead derived pharmacophore seemed rather favourable. It had five features but returned more hits than the PyRod pharmacophores. This might be because it was organically retrieved from an existing compound and the distances and angles between features are, without doubt, structurally sensible.

Apo Structure Pharmacophore Screening

With the apo protein approach, we broadened the chemical space and searched for novel possibilities to target the more open space of the binding pocket. Extending the binding position towards that space turned out to be challenging.

PyRod returned an abundance of features and selecting a sensible set, which differs from the lead and generates a reasonable amount of hits, was the main concern.

Pharmacophores apoPharm-1A and 1B with three features each returned to many hits. From the two, the ATP inspired apoPharm-2B was closer to a reasonable hit number and we modified it. When we added a feature across the pocket for apoPharm-2, the numbers dropped below a feasible amount. So instead we added a donor next to the acceptor at MET240 (apoPharm-3A) or an acceptor between double donor and aromatic interaction (apoPharm-3B). Both returned promising hit numbers from the screening, but apoPharm-3A hits failed to meet the thresholds during scoring. Thus we chose candidates only from the apoPharm-3B screening.

In apoPharm-3B the the acceptor close-by seemed to harmonize better with the aromatic interaction. We did not probe the backbone further away from the core than the first hydrogen bond with MET240. Subsequently, most hits follow a binding mode orthogonal to the lead position. The aromatic feature is barely recognized as such in the features, because the rings are often rotated after minimization. An hydrophobic feature takes its place interacting most commonly with VAL196 and LEU211. Most candidates do not extend much beyond it with the largest structures being A1 and A2. The screening did return some larger structures, but they were filtered in the subsequent steps, because they were prone to get unstable and express high RMSF values. This tendency can already be seen in A1, which has the best RMSD of the candidates but almost exceeds the threshold with its RMSF.

We retrieved sensible and promising sets from both approaches. In spite of all selected compounds having a worse docking score than our lead, they might still be active *in-vitro* and build towards a library of active compounds for PPIP5K2.

6 Conclusion

With this work, we aim to support the discovery of inhibitors for PPIP5K2. We created a pipeline centered around a virtual screening to present an enriched library of estimated active compounds for PPIP5K2.

The pipeline combines CADD methods and visual inspections and was used in this thesis to retrieve a set of 17 candidate compounds from a database containing 16 Million compounds.

Virtual screening campaigns often have access to libraries of known active compounds and are able to base their pharmacophores on multiple ligands. We only had one lead compound, which yet proved to be a good premise. Nevertheless we introduced a second approach based on the apo structure.

For the lead-based approach we started with extrapolating the SAR of the lead to the binding pocket. Already during docking we detached from the orientation of ATP, which facilitated exploration of the new binding mode suggested by the lead compound. We were able to discover potential inhibitors straight from the first lead pharmacophore leadPharm, which was quite restrictive, but generated the desired number of hits from the screening. As expected, the candidates follow the same binding mode as the lead thus interacting with the backbone without utilizing the spacious and charged part of the pocket targeted by the ATP phosphate groups. The compounds still exhibit a variety of different core structures. We expect some *in-vitro* active compounds among the lead-derived candidates, as we are confident in our binding hypothesis and the resulting pharmacophore.

Nonetheless, the ATP-grasp fold in its variance and size is fascinating and we wished to delve deeper. We included the apo protein structure-based screening to explore a different part of the pocket and other binding modes. The unbiased nature of this approach offers unused potential, albeit requiring more rounds of trial and error. If no lead compound is available, this approach is viable for any screening campaign. We mainly performed it to investigate the binding pocket of the kinase PPIP5K2 further. Interestingly, some of the apo derived structures share a similar binding mode as the lead, but we were also able to generate variations. A success or failure to identify active compounds with

the hereby created pharmacophores will reveal a lot more information about the ATP-grasp fold and might further specify opportunities of inhibition thereof.

Candidates were handed over to our collaboration partner at FMP for experimental testing and evaluation of activity *in-vitro*. If sufficient affinities are reached, the respective compounds may already be used in research elevating the efforts to study PPIP5K2s role in phosphate homeostasis. In any case, any active compound we identify will improve future virtual screening campaigns, as a library of known actives can boost pharmacophore creation. It also enables the evaluation of potential library enrichment already during screening. We merely continued the just started search for PPIP5K2 inhibitors, which will likely be carried on in the future.

We hope that our pipeline as well as the proposed candidates will aid to explain the complicated but crucial processes of phosphate homeostasis in the future.

Bibliography

[1] J.-P. Bonjour, "Calcium and phosphate: A duet of ions playing for bone health," *Journal of the American College of Nutrition*, vol. 30, no. sup5, 438S–448S, 2011. DOI: 10.1080/07315724.2011.10719988.

[2] P. Sadava *et al.*, *Purves Biologie*, 9th ed. Spektrum Akademischer Verlag, 2011, 201f, ISBN: 978-3-8274-2650-5.

[3] M. G. Penido and U. S. Alon, "Phosphate homeostasis and its role in bone health," *Pediatric nephrology*, vol. 27, pp. 2039–2048, 11 2017. DOI: 10.1007/s00467-012-2175-z.

[4] N. W. Brown, A. M. Marmelstein, and D. Fiedler, "Chemical tools for interrogating inositol pyrophosphate structure and function," *Chem. Soc. Rev.*, vol. 45, pp. 6311–6326, 22 2016. DOI: 10.1039/C6CS00193A.

[5] S. B. Shears, "Intimate connections: Inositol pyrophosphates at the interface of metabolic regulation and cell signaling," *Journal of Cellular Physiology*, vol. 233, no. 3, pp. 1897–1912, 2018. DOI: 10.1002/jcp.26017.

[6] P. Draškovič *et al.*, "Inositol Hexakisphosphate Kinase Products Contain Diphosphate and Triphosphate Groups," *Chemistry Biology*, vol. 15, no. 3, pp. 274–286, 2008, ISSN: 1074-5521. DOI: 10.1016/j.chembiol.2008.01.011.

[7] C. Gu *et al.*, "The Significance of the Bifunctional Kinase/Phosphatase Activities of Diphosphoinositol Pentakisphosphate Kinases (PPIP5Ks) for Coupling Inositol Pyrophosphate Cell Signaling to Cellular Phosphate Homeostasis*," *Journal of Biological Chemistry*, vol. 292, no. 11, pp. 4544–4555, 2017, ISSN: 0021-9258. DOI: 10.1074/jbc.M116.765743.

[8] J. Zhu *et al.*, "Two bifunctional inositol pyrophosphate kinases/phosphatases control plant phosphate homeostasis," *eLife*, vol. 8, e43582, 2019. DOI: 10.7554/eLife.43582.

[9]   N. Gokhale, A. Zaremba, and S. Shears, "Receptor-dependent compartmentalization of PPIP5K1, a kinase with a cryptic polyphosphoinositide binding domain," *Biochemical Journal*, vol. 434, no. 3, pp. 415–426, Feb. 2011, ISSN: 0264-6021. DOI: 10.1042/BJ20101437.

[10] P. C. Fridy *et al.*, "Cloning and characterization of two human vip1-like inositol hexakisphosphate and diphosphoinositol pentakisphosphate kinases*," *Journal of Biological Chemistry*, vol. 282, no. 42, pp. 30754–30762, 2007, ISSN: 0021-9258. DOI: 10.1074/jbc.M704656200.

[11] R. Yousaf *et al.*, "Mutations in diphosphoinositol-pentakisphosphate kinase PPIP5K2 are associated with hearing loss in human and mouse," *PLOS Genetics*, vol. 14, e1007297, Mar. 2018. DOI: 10.1371/journal.pgen.1007297.

[12] Z. Chen *et al.*, "PPIP5K2 and PCSK1 are Candidate Genetic Contributors to Familial Keratoconus," *Scientific Reports*, vol. 9, no. 1, 2019. DOI: 10.1038/s41598-019-55866-5.

[13] T. Akoto *et al.*, "Identifying the target genes of PPIP5K2 in relation to keratoconus," *Investigative Ophthalmology & Visual Science*, vol. 62, no. 5, p. 781, 8 2021, ISSN: 1552-5783. DOI: 10.1016/j.chembiol.2014.03.009.

[14] C. Gu *et al.*, "The Significance of the Bifunctional Kinase/Phosphatase Activities of Diphosphoinositol Pentakisphosphate Kinases (PPIP5Ks) for Coupling Inositol Pyrophosphate Cell Signaling to Cellular Phosphate Homeostasis," *Journal of Biological Chemistry*, vol. 292, no. 11, 4544f, 2017. DOI: 10.1074/jbc.M116.765743.

[15] H. Wang *et al.*, "Structural basis for an inositol pyrophosphate kinase surmounting phosphate crowding," *Nature chemical biology*, vol. 8, pp. 689–699, 1 2011, ISSN: 1074-5521. DOI: 10.1038/nchembio.733.

[16] H. Wang *et al.*, "Synthetic Inositol Phosphate Analogs Reveal that PPIP5K2 Has a Surface-Mounted Substrate Capture Site that Is a Target for Drug Discovery," *Chemistry & Biology*, vol. 21, no. 5, pp. 689–699, 2014, ISSN: 1074-5521. DOI: 10.1016/j.chembiol.2014.03.009.

[17] M. S. Wilson, H. J. Jessen, and A. Saiardi, "The inositol hexakisphosphate kinases IP6K1 and -2 regulate human cellular phosphate homeostasis, including XPR1-mediated phosphate export," *Journal of Biological Chemistry*, vol. 294, no. 30, pp. 11597–11608, 2019, ISSN: 0021-9258. DOI: 10.1074/jbc.RA119.007848.

[18] F. D. Prieto-Martínez *et al.*, "Chapter 2 - computational drug design methods—current and future perspectives," in *In Silico Drug Design*, K. Roy, Ed., Academic Press, 2019, pp. 19–44, ISBN: 978-0-12-816125-8. DOI: 10.1016/B978-0-12-816125-8.00002-X.

[19] A. C. Anderson, "The Process of Structure-Based Drug Design," *Chemistry & Biology*, vol. 10, no. 9, pp. 787–797, 2003, ISSN: 1074-5521. DOI: 10.1016/j.chembiol.2003.09.002.

[20] D. Mendez *et al.*, "ChEMBL: towards direct deposition of bioassay data," *Nucleic Acids Research*, vol. 47, no. D1, pp. D930–D940, Nov. 2018, ISSN: 0305-1048. DOI: 10.1093/nar/gky1075.

[21] *MolPort*. [Online]. Available: https://www.molport.com/ (visited on 06/28/2021).

[22] D. Schaller *et al.*, "Next generation 3D pharmacophore modeling," *WIREs Computational Molecular Science*, vol. 10, no. 4, e1468, 2020. DOI: 10.1002/wcms.1468.

[23] J. L. Pederick *et al.*, "d-Alanine–d-alanine ligase as a model for the activation of ATP-grasp enzymes by monovalent cations," *Journal of Biological Chemistry*, vol. 295, no. 23, pp. 7894–7904, 2020, ISSN: 0021-9258. DOI: 10.1074/jbc.RA120.012936.

[24] C. Wu, D. Dunaway-Mariano, and P. S. Mariano, "Design, Synthesis, and Evaluation of Inhibitors of Pyruvate Phosphate Dikinase," *J. Org. Chem.*, vol. 78, 1910–1922, 2013. DOI: 10.1021/jo3018473.

[25] V. Škedelj *et al.*, "Discovery of the first inhibitors of bacterial enzyme d-aspartate ligase from Enterococcus faecium (Aslfm)," *European Journal of Medicinal Chemistry*, vol. 67, pp. 208–220, 2013, ISSN: 0223-5234. DOI: 10.1016/j.ejmech.2013.06.017.

[26] I. Mochalkin *et al.*, "Discovery of Antibacterial Biotin Carboxylase Inhibitors by Virtual Screening and Fragment-Based Approaches," *ACS Chem. Biol.*, vol. 4, no. 6, 473–483, 2009. DOI: 10.1021/cb9000102.

[27] H. Alonso, A. A. Bliznyuk, and J. E. Gready, "Combining docking and molecular dynamic simulations in drug design," *Medicinal Research Reviews*, vol. 26, no. 5, pp. 531–568, 2006. DOI: 10.1002/med.20067.

[28] D. Schaller, S. Pach, and G. Wolber, "PyRod: Tracing Water Molecules in Molecular Dynamics Simulations," *Journal of Chemical Information and Modeling*, vol. 59, 6 2019. DOI: 10.1021/acs.jcim.9b00281.

[29] A. Fischer *et al.*, "Decision Making in Structure-Based Drug Discovery: Visual Inspection of Docking Results," *Journal of Medicinal Chemistry*, vol. 64, no. 5, pp. 2489–2500, 2021, PMID: 33617246. DOI: 10.1021/acs.jmedchem.0c02227.

[30] G. M. Morris and M. Lim-Wilby, "Molecular Docking," in *Molecular Modeling of Proteins*, A. Kukol, Ed. Totowa, NJ: Humana Press, 2008, pp. 365–382, ISBN: 978-1-59745-177-2. DOI: 10.1007/978-1-59745-177-2_19.

[31] M. Xuan-Yu *et al.*, " Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery," *Current Computer-Aided Drug Design*, vol. 7, 2 2011. DOI: 10.2174/157340911795677602.

[32] Santa Fe, NM, *OEDOCKING 4.0.0.2: OpenEye Scientific Software*. [Online]. Available: http://www.eyesopen.com/oedocking.

[33] M. McGann, "Fred pose prediction and virtual screening accuracy," *Journal of Chemical Information and Modeling*, vol. 51, no. 3, pp. 578–596, 2011, PMID: 21323318. DOI: 10.1021/ci100436p.

[34] ——, "FRED and HYBRID docking performance on standardized datasets," *Journal of Computer-Aided Molecular Design*, vol. 26, pp. 897–906, 2012. DOI: 10.1007/s10822-012-9584-8.

[35] OpenEye Scientific Software. (2021). "HYBRID Theory," [Online]. Available: https://docs.eyesopen.com/applications/oedocking/theory/hybrid_theory.html (visited on 10/03/2021).

[36] S. Riniker, "Fixed-Charge Atomistic Force Fields for Molecular Dynamics Simulations in the Condensed Phase: An Overview," *Journal of Chemical Information and Modeling*, vol. 58, no. 3, pp. 565–578, 2018. DOI: 10.1021/acs.jcim.8b00042.

[37] A. D. Mackerell Jr., "Empirical force fields for biological macromolecules: Overview and issues," *Journal of Computational Chemistry*, vol. 25, no. 13, pp. 1584–1604, 2004. DOI: 10.1002/jcc.20082.

[38] *Openforcefield: An open and collaborative approach to better force fields*. [Online]. Available: https://openforcefield.org/.

[39] A. D. Vogt and E. Di Cera, "Conformational Selection or Induced Fit? A Critical Appraisal of the Kinetic Mechanism," *Biochemistry*, vol. 51, no. 30, pp. 5894–5902, 2012, PMID: 22775458. DOI: 10.1021/bi3006913.

[40] M. De Vivo *et al.*, "Role of molecular dynamics and related methods in drug discovery," *Journal of Medicinal Chemistry*, vol. 59, no. 9, pp. 4035–4061, 2016, PMID: 26807648. DOI: `10.1021/acs.jmedchem.5b01684`.

[41] G. Cruciani, "Molecular Interaction Fields: Applications in Drug Discovery and ADME Prediction," in, G. Cruciani, Ed. Wiley-VCH Verlag GmbH & Co. KGaA, 2005, ISBN: 978-3-52731-087-6. DOI: `10.1002/3527607676`.

[42] A. Artese *et al.*, "Molecular interaction fields in drug discovery: recent advances and future perspectives," *WIREs Computational Molecular Science*, vol. 3, no. 6, pp. 594–613, 2013. DOI: `10.1002/wcms.1150`.

[43] W. Walters, M. T. Stahl, and M. A. Murcko, "Virtual screening—an overview," *Drug Discovery Today*, vol. 3, no. 4, pp. 160–178, 1998, ISSN: 1359-6446. DOI: `10.1016/S1359-6446(97)01163-X`.

[44] R. Abagyan and M. Totrov, "High-throughput docking for lead generation," *Current Opinion in Chemical Biology*, vol. 5, no. 4, pp. 375–382, 2001, ISSN: 1367-5931. DOI: `10.1016/S1367-5931(00)00217-9`.

[45] B. K. Shoichet, "Virtual screening of chemical libraries," *Nature*, vol. 432, 7019 2004. DOI: `10.1038/nature03197`.

[46] E. H. B. Maia *et al.*, "Structure-Based Virtual Screening: From Classical to Artificial Intelligence," *Frontiers in Chemistry*, vol. 8, p. 343, 2020, ISSN: 2296-2646. DOI: `10.3389/fchem.2020.00343`.

[47] K.-H. Kim, N. D. Kim, and B.-L. Seong, "Pharmacophore-based virtual screening: a review of recent applications," *Expert Opinion on Drug Discovery*, vol. 5, no. 3, pp. 205–222, 2010. DOI: `10.1517/17460441003592072`.

[48] C. G. Wermuth *et al.*, "Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998)," *Pure and Applied Chemistry*, vol. 70, no. 5, pp. 1129–1143, 1998. DOI: `10.1351/pac199870051129`.

[49] J. H. Van Drie, "Generation of three-dimensional pharmacophore models," *WIREs Computational Molecular Science*, vol. 3, no. 5, pp. 449–464, 2013. DOI: `10.1002/wcms.1129`.

[50] A. Artese *et al.*, "Molecular interaction fields in drug discovery: recent advances and future perspectives," *WIREs Computational Molecular Science*, vol. 3, no. 6, pp. 594–613, 2013. DOI: 10.1002/wcms.1150.

[51] H. M. Berman *et al.*, "The Protein Data Bank," *Nucleic Acids Research*, vol. 28, pp. 235–242, 2000. DOI: 10.1093/nar/28.1.235. [Online]. Available: rcsb.org.

[52] H. Berman, K. Henrick, and H. Nakamura, "Announcing the worldwide Protein Data Bank," *Nature Structural & Molecular Biology*, vol. 10, p. 980, 2003. DOI: 10.1038/nsb1203-980. [Online]. Available: www.wwpdb.org.

[53] H Wang and S. B. Shears, *Crystal structure of the catalytic domain of PPIP5K2 in complex with AMPPNP and 2-O-BN-5-PA-INSP4*, 2014. DOI: 10.2210/pdb4NZN/pdb.

[54] OpenEye Scientific Software. (2021). "SPRUCE Introduction," [Online]. Available: https://docs.eyesopen.com/applications/spruce/theory/introduction.html#spruce-introduction (visited on 10/03/2021).

[55] J. Rodríguez-Guerra *et al.*, *KinoML*. [Online]. Available: https://volkamerlab.org/projects/kinoml/.

[56] T. A. Halgren, "Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94," *Journal of Computational Chemistry*, vol. 17, no. 5-6, pp. 490–519, 1996. DOI: 10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P.

[57] G. Wolber and T. Langer, "LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters," *Journal of Chemical Information and Modeling*, vol. 45, 1 2005. DOI: 10.1021/ci049885e.

[58] P. Eastman *et al.*, "OpenMM 7: Rapid development of high performance algorithms for molecular dynamics," *PLOS Computational Biology*, vol. 13, no. 7, pp. 1–17, Jul. 2017. DOI: 10.1371/journal.pcbi.1005659.

[59] D. Sydow *et al.*, "TeachOpenCADD: a teaching platform for computer-aided drug design using open source packages and data," *J. Cheminform.*, vol. 11, no. 1, p. 29, 2019. DOI: 10.1186/s13321-019-0351-x.

[60] J. A. Maier *et al.*, "ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB," *Journal of Chemical Theory and Computation*, vol. 11, no. 8, pp. 3696–3713, 2015, PMID: 26574453. DOI: 10.1021/acs.jctc.5b00255.

[61] W. L. Jorgensen *et al.*, "Comparison of simple potential functions for simulating liquid water," *Journal of Chemical Physics*, vol. 79, pp. 926–935, 1983.

[62] J. Wagner *et al.*, *openforcefield/openforcefields: Version 1.3.0 "Parsley" Update*, version 1.3.0, Oct. 2020. DOI: `10.5281/zenodo.4118484`. [Online]. Available: `10.5281/zenodo.4118484`.

[63] D. Schaller, S. Pach, and G. Wolber, *PyRod on GitHub*. [Online]. Available: `https://github.com/wolberlab/pyrod`.

[64] H. W. Horn *et al.*, "Development of an improved four-site water model for biomolecular simulations: Tip4p-ew," *The Journal of Chemical Physics*, vol. 120, no. 20, pp. 9665–9678, 2004. DOI: `10.1063/1.1683075`.

[65] OpenEye Scientific Software. (2021). "Quacpac Tautomers," [Online]. Available: `https://docs.eyesopen.com/toolkits/python/quacpactk/tautomerstheory.html`.

[66] V. L. Guilloux, P. Schmidtke, and P. Tuffery, "Fpocket: an open source platform for ligand pocket detection," *BMC Bioinformatics*, vol. 10, 2009. DOI: `10.1186/1471-2105-10-168`.

[67] Schrödinger, LLC, "The PyMOL Molecular Graphics System, Version 2.0," [Online]. Available: `https://pymol.org/`.

[68] N. Michaud-Agrawal *et al.*, "MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations," *J. Comput. Chem.*, vol. 32, 2011. DOI: `10.1002/jcc.21787`.

[69] Richard J. Gowers *et al.*, "MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations," in *Proceedings of the 15th Python in Science Conference*, Sebastian Benthall and Scott Rostrup, Eds., 2016, pp. 98 –105. DOI: `10.25080/Majora-629e541a-00e`.

[70] L. Bennett, B. Melchers, and B. Proppe, *Curta: A General-purpose High-Performance Computer at ZEDAT, Freie Universität Berlin*, 2020. DOI: `10.17169/refubium-26754`.

[71] Z. Wang *et al.*, "Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power and scoring power," *Physical Chemistry Chemical Physics*, vol. 18, pp. 12 964–12 975, 18 2016. DOI: `10.1039/C6CP01555G`.

[72] G. L. Warren *et al.*, "A Critical Assessment of Docking Programs and Scoring Functions," *Journal of Medicinal Chemistry*, vol. 49, no. 20, pp. 5912–5931, 2006, PMID: 17004707. DOI: `10.1021/jm050362n`.

[73] F. A. D. M. Opo *et al.*, "Structure based pharmacophore modeling, virtual screening, molecular docking and ADMET approaches for identification of natural anti-cancer agents targeting XIAP protein.," *Scientific Reports*, vol. 11, 1 2021. DOI: 10.1038/s41598-021-83626-x.

List of Figures

List of Tables

Abbreviations

ATP	Adenosine triphosphate
$PPIP5K$	Diphosphoinositol pentakisphosphate kinase
$CADD$	Computer-aided drug design
IP	Inositol phosphate
$PP-IP$	Inositol pyrophosphate
$IP6K$	Inositol hexakisphosphate kinase
FMP	Leibniz-Forschungsinstitut für Molekulare Pharmakologie
HTS	High-throughput screening
IC_{50}	Half maximal inhibitory concentration
MD	Molecular Dynamics
$PPDK$	Pyruvate phosphate dikinase
MIF	Molecular interaction field
VS	Virtual screening
PDB	Protein Data BanK
$AMP-PNP$	Adenylyl-imidodiphosphate
$MMFF94$	Merck molecular force field 94
$RMSF$	Root mean-square fluctuation
$RMSD$	Root-mean-square deviation
HPC	High-performance cluster
SAR	Structure activity relationship

Appendix

A Structures of Comerciallly Available Analogues for 105229
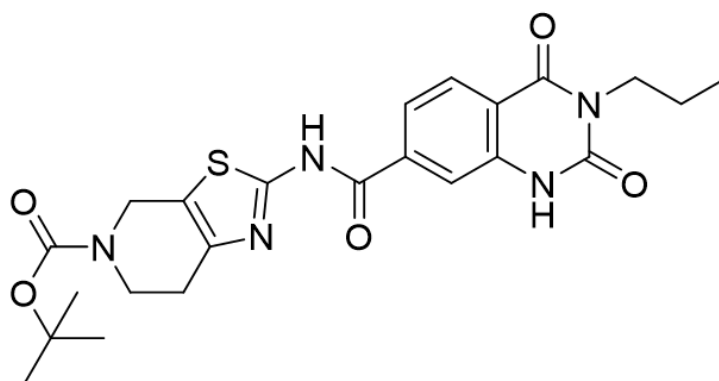
A.1 Autogrammin-1

Figure A.1: **Structure of Autogrammin-1.** Commercially available analogue to the lead with minimal inhibitory activity to PPIP5K2.
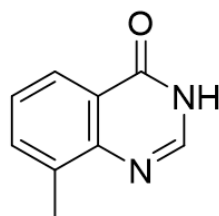
A.2 Methyl-Quinanzolinone

Figure A.2: **Structure of Methyl-Quinanzolinone.** Commercially available analogue to the lead with no inhibitory activity to PPIP5K2.

B Activities of Analogues of 105229

Compound	SMILES	IC_50
Cyclohexyl	O=C(NC=N1)C2=C1C=C(C(NC3CCCCC3)=O)C=C2	3197 µM
Phenyl	O=C(NC=N1)C2=C1C=C(C(NC3=CC=CC=C3)=O)C=C2	21669 µM
Benzothiazol	O=C(NC=N1)C2=C1C=C(C(NC3=NC(C=CC=C4)=C4S3)=O)C=C2	314.4 µM
4-Phenyl, 5-Methyl	O=C(NC=N1)C2=C1C=C(C(NC3=NC(C4=CC=CC=C4)=C(C)S3)=O)C=C2	249.8 µM
5-Methyl	O=C(NC=N1)C2=C1C=C(C(NC3=NC=C(C)S3)=O)C=C2	160.7 µM
4-Methyl	O=C(NC=N1)C2=C1C=C(C(NC3=NC(C)=CS3)=O)C=C2	8.2 µM
4-Phenyl	O=C(NC=N1)C2=C1C=C(C(NC3=NC(C4=CC=CC=C4)=CS3)=O)C=C2	255.1 µM
4-Tolyl	O=C(NC=N1)C2=C1C=C(C(NC3=NC(C4=CC=C(C)C=C4)=CS3)=O)C=C2	—
4-p-Chlorophenyl	O=C(NC=N1)C2=C1C=C(C(NC3=NC(C4=CC=C(Cl)C=C4)=CS3)=O)C=C2	141.7 µM
4-Difluorophenyl	O=C(NC=N1)C2=C1C=C(C(NC3=NC(C4=CC=C(F)C=C4F)=CS3)=O)C=C2	1204 µM
4-Carboxymethyl	O=C(NC=N1)C2=C1C=C(C(NC3=NC(C(OC)=O)=CS3)=O)C=C2	1.3e+6 µM
4-Carboxyethyl	O=C(NC=N1)C2=C1C=C(C(NC3=NC(C(OCC)=O)=CS3)=O)C=C2	2951 µM
4-Ethyl	O=C(NC=N1)C2=C1C=C(C(NC3=NC(CC)=CS3)=O)C=C2	8.1 µM
4-Cyclopropyl	O=C(NC=N1)C2=C1C=C(C(NC3=NC(C4CC4)=CS3)=O)C=C2	40.6 µM
N-Methylquinazolinone	O=C(N(C)C=N1)C2=C1C=C(C(NC3=NC(C)=CS3)=O)C=C2	259.7 µM
2-Methylquinazolinone	O=C(NC(C)=N1)C2=C1C=C(C(NC3=NC(C)=CS3)=O)C=C2	39671 µM
Quinoline	O=C(NC1=NC(C)=CS1)C2=CC3=C(C=CC=N3)C=C2	3512 µM
Amidin	O=C(NC1=NC(C)=CS1)C2=CC3=C(C(N)=NC=N3)C=C2	258 µM

Table B.1: SMILES and activity values for analogues synthesised and tested by Bartsch and Fiedler (unpublished)

C Candidate SMILES

C.1 Lead candidates

Candidate	SMILES
L1	Nc1nc2CN(CCc2c(=O)[nH]1)C(=O)c1n[nH]c2CCCc12
L2	COc1ccc(NC(=O)N2CCc3c(C2)[nH]cnc3=O)cc1Cl
L3	Cc1ccc(NC(=O)N2CCc3c(C2)nc[nH]c3=O)cc1
L4	Cc1cc(NC(=O)CCn2c(S)nnc2-c2cccs2)no1
L5	Cc1c(sc2nc[nH]c(=O)c12)C(=O)Nc1ccc(C)cc1C
L6	Cc1c(sc2nc[nH]c(=O)c12)C(=O)Nc1ccc2[nH]ccc2c1
L7	NC(=O)c1ccc(NC(=O)c2cccc(F)c2Cl)cn1
L8	O=C(NCc1nnc[nH]1)c1cc([nH]n1)-c1ccco1
L9	Cn1cc(C2C3C(Sc4[nH]c(=O)sc24)C(=O)NC3=O)c2ccccc12

Table C.1: SMILES as given by MolPort from the candidates derived from the pipeline based on lead-target activities.

C.2 Apo candidates

Candidate	SMILES
A1	COC(=O)c1ccc(cc1)C1C2=C(CCCC2=O)Nc2c(cnn12)C(N)=O
A2	COc1cc(ccc1O)C1C2=C(CC(C)(C)CC2=O)Nc2c(cnn12)C(N)=O
A3	Cc1c(sc2nc[nH]c(=O)c12)C(=O)Nc1ccc(C)cc1C
A4	NC(=O)c1ccc(NC(=O)c2cccc(F)c2Cl)cn1
A5	CC(=O)c1ccc(cc1)-c1ccc(cn1)C(N)=O
A6	Cc1cccc(C)c1Nc1snc(N)c1C#N
A7	NC1=C(C#N)C(C(C#N)=C2SCC(=O)N12)c1ccccc1Cl
A8	Nc1c(C#N)c(cc(-c2ccc3OCOc3c2)c1C#N)-c1ccc(F)cc1

Table C.2: SMILES as given by MolPort from the candidates derived from the pipeline based on apo protein structure.