

## RESEARCH ARTICLE

# AI-Assisted chemical probe discovery for the understudied Calcium-Calmodulin Dependent Kinase, PNCK

Derek J. Essegian<sup>1\*</sup>, Valery Chavez<sup>2</sup>, Rabia Khurshid<sup>1</sup>, Jaime R. Merchan<sup>2</sup>, Stephan C. Schürer<sup>1,2</sup>

**1** Department of Molecular and Cellular Pharmacology, University of Miami Miller School of Medicine, Miami, Florida, United States of America, **2** Department of Medicine, Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, Florida, United States of America

\* [dje23@med.miami.edu](mailto:dje23@med.miami.edu)



## Abstract

PNCK, or CAMK1b, is an understudied kinase of the calcium-calmodulin dependent kinase family which recently has been identified as a marker of cancer progression and survival in several large-scale multi-omics studies. The biology of PNCK and its relation to oncogenesis has also begun to be elucidated, with data suggesting various roles in DNA damage response, cell cycle control, apoptosis and HIF-1-alpha related pathways. To further explore PNCK as a clinical target, potent small-molecule molecular probes must be developed. Currently, there are no targeted small molecule inhibitors in pre-clinical or clinical studies for the CAMK family. Additionally, there exists no experimentally derived crystal structure for PNCK. We herein report a three-pronged chemical probe discovery campaign which utilized homology modeling, machine learning, virtual screening and molecular dynamics to identify small molecules with low-micromolar potency against PNCK activity from commercially available compound libraries. We report the discovery of a hit-series for the first targeted effort towards discovering PNCK inhibitors that will serve as the starting point for future medicinal chemistry efforts for hit-to-lead optimization of potent chemical probes.

## OPEN ACCESS

**Citation:** Essegian DJ, Chavez V, Khurshid R, Merchan JR, Schürer SC (2023) AI-Assisted chemical probe discovery for the understudied Calcium-Calmodulin Dependent Kinase, PNCK. *PLoS Comput Biol* 19(5): e1010263. <https://doi.org/10.1371/journal.pcbi.1010263>

**Editor:** Roland L. Dunbrack, Jr., Fox Chase Cancer Center, UNITED STATES

**Received:** May 31, 2022

**Accepted:** April 13, 2023

**Published:** May 26, 2023

**Copyright:** © 2023 Essegian et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper, its [Supporting Information](#) files, and on Zenodo at <https://doi.org/10.5281/zenodo.7779371>.

**Funding:** This work was supported by NIH grants U54HL127624 (BD2K LINCS Data Coordination and Integration Center, DCIC), U24TR002278 (Illuminating the Druggable Genome Resource Dissemination and Outreach Center, IDG-RDOC), U01LM012630 (BD2K, Enhancing the efficiency and effectiveness of digital curation for biomedical

## Author summary

Machine learning and virtual screening are powerful tools in the pharmacologist's arsenal for accelerating the process of drug discovery. When targeted lesser-known proteins, however, it is important to first develop a potent, selective chemical probe. The chemical probe allows for pharmacological inhibition of protein activity to be used in addition to genetic knock-down or knock-out assays for studying the biological function of your protein in interest. In a previous multi-omics study of patient tumors, we had identified PNCK as a target of interest in kidney cancer. However, the function of PNCK is largely unknown as it has been designated as an understudied kinase. As such, we utilized kinase activity data for compounds known to target structurally similar kinases to develop machine learning models to predict small molecule binding to PNCK. Additionally, we

'big data') and Sylvester Comprehensive Cancer Center Support Grant P30CA240139 for S.C.S. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

simply used shape screening to find compound similar in shape and electronics to ATP when bound to the active site of PNCK in our multiple homology models. Using a combination of virtual methods, we were able to identify several hit compounds with favorable, tractable scaffolds, to move forward in a hit-to-lead campaign to develop the first in class, selective PNCK chemical probe. This work will lead to the elucidation of PNCKs function in several cancer models and subsequently lead to the development of a novel drug.

## Introduction

Understudied kinases, a designation assigned by the common-fund NIH project, Illuminating the Druggable Genome (IDG), comprise 24% of the entire human kinome [1–3]. This number reflects our collective gap in knowledge and reveals biases in research towards certain classes or families of kinases. For instance, the CAMK family (Calcium Calmodulin Dependent Kinase) is understudied as a whole, with no targeted small molecule inhibitors to CAMKs in clinic or under clinical investigation despite an abundance of preclinical evidence suggesting pervasive roles of CAMKs in disease progression [4–11]. Conversely, kinases of the TK (Tyrosine Kinase) family represent only 2% of the understudied kinases [1]. Understudied kinases are denoted as such due to their lack of biological annotations, experimentally derived crystal structures, potent small molecule inhibitors and bibliographic references [3]. Such dearth of information presents significant hurdles to the discovery of novel drug targets and subsequently novel small molecule therapeutics. However, analysis of large-scale multi-omics data has allowed for the identification and prioritization of understudied kinases whose expression and activity may correlate with disease activity or disease prognosis. Using data from The Cancer Genome Atlas (TCGA), we have identified several CAMKs, in particular, PNCK (CAMK1b), to be clinically relevant in several solid-tumor cancers [1,12]. Expression of PNCK was on average 6-log<sub>2</sub>-fold greater in tumor samples compared to adjacent normal tissue in clear cell renal cell carcinoma (ccRCC) patients. Further, high expression correlated significantly with survival, tumor size and histological grade in all kidney cancer cohorts. We have previously shown PNCK to play roles in angiogenesis, apoptosis, DNA damage and cell cycle control pathways [13]. To further study the biological relevance of PNCK in vitro and validate this kinase as a novel target, an AI-assisted chemical probe discovery campaign was initiated.

Due to lack of chemical and biological data, machine learning models and homology models were generated to enrich libraries of over 7 million commercially available small-molecule compounds to test for inhibitory activity against PNCK.

## Materials and methods

### Homology modeling

Protein models and homology models were generated using Prime (Schrödinger Release 2019–4: Prime, Schrödinger, LLC, New York, NY, 2020) [14]. For protein models, crystal structures were imported into Prime by PDB ID (4FG7, 4FG8, 4FG9). Using Prime Protein Preparation Wizard, missing loops and side chains were filled in, missing hydrogens were added, and the protein underwent restrained energy minimization. Water molecules with less than 3 contacts with the protein were removed. For homology modeling of PNCK, the canonical amino acid sequence was obtained from UniProt (Q6P2M8). All homology models underwent loop refinement using serial loop sampling. For all models, extended molecular dynamic

simulations (500ns-1us) were run to assess model validity and stability. All models contained co-crystallized ligands and as such, ligands were manually removed and then re-docked into the binding site to assess for model accuracy in predicting the experimentally derived binding pose and residue contacts and interactions. Homology models were assessed using SWISS-MODEL Structure Assessment tool [15,16] for structural reliability which generates a MolProbity score of global structural assessment and Ramachandran plots to ascertain allowed and disallowed dihedral angles of residues in comparison to resolved parent crystal structures. Additionally, our generated homology models were compared to those predicted by Alpha Fold via structural alignment and RMSD [17]. "[The MolProbity score] is a log-weighted combination of the clashscore, percentage Ramachandran not favored and percentage bad side-chain rotamers, giving one number that reflects the crystallographic resolution at which those values would be expected. Therefore, a structure with a numerically lower MolProbity score than its actual crystallographic resolution is, quality-wise, better than the average structure at that resolution" [16]. Residue-by-residue overlay analysis and analysis of binding site interactions between both models showed considerable overlap in structure and binding mode of ATP. Variability existed mainly in the alpha-c loop. (**Fig A in S1 File**).

### Docking Grid Generation

Receptor docking grids were generated for virtual screening with Glide Receptor Grid Generation (Schrödinger Release 2019–1: Glide, Schrödinger, LLC, New York, NY, 2019) and OpenEye's Make\_Receptor function [18] (OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>.) All proteins were initially prepared using Schrödinger Protein Prep Wizard prior to grid generation. Grids were then generated by using the co-crystallized ligand as the centroid for docking similarly sized ligands. Therefore, de novo grids in the active sites did not need to be created. Docking restrictions were made in the grid to ensure certain hydrogen-bond interactions were maintained during docking, particularly interactions with the hinge residues. Default settings were used in both OpenEye and Schrodinger grid generating modalities.

### Molecular Dynamics (MD) analyses

All-atom explicit solvent Molecular Dynamic (MD) simulations were run using the Desmond GPU Accelerated suite of Schrodinger (**Schrödinger Release 2019–1** Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2020. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2020.) Simulations used the multi-sim Desmond implementation and the simulation times varied between 300ns and 3μs, with a maximum of 3000 frames and a recording interval of 1/1000th of the overall simulation time. NPT (Isothermal-Isobaric) ensemble was used for the simulation, which most accurately represents laboratory conditions at ambient temperature and pressure (300K and 1.01325 bar). System boundary distances were specified as 10.0 Å in each dimension in an orthorhombic box and each model system was relaxed before simulation for several nanoseconds. Systems were built also using Desmond with the SPC solvent model and OPLS3e forcefield with an addition of 0.15 M NaCl to balance the charges. Simulations were computed across dedicated GPU cores. Trajectories were clustered using default settings for the "Cluster Trajectory" function.

### Machine learning and compound enrichment

3 unique virtual screening campaigns were used to prioritize commercially available compounds to purchase for biochemical screening: A naïve Bayesian classifiers model, a neural net model, and a Tanimoto color and shape screen. Our group has previously developed ligand-

based Laplacien-modified Naïve Bayesian classifiers to predict small molecule kinase activity trained on small molecule activity data across the entire human kinome [19]. A list of homologous proteins was obtained via Clustal Omega multiple sequence alignment tool using the primary sequence of the PNCK kinase domain obtained from Uniprot [20,21]. The top 31 results had sequence identity homology which ranged from 70% to 30%, with annotated active compounds with IC50s <100nM (AURKA, AURKB, AURKC, BRSK1, CAMK2D, CAMK2G, CHEK2, DAPK3, FER, MAPKAPK2, MKNK2, MYLK, NUA1, PDPK1, PLK4, PRKAA1, PRKACA, PRKACB, PRKACG, PRKD1, PRKD2, PRKD3, PRKG2, PRKX, RPS6KA1, RPS6KA2, RPS6KA3, RPS6KA5, SGK1, STK17A, TSSK1B). Amongst selected kinases, there were 9,693 active small molecule compounds in the training set. The model was then tested using over **7 million** commercially available compounds from the eMolecules library. Compound predictions from eMolecules were then scored based on their probability of binding to said target kinase- using a score known as EstPGood. A probability cut-off of 0.1 was used to enrich the virtual screening library. Additionally, compounds with multiple target predictions were prioritized while those with only one target prediction were filtered out. Compounds with a molecular weight <250 g/mol or >700 g/mol were also filtered. In total, 49,512 unique compounds were prioritized for the first virtual screen.

A second machine learning effort, a neural net-based approach, was also used which compiled data from the KKB (Kinase Knowledge Base) [22] and the eMolecules library of commercially available compounds [23]. For the neural net machine learning predictions, 314 kinases with 331,985 known active compounds were used for building the model. Compounds were featurized with Deepchem and ECFP4 fingerprints with a size of 1024 bits. The dataset was split randomly into 10% test, 10% verification and 80% training. There were 4 layers in a multi-class tensorflow network with the first layer being the 1024-bit input layer, the second layer having 2000 nodes, the third layer having 500 and the final layer being the output layer of 314 kinases. Prediction results came from the eMolecules library of compounds that were then predicted for their activity in this network and stored for later retrieval. The same homologous kinases used in the Naïve Bayesian model were used to retrieve predictions from the neural net. 289,008 unique compounds were identified as predictions to be used in the virtual screen. Compounds were filtered with a probability cutoff of >0.1, compounds active toward >1 kinases, and with molecular weights ranging from 250-700g/mol.

A third approach took advantage of the binding poses of ATP in the active site of PNCK homology models. A Tanimoto shape and color screen was employed using OpeneEye's fastROCS GPU implementation. Compounds were screened in Enamine, MolportDB and eMolecules libraries. Libraries were prepared in OpenEye first using QUACPAC which involves tautomer and protomer enumeration at physiological pH ranges. OMEGA was then used to generate up to 150 conformers for each ligand. The eMolecules dataset was filtered before the shape screen using the OpenEye "Lead-Like" filter, generating 1,169,867 unique compounds. For Enamine, the advanced HTS screening library was used which contained 97,410 unique compounds while MolportDB Screening Compound Library had 287,773 compounds. A Tanimoto combination score cutoff of 0.5 (which is the sum of a color score and shape score) was used to prioritize compounds for docking using HYBRID.

## Ligand Library Preparation

Compounds selected for virtual screen were populated as SMILES in a comma-separated file along with a unique identifier or catalog number obtained from the vendor. Using Pipeline Pilot (BIOVIA **Pipeline Pilot**, Release 2018, San Diego: **Dassault Systèmes**), SMILES were canonicalized and the file was transformed to a.SDF structure file to be used as input for

LigPrep (**Schrödinger Release 2019–3**: LigPrep, Schrödinger, LLC, New York, NY, 2019) or for OpenEye QUACPAC and OMEGA (OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>). For large-scale virtual screens with over 100,000 compounds, OpenEye Hybrid docking was utilized. Following enrichment by Hybrid Docking Score, smaller subsets of ligands were then used to dock using Glide (Schrödinger) for further binding pose visualization and MD simulations. Ligand libraries were obtained from Enamine's Advanced HTS library, MolportDB and eMolecules. The Enamine library was curated to have compounds with molecular weights between 300 and 500g/mol. The MolPortDB had no molecular weight filters, with compounds ranging from 150 to 750 g/mol. The eMolecule library was filtered into screening subsets using both lead-like and drug-like filters provided by OpenEye because the library was too large to screen as a whole. For ligand preparation, default settings were used in both programs. Compounds to be used in OpenEye were prepared first using QUACPAC which involves tautomer and protomer enumeration. OMEGA was then used to generate up to 150 conformers for each ligand. For GLIDE docking, ligands were prepared using LigPrep (**Schrödinger Release 2021–1**: LigPrep, Schrödinger, LLC, New York, NY, 2021) which generates all conformers, tautomers and ionization states at physiologic pH ranges.

### Ligand shape screen

Shape screens were conducted using fastROCS GPU (OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>), which can screen millions of compounds in seconds. Structure files of ATP exported from docking models of PNCK were used as input for the screens. A combined Tanimoto shape and color score was used to filter the most similar compounds in eMolecules, MolportDB and Enamine prepared compound libraries.

### Ligand docking

For large scale docking studies utilizing over 100,000 structures, OpenEye's HYBRID Dock was used. HYBRID docking was done in parallel on all PNCK grids in one job thus aggregation of scores was not necessary as HYBRID only reports the highest scoring pose per compound per grid. The top 25,000 docking poses were exported for final analysis. For smaller scale docking jobs or confirmatory docking studies prior to MD simulation, Glide Docking was used (**Schrödinger Release 2019–1**: Glide, Schrödinger, LLC, New York, NY, 2019). All default settings were used. Glide standard precision docking (SP) was performed using 3D ligands generated from LigPrep. Up to 25 poses were generated per ligand with the top 5 poses being selected for output. Glide scores were exported for aggregation across various grids using Pipeline Pilot.

### Compound storage and cataloging

All compounds were purchased as powders in quantities less than 5mg. Compounds were dissolved in various organic solvents (Methanol, Dichloromethane, ethyl acetate) and were aliquoted out in separate vials. Solvents were evaporated using a rotovap and high vacuum and aliquoted powders were stored at -20C. Compounds were dissolved in DMSO to a concentration of 10mM prior to use in assays. Compounds were stored in DMSO at -80C no longer than 1 week. All purchased compounds were catalogued virtually using the Collaborative Drug Discovery (CDD) Vault (<https://www.collaborativedrug.com/>).

### Biochemical PNCK Inhibitor Screen

ADP-Glo Kinase Assay (**V6930, Promega**) was used to screen the first round of compounds against PNCK. GST-tagged PNCK recombinant protein was purchased from Novus (NBP1-



99886). Autocamtide-II (GenScript RP10271) was used as the peptide substrate. This peptide was dissolved in water and prepared immediately prior to the assay. Calcium/Calmodulin 10X (Abcam ab189137) solution was used to create a custom kinase buffer: 40mM Tris,pH 7.5; 20mM MgCl<sub>2</sub>; 0.1mg/ml BSA; 50μM DTT and Ca<sup>2+</sup>/Calmodulin solution (0.03μg/μl, Calmodulin, 1mM Tris,pH 7.3, 0.5mM CaCl<sub>2</sub>). The assay was run according to protocol using 25uM of ultra-pure ATP. A standard curve was generated per run which displayed the linear range of luminescence at various ratios of ATP to ADP. A standard curve was then done to determine the amount of protein needed to generate a signal in the linear range of the ATP curve. 100ng of PNCK and 2ng of substrate was used per well in the 96-well plate format of this assay. Compounds were screened at 10uM and incubated at RT for 1-2h. Plates were read using the Clariostar Plus. Data was normalized to a DMSO and staurosporine controls.

### Cell-based PNCK inhibitor Screen

The NanoBRET Target Engagement Intracellular Kinase Assay (Promega **N2520**) is a cell-based binding screen which reads loss of BRET signal as a result of target binding. A PNCK Nano-luciferase fusion vector plasmid was generated custom by Promega. The K-10 tracer was used for the assay, as determined by Promega. The plasmid was maxiprepmed by Genewiz for use in all future assays. HEK293 cells were seeded in a T175 flask in DMEM with 10% FBS. The following day, cells were counted to 10,000, resuspended in assay media (Optimem with no phenol red, 1% FBS). Plasmid and transfection reagent (Fugene HD, Promega **E2311**) were added to the mixture and cells were seeded in a white 96-well plate. After 20–24 hours, media was changed, and the assay proceeded according to protocol. K10 tracer was added at the recommended concentration of 1uM while compounds were initially screened at both 10uM and 1uM. Cells incubated with compound at 37 degrees Celsius for a maximum of 2 hours before the plate was read using Clariostar Plus. Data was normalized to Staurosporine and DMSO controls. For compounds selected for further analysis, dose response curves were generated using the same testing protocol but at 8 different concentrations: 24uM, 12uM, 6uM, 3uM, 1.5uM, .75uM 0.375uM and 0.1875uM. UM195 was tested with concentrations from 12uM to 0.09375. Similarly, compounds from the analog by catalog analysis were tested at concentrations from 24uM to 0.1875uM.

### Data analysis

Data analysis of ADP Kinase Glo results were carried out per the manufacturer's instructions. Results were represented as a percent decrease in activity. Data was normalized to staurosporine (positive control) as 100% decrease and DMSO (negative control) as a 0% decrease. Data for NanoBRET was analyzed by dividing the acceptor by the donor emission and dividing by 1000 to obtain raw millibret units (mBU). Background was subtracted using control emission averages. Data was normalized to positive and negative controls of staurosporine and DMSO. Concentration-response curves were plotted in GraphPad Prism 8 and IC<sub>50</sub>s were extrapolated using non-linear fit with concentrations transformed to logarithmic scale. Baseline was held constant and represented the average mBU of staurosporine controls while the top was constrained to represent the average mBU of DMSO controls. Error bars on dose-response curves represent standard deviations.

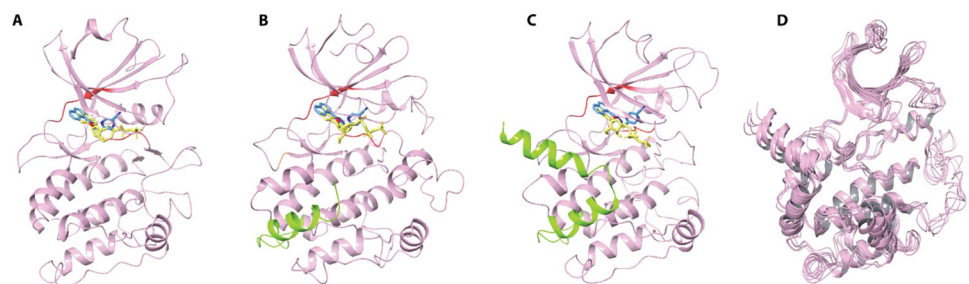
## Results

### Building and Assessing PNCK Homology Models

No experimentally derived crystal structure exists for PNCK. However, several related structures in the CAMK family have been resolved which could be used as a structural template to

generate a homology model. To determine the CAMK member with greatest homology to PNCK, the amino acid sequence of PNCK (isoform 1) was used in pBLAST (NCBI). PNCK canonical amino acid sequence was obtained from Uniprot [20] (ID: Q6P2M8) and the kinase domain is predicted to be from residues 15–270. CAMK1a was identified as the most homologous structure in the PDB with sequence identity of >70% in the kinase domain [24]. Three individual iterations of CAMK1a structures were selected as templates. All structures contained an N-terminal lobe of 5 anti-parallel beta-sheets and the canonical alpha-helix with a C-terminal lobe composed of primarily alpha helices. All native CAMKs contain a C-terminal region with auto-inhibitory and CaM binding sequences, assuming a helix-loop-helix structure [6,24–27]. CAMK1a (1–320) (PDB: 4FG9, 2.40 Å) represents a full-length version of the calmodulin kinase with the regulatory auto-inhibitory domain intact. Therefore, this kinase was experimentally shown to be inactive despite being co-crystallized with ATP in the active site. PDB structures 4FG8 (1–305, 2.20 Å) and 4FG7 (1–293, 2.70 Å) are also co-crystallized with ATP but represent two distinct forms of the kinase. 4FG8 is a truncated structure with no calmodulin binding site, thus it is in an inactive state. 4FG7 has the entire regulatory domain removed and thus represents a constitutively active form. It has been determined that the regulatory domain of CAMK1a inhibits kinase activity by interacting with the N-terminal lobe and masks the Thr177 phosphorylation site, restraining the kinase in an inactive conformation [24]. The regulatory domain, however, does not occlude nucleotide binding site, thus 4FG9 (1–320) is able to bind ATP. The activation loop still adopts an inactive conformation, with a “DFG-up” like state whereby the phenylalanine is not completely rotated out of the binding pocket, representative of a distinct inactive conformation [24,28]. 4FG8 (1–315) similarly adopts an inactive conformation, with no magnesium detected in the active site. 4FG7 (1–293) displays the structural elements of a kinase in active conformation, particularly the alpha-helix displays interactions with its conserved glutamate, forming salt bridges with ATP. This structure has a true, active, “DFG-in” conformation. Conformational analyses of the DFG structures in our homology models with KinCore confirmed 4FG7, was indeed DFG-In while 4FG8 and 4FG9 were classified as “DFG-Intermediate” [29] (**Fig B in S1 File**). In all three ATP-bound structures, the activation loop is largely disordered; contrasted with the Apo-structure, whereby the activation segment adopts a helical formation. The use of three unique templates of CAMK1a allowed for a broad sampling of conformational states in the virtual screen.

Each homology model was created using PRIME and included the entire protein with ATP in the binding site (**See Methods**) (**Fig 1A–1C**). Loops were extensively refined, and the



**Fig 1. PNCK Homology Models with ATP Docked.** ATP is docked in the active site of each model (displayed in yellow) with the hinge and gatekeeper residues colored in red. Compound UM195 is also overlaid in the active site of each model (displayed in blue). A) 4FG7 Homology model with no CAM-binding site or auto-inhibitory domain. B) 4FG8 Homology Model with CAM-binding site (Green helix) and no autoinhibitory domain. C) 4FG9 Homology model with full length PNCK including autoinhibitory domain and CAM-binding site (Green helix). D) Overlay of 5 representative poses extracted from the 500ns molecular dynamics (MD) simulation trajectory of the 4FG9 homology model.

<https://doi.org/10.1371/journal.pcbi.1010263.g001>

Table 1. SwissModel Structure Assessment.

Model	MolProbity Score	Ramachandran Favored %	Ramachandran Outliers %
4FG7-Homology	1.39	93.31%	0.00%
4FG7 Parent	2.07	95.87%	0.00%
4FG8-Homology	1.35	92.44%	0.69%
4FG8 Parent	2.31	95.38%	0.00%
4FG9-Homology	1.66	93.38%	0.66%
4FG9 Parent	2.05	96.88%	0.00%

<https://doi.org/10.1371/journal.pcbi.1010263.t001>

structure went through multiple rounds of energy minimization to produce the most stable model. To determine if the model is valid, the co-crystallized ATP was removed from the binding site and was docked back in using Glide. A model was deemed valid if it accurately predicted the pose and binding mode which was present in the experimentally derived structure within 1–2 Angstroms. The PDB structures all have a conserved interaction with the adenine amino group and a glutamate or valine in the hinge. Additionally, there are hydrogen bonds between the catalytic lysine (Lys44) to the gamma phosphate group and hydroxyl groups on the sugar. For PNCK, the predicted hinge residues included Val93 and Glu91 with the gate-keeper residue being Met90 and the catalytic lysine at Lys44. Models were finally evaluated for stability through extended all-atom explicit water molecular dynamic (MD) simulations (**Fig C in S1 File**). MD simulations were run using Desmond (GPU implementation) for 500ns (**see Methods**). An analysis of the RMSD of the C-alpha backbone as a function of time was used to assess overall structure integrity; as the change of RMSD approaches zero, the structure “converges” to a local stable conformation. Each PNCK model converges at about 150–300ns to an average RMSD of about 3–5 angstroms (**S3 Fig**). Models were additionally assessed using SWISS-MODEL Structure Assessment to ensure MolProbity scores at or less than the resolution of the parent structures used to generate the homology models and Ramachandran allowed residue percentages greater than 90% (**Fig D in S1 File and Table 1**). Finally, structures were compared to those predicted by Alpha Fold via structural alignment and RMSD (**Table 2**). To further increase the conformational space sampled during docking, the trajectories of the 500ns molecular dynamic simulations were clustered and 5 structurally diverse poses for each of the three homology models were extracted (**Fig 1D**). Thus, in total, 15 unique representations of PNCK were used in parallel for the virtual screening docking campaign.

### Curating compound libraries for scalable Ligand-based Virtual screening

In total, three unique virtual screening campaigns were used to prioritize commercially available compounds for *in vitro* screening. Each combined a scalable ligand-based approach with docking and molecular dynamics. Specifically, we used Naïve Bayesian classification models, a multi-task deep neural network model [23], and 3D shape-based screening (**Fig 2**) (**see Methods**).

Our group has previously developed ligand-based Laplacien-modified Naïve Bayesian classifiers to predict small molecule kinase activity. These models have also been used to predict novel small molecule inhibitors of EGFR and BRD4 [19,30]. The predictors were trained on small molecule activity data across the entire human kinome. Although there are no published selective compounds that target PNCK, similarity amongst kinase binding sites and cross-kinase activity of many kinase inhibitors facilitate the enrichment of PNCK inhibitors using predictive models of the most similar kinases. A list of homologous kinase domains was obtained by multiple sequence alignment with Clustal Omega using the primary sequence of



Table 2. Structural Alignment of Homology Models using AlphaFold.

Homology Model	Alignment Score	RMSD
4FG7 (1–293)	0.199	2.715 Å
4FG8 (1–308)	0.184	2.147 Å
4FG9 (1–308)	0.295	2.198 Å

<https://doi.org/10.1371/journal.pcbi.1010263.t002>

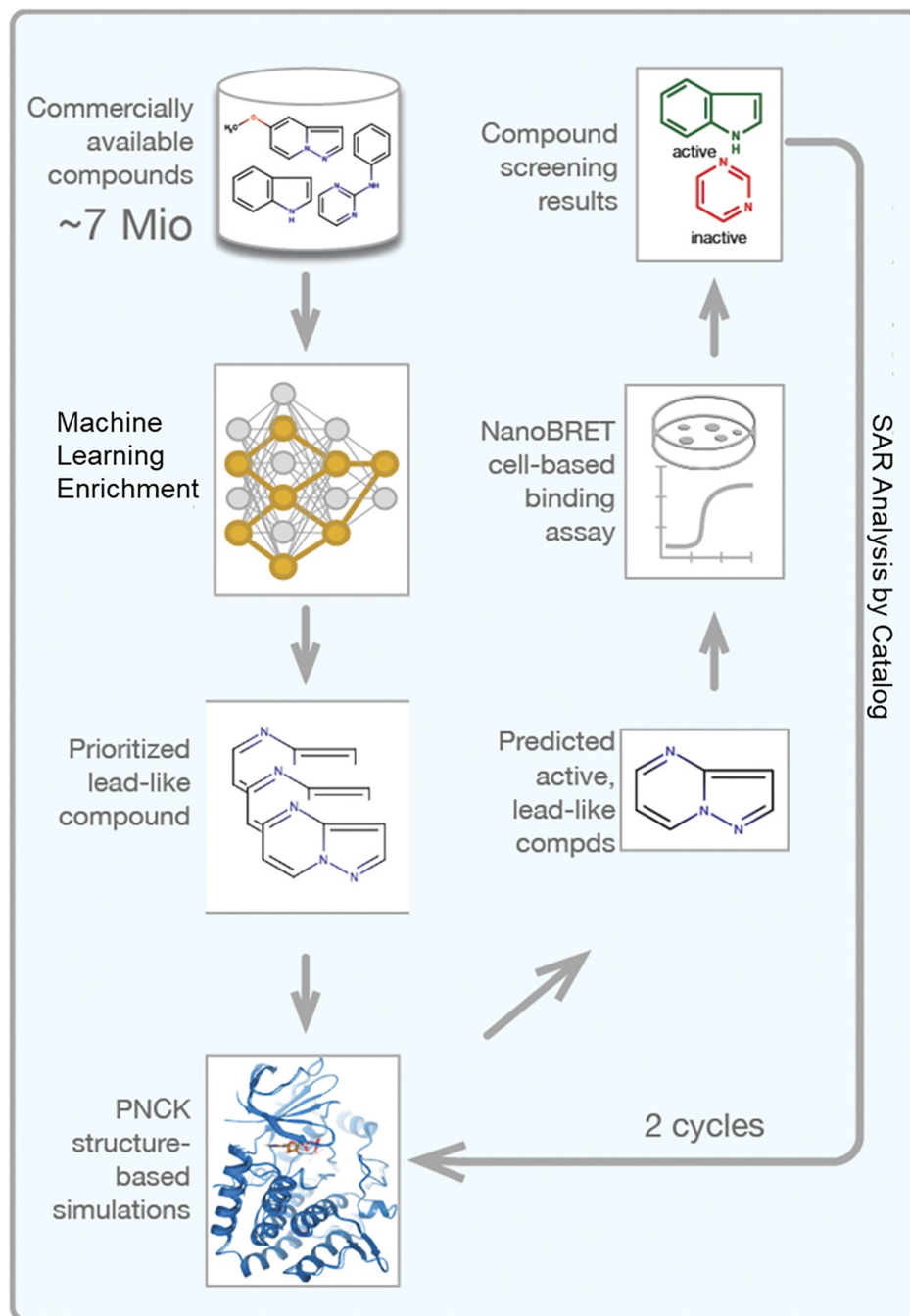
PNCK kinase domain obtained from Uniprot [20,21] (See [Methods](#)). Most of the homologous kinases with sufficient small molecule data for high-quality predictors belonged to the CAMK family, including DAPK3 and BRSK1; other kinases included PRKACA, AURKA and AURKB. The model was then employed toward a dataset with over 7 million commercially available compounds from the eMolecules library. In total, 44,736 unique compounds were prioritized for the first virtual screen.

More recently, we developed a kinome-wide multi-task deep neural network classifier trained on aggregated curated data from ChEMBL [31] and the KKB (Kinase Knowledge Base) [22,23]. Overall, 331,985 known active compounds across 314 kinases were used for building this model. Predictions were made for the entire eMolecules library. Similar to the previous approach, compounds predicted active for homologous kinases were prioritized (See [Methods](#)). Aggregating the predictions, a total of 289,008 unique compounds were prioritized. PRKACA was the kinase that was most represented, with 13,780 compounds predicted active.

A third approach took advantage of the three-dimensional predicted binding poses of ATP in each of the active sites of the PNCK homology models. A Tanimoto shape and color screen was employed using OpenEye's fastROCS (GPU implementation). Compounds from the Enamine, MolDB and eMolecules libraries were screened after generating 3D conformers. ROCS (Rapid Overlay of Chemical Structures) is a powerful method for shape similarity and pharmacophore screening which uses Gaussian overlay functions to measure both shape and "color" similarity between compounds. ROCS utilizes both Tanimoto and Tversky functions to score overlap similarities. Volume overlap generates a shape score while color Tanimoto scores are derived from scoring similarities between electrostatic forces for functional atoms or chemical groups. The eMolecules dataset was filtered before the shape screen using the OpenEye "Lead-Like" filter, generating 1,169,867 unique compounds. For Enamine, the advanced HTS screening library was used which contained 97,410 unique compounds while MolportDB had 281,987 compounds. A Tanimoto combination score cutoff of 0.5 (which is the sum of a color score and shape score), was used to prioritize compounds for docking using HYBRID. 163,493 compounds with Tanimoto scores above 0.5 were docked. The top 25,000 scoring compounds from each dataset were then evaluated further.

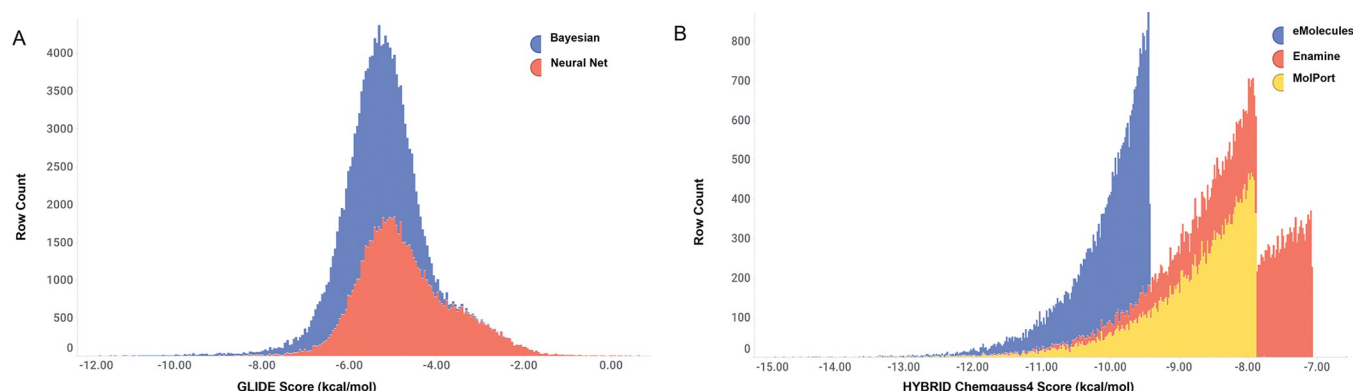
### Virtual screening campaign 1: Naïve Bayesian Classifiers and Docking

Compounds prioritized using the Laplacien modified Naïve Bayes models were prepared for docking using Schrodinger LigPrep. Compounds were then docked using Glide [32] into each of the 15 docking grids generated from the 3 homology models. Docking scores were then aggregated across all 15 receptors (See [Methods](#)). The top Glide Scores and mean Glide scores for each compound were used to rank and prioritize. As a reference, ATP received a glide average glide score of -8.59 kcal/mol. Docking scores for the machine-learning predicted compounds had an average of -5.52 kcal/mol and a standard deviation of -0.83. The top 1% of docking scores scored below (better than) -8.26 kcal/mol ([Fig 3](#)). Compounds were first clustered using extended connectivity fingerprint methods with a radius of 6 (ECFP\_6). Clusters were then ranked using the average glide score and top clusters were evaluated individually



**Fig 2. Workflow of AI-Assisted PNCK Chemical Probe Study.** Compounds from commercially available sources totaled over 7 million (from eMolecules, Enamine and MolDB). Enrichment via machine learning and shape screening led to over 358K compounds to be docked. Using the top docking scores and clustering scaffolds which scored similarly, 64 compounds were selected to use in chemical screens. 16/64 compounds were considered “hits” (>50% inhibition at 10 micromolar in the NanoBRET screen) and 5 of those were selected to further analysis and improvement by SAR. The cycle of docking, MD and cell-based screening was repeated until compound binding was improved both in silico and in vitro.

<https://doi.org/10.1371/journal.pcbi.1010263.g002>



**Fig 3. Distribution of Docking Scores in Different Machine Learning Models.** A) The naïve bayesian model and neural net model were both used to screen for predictions of inhibitory compounds towards kinases homologous to PNCK in the eMolecules library. The distribution of docking scores was very similar suggesting the neural net did not offer many new predictions. This was confirmed by manual inspection of clustered scaffolds B) vROCs shape screening and HYBRID docking using OpenEYE was performed using three small molecule libraries. The composition of each library offered much more diversity and subsequently was reflected in the distribution of docking scores, with eMolecules compounds scoring the best.

<https://doi.org/10.1371/journal.pcbi.1010263.g003>

based on synthetic tractability of the scaffold, likely binding pose, and ability for derivatization. One to two compounds from each cluster were then selected for purchase to test in *in vitro* screening. In total, 25 compounds were purchased from the eMolecules catalogue (S1 Table).

Among compounds selected for biochemical screening, the docking scores ranged from -7.37 kcal/mol (UM\_027) to -10.99 kcal/mol (UM\_039) and molecular weights ranged from 315 (UM\_019) to 410 (UM\_027) g/mol. All compounds passed The REOS filter, (rapid elimination of swill) and PAIN filter screening (pan-assay interference) except for UM\_036, as the presence of a 1,3-diazinane-2,4,6-trione may react with reagents used in such screens. Compounds UM\_027, UM\_029, UM\_031, UM\_034, and UM\_039, all had at least one “Rule of Three” violation, while none of the compounds were predicted to be exceptionally cell-permeable (Exceptional Cell permeability has a QPPCaCo score >500) [33]. In fact, 4 compounds are predicted to be considered poorly cell permeable (QPPCaCo <25) (UM\_021, UM\_031, UM\_034 and UM\_039) from QikProp analysis (S2 Table). Thus, compounds from this cohort were more “drug like” (MW <500, cLogP <5, Molar Refractivity between 40–130) than “lead like”.

## Virtual screen campaign 2: MTDNN (Multi-task Deep Neural Net) and Docking

As described above, MTDNN predictions were obtained using the same homologous kinases as campaign 1 and 289,008 unique compounds were identified as predictions to be used in the virtual screen. Compounds were filtered with a probability cutoff of >0.5, activity to >1 kinase and with molecular weights ranging from 250–700g/mol leading to a final docking library of 39,787 compounds. Although PRKACA was the most overrepresented kinase, distributions of docking scores amongst the various kinase predictions did not vary significantly. The overall distribution of docking scores was slightly different compared to the naïve Bayesian classifier model, with the average docking scores shifting to -4.70 kcal/mol, with the best docking score being -9.69 kcal/mol (Fig 3). In fact, many of the same scaffolds as in campaign 1 were similarly identified in this method. For example, the cyclopropyl-pyrazole-carbonyl moiety in UM\_019 and UM\_034 was present in many of the top clusters of the MTDNN campaign. Additionally, the thieno [2,3-d] pyrimidine scaffold found in UM\_027, UM\_028, UM\_033 and UM\_035 was also highly represented in these predictions campaign. Therefore, this method

did not greatly increase the diversity of scaffolds, likely due to the fact that compounds were screened against the same eMolecules library and the two machine learning methods did not identify significantly different compounds. Therefore, no compounds identified from this method were purchased for further analysis in biochemical or cell-based screens.

### Virtual screen campaign 3: Tanimoto ROCS Shape Screen and Docking

By far, the greatest and most diverse docking results came from a third approach which simply consisted of finding molecules that are of similar shape, size, and electronic configuration as our computational model of ATP docked in the PNCK active site. Using fastROCS GPU of OpenEye, millions of compounds from eMolecules, Enamine and MolDB were queried against the 15 poses of ATP obtained from the molecular dynamic trajectories of the 3 PNCK homology models (See [Methods](#)). All conformers and ionization states generated from the compounds with Tanimoto combo scores >0.5 were then docked across all 15 receptors using OpenEye Hybrid and the top 25,000 scoring representations across all docking models were selected for further evaluation. Compounds were clustered using ECFP6 fingerprints as described before and top scoring clusters were evaluated for synthetic tractability and binding mode. As with the Naïve-Bayesian campaign, representative structures with unique scaffolds from the top scoring clusters were purchased for analysis in *in vitro* screens. Significant differences were noted in distribution of Tanimoto and docking scores amongst the three databases. Compounds from the eMolecules database had an average HYBRID docking score of -9.9 kcal/mol, followed by MolPortDB with an average of -8.56 kcal/mol and Enamine with an average of -7.92 kcal/mol ([Fig 3](#)).

When comparing the three homology models for PNCK, docking results from the Tanimoto screen using HYBRID show a strong preference for the 4FG9 grid (the inactive form of PNCK), with the best docking scores from these compounds coming from poses of the auto-inhibited state. Conversely, 42% of compounds selected from the Naïve Bayesian cohort (UM\_019-UM\_039) scored best with GLIDE in the 4FG7, constitutively active PNCK model.

### Target engagement assays

The 25 Compounds prioritized by Naive Bayesian classifiers and docking (Campaign 1) were initially tested in a cell-free biochemical kinase activity screen, ADP Kinase Glo (See [Methods](#)). Six compounds were identified as hits (>50% inhibition of kinase activity at 10uM): UM\_026, UM\_029, UM\_032, UM\_035, UM\_038 and UM\_039 ([Fig E in S1 File](#)). However, it was determined that PNCK is not a highly active kinase and likely sub-optimal activity without phosphorylation at Thr171 (equivalent of Thr177 on CAMK1a) [[34,35](#)]. Thus, significant amounts of PNCK (>100ng per reaction) were needed to obtain signals in the linear range of the standard curve to extrapolate kinase inhibitory activity. Therefore, NanoBRET target engagement assay was used for future screens, including re-testing all compounds tested in the ADP-Glo Kinase Screen. NanoBRET is more sensitive than the ADP-Glo and additionally selects for compounds that are cell permeable (see [Methods](#)).

From the shape screen and docking campaign, compounds UM\_194 to UM\_244 were prioritized for testing in the NanoBRET target engagement assay. All compounds chosen for binding assays were analyzed for their putative binding mode by GLIDE docking and molecular dynamics (MD) simulations using the homology model for which it scored the highest. 55 compounds were initially tested in the NanoBRET assay, including compounds UM\_019-UM\_039 from the first campaign. Compounds UM\_192, UM\_193, UM\_242, UM\_243 and UM\_244 were not cell permeable while UM\_234 and UM\_245 had poor aqueous solubility, which was properly indicated by QikProp predictions. Additionally, UM\_234 and

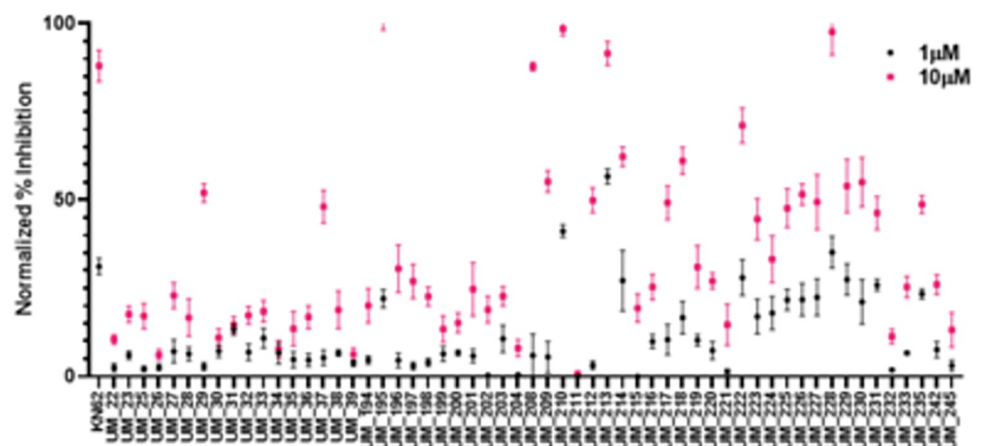
Table 3. Physiochemical Properties of Compounds From Each Virtual Campaign.

Property	Naïve Bayesian average (+/- SD)	Range	Tanimoto average (+/- SD)	Range
Molecular weight (g/mol)	368.203 (+/- 24.95)	315.329–410.54	343.52 (+/- 36.18)	262.28–420
Hydrogen-bond donors	2.095 (+/- 0.44)	1–3	1.093 (+/- 1.09)	0–4
Hydrogen-bond acceptors	6.557 (+/- 0.99)	5–8.5	6.509 (+/- 1.997)	2.8–12
QlogPo/Pw	2.707 (+/- 0.68)	1.664–4.226	2.336 (+/- 1.23)	-1.422–4.2
QPPCaco	88.334 (+/- 82.27)	8.93–320.013	674.537 (+/- 754.9)	9.677–3567.688
SASA	629.361 (+/- 44.1)	553.358–720.076	598.04 (+/- 42.38)	503.04–704.465
PSA	111.89 (+/- 17.91)	84.896–146.138	87.548 (+/- 25.61)	27.488–142.687

<https://doi.org/10.1371/journal.pcbi.1010263.t003>

UM\_243 were flagged by the PAINS filter due to the presence of a thiophene structure and a catechol structure, respectively. As such, these compounds were not subject to further analyses. Compounds from each campaign were similar in physiochemical properties but compounds from the Tanimoto shape screen were significantly more cell-permeable (Table 3). The NanoBRET screen was performed using two concentrations for compounds: 10 $\mu$ M and 1 $\mu$ M. Data was normalized to DMSO and staurosporine controls (Fig 4). Hits were defined as having greater than 50% inhibitory activity at 10 $\mu$ M. This assay confirmed actives and inactives that were identified via ADP kinase GLO with the exception of UM\_037. UM\_037 was only shown to inhibit ~31% of PNCK activity at 10 $\mu$ M in the ADP GLO assay. However, in the NanoBRET screen, UM\_037 was shown to be a potent binder at 10 $\mu$ M with 56% inhibition. Conversely, top compounds in the ADP GLO assay failed in NanoBRET, likely due to poor cell-permeability. In particular, UM\_039 had 91% inhibition of kinase activity in the biochemical screen with negligible activity in the cell-based screen. QPPCaco predictors from QikProp (Schrodinger) score UM\_039 at 8.374, which predicts poor cell permeability.

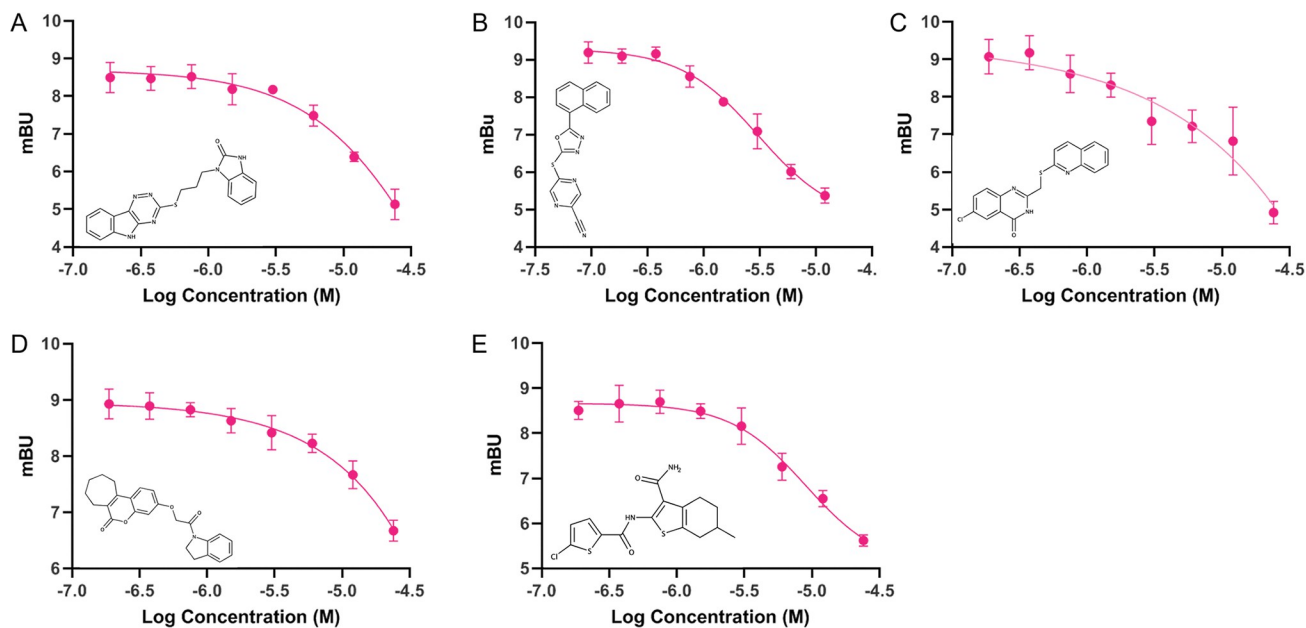
The top compounds chosen for follow-up validation in concentration-response curves were selected based on normalized percent inhibition at 10 $\mu$ M and 1 $\mu$ M, synthetic tractability and kinase inhibitor scaffold diversity or novelty (UM\_037, UM\_195, UM\_210, UM\_213 & UM\_228). Concentration-response curves were generated for each compound using the NanoBRET system at 8 concentrations ranging from 24 $\mu$ M to 0.175 $\mu$ M. Nonlinear curve



**Fig 4. Cell-Based Screening of Prioritized Compounds.** All compounds from the naïve Bayesian model in addition to compounds identified in the vROCs shape screen were tested in the cell-based target engagement assay, NanoBRET. Compounds were tested at concentrations of 10 $\mu$ M and 1 $\mu$ M. Activity was reported as normalized percent inhibition using DMSO and staurosporine as controls. Pan-CAMK inhibitor KN-62 was also used as a reference. Hits were identified as compounds with at least 50% inhibition at 10 $\mu$ M.

<https://doi.org/10.1371/journal.pcbi.1010263.g004>





**Fig 5. Dose-Response Curves for Top 5 Hit Compounds.** A) UM\_37, B) UM\_195, C) UM\_210, D) UM\_213, E) UM\_228.

<https://doi.org/10.1371/journal.pcbi.1010263.g005>

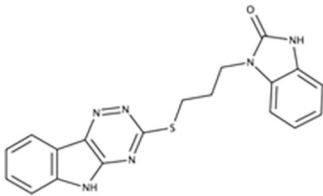
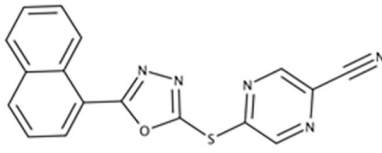
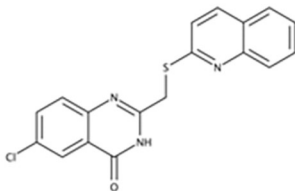
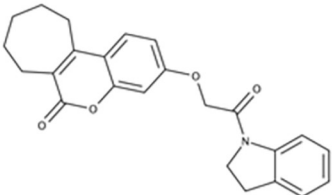
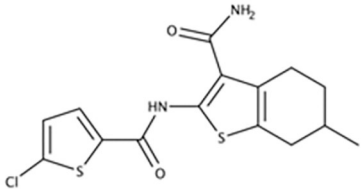
fitting after log-transforming of the concentrations allowed for calculation of IC<sub>50</sub>s (Figs 5 and 6).

Physiochemical properties for compounds were predicted from 3D structures in Schrödinger using the QikProp tool. QLogPo/Pw is the predicted octanol/water partition coefficient with higher, positive values representing increased hydrophobicity. QPPCaco is the predicted apparent Caco-2 cell permeability in nm/sec. Caco-2 cells are a model for the gut-blood barrier. QikProp predictions are for non-active transport; <25 is considered poor while >500 is great. SASA is the total solvent accessible surface area (SASA) in square angstroms using a probe with a 1.4 Å radius. PSA (polar surface area) represents the Van der Waals surface area of polar nitrogen and oxygen atoms.

In conclusion, one top scaffold was identified from Naïve-Bayesian Classifier / docking campaign while the other 4 were identified through the Tanimoto nucleotide ligand shape and color screen & docking.

### Structure Activity Relationship (SAR)–Analogue by catalogue

Each hit compound was studied again extensively to develop the most accurate binding hypothesis based on classical binding poses kinase inhibitors. Using the most likely predictive pose from initial docking studies (Fig 7) and MD simulations (Fig F-Z in S1 File), analogs were identified to potentially maximize hydrogen bonding interactions in the binding pocket, particularly at the hinge residues and to gain evidence for its true binding mode. Parent hit compounds were queried in ChemSPACE to find analogs among billions of chemicals in “REAL space” that had variation at one or more R-group sites (Fig F-Z in S1 File) [36]. Analogs were then docked in the PNCK models to analyze any improvement in docking scores based off additional hydrogen bond and hydrophobic interactions at the hinge, gatekeeper, or catalytic lysine residues. Analog compounds with improvement in docking scores or with modifications congruent with a structure-activity relationship hypothesis were ordered for synthesis on demand. Several custom analogs were ordered for each initial hit (limited by

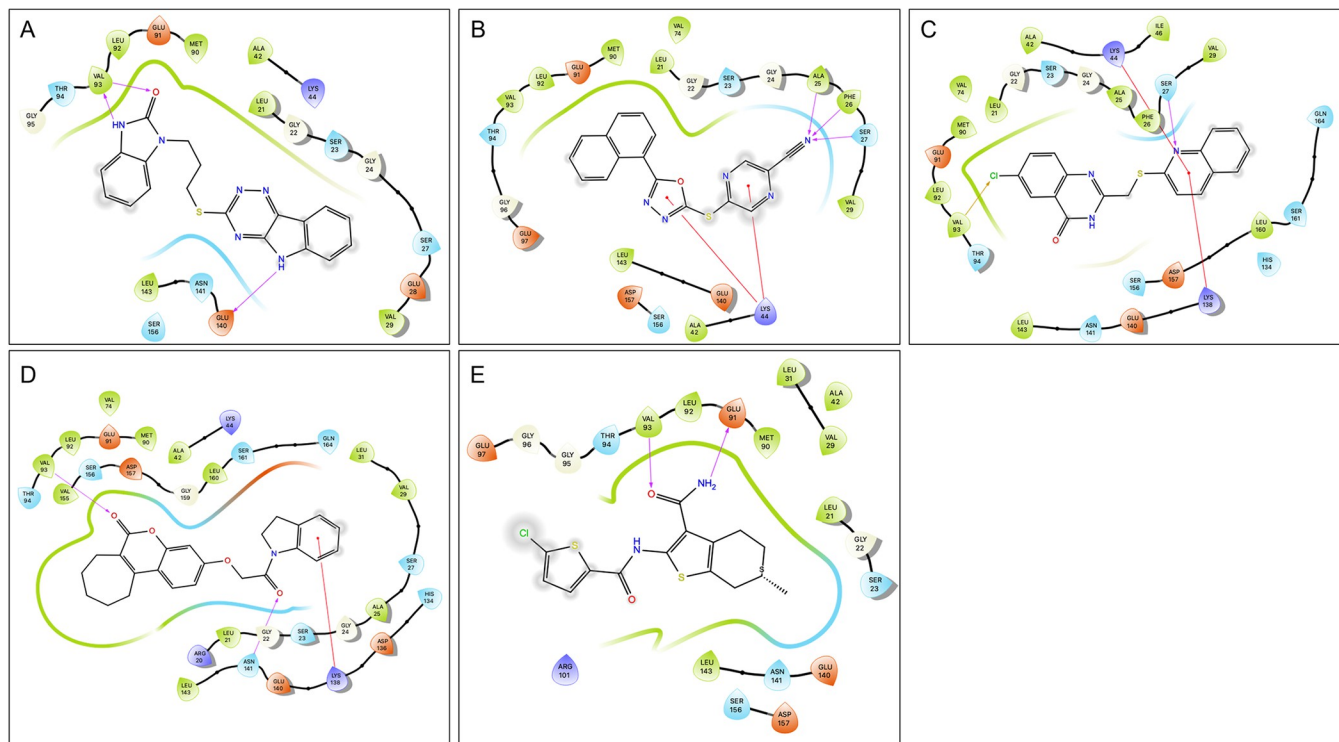
Compound	Structure	IC <sub>50</sub> (μM)	MW (g/mol)
UM 37		8.78 (+/- 1.32)	376.44
UM 195		3.36 (+/- 0.91)	331.35
UM_210		8.09 (+/- 2.19)	353.82
UM 213		9.49 (+/- 3.58)	389.451
UM 228		8.32 (+/- 2.02)	354.87

**Fig 6. IC<sub>50</sub>s of Top 5 Hit Compounds.**

<https://doi.org/10.1371/journal.pcbi.1010263.g006>

commercial availability) and were tested again in the NanoBRET screen at 1μM and 10μM (Fig 8). Compounds with at least 50% activity at 10μM were analyzed further with concentration-response curves to generate IC<sub>50</sub> values (Fig 9). These data then allowed for preliminary structure-activity relationships for each of the scaffolds.

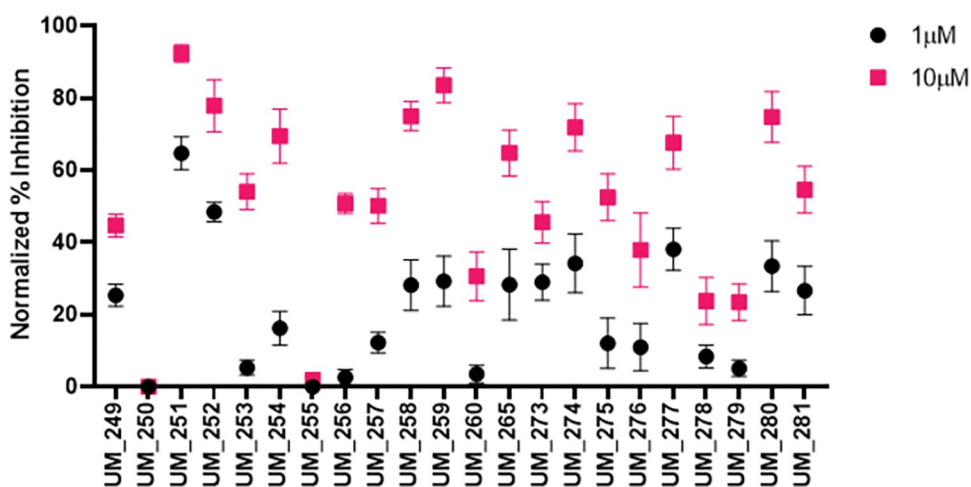
**UM\_228**, (6R)-2-(5-chlorothiophene-2-amido)-6-methyl-4,5,6,7-tetrahydro-1-benzothio-phen-3-carboxamide: Molecular modeling of UM\_228 led to several binding hypotheses. As we prioritized poses with hydrogen bonds at the hinge, the predicted pose with the chlorine atom solvent exposed, the thiophene group deep in the binding pocket and the amide group positioned for hydrogen bonding with hinge residues Glu91 and Val93 was used for selecting analogs. Additional poses reversed the orientation by 180 degrees with the methyl group being solvent exposed and the chlorine participating in bonding interactions at the catalytic lysine. In selecting analogs, various parts of the molecule were considered to determine which groups of the scaffold were important for binding, determined by analyzing docking studies and MD



**Fig 7. 2D Representation of Predicted Binding Poses of Hit Compounds.** A) UM\_37, B) UM\_195, C) UM\_210, D) UM\_213, E) UM\_228 (Add info about hinge residues).

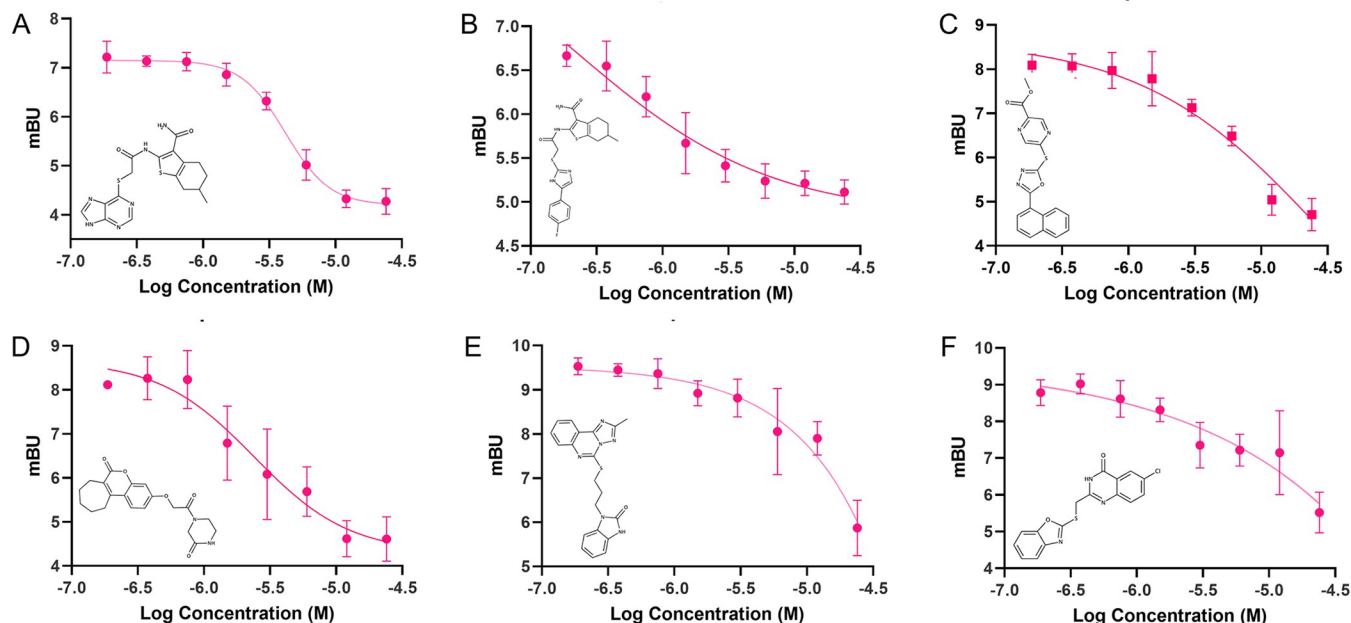
<https://doi.org/10.1371/journal.pcbi.1010263.g007>

simulations (**Fig F-H in S1 File**). The methyl group was removed in one analog or replaced with bulkier groups to maximize hydrophobic interactions in the hinge. Chlorine was removed or was replaced with various hydrogen bond donors or acceptors such as a nitro group or various heterocycles. Docking studies of the analogs supported the hypothesis that the chlorine-substituted thiophene was likely situated towards the hinge, as replacing this with a purine



**Fig 8. Cell-Based Screening of Analog Compounds.** Analogs of the 5 hit compounds were identified from ChemSpace and prioritized via docking and SAR hypothesis. Compounds were tested using the NanoBRET assay at concentrations of 10 μM and 1 μM.

<https://doi.org/10.1371/journal.pcbi.1010263.g008>



**Fig 9. Dose Response Curves of Top Analog Compounds.** A) UM\_251, B) UM\_252, C) UM\_254, D) UM\_259, E) UM\_277, F) UM\_280.

<https://doi.org/10.1371/journal.pcbi.1010263.g009>

group as in UM\_251 greatly improved docking scores with sustained interactions with Val93 (**Fig I in S1 File**). Molecular dynamics analyses of UM\_251 in the 4FG9 model demonstrates various hydrogen bond interactions along with significant mobility of the auto-inhibitory domain to create a new binding pocket, with new interactions gained at Serine156 (**Fig J in S1 File**). In UM\_252, thiophene in UM\_228 is replaced with a fluorophenyl-pyrazole. Binding hypotheses situate the fluorophenyl group in the hinge with sustained hydrophobic interactions during molecular dynamic simulations while the pyrazole and amide groups participated in multiple hydrogen bonds and water bridges to Val93 and Ser156; more than the parent compound 228 (**Fig K in S1 File**). This was supported via NanoBRET assays with analogs UM\_251 and UM\_252 improving from the IC<sub>50</sub> of UM\_228 by one order of magnitude (**Fig 8**).

**UM\_213**, 3-[2-(2,3-dihydro-1H-indol-1-yl)-2-oxoethoxy]-6H,7H,8H,9H,10H,11H-cyclohepta [c]chromen-6-one: All 3 docking models oriented UM\_213 into the binding pocket of PNCK with the pyrone carbonyl hydrogen bonding with Val93 and the cycloheptane ring pointing towards either the solvent or towards to the gatekeeper, Met90. Additionally, there were hydrogen bonds between Lys44 and Ser156 with the amide nitrogen and oxygen. Molecular dynamic simulations supported hypothesis that these hydrogen bond interactions are indeed sustained; and in this pose, UM\_213 was very stable with an average RMSD of <1 Å (**Fig L-N in S1 File**). Analogs were designed to assess the importance of the large hydrophobic cycloheptane group and to optimize hydrogen bonding interactions at the opposite end of the molecule, changing the indole group for an oxo-piperazine (**UM\_259**) or an imidazolidine-2,4-dione (**UM\_258**) (**Fig L in S1 File**). Docking of analogs supported the hypothesis that the chromenone-cycloheptanyl group fits in the binding pocket by the hinge and gatekeeper while increased hydrogen bonding interactions could be observed via Val93, Lys44 and Ser156 with additional bonds at Asp136 in the 4FG9 model for UM\_259 (**Fig O,P in S1 File**). **UM\_259** was experimentally determined to have a higher affinity for PNCK than the parent, UM\_213.

**UM\_195**, 5-[5-(naphthalen-1-yl)-1,3,4-oxadiazol-2-yl]sulfanylpiperazine-2-carbonitrile: While UM\_195 was one of the most potent ligands as determined via the NanoBRET assay,

modeling of UM\_195 proved to be the most difficult. Kinase inhibitors classically bind with strong hydrogen bond interactions at the hinge. However, UM\_195 docking poses in all 15 grids did not predict any such interactions. The naphthalene group was predicted to be deep in the binding pocket with hydrophobic interactions at the gatekeeper residue (Fig 7). Some hydrogen bonding to the oxadiazol, nitrile or pyrazine groups may occur with the catalytic lysine. Molecular dynamics interactions shifted the compound into the binding pocket so that the nitrogen atoms on the oxadiazol interacted with Val93 of the hinge, leaving the nitrile group solvent exposed (Fig Q,R in S1 File). Analogs replacing the naphthalene group with heterocycles containing hydrogen bond donors and acceptors were virtually screened. Additionally, the nitrile group was replaced with various groups to improve hydrogen bonding interactions with Ser156, Lys44, Ser27 (Fig S in S1 File). Docking of analogs, however, supported the initial binding hypothesis. Analogs tested (UM\_242, UM\_244, UM\_245, UM\_253, UM\_254) represented these changes. UM\_254 proved to be the most potent analog (which substituted the terminal nitrile group with a methoxy-ester) with an IC<sub>50</sub> of 2.68 μM. However, molecular dynamics analysis of UM\_254 in any of the three models did not predict stable binding in the active site.

**UM210**, 6-chloro-2-[(quinolin-2-ylsulfanyl)methyl]-3,4-dihydroquinazolin-4-one: Initial molecular modeling predicted that UM\_210 binds to PNCK with hydrogen bonding interactions to the hinge residue, Val93, via the chlorine atom on the fused pyrimidine heterocycle (Fig T,U in S1 File). It was considered more likely that the chlorine residue is solvent exposed and the heterocycle oxygen forming a hydrogen bond with either Lys44 or Val93. Indeed, molecular dynamics simulations of UM\_210 in both the 4FG8 and 4FG9 models situated the compound in the active site with conserved hinge interactions at the pyrimidine scaffold, sustaining contacts with Val93 and Glu91. Analogs were selected for testing to assess the likely binding mode of the compound (Fig V in S1 File). For example, the quinoline ring was changed to an indazole or pyridine group. Additionally, the chlorine was removed to determine if this atom in this position contributed significantly to binding. Docking of the analogs, however, supported another binding hypothesis whereby the quinoline ring was situated deep in the binding pocket with the benzene ring extending into the gatekeeper region while the nitrogen on the quinoline participated in hydrogen bonding with hinge residues. Analogs tested, UM\_278, UM\_279 and UM\_280 did not demonstrate significantly improved binding in NanoBRET assays.

**UM\_037**, 1-(3-(5H-[1,2,4] triazino [5,6-b]indol-3-ylsulfanylpropyl)-2,3-dihydro-1H-1,3-benzodiazol-2-one: UM\_037 was the highest scoring compound in every docking model of PNCK. It was initially screened in ADP-KinaseGLO where it had minimal inhibitory activity. However, it showed significant binding in NanoBRET with an IC<sub>50</sub> of 8.79 μM. UM\_037 contains two fused heterocycle moieties, both of which are common structures among hinge-interacting kinase inhibitors. The benzimidazolinone is predicted to be situated towards the hinge with interactions between the nitrogen NH and carbonyl oxygen with the Val93 backbone. Molecular dynamics also suggested additional interactions of the benzimidazolinone with Ser156 while the triazino group had hydrogen bonds with Lys44 and Ser156 and the indole group interacted with Glu140 (Fig W,X in S1 File). Additionally, there is likely pi-stacking interactions with these groups with Phe158. One model did orient the molecule in the opposite direction, with the indole group making interactions at the hinge. To determine the proper pose, analogs were designed that altered the triazino/indol group, including adding or removing bulky groups or adding carbonyl groups to alter hydrogen bonding (Fig Y in S1 File). Molecular modeling of analogs supported the initial binding hypothesis whereby substituting the triazino-indol group for more flexible groups which retained hydrogen bonding ability resulted in better docking results. Despite improvement in docking scores, the analogs were not significantly better inhibitors of PNCK *in vitro*.



## Discussion

Computational methods are increasingly being employed in the preclinical hit discovery and lead optimization process to assist in the identification, prioritization, and optimization of active kinase inhibitor chemotypes. With the wide-spread availability of high-resolution crystals structures, large-scale chemical databases, and a slew of virtual screening platforms, *in silico* methods have contributed to the discovery of many novel kinase inhibitors in recent years. For example, several studies have used virtual screening to discover inhibitors for EGFR [154], BCR-ABL [155], PI3K [156], AURKA [157] and PKC [158]. Even understudied kinases were able to be “drugged” via the use of homology modeling and machine learning as in the case of VEGFR2 (for which no crystal structure exists) [159]. Thus, there is strong precedent for the successes of such computational pipelines to generate novel discoveries, which otherwise likely would have taken much longer and at significantly higher cost [1]. The goal of such computational screening is the identification of chemically tractable starting points to use for medicinal chemistry optimizations towards a potent molecular probe or preclinical drug candidate. Computer aided drug discovery can result in a 1000-fold enrichment of hit discovery compared to traditional screening methods. Whereby HTS is limited to hundreds of thousands to a million compounds at enormous cost, a virtual screening campaign can readily screen tens to hundreds of million and, more recently even billions of compounds to prioritize a small subset of compounds to test in biochemical screens, dramatically increasing the likelihood of obtaining starting points for lead optimization [157]. Typical hit rates from experimental HTS can range between 0.01% and 0.14%, while hit rates for prospective virtual screen can be as high as 50% [160]. Various rounds of enrichment, including applying certain structure filters (PAINS, lead-like, drug-like), machine-learning predictions, 3D shape overlay, molecular docking, and molecular dynamics, contribute to the improved success of computational drug discovery efforts. There are two major methodologies for computer aided drug discovery: structure-based methods and ligand-based methods [161]. Structure based methods are built upon the foundation of an experimentally derived, high resolution crystal structure of the target protein and active site. Ligand based methods, conversely, are built upon the foundation of a known ligand or ligands for a particular target. Due to the fact that this ambitious project sought to drug an understudied kinase, there exists no crystal structure and no known exogenous ligands to use as a foundation. Thus, several methods from both structure-based and ligand-based approaches were combined to produce a pipeline capable of predicting first-in-class active PNCK inhibitors.

The process of this computational drug discovery project involves multiple essential components. First, a reliable and highly predictive homology model of the target protein, PNCK, needed to be built. As PNCK is an understudied kinase, there are no reported crystal structures of this protein. However, structures of homologous kinases were readily available to use as a structural template for building such a model. Multiple protein ensembles were evaluated to most accurately and broadly sample the kinase conformational and energetic space. Second, a library of ligands needed to be curated for use in a virtual screening campaign. The selection of compounds to use in a screen is one of the largest contributors to success in hit identification [160, 162]. Compound curation and selection determines the size of the screening library, the scaffold diversity, and quality of positive and negative data [162]. Various computational methods are employed to filter molecules for favorable factors that have been noted in various large-scale studies of “drug-like” and “lead-like” properties. Most notable, “Rule of Five” [163] or “Rule of Three” [164] filters select for compounds with desirable molecular weights, lipophilicity and hydrogen bond interactors. More sophisticated curation efforts involve the use of pharmacophore searching, similarity screening, and AI-assisted machine learning. While it is

impossible to sample the entirety of the chemical space, the use of several diverse, commercially available databases is a reasonable starting point to identify active chemical structures as these can be purchased off-the-shelf and tested rapidly. Here we used two distinct machine learning algorithms to survey >7million compounds collectively in eMolecules, MolDB and Enamine compound libraries. Through initial rounds of filtering based on predicted probabilities of activities as well as physiochemical properties, the number of compounds chosen for use in molecular docking studies was greatly reduced from millions to thousands. Following molecular docking, compounds were scored, ranked and prioritized for further analysis. Compounds were clustered using topological features to identify chemotypes that would be enriched for potent binding and inhibition. Docking poses and binding hypotheses were evaluated manually for the top scoring compounds. Scaffolds were additionally manually curated based on medicinal chemistry considerations, evaluating for synthetic tractability and preliminary ADMET properties or liabilities. Extended molecular dynamic simulations were run for select protein-compound complexes to further refine predicted protein-ligand complexes and to study the energetics of binding. Top compounds from each cluster were purchased for testing in various binding assays or screens. Data from those screens informed the selection or design of analogs (virtually or synthetically) which allowed for further rounds of preliminary optimizations.

In conclusion, 1 top Scaffold was identified from Naïve-Bayesian classification followed by docking while the other 4 were identified through the nucleotide shape and color screen followed by docking. The docking scores could not be directly compared across the two methods (GLIDE and HYBRID) due to varying algorithms and there was no clear correlation between the scores, with each model ranking and prioritizing compounds differently. Few compounds scored highly in both models, making it difficult to draw generalized conclusions. Interestingly, all compounds preferably bound to the autoinhibited structure, 4FG9, and occupy a space that is distinct from that of ATP; while still making interactions at the hinge and gate-keeper residues. Further research on advanced compounds from these selected scaffolds, such as x-ray crystallography, may indeed identify such compounds as type-II inhibitors. As all virtual screening methods, docking scores should not be interpreted as relative binding free energies, but used as a tool for enrichment. UM\_039 is the highest ranked kinase in campaign 1 and subsequently had the highest level of inhibitory activity in the ADP Glo assay. Conversely, UM\_228 and UM\_195, two of the most potent scaffolds in the NanoBRET assay, rank quite low when scored by GLIDE. For the HYBRID docking used in campaign 3, UM\_195 is amongst the top scoring compounds and subsequently is a top inhibitory scaffold. The overall success of the entire campaign is therefore likely due to the combination of diverse methods and compound libraries, which can result in synergies.

PNCK is an understudied kinase for which there currently exists no published crystal structure and no known exogenous ligands. However, using a combination of homology modeling, machine learning and shape based virtual screening, we report a series of hit compounds with inhibitory low molecular binding potency toward PNCK. While other compounds have recently been reported to have activity against CAMK1b, they were part of a CAMK1D inhibitor campaign with non-selective activity against the family of CAM Kinases [37]. None of the compounds reported in this study, although commercially available, have any target annotation data and thus have not been previously characterized as kinase inhibitors. The overall “hit rate” of 27% is exceptionally high due to rounds of iterative enrichment using machine learning and shape-based scoring with molecular docking and molecular dynamics studies. Specifically, all the compounds discovered have IC50s in the low micromolar range and molecular weights under 400 Daltons, making them a good starting point for hit-to-lead optimization. Future goals are to evaluate the target specificity of these initial hit compounds in a kinome-

wide screen, with particular focus on developing isoform specific inhibitors of PNCK (vs other CAMK proteins). The most selective scaffold(s) will then be chosen for lead optimization using medicinal chemistry. With preliminary SAR by catalog studies, we have improved the IC50s of UM\_228, UM\_213 and UM\_195. UM\_251 and UM\_252 are both analogs of UM\_228 with IC50s of 2.13 and 1.09  $\mu$ M respectively- making them the most potent PNCK inhibitors in the series. Additional rounds of optimizations will be needed to optimize their activity while improving ADMET properties. In particular, we hope to replace all thioethers with oxygen isosteres and eliminate any furan rings, which are known to be metabolic liabilities. Once an isoform specific and potent chemical probe is generated, with an IC50 in the nanomolar range, we will move forward for testing in cells and in animal models. Previous work has been done to develop potent CAMK1D inhibitors. While the authors achieved specificity to CAMKs, they were unable to obtain isoform specificity.

## Supporting information

**S1 Table. List of Compounds with Screening Data.**  
(XLSX)

**S2 Table. QikProp Properties of all Compounds.**  
(CSV)

**S3 Table. Dose-Response Curve Data.**  
(XLSX)

**S1 File. Fig A.** Comparison of PNCK Schrödinger Prime homology model with AlphaFold model shows some differences but similar orientation with key interactions well aligned. A) PNCK Prime homology model (pink with bound ATP in yellow) aligned with AlphaFold model (blue with bound ATP in grey) in Schrödinger Maestro v3.0.137 using Protein Structure Alignment tool; shows considerable similarity in orientation of both the models; B) PNCK binding site shows key interactions are fairly aligned between both Prime and AlphaFold models; C) Hinge residues (Glu 91, Leu 92, Val 93) show similar orientation and interactions with ligand (ATP); D) DFG residues (Asp 157, Phe 158, Gly 159) show slight difference in orientation; E) Gatekeeper (Met 90) shows similar orientation; F) Key residue, Lys 44, shows similar orientation and interaction with the ligand (ATP); G) Key residues Asn 141 shows slightly different orientation and makes interaction with ligand in AlphaFold model; H)  $\alpha$ -c loop is somewhat different among the two models; I) 2D interactions of PNCK AlphaFold model with bound ATP; J) 2D interactions of PNCK Schrödinger Prime model with bound ATP. **Fig B.** PNCK Schrödinger Prime homology model conformation identification using KinCoRe. **Fig C. Molecular Dynamics (MD) Analysis of 3 PNCK Homology Models.** (A-C), Molecular Dynamic RMSD analysis of protein (C-alpha) and ligand fit on protein (4FG7, 4FG8, 4FG9). (D-F) Interaction Fraction of ATP to Residues in PNCK Active Site (4FG7, 4FG8, 4FG9). **Fig D. SwissModel Ramachandran Plots for PNCK Homology Models.** A) 4FG7 Homology Model, B) 4FG7 Original, C) 4FG8 Homology Model, D) 4FG8 Original, E) 4FG9 Homology Model, F) 4FG9 Original. **Fig E.** ADP-Kinase GLO Results from Naïve Bayesian Classifier Campaign. **Fig F. UM\_228 Binding Mode Hypothesis After MD, 4FG8 Model.** A) 2D representation, B) 3D Representation C) 3D Surface Representation of Binding Pocket. **Fig G. Molecular Dynamics Analysis of UM\_228.** A) Protein-Ligand RMSD plot of 500ns MD simulation as a function of time. B) Protein-Ligand Contact Interactions as Fraction of total MD simulation time. C) Protein-Ligand Contact plot as function of time. Darker lines indicate more contacts between the compound and residue. **Fig H. Structure Activity Relationship (SAR) Table for Analogs of UM\_228.** **Fig I. 2D Binding Hypothesis of UM\_228 Analogs.**

A) UM\_251, B) UM\_252. **Fig J. Molecular Dynamics Analysis of UM\_251.** A) Protein-Ligand RMSD plot of 500ns MD simulation as a function of time. B) Protein-Ligand Contact Interactions as Fraction of total MD simulation time. C) Protein-Ligand Contact plot as function of time. Darker lines indicate more contacts between the compound and residue. **Fig K. Molecular Dynamics Analysis of UM\_252.** A) Protein-Ligand RMSD plot of 500ns MD simulation as a function of time. B) Protein-Ligand Contact Interactions as Fraction of total MD simulation time. C) Protein-Ligand Contact plot as function of time. Darker lines indicate more contacts between the compound and residue. **Fig L. UM\_213 Binding Mode Hypothesis After MD, 4FG9 Model.** A) 2D representation, B) 3D Representation C) 3D Surface Representation of Binding Pocket. **Fig M. Molecular Dynamics Analysis of UM\_213.** A) Protein-Ligand RMSD plot of 500ns MD simulation as a function of time. B) Protein-Ligand Contact Interactions as Fraction of total MD simulation time. C) Protein-Ligand Contact plot as function of time. Darker lines indicate more contacts between the compound and residue. **Fig N. Structure Activity Relationship (SAR) Table for Analogs of UM\_213.** **Fig O. 2D Binding Hypothesis of UM\_213 Analog, UM\_259.** **Fig P. Molecular Dynamics Analysis of UM\_259.** A) Protein-Ligand RMSD plot of 500ns MD simulation as a function of time. B) Protein-Ligand Contact Interactions as Fraction of total MD simulation time. C) Protein-Ligand Contact plot as function of time. Darker lines indicate more contacts between the compound and residue. **Fig Q. UM195 Binding Mode Hypothesis After MD, 4FG7 Model.** A) 2D representation, B) 3D Representation C) 3D Surface Representation of Binding Pocket. **Fig R. Molecular Dynamics Analysis of UM\_195.** A) Protein-Ligand RMSD plot of 500ns MD simulation as a function of time. B) Protein-Ligand Contact Interactions as Fraction of total MD simulation time. C) Protein-Ligand Contact plot as function of time. Darker lines indicate more contacts between the compound and residue. **Fig S. Structure Activity Relationship (SAR) Table for Analogs of UM\_195.** **Fig T. UM210 Binding Mode Hypothesis After MD, 4FG9 Model.** A) 2D representation, B) 3D Representation C) 3D Surface Representation of Binding Pocket. **Fig U. Molecular Dynamics Analysis of UM\_210.** A) Protein-Ligand RMSD plot of 500ns MD simulation as a function of time. B) Protein-Ligand Contact Interactions as Fraction of total MD simulation time. C) Protein-Ligand Contact plot as function of time. Darker lines indicate more contacts between the compound and residue. **Fig V. Structure Activity Relationship (SAR) Table for Analogs of UM\_210.** **Fig W. UM37 Binding Mode Hypothesis After MD, 4FG8 Model.** A) 2D representation, B) 3D Representation C) 3D Surface Representation of Binding Pocket. **Fig X. Molecular Dynamics Analysis of UM\_37.** A) Protein-Ligand RMSD plot of 500ns MD simulation as a function of time. B) Protein-Ligand Contact Interactions as Fraction of total MD simulation time. C) Protein-Ligand Contact plot as function of time. Darker lines indicate more contacts between the compound and residue. **Fig Y. Structure Activity Relationship (SAR) Table for Analogs of UM\_37.** (PDF)

## Acknowledgments

The authors would like to acknowledge Ron Lampert and Sarah Simon for their editing of review of manuscript and creation of figures and tables.

## Author Contributions

**Conceptualization:** Derek J. Essegian, Valery Chavez, Stephan C. Schürer.

**Data curation:** Derek J. Essegian.

**Formal analysis:** Derek J. Essegian, Valery Chavez.

**Funding acquisition:** Jaime R. Merchan, Stephan C. Schürer.

**Investigation:** Derek J. Essegian, Valery Chavez, Jaime R. Merchan, Stephan C. Schürer.

**Methodology:** Derek J. Essegian, Valery Chavez, Jaime R. Merchan, Stephan C. Schürer.

**Project administration:** Jaime R. Merchan, Stephan C. Schürer.

**Resources:** Jaime R. Merchan, Stephan C. Schürer.

**Software:** Stephan C. Schürer.

**Supervision:** Jaime R. Merchan.

**Validation:** Valery Chavez.

**Visualization:** Derek J. Essegian, Valery Chavez, Rabia Khurshid.

**Writing – original draft:** Derek J. Essegian.

**Writing – review & editing:** Derek J. Essegian, Valery Chavez, Stephan C. Schürer.

## References

- Essegian D, Khurana R, Stathias V, Schurer SC. The Clinical Kinase Index: A Method to Prioritize Understudied Kinases as Drug Targets for the Treatment of Cancer. *Cell Rep Med*. 2020; 1(7):100128. <https://doi.org/10.1016/j.xcrm.2020.100128> PMID: 33205077
- Nguyen DT, Mathias S, Bologa C, Brunak S, Fernandez N, Gaulton A, et al. Pharos: Collating protein information to shed light on the druggable genome. *Nucleic Acids Res*. 2017; 45(D1):D995–D1002. <https://doi.org/10.1093/nar/gkw1072> PMID: 27903890
- Oprea TI, Bologa CG, Brunak S, Campbell A, Gan GN, Gaulton A, et al. Unexplored therapeutic opportunities in the human genome. *Nat Rev Drug Discov*. 2018; 17(5):317–32. <https://doi.org/10.1038/nrd.2018.14> PMID: 29472638
- Swulius MT, Waxham MN. Ca(2+)/calmodulin-dependent protein kinases. *Cell Mol Life Sci*. 2008; 65(17):2637–57. <https://doi.org/10.1007/s00018-008-8086-2> PMID: 18463790
- Rodriguez-Mora OG, LaHair MM, McCubrey JA, Franklin RA. Calcium/calmodulin-dependent kinase I and calcium/calmodulin-dependent kinase kinase participate in the control of cell cycle progression in MCF-7 human breast cancer cells. *Cancer Res*. 2005; 65(12):5408–16. <https://doi.org/10.1158/0008-5472.CAN-05-0271> PMID: 15958590
- Rodriguez-Mora O, LaHair MM, Howe CJ, McCubrey JA, Franklin RA. Calcium/calmodulin-dependent protein kinases as potential targets in cancer therapy. *Expert Opin Ther Targets*. 2005; 9(4):791–808. <https://doi.org/10.1517/14728222.9.4.791> PMID: 16083343
- Liu Z, Han G, Cao Y, Wang Y, Gong H. Calcium/calmodulin-dependent protein kinase II enhances metastasis of human gastric cancer by upregulating nuclear factor kappaB and Akt-mediated matrix metalloproteinase9 production. *Mol Med Rep*. 2014; 10(5):2459–64.
- Pellicena P, Schulman H. CaMKII inhibitors: from research tools to therapeutic agents. *Front Pharmacol*. 2014; 5:21. <https://doi.org/10.3389/fphar.2014.00021> PMID: 24600394
- Sang LJ, Ju HQ, Liu GP, Tian T, Ma GL, Lu YX, et al. LncRNA CamK-A Regulates Ca(2+)-Signaling-Mediated Tumor Microenvironment Remodeling. *Mol Cell*. 2018; 72(1):71–83 e7. <https://doi.org/10.1016/j.molcel.2018.08.014> PMID: 30220561
- Schulman H. The multifunctional Ca2+/calmodulin-dependent protein kinases. *Curr Opin Cell Biol*. 1993; 5(2):247–53. [https://doi.org/10.1016/0955-0674\(93\)90111-3](https://doi.org/10.1016/0955-0674(93)90111-3) PMID: 8507497
- Liskovych M, Goncharov NV, Petrov N, Aksanova V, Pegoraro G, Ozbun LL, et al. A novel assay to screen siRNA libraries identifies protein kinases required for chromosome transmission. *Genome Res*. 2019; 29(10):1719–32. <https://doi.org/10.1101/gr.254276.119> PMID: 31515286
- Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*. 2013; 45(10):1113–20. <https://doi.org/10.1038/ng.2764> PMID: 24071849
- Essegian DJ, Chavez V, Bustamante F, Schurer SC, Merchan JR. Cellular and molecular effects of PNCK, a non-canonical kinase target in renal cell carcinoma. *iScience*. 2022; 25(12):105621. <https://doi.org/10.1016/j.isci.2022.105621> PMID: 36465101



14. Jacobson MP, Pincus DL, Rapp CS, Day TJ, Honig B, Shaw DE, et al. A hierarchical approach to all-atom protein loop prediction. *Proteins*. 2004; 55(2):351–67. <https://doi.org/10.1002/prot.10613> PMID: 15048827
15. Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T. Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc*. 2009; 4(1):1–13. <https://doi.org/10.1038/nprot.2008.197> PMID: 19131951
16. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*. 2010; 66(Pt 1):12–21. <https://doi.org/10.1107/S0907444909042073> PMID: 20057044
17. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021; 596(7873):583–9. <https://doi.org/10.1038/s41586-021-03819-2> PMID: 34265844
18. McGann M. FRED and HYBRID docking performance on standardized datasets. *J Comput Aided Mol Des*. 2012; 26(8):897–906. <https://doi.org/10.1007/s10822-012-9584-8> PMID: 22669221
19. Allen BK, Mehta S, Ember SW, Schonbrunn E, Ayad N, Schurer SC. Large-Scale Computational Screening Identifies First in Class Multitarget Inhibitor of EGFR Kinase and BRD4. *Sci Rep*. 2015; 5:16924. <https://doi.org/10.1038/srep16924> PMID: 26596901
20. The UniProt C. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2017; 45(D1):D158–D69. <https://doi.org/10.1093/nar/gkw1099> PMID: 27899622
21. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, et al. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res*. 2010; 38(Web Server issue):W695–9. <https://doi.org/10.1093/nar/gkq313> PMID: 20439314
22. Brooijmans N, Chang YW, Mobilio D, Denny RA, Humblet C. An enriched structural kinase database to enable kinome-wide structure-based analyses and drug discovery. *Protein Sci*. 2010; 19(4):763–74. <https://doi.org/10.1002/pro.355> PMID: 20135687
23. Allen BK, Ayad NG, Schürer SC. Kinome-wide activity classification of small molecules by deep learning. *bioRxiv*. 2019.
24. Zha M, Zhong C, Ou Y, Han L, Wang J, Ding J. Crystal structures of human CaMKIalpha reveal insights into the regulation mechanism of CaMKI. *PLoS One*. 2012; 7(9):e44828.
25. Wilmann M, Gautel M, Mayans O. Activation of calcium/calmodulin regulated kinases. *Cell Mol Biol (Noisy-le-grand)*. 2000; 46(5):883–94. PMID: 10976872
26. Gardner HP, Ha SI, Reynolds C, Chodosh LA. The caM kinase, Pnck, is spatially and temporally regulated during murine mammary gland development and may identify an epithelial cell subtype involved in breast cancer. *Cancer Res*. 2000; 60(19):5571–7. PMID: 11034105
27. Gardner HP, Rajan JV, Ha SI, Copeland NG, Gilbert DJ, Jenkins NA, et al. Cloning, characterization, and chromosomal localization of Pnck, a Ca(2+)/calmodulin-dependent protein kinase. *Genomics*. 2000; 63(2):279–88. <https://doi.org/10.1006/geno.1999.6091> PMID: 10673339
28. Vijayan RS, He P, Modi V, Duong-Ly KC, Ma H, Peterson JR, et al. Conformational analysis of the DFG-out kinase motif and biochemical profiling of structurally validated type II inhibitors. *J Med Chem*. 2015; 58(1):466–79. <https://doi.org/10.1021/jm501603h> PMID: 25478866
29. Modi V, Dunbrack RL. Kincore: a web resource for structural classification of protein kinases and their inhibitors. *Nucleic Acids Res*. 2022; 50(D1):D654–D64. <https://doi.org/10.1093/nar/gkab920> PMID: 34643709
30. Schurer SC, Muskall SM. Kinome-wide activity modeling from diverse public high-quality data sets. *J Chem Inf Model*. 2013; 53(1):27–38. <https://doi.org/10.1021/ci300403k> PMID: 23259810
31. Bento AP, Gaulton A, Hersey A, Bellis LJ, Chambers J, Davies M, et al. The ChEMBL bioactivity database: an update. *Nucleic Acids Res*. 2014; 42(Database issue):D1083–90. <https://doi.org/10.1093/nar/gkt1031> PMID: 24214965
32. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem*. 2004; 47(7):1739–49. <https://doi.org/10.1021/jm0306430> PMID: 15027865
33. Jorgensen WL, Duffy EM. Prediction of drug solubility from structure. *Adv Drug Deliv Rev*. 2002; 54(3):355–66. [https://doi.org/10.1016/s0169-409x\(02\)00008-x](https://doi.org/10.1016/s0169-409x(02)00008-x) PMID: 11922952
34. Haribabu B, Hook SS, Selbert MA, Goldstein EG, Tomhave ED, Edelman AM, et al. Human calcium-calmodulin dependent protein kinase I: cDNA cloning, domain structure and activation by phosphorylation at threonine-177 by calcium-calmodulin dependent protein kinase I kinase. *EMBO J*. 1995; 14(15):3679–86. <https://doi.org/10.1002/j.1460-2075.1995.tb00037.x> PMID: 7641687

35. Deb TB, Zuo AH, Wang Y, Barndt RJ, Cheema AK, Sengupta S, et al. Pnck induces ligand-independent EGFR degradation by probable perturbation of the Hsp90 chaperone complex. *Am J Physiol Cell Physiol.* 2011; 300(5):C1139–54. <https://doi.org/10.1152/ajpcell.00167.2010> PMID: 21325639
36. Grygorenko OO, Radchenko DS, Dziuba I, Chuprina A, Gubina KE, Moroz YS. Generating Multibillion Chemical Space of Readily Accessible Screening Compounds. *iScience.* 2020; 23(11):101681. <https://doi.org/10.1016/j.isci.2020.101681> PMID: 33145486
37. Fromont C, Atzori A, Kaur D, Hashmi L, Greco G, Cabanillas A, et al. Discovery of Highly Selective Inhibitors of Calmodulin-Dependent Kinases That Restore Insulin Sensitivity in the Diet-Induced Obesity in Vivo Mouse Model. *J Med Chem.* 2020; 63(13):6784–801. <https://doi.org/10.1021/acs.jmedchem.9b01803> PMID: 32433887