ITÄ-SUOMEN YLIOPISTO

Scoring assessment of three machine learning models based on molecular docking at multiple conformations of two kinase targets

Alejandro Díaz Holguín
Master of Science Thesis
Master's Degree Programme
in Biomedicine
University of Eastern Finland
Faculty of Health Sciences
School of Medicine
19th of May 2020

University of Eastern Finland, Faculty of Health Sciences
Master's Degree Programme in Biomedicine
Student, Alejandro Díaz Holguín: Scoring assessment of three machine learning models based on molecular docking at multiple conformations of two kinase targets
Master of Science Thesis; 36p.
Supervisors: Professor Antti Poso, Adj. Professor Tuomo Laitinen, PhD Thales Kronenberger, PhD student Arun Tonduru
19th of May 2020

Keywords: Docking, Machine Learning, ROC curves, classification, AURKA, PKACA

Abstract

The drug discovery pipeline is time consuming and expensive for both industry and academia. Molecular docking, is an efficient method which succeeds when generating reasonable binding poses. However, conventional scoring functions are limited when predicting binding energies or classifying molecules as actives or inactives. To address the latter, herein three machine learning models were implemented to predict the pKi of the ChEMBL benchmark for protein kinase A catalytic subunit α (PKACA) and Aurora Kinase A (AURKA).The classification performance of the models was assessed together with XP GlideScore from Maestro software. Multiple conformational states of PKACA and AURKA were selected as docking targets. In average for all the studied conformations, the ML model predictions showed up to 17 and 35 $\%$ enhancement in the ROC AUC (receiver operator characteristic, area under the curve) when compared to the XP Scores reported for PKACA and AURKA, respectively. Also, it was found that ML models built for the inhibitor bound conformations of both kinase targets reported an enhanced regression and classification performance compared to the other states.

Introduction

Drug development has actively evolved throughout history. 60,000 years old fossil records show the burial of a Neanderthal individual alongside herbs with medicinal properties (Lietava 1992). These archaic records are not only of anthropological interest, as they pinpoint the first evidence of herb usage for medicinal purposes. Traditional medicine originated independently in every corner of the world. This is a consequence of the verbal transfer of knowledge gathered from observations made after the consumption or appliance of substances derived from plant, animal or mineral source.

Until the 19th century the pharmaceutical era is known as the age of botanicals and it is characterized by the elaboration of drugs starting from plant extracts (Schmidt et al. 2008). Even though most of these antique therapies could not be validated by the modern definition of drug, this era provided the cornerstone for the understanding of the biological activities of substances found in nature.

The development of novel analytical chemistry techniques during the Industrial Revolution allowed the isolation and purification of plant secondary metabolites such as morphine by Setürner and papaverine by Merck (Drews 2000). Moreover, coal tar, a byproduct of early industrialization became of high interest for chemists as it contains a plethora of aromatic molecules like aniline, benzene, naphthalene. Bayer company succeed developing the first synthetic antipyretic Antifebrin (acetanilide) derived from aniline and afterward turned its focus on the hunt for new molecules found in this industrial waste product (Jones 2011).

During late 19th century and the first half of the 20th century, advances in the fields of biochemistry and chemistry led to the sophistication of the drug discovery process. The introduction of key concepts such as receptor by Langley and quantitative structure-property relationship (QSPR) by Crum-Brown started the transition to a more rational drug discovery process (Fraser, & Crum Brown 1868). This latter idea of QSPR was refined and adapted to the needs of medicinal chemists mainly due to the research of Corwin Hansch in early 60's. Nowadays, QSAR is part of the medicinal chemist toolbox for relating physical characteristics with biological activities for a set of molecules with high structural similarity (Hansch et al. 1962).

In 1981 the Fortune magazine cover highlighted "Next Industrial Revolution: Designing Drugs by Computer at Merck" (Van Drie 2007). This was the beginning of the interest in computer aided drug design (CADD). Parallel to this, the development of the first version of high throughput screening (HTS), an automated robotic system that tests a large number of compounds against potential targets was developed.

Later, in 1982 the molecular docking technique was developed in order to geometrically fit a ligand into the binding site of a protein aiming to simulate a protein-ligand complex (Kuntz et al. 1982). Docking can be divided in two independent stages: the fitting of ligand in the receptor binding site (pose generation) and the scoring of this pose based on the protein-ligand interactions. Docking scoring is a mathematical function which aims to predict the binding affinity of a small-molecule to its drug target. Existing scoring functions can be divided in four classes: force field based, empirical, knowledge-based and machine-learning (ML) based. The latter type performs better than classical scoring functions when predicting binding affinities (H. Li et al. 2015; Ain et al. 2015; Kundu et al. 2018; Fergus et al. 2019).

Moreover, the docking technique is implemented in the drug discovery pipeline mainly due to its capacity to virtually screen and rank up to millions of compounds from chemical libraries at a relatively low computational cost. The previous is particularly useful at the moment of hit finding, the process of identifying promising active molecules that can be further modified to generate lead molecules (Enyedy, & Egan 2008).

In order to achieve high efficiency, molecular docking technique applies many simplifications in its methodology. These restrictions involve mainly the neglection of protein and solvent motions during the binding process together with a poor description of H-bonding. As a result of these limitations, the docking method is computationally efficient but not optimal for accurately predicting binding affinities (Pantsar, & Poso 2018).

On the contrary, free energy calculation methods (i.e. thermodynamic integration, free energy perturbations) provide a more accurate but less efficient prediction towards binding energies of ligands (Reddy et al. 2014). However, due to their computational expense, as it is needed to perform short molecular dynamic simulations, its applicability is greatly reduced. Thus, nowadays there is an existing need for developing efficient methods that can outperform docking

scoring when predicting drug candidates.

Last year 2019, Nature journal published an article titled "Rethinking drug design in the artificial intelligence era" concluding the positive boost that new technologies will have in the drug discovery process. For example, ML is a branch of artificial intelligence where a computer could be programmed to predict physical and biological activities of a drug candidate (Schneider et al. 2019). Nonetheless, there are still limitations to machine learning (ML) applications in drug discovery as the experimental error present in public large data sets might affect the utility of build models. Also, the lack of standardized measurements due to the heterogeneity in the experimental procedures makes a challenge to develop high-input ML models with any application (Shepperd et al. 2019).

In this thesis, the need for developing efficient methods for identifying promising drug candidates is addressed. Furthermore, the classification performance of docking scoring and its sensitivity towards different protein conformation is evaluated. Mainly, as this technique is considered unreliable for its ability to differentiate between active and inactive molecules, and as its performance is sensitive to protein conformation (Enyedy, & Egan 2008; Pagadala et al. 2017). Hence, it is hypothesized that by using molecular and docking descriptors as inputs for ML models, it could possible to achieve scoring functions with an increased diagnosis performance when compared to commercial tools like extra precision (XP) Docking from Schrödinger Maestro software.

The degree of separation (between active and decoys) of XP Glide Score and ML regression models is assessed by calculating the area under the curve (AUC) of the receiver operating characteristics (ROC) curve. This ROC curve plots the sensitivity and specificity performance at several cut-off values obtained from the distribution of the given scores. The idea of ROC curves is to identify the cut-off point where a desired sensitivity and specificity is achieved, minimizing undesired failed predictions.

Interestingly, ROC curves were developed by British radar engineers during the World War II to differentiate between enemy aircraft and signal noise derived from birds (Lusted 1971). The diagnostic virtues of this plot are reflected by the diverse fields including psychology, meteorology and radiology that include it as an analysis tool for different methodologies (Swets 1973).

In summary, molecular docking techniques succeed when enriching huge data sets of chemicals with high efficiency, successfully generating reasonable binding poses and identifying key residues for ligand interaction (Pinzi, & Rastelli 2019). Nonetheless, several well-documented reports point out the fact that docking methods fail to predict accurately binding energies of ligands, generally due to poor scoring functions (Ramírez, & Caballero 2016; Ramírez, & Caballero 2018; Pagadala et al. 2017). The latter is addressed in this thesis, using ML regression aiming to improve the docking scoring in the context of two relevant protein kinase targets.

Kinases are considered a major drug target as protein phosphorylation regulates vast aspects of cell physiology and signalling (Cohen 2002). A correct balance between kinase and phosphatase activity is essential for a proper homeostasis. A considerable amount of types of cancers (Cicenas et al. 2018), neurodegenerative (Tenreiro et al. 2014; Gong et al. 2006), autoimmune and inflammatory diseases (Szilveszter et al. 2019) are due to an imbalance in the phosphorylation states, often due to deregulated kinase activity. In this thesis we focus on two kinases: Protein kinase A catalytic subunit α (PKACA) and Aurora kinase A (AURKA).

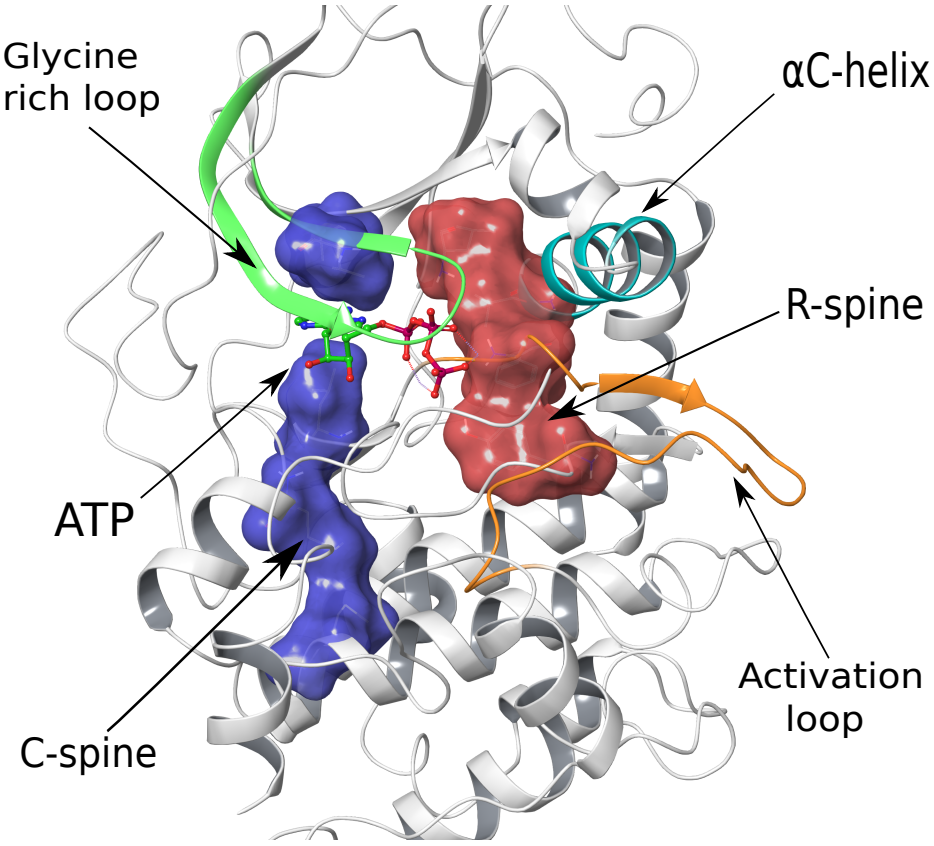

Figure 1: Kinase structural hallmarks

Kinases structure can be summarized in two major lobes (N-terminal and C-terminal) connected by a peptide hinge region where ATP binds. N-terminal lobe consists of five stranded β sheets, the αC-helix and the glycine rich loop (GRL). C-terminal lobe is comprised principally by six α-helices. After the hinge region, inside the C-terminal lobe is located the activation loop, containing a phosphorylation site and the highly conserved DFG motif, responsible of coordinating two magnesium ions and ATP's γ phosphate needed for kinase substrate phosphorylation (Modi, & Dunbrack 2019; Treiber, & Shah 2013; Zheng et al. 1993).

Moreover, the research team lead by Susan Taylor identified two conserved major structural features common to the kinase family. These kinase hallmarks are known as the regulatory R-spine and the catalytic C-spine containing hydrophobic residues of both major lobes from the kinases. In the one hand, R-spine consists of four residues including two aliphatic residues in β4-strand and αC-helix in the N-lobe, phenyalanine from the DFG motif and histidine residue from the HRD motif in the C-lobe. On the other hand, the C-spine is completed after ATP binding, allowing communication between the two major lobes enabling the closure of the binding pocket (A. Kornev et al. 2008; Taylor, & A. P. Kornev 2011).

Protein kinase A (PKA), a member of AGC kinase family, was firstly studied by pioneer scientists Edmond Fischer and Edwin Krebs for its role in glycogen metabolism. After reporting the existence of a cAMP dependent protein kinase which triggers glycogen phosphorylase activity these researchers were awarded with Nobel Prize in Physiology in 1992 (Turnham, & J. D. Scott 2016). Their studies concluded that PKA is involved in a plethora of signalling pathways dependent upon cellular compartment elevating highly the interest towards this enzyme (Sim, & J. Scott 1999; Carnegie et al. 2009).

Further research performed by Edwin Krebs and Susan Taylor was aimed to identify the stoichiometric ratio and the proteins partners involved in the PKA complex. It was found that PKA exists as a tetramer between a pair of inactive catalytic subunits (C) and a pair regulatory subunits (R). Moreover, upon the increase cAMP concentration, two molecules of this second messenger bind into each R subunit, leading to the release of the of the C subunits (Ringheim, & Taylor 1990; Rannels, & Corbin 1980). Later on, it was understood that at the same time, PKA is found docked to cellular organelles through A-kinase anchoring protein (AKAP).

Once the C subunits are released from the PKACA holoenzyme complex, they can be activated by phosphorylation in the activation loop by upstream PDK1. As a consequence of this activation and upon ATP binding, a conserved glutamate (Glu) within the αC-helix repositions to form a salt bridge with the ceiling lysine (Lys) in the β3-strand from the N-lobe (Kannan et al. 2008). While this lysine coordinates the α and β phosphates from ATP, the Asp from the DFG motifs aids in the γ phosphate transfer to downstream protein substrates (Johnson et al. 2001).

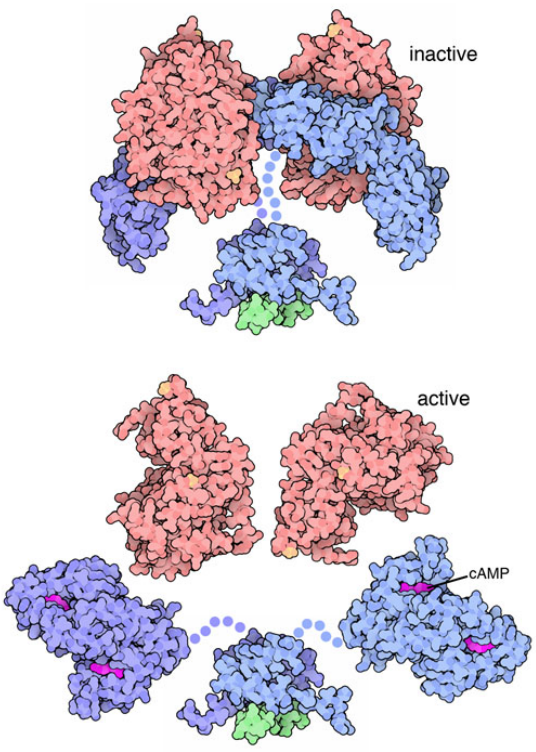

Figure 2: Inactive (top) and active (bottom) forms of PKACA (pink) (RCSB PDB 2012)

Interestingly, this phosphotransfer mechanism is regulated by a small 20 amino acid termostable peptide (PKI) which induces the recruitment of PKACA to the nuclear export complex and inhibits protein substrates from binding into PKACA. Moreover, energy landscape studies lead by Susan Taylor point out that PKI binding has a greater contribution towards complex stabilization compared to ATP (Walker et al. 2019)). In the line of the previous, it is believed that similar to the regulatory subunit of PKA, PKI is involved modulating the catalytic subunit effect by influencing its cellular localization and disrupting is kinase activity (Dalton, & Dewey 2006).

Regarding the second target of study, AURKA plays a major physiological role in the regulation of centrosome maturation, spindle formation (Fu et al. 2007; Lukasiewicz, & Lingle 2009) and aiding in the cellular migration process (Nikonova et al. 2013). Importantly, AURKA is the center of an array of oncological studies as it is found overexpressed in a wide range of tumors (Tsuchiya et al. 2020). Moreover, AURKA is activated by two means: phosphorylation in the activation loop or microtubule associated TPX2 protein binding. This latter mechanism, induces the localization of AURKA towards the spindle microtubules (Bayliss, Sardon, et al. 2003).

AURKA kinase, another member of the AGC family, has low enzymatic activity but upon phoshorylation of Thr288 member of the activation loop, this catalysis rate is bolstered. Moreover, it is known that activation via TPX2 binding induces AURKA autophosphorylation and as well protects from dephosphorylation by PP1 phosphatase. Studies point out that the effect of these two activation steps can combine forming an abnormally active kinase (Bayliss, Burgess, et al. 2017).

Furthermore, by comparing many unbound and bound structures of AURKA, several structural differences can be found. For instance, CD532-bound AURKA (PDB ID: 4J8M) exhibits the activation loop in the inactive conformation consistent with the fact that the urea moiety of CD532 ligand induces the inactive "in" conformation of the DFG motif for AURKA. Additionally, inhibitor bound crystal structures display structural rearrangements when compared to unbound structures. These structural changes include the tilting of the N-lobe in reference to C-lobe, the increase of glycine rich loop (GRL) angle and an outward motion of the αC-helix relative to the ATP binding site (Gustafson et al. 2014).

There is a growing need to provide the pharmaceutical industry and academia with improved computational methods to accelerate the drug discovery pipeline. Popular tools like molecular docking are computationally efficient but its scoring are unable to accurately predict binding affinities for drug candidates. Here, the previous is addressed by developing a ML based scoring function, specifically evaluating the classification performance (between active and inactive molecules) via ROC curves and comparing it to XP Glide Score from Schrödinger software. Also, the regression performance of the training and test procedures part of the ML are assessed with typical metrics such as R^2 and RMSE.

Materials and Methods

Compilation of the protein/ligand sets

Crystal structures from representative biological conformations for PKACA and AURKA were retrieved from the RCSB PDB (Berman 2000) with the following IDs: (PKACA: 1J3H, 1BKX, 4UJB, 3L9L, 3FJQ, 1YDR; AURKA: 4J8N, 4JBQ, 1OL5, 1OL7, 4C3P, 4C3R and 3H10). Known ligand and decoy sets for each target were downloaded from ChEMBL database (Bento et al. 2013), filtering was based on the respective Uniprot entry code for each target and only ligands with a stored Ki value were selected. Both protein and ligand sets were downloaded on December 2019.

Ligand preparation

We initially obtained the ligands as a list of SMILES (simplified molecular-input line-entry system) which consist of a one-dimensional array of strings without any 3D topology information for a molecule. We processed the SMILES using open-source version of OpenBabel 3.0.0 (O'Boyle et al. 2011) having as an output 2D structures in .sdf format. As a final step, Ligprep from Schrödinger 2019-2 release was configured to generate possible 3D tautomers and protomers at pH 7.0 ± 2.0 using OPLS3e force field and Epik. For each ligand the most likely conformer was chosen.

Protein preparation

The chosen PDB crystal structures were processed with the Protein Preparation Wizard (Sastry et al. 2013) from the Schrödinger 2019-2 release. Bond orders were adjusted for both co-crystallized ligands and amino acid residues using CCD database. Hydrogen atoms were added followed by the deletion of water molecules beyond 5 Å of any het group including ions. Protonation states for co-crystallized ligands, co-factors and metals was performed with Epik at pH 7.0 ± 2.0. Sampling of hydrogen-bond network was carried out with PROPKA at pH 7.0. Finally, in order to refine the geometry of the protein models they were subjected to a restrained minimization using OPLS3e force field with a stopping threshold set to a convergence

of heavy atoms of 0.3 Å between successive iterations. All water molecules, ions and metals were deleted.

Receptor grid generation

Grid generation for docking was carried on with the receptor grid generation tool from Glide package using Schrödinger 2019-2 release. For the grid generation, OPLS3e force field was chosen, Van der Waals radius scaling default settings were maintained. For holo-structures we defined the receptor box by picking the co-crystallized ligand and for apo structures annotated residues (5 Å from the co-crystalized ligand) from holo structures were used.

Molecular docking

Extra Precision (XP) docking (Friesner, Murphy, et al. 2006) was performed with Glide package (Friesner, Banks, et al. 2004) using Schrödinger 2019-2 release. Default settings for scaling of the van der Waals radii of non-polar ligand moieties were kept. Expanded sampling was chosen for the selection of initial poses prior to Glide screening and the number of poses per ligand to include for post-docking minimization was set to 20. Docking was performed with OPLS3e force field writing per-residue interaction scores for amino acids 12 Å away of the grid centre.

Building the machine learning models

The ML models used molecular descriptors obtained with the chemoinformatics toolkit RDKit (https://www.rdkit.org/, accessed March 2020), ChEMBL (Gaulton et al. 2011) and docking features from Maestro software. Thus, the descriptors used herein are based on topological, physical and chemical properties of the data set molecules and the predicted ligand-protein complex by docking XP.

We implemented random forest regression and ridge regression models part of the Python ML library scikit-learn (Pedregosa et al. 2011), also a neural network model built with TensorFlow library (Abadi et al. 2016). The models were trained to predict the pKi of each ligand estimated as the -$\log_{10} Ki$. Ligands were classified as active using a threshold of 1 μM based on their

inhibition (Ki) values, molecules not satisfying this threshold were classified as decoys. Stratified training and test set splitting was performed to achieve same proportion of inhibitors and decoys in training and testing sets.

Unbiased estimates of the regression performance of the models were gathered by using k-fold cross validation (k=10), were at each iteration one fold is used for validation while the remaining k-1 are used for training. Testing was performed on 20% of the data set. The previous process was repeated five times in order to generate a distribution of the regression metrics. The predictions of the models were evaluated with (R^2 and RMSE), maximum enrichment factor and their diagnostic performance was assessed by analysis of receiver operator characteristic (ROC) curves.

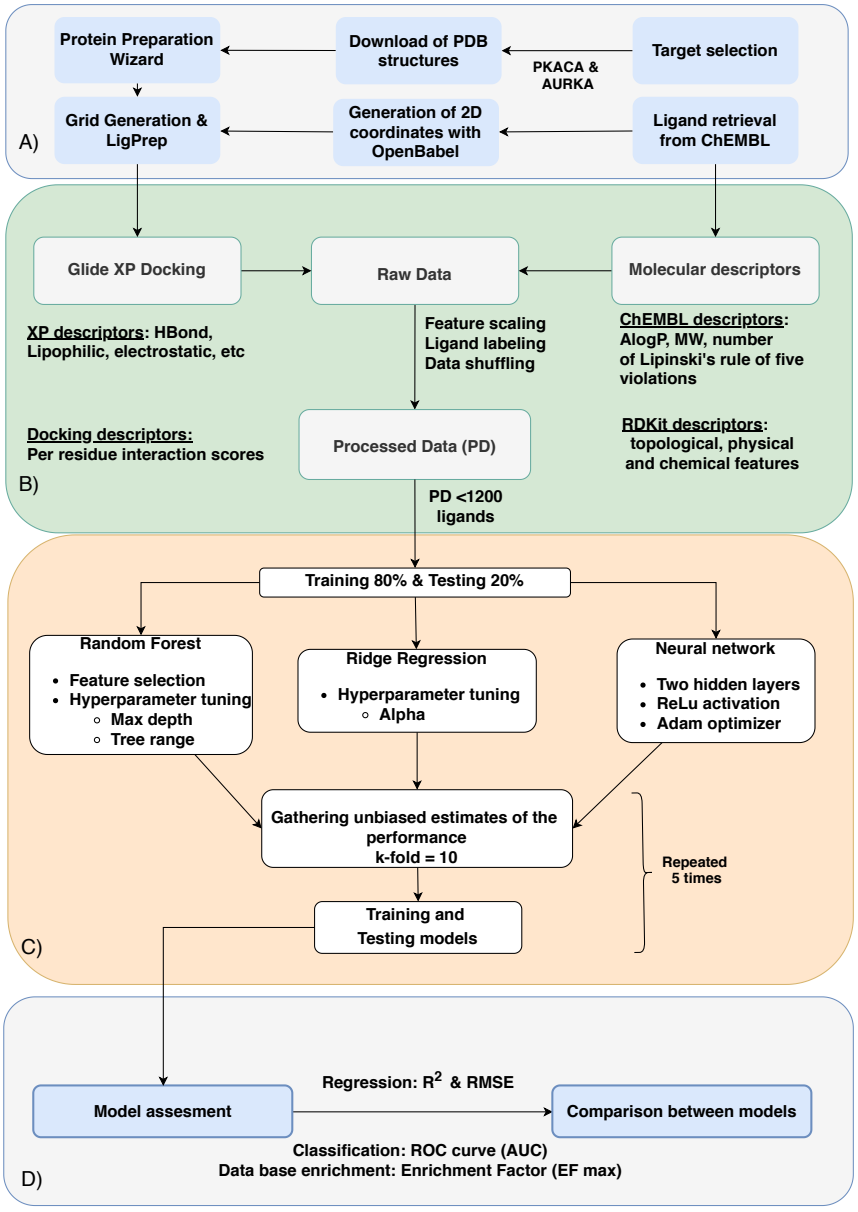

Figure 3: Master Thesis pipeline

Illustration of the pipeline implemented for the development of the ML models. A) Target selection, followed by PDB and ligand retrieval with their preparation steps prior to docking. B) Molecular docking using XP, subsequently the raw data is created by putting together the XP descriptors, interaction scores and the molecular descriptors obtained from ChEMBL and RDKit. The features are scaled, ligands are labeled and data is shuffled. C) The processed data is split in a stratified fashion so the same proportion of ligands and decoys is in training and test sets. Hyperparameter tuning and feature selection is done prior to gathering unbiased estimates of the performance with k-fold = 10. D) ML models are assessed by its regression, classification and database enrichment.

Results

Each ML model had as an input the ChEMBL and RDKit molecular features, the estimated XP descriptors and the per residue interaction scores for each ligand. This data was gathered for each docking at a specific conformation of either PKACA or AURKA. After splitting the data into training and testing sets, unbiased estimates of the regression performance were gathered with k-fold (k=10) in order to evaluate the performance of the model against unseen data. Subsequently, models were tested against 20% of the data set. R^2 and RMSE were the selected metrics for evaluating the regression performance and the diagnostic capability of the models was evaluated by ROC curves.

PKA-Cα and AURKA protein structure evaluation

Protein structures reflecting the ligand binding process or different physiological states were chosen for docking. Most protein kinases depict structural rearrangement upon substrate or inhibitor binding. To tackle this issue, we included bound and unbound kinase structures depicting evident differences at kinase structural hallmarks such as the G-rich loop (GRL), activation loop, DFG motif and αC-helix. For the chosen structures, G-rich loop and activation loop hallmarks are responsible for main structural dissimilarities in the active site of the chosen models.

Angle measurements and crystal structure features were gathered from KLIFS database (Kooistra et al. 2016) (Table 1). PKACA structures differ in the opening angle of the G-rich loop and are generally found at the conformationally active “DFG-in” state. AURKA models main differences relay on different phosphorylation states and the binding of TPX2 factor for spindle assembly (Zorba et al. 2014).

Protein Target	PDB ID	Complex type	Resolution (Å)	DFG conf. / αC-helix conf.	GRL angle	GRL rotation
PKACA	1J3H	Apo	2.9	in / in	70.1°	56.7°
	1BKX	Binary	2.6	in / in	56.1°	70.8°
	4UJB	Ternary	1.95	in / in	56.2°	70.3°
	3L9L	Ternary	2.0	in / in	57.2°	40°
	3FJQ	Ternary	1.6	in / in	48.4°	58°
	1YDR	Ternary	2.2	in / in	--	--
AURKA	4J8N	Apo	3.13	in / in	58.1°	75.5°
	4C3P	deP – Ternary	2.69	in / in	61.6°	75.8°
	4C3R	deP – Binary	2.79	out-like / in	59.6°	80.9°
	1OL5	P – Ternary	2.5	in / in	59.3°	80.8°
	1OL7	P – Binary	2.75	in / in	57.1°	79.2°
	3H10	Inhibitor bound	2.2	out-like / in	69°	76.2°

Table 1: Summary of crystal structure features

Regression and diagnostic performance of PKACA based ML models

Three representative stages of PKACA catalytic cycle, open (1J3H), intermediate (1BKX) and closed (4UJB, 3L9L, 1YDR, 3FJQ) conformations were used for molecular docking. 1157 ligands withdrawn from ChEMBL database were docked using extra precision (XP) docking to the conformations of interest.

Figures 4 and 5 reflect the performance of the ML models during five repetitions of the training and testing processes. The violin plots depict the probability density of the gathered unbiased estimates of the regression metrics in both training and test sets. It can be appreciated how the distribution of the regression metric values for the testing process lies within the training distribution values.

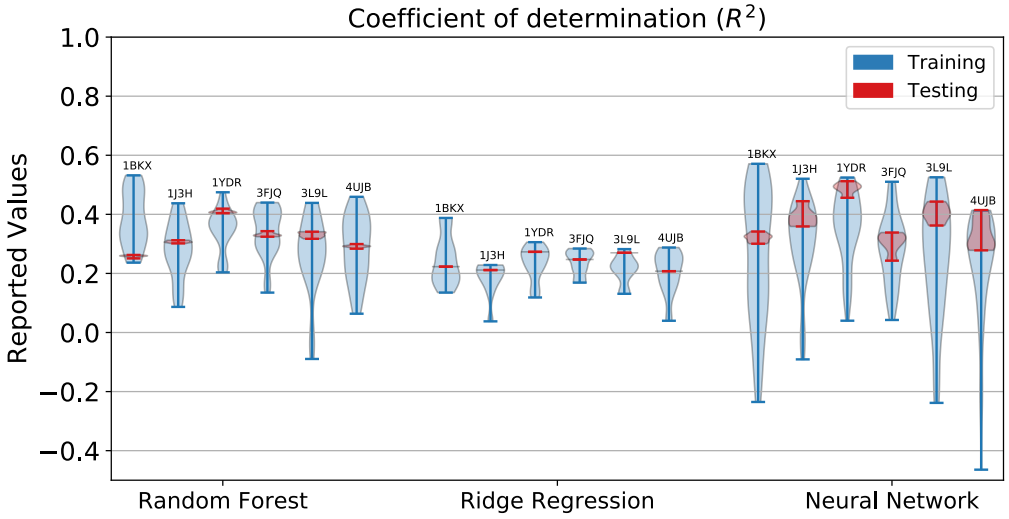

Figure 4: R^2 scores distribution for training and testing procedures for PKACA models

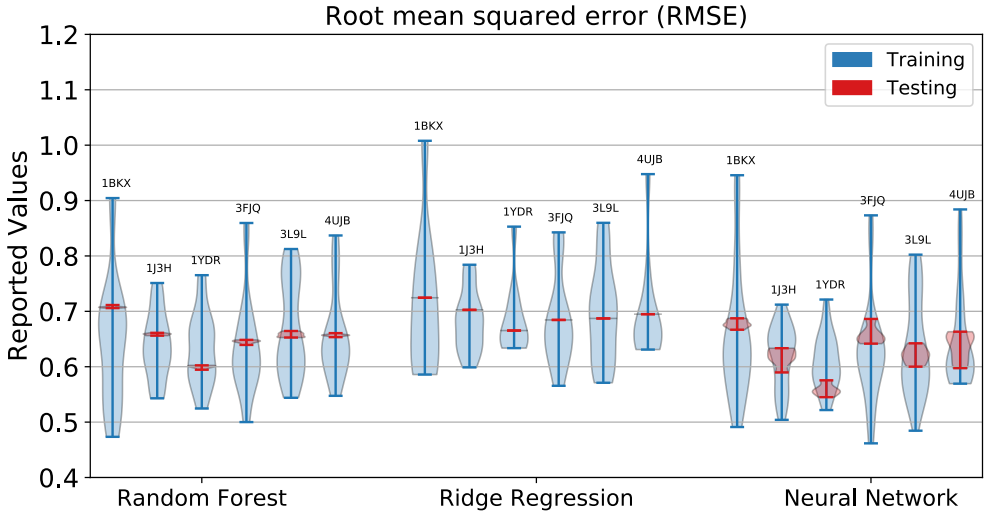

Figure 5: RMSE distributions for training and testing procedures for PKACA models

The predictions obtained with random forest and neural network models achieve the highest accuracy in terms of RMSE (Fig.4). Moreover, the neural network scores obtained the highest explanation of the variance found in the target pKi values in terms of R^2 (Fig. 5). Interestingly, for all three built models, the inhibitor bound conformation (PDB 1YDR) achieved a better performance compared to the other studied conformations.

PDB ID	Sensitivity				Specificity			
	XP	RF	RR	NN	XP	RF	RR	NN
1J3H	0.83	0.89	0.78	0.83	0.59	0.57	0.70	0.77
1BKX	0.57	0.80	0.5	0.86	0.74	0.67	0.9	0.64
1YDR	0.89	0.84	0.76	0.80	0.54	0.67	0.77	0.79
3FJQ	0.52	0.89	0.68	0.74	0.76	0.59	0.74	0.74
4UJB	0.53	0.91	0.78	0.77	0.66	0.43	0.64	0.70
3L9L	0.76	0.95	0.88	0.81	0.53	0.4	0.58	0.77
Average	**0.68**	**0.88**	**0.73**	**0.8**	**0.64**	**0.55**	**0.72**	**0.73**

Table 2: Sensitivity and Specificity for PKACA models

Table 2 reports the classification performance in terms of sensitivity and specificity at the chosen threshold for each ML model including XP scoring. The threshold for each model was selected using the Youden's J statistic in order that the sum between the true positive rate (TPR) and 1 - false positive rate (FPR) was maximized. That is, for all thresholds, the maximum value for $TPR - 1 + FPR$. Based on the selected thresholds, in average, for all the PKACA conformations the predictions of the ML models achieve an improvement of 20 and 9% in terms of sensitivity and specificity, compared to XP scoring.

PDB ID	Area under the curve (AUC)				Enrichment Factor (E_{max})			
	XP	RF	RR	NN	XP	RF	RR	NN
1J3H	0.72	0.77	0.81	0.85	3.65	4.8	4.8	4.8
1BKX	0.64	0.78	0.76	0.79	2.19	3.89	4.04	4.38
1YDR	0.69	0.79	0.79	0.86	4.67	4.67	4.67	4.67
3FJQ	0.63	0.77	0.77	0.78	2.13	4.62	4.62	4.62
4UJB	0.59	0.73	0.75	0.8	4.7	4.7	4.7	4.7
3L9L	0.64	0.71	0.77	0.84	3.83	4.60	4.6	4.6
Average	**0.65**	**0.76**	**0.77**	**0.82**	**3.53**	**4.55**	**4.57**	**4.63**

Table 3: AUC and Max enrichment factor for PKACA models

Table 3 reports the database enrichment performance of the scoring methods and the measure of separability in conformity with the AUC values. ML scoring functions slightly outperform XP scoring in regard to the enrichment of the docked ligands. This E_{max} does not represents the enrichment of the whole data set, but rather the testing set. Moreover, the neural network model, in average, is able to enhance up to 17% the classification of actives and decoys compared to XP scoring.

Docking at apoenzyme with open conformation (1J3H)

The apoform of PKACA represents the first step in the dynamic catalytic cycle of this enzyme. During this stage, the GRL angle and the distance between C-terminal and N-terminal lobes is the maximum. It is believed the latter can enable the entrance of the substrate and solvent to the active site (Akamine et al. 2003). Moreover, this unbound state conformation exhibits a preformed substrate binding site inherent of ligand binding process.

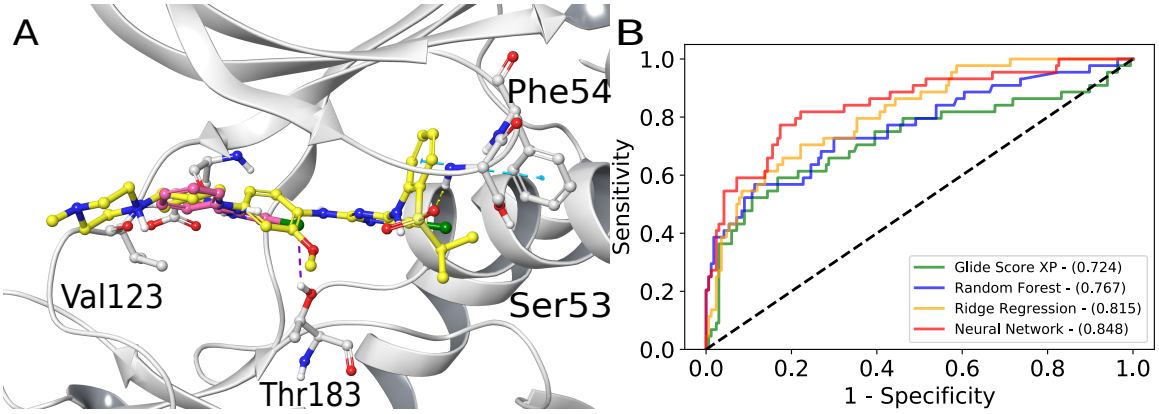

Figure 6: Example of docking poses and ROC curve for 1J3H model

A) Representative docking poses for active ligand CHEMBL509632 (yellow) and decoy CHEMBL2007138 (pink) can be seen in Fig. 6. Active ligand makes a pi-pi interaction with Phe54 (teal dashes) and decoy makes a halogen bond with Thr183. While the XP scores are -3.99 and -6.63, neural network is able to establish a score where the both molecules rank in accordance to their biological activity with 6.20 and 4.40, respectively. B) ROC curve for the assessed scoring methods, each colored line represents a different scoring function. Moreover, it can be seen how the ML models, specially the neural network and ridge regression, outperform XP docking in the classification of the data set. Numbers in the legend represent the AUC for each model.

Docking at intermediate conformation (1BKX)

1BKX is the first crystal structure of the catalytic subunit of PKACA in absence of PKI binding. Moreover, after comparing intermediate state bound to AMP (1BKX) and closed conformation bound to ATP and PKI (3FJQ) an 8° decrease in the GRL angle is evidence of the conformational changes undergone prior to catalysis of the transferable γ phosphate of ATP. This conformation reflects the transitory intermediary stage in the catalytic cycle of PKACA between closed and open conformations (Narayana et al. 1997).

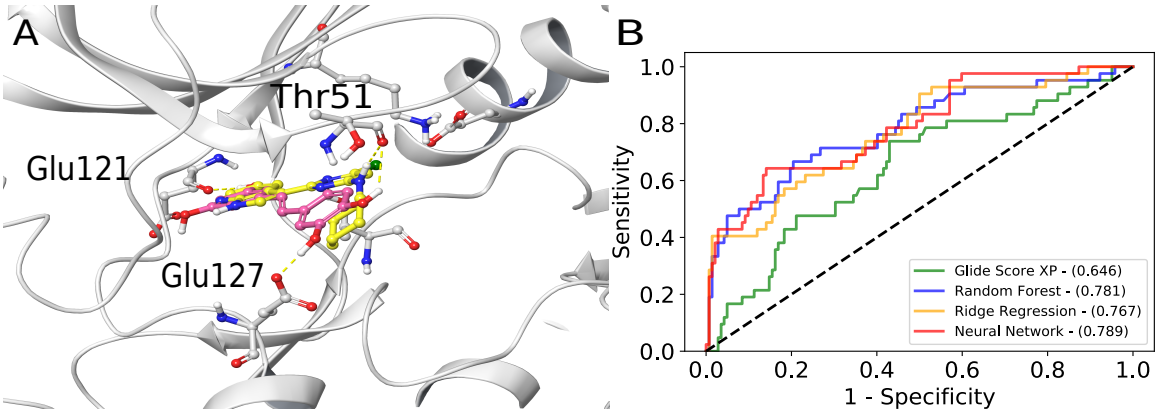

Figure 7: Example of docking poses and ROC curve for 1BKX model

A) Fig. 7 depicts the docking poses for active ligand CHEMBL1997846 (yellow) and decoy CHEMBL1975208 (pink). The shown poses of both molecules share a hydrogen bond with Thr51 from the GRL, besides additional hydrogen bonds with Glu121 and Glu127 for the decoy and Thr183 for the active ligand (yellow dashes). While XP scores -4.88 for active and -8.04 for decoy, the neural network scores 6.84 for active and 4.95 for decoy. B) ROC curve for the evaluated methods, each colored line represents a different scoring function. Numbers in the legend represent the AUC for each model.

Docking at closed conformation (4UJB, 3L9L, 1YDR and 3FJQ)

Closed conformations can be further divided into ATP-induced and inhibitor-induced closed states. Interestingly, upon nucleotide binding into PKACA (3FJQ), the distance between the major and minor lobe shrinks which can be appreciated as a decrease in the GRL angle, as reported in Table 1. Also, as a result of the rearrangement of the lobes, the GRL buries into the active site in order to relocate the non-transferable α and β phosphates (Lauber et al. 2016; Terasawa et al. 2006).

Moreover, both ATP and inhibitor induced conformations are characterized by the binding of the PKI peptide which acts as a pseudo-substrate that modulates PKACA catalytic activity, prevents the access of substrates and induces the nuclear transport of the protein (Taylor, Ilouz, et al. 2012).

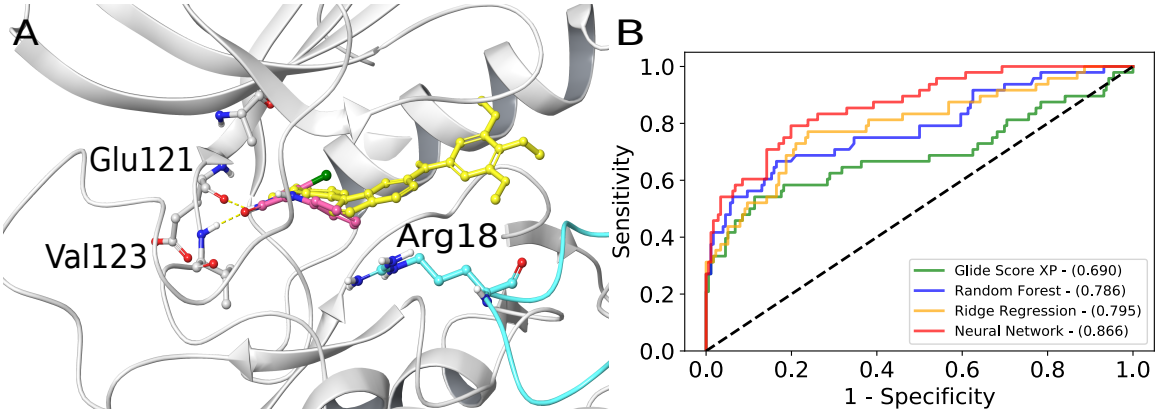

Figure 8: Example of docking poses and ROC curve for 1YDR model

A) Docking poses for active ligand CHEMBL1975927 (yellow) and decoy CHEMBL2007138 (pink) can be appreciated in Fig. 8. Both ligand poses share a hydrogen bond with either Glu121 or Val123 from the hinge region. Also, it can be appreciated how pseudo-substrate PKI invades the cleft of PKACA modifying the chemical environment in the binding site and restricting the entrance towards the catalytic site. While XP scores -4.97 for active and -7.9 for decoy, the neural network scores 6.0 for active and 4.7 for decoy. Numbers in the legend represent the AUC for each model.

Regression and diagnostic performance of AURKA based ML models

Multiple crystal structures of AURKA depicting four physiological states were selected for docking. Apoenzyme (4J8N), dephosphorilated (4C3P, 4C3R), phosphorilated (1OL5, 1OL7) and inhibitor bound (3H10) were docked with 1129 ligands withdrawn from ChEMBL database using extra precision (XP) docking.

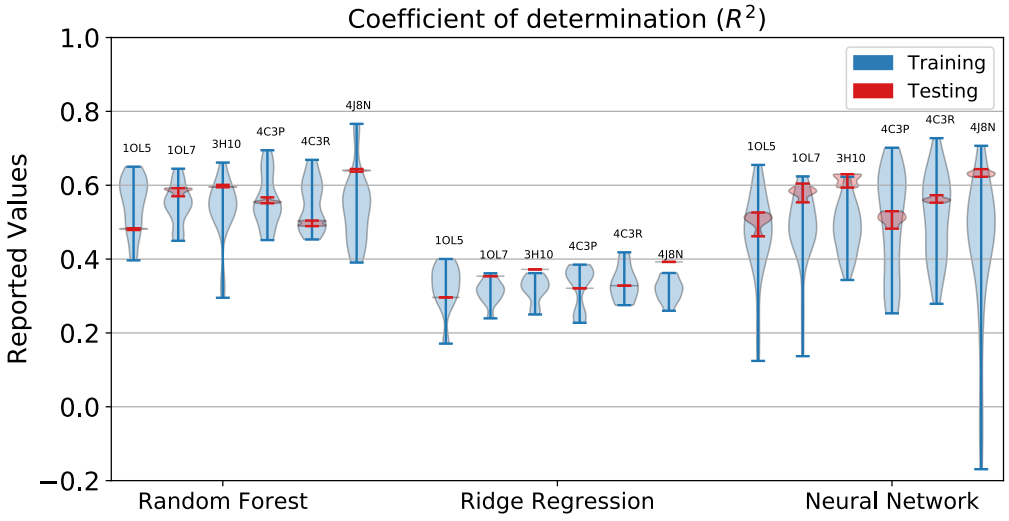

Figure 9: R^2 score for train and test procedures for AURKA models

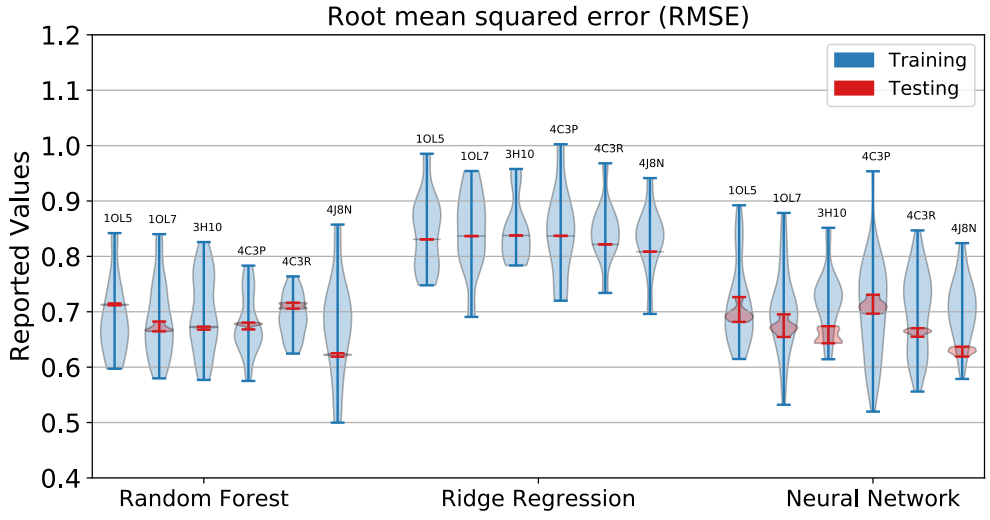

Figure 10: RMSE for training and test procedures for AURKA models

Figures 9 and 10 reflect the performance of the ML models during five repetitions of the training and testing processes for all the selected AURKA conformations. Interestingly, it can be appreciated in Figs. 9 and 10, how models built with the output data from the docking at apoform (4J8N) and inhibitor bound (3H10) conformations, achieve a higher regression performance with smaller errors in term of RMSE.

PDB ID	Sensitivity				Specificity			
	XP	RF	RR	NN	XP	RF	RR	NN
4J8N	0.99	0.81	0.90	0.86	0.05	0.76	0.65	0.77
4C3P	0.96	0.85	0.76	0.77	0.13	0.63	0.65	0.72
4C3R	0.98	0.85	0.86	0.73	0.08	0.66	0.63	0.79
1OL5	0.06	0.81	0.81	0.83	0.98	0.69	0.68	0.8
1OL7	0.30	0.81	0.74	0.85	0.74	0.73	0.76	0.71
3H10	0.45	0.91	0.85	0.86	0.64	0.64	0.71	0.74
Average	**0.62**	**0.84**	**0.82**	**0.82**	**0.44**	**0.69**	**0.68**	**0.76**

Table 4: Sensitivity and Specificity for AURKA models

The diagnostic performance in terms of sensitivity and specificity at the chosen threshold using Youden's J statistic for each ML model including XP docking is reported in Table 4. In average, for all the selected AURKA conformations, the ML models achieved an enhancement of up to 22 and 32% in terms of sensitivity and specificity, respectively, when compared to Maestro's XP score.

PDB ID	Area under the curve (AUC)				Enrichment Factor (E_{max})			
	XP	RF	RR	NN	XP	RF	RR	NN
4J8N	0.44	0.82	0.81	0.86	2.76	2.76	2.76	2.76
4C3P	0.5	0.78	0.74	0.78	2.57	2.57	2.57	2.57
4C3R	0.49	0.77	0.77	0.83	2.57	2.57	2.57	2.57
1OL5	0.49	0.80	0.80	0.86	2.61	2.61	2.32	2.61
1OL7	0.48	0.84	0.80	0.85	2.61	2.24	2.61	2.61
3H10	0.53	0.83	0.83	0.83	2.56	2.56	2.56	2.56
Average	**0.49**	**0.81**	**0.79**	**0.84**	**2.61**	**2.55**	**2.57**	**2.61**

Table 5: AUC and Max enrichment factor for AURKA models

Table 5 reports the classification performance based on the AUC of the ROC curve and the database enrichment efficiency for the testing set for the evaluated scoring methods. Interestingly, in average for all the AURKA conformations, XP scoring achieves no improvement in the classification of ligands compared to chance. In contrast, ML models reach at least 29% improvement compared to chance in terms of ROC AUC. Moreover, all the scoring methods achieve similar enrichment properties for the ligands in the testing set.

Docking at AURKA apoenzyme (4J8N)

Upon ligand binding AURKA undergoes major structural rearrangement at kinase structural hallmarks such as the activation loop, GRL and αC-helix. For instance, when comparing AURKA apoenzyme (4J8N) to bound state with inhibitor CD532 (4J8M), the activation loop is twisted 180° away from the active site and the GRL and αC-helix are packed towards the binding pocket (Gustafson et al. 2014).

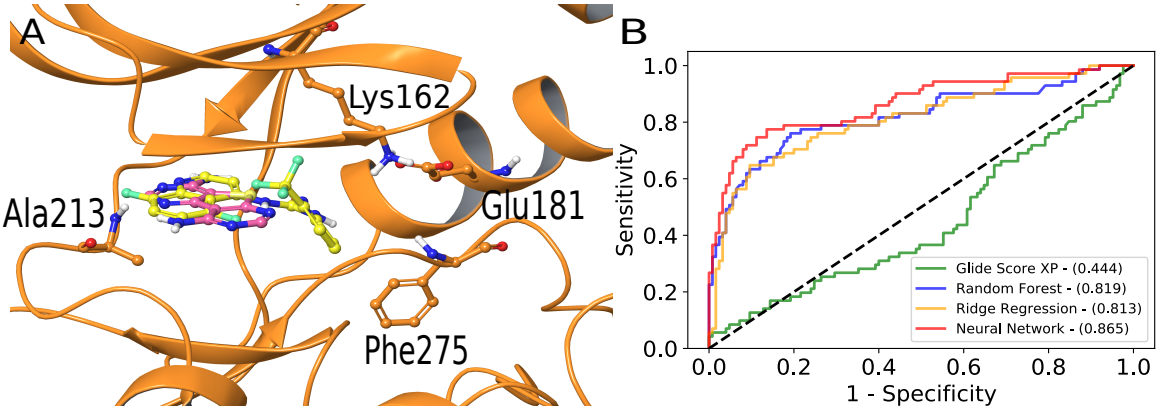

Figure 11: Example of docking poses and ROC curve for 4J8N model

A) Fig. 11 depicts the docking poses for active ligand CHEMBL2001987 (yellow) and decoy CHEMBL1976090 (pink). The conserved salt bridge between Lys162 and αC-helix Glu181 can be appreciated together with the Phe275 from DFG motif tilted outwards from the binding site. While XP scores with -4.95 (ac

random forest scores are 6.04 (active) and 5.02 (decoy), XP scores are -6.15 (active) and -7.98 (decoy). B) ROC curve for the evaluated scoring functions, each colored line represent a different scoring function. Docking scoring for this particular conformation is slightly better than chance diagnosing the ligands. Numbers in the legend represents the AUC for each model.

Docking at dephosphorilated AURKA (4C3R and 4C3P)

Catalytic activation of AURKA can be achieved in two ways, either TPX2 binding or adding a phosphate group to the activation loop. Additionally, TPX2 binding protects from PP1 phosphatase activity. Thus, 4C3R represents the inactive version of AURKA, as it is unphosphorilated and in abscence of TPX2. This state of AURKA is characterized for the insertion of the Phe275 between the salt bridge formed by Glu181 and Lys162, this leads to a distorted active site with the αC-helix leaning out from the active site (Bayliss, Sardon, et al. 2003).

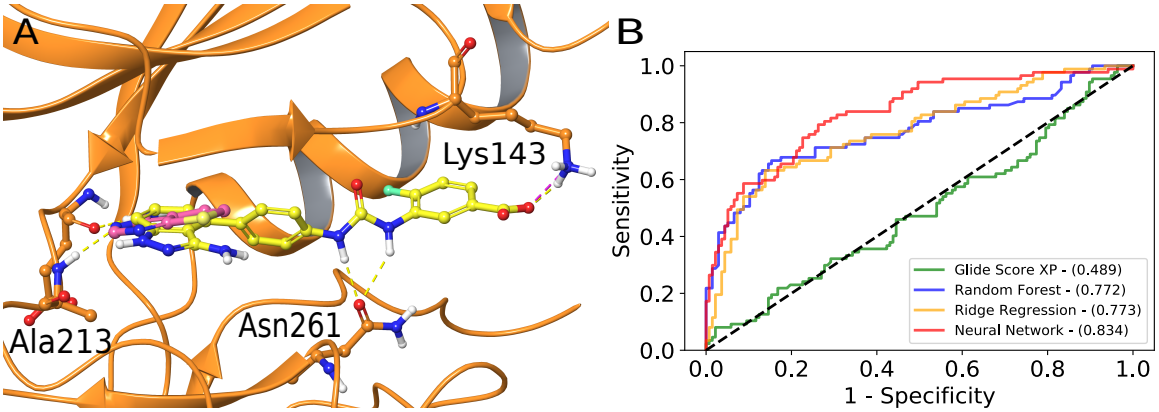

Figure 12: Example of docking poses and ROC curve for 4C3R model

A) Docking poses for active ligand CHEMBL1993722 (yellow) and decoy CHEMBL1972934 (pink) (Fig. 12). Both example poses have a hydrogen bond with either Glu211 and Ala 213 from the hinge region. Moreover, active CHEMBL1777842 also donates two hydrogens from its urea moiety towards Asn261. While XP scores are -5.43 for active and -6.82 for decoy, neural network scores with 6.51 and 3.74, respectively. B) ROC curve for the assessed scoring methods, each colored line represents a different scoring function. Numbers in the legend represent the AUC for each model.

Docking at phosphorilated AURKA (1OL7 and 1OL5)

Similar to other kinases, AURKA undergoes autophosphorilation in Thr288 of the activation loop in order to compensate its low catalytic activity rate. The mechanism of this process is still not clear. It is not understood whether one single AURKA, adds a phosphate to its own activation loop or whether AURKA gets activated by a second one. Interestingly, phosphorilated and TPX2 bound AURKA (1OL5) has a conformational state similar to PKACA (RMSD 1.3Å) with conserved active site residues oriented for catalysis. This residues, involved in coordination of the ATP phosphates include PKACA's positively charged Lys72, DFG motif Asp274 and Asp166 (Bayliss, Sardon, et al. 2003).

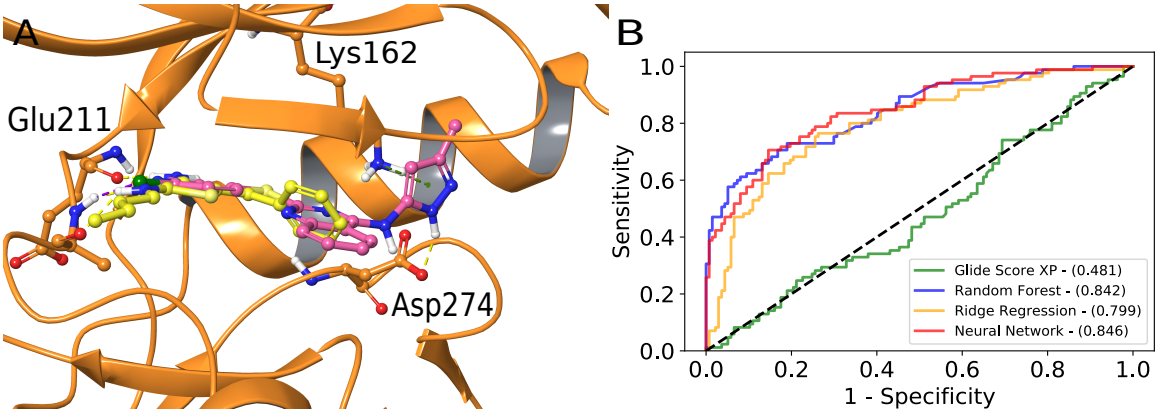

Figure 13: Example of docking poses and ROC curve for 1OL7 model

A) Representative docking poses for active ligand CHEMBL562157 (yellow) and decoy CHEMBL2007479 (pink). Both active and decoy ligands have a hydrogen bond with Ala213 or Glu 211 (yellow dashes). While XP scores are -5.3 for active and -7.07 for decoy, neural network model scores with 6.51 and 4.96, respectively. B) ROC curve for the evaluated scoring functions, each colored line represents a different scoring function. Additionally, neural network and random forest achieve similar AUC values. Numbers in the legend represent the AUC for each model.

Docking at inhibitor induced conformation (3H10)

In comparison to unbound AURKA state (4J8N), inhibitor bound conformation depicts major structural rearrangement induced after inhibitor binding. Mainly, it can be appreciated the transformation of the protein surface after 180° twist of the activation loop towards the catalytic site. Moreover, other bound AURKA structures are found with a 6Åtilt of the N-terminal domain in respect to the C-terminal domain (Gustafson et al. 2014; Aliagas-Martin et al. 2009). The conformational changes described previously are also translated to the GRL angle difference for 3H10 reported in Table 1.

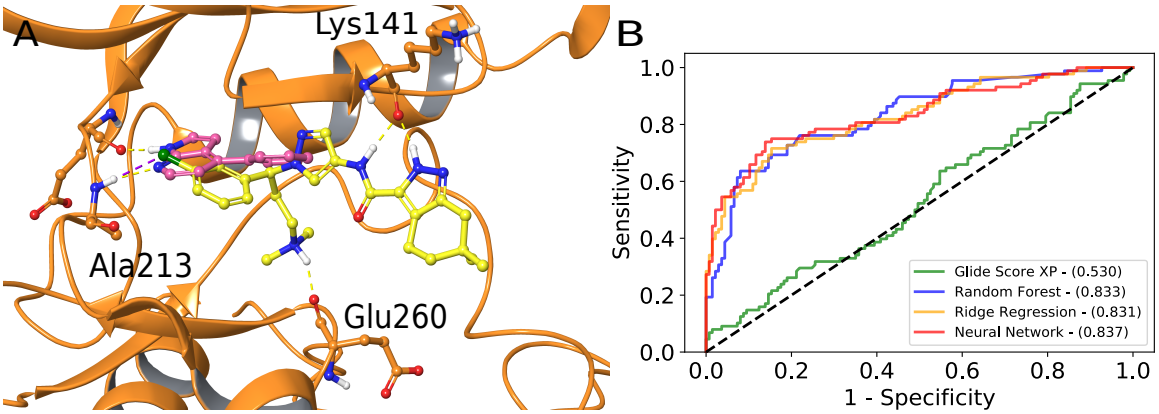

Figure 14: Example of docking poses and ROC curve for 3H10 model

A) Representative docking poses for active ligand CHEMBL3298376(yellow) and decoy CHEMBL482538 (pink) can be appreciated (Fig. 14). Decoy pose have two hydrogen bonds with Glu211 and Ala213 (dashed yellow lines). Active binding pose has a halogen bond with Ala213, donates two hydrogen bonds to Lys141 and one to Glu260 backbone carbonyl. XP scores are .6 and 4.87, respectively. B) ROC curve for the assessed scoring methods, each colored line represents a different scoring function. Additionally, neural networkand random forest achieve similar AUC values. Numbers in the legend represent the AUC for each model.

Discussion

This thesis is an attempt to address the need for developing efficient methods for identifying promising drug candidates. Herein, three ML models were developed aiming to predict the pKi of the molecules in the ChEMBL benchmark sets for PKACA and AURKA protein targets. In the line of the formulated hypothesis, the results support the idea that molecular and docking descriptors provide useful information which, when included in ML regression models lead to an improved diagnosis (whether active or inactive) of drug candidates.

Principally, the results indicate that for both protein targets and their different studied conformations, the ML scoring functions were able to establish a higher degree of separation between the scores of active and inactive molecules when compared to the XP score from Maestro software. For all the studied conformations of the two target kinases, an enhancement of at least 9% for PKACA and 30% of AURKA was achieved in terms of the AUC of the ROC curve. Surprisingly, in the case of AURKA, XP score was not better than chance classifying between active and decoy molecules based on their score (Table 5).

Additionally, the specificity and sensitivity of the scoring methods was assessed for the chosen thresholds according to the Youden's J statistic. For both protein targets, in average XP score displayed a decreased specificity and sensitivity when compared to the ML scoring functions (Tables 2 and 3). This finding indicates that for both protein targets of study, XP score had an increased number of both false positive and negatives when compared to the implemented ML models.

The E_{max} measures the maximum enrichment of active molecules upon the top scoring positions of a data set. As a consequence, this is dependent on the sensitivity of the scoring method and its capacity to accurately rank the binding affinities for the top scoring ligands. ML models slightly outperformed docking when enriching training set (200 ligands) based on the scores and showed similar E_{max} values for AURKA.

The previous results support the idea that ML scoring functions are better than XP score when classifying molecules as active or inactives. This statement goes in accordance to a study that analyzes docking scoring capacity for hit-to-lead optimization (Enyedy, & Egan 2008). This

study concludes that there are no significant differences observed when using XP score at the moment of differentiating between active and inactive molecules. Similarly, the results reported in this thesis support the same conclusion, specially for all the evaluated conformations of AURKA (Table 5). Furthermore, it is interesting the fact that the ML models trained with the output data of the docking at the inhibitor bound conformations reported the highest performance among all the studied conformations. This observation was found for both PKACA and AURKA.

Another study tried to assess the classification between active and decoys using a molecular dynamics and the evaluation of ligand-protein complex stability across the simulation. This study involving 56 different protein targets reported an improvement of 22% in terms of ROC AUC compared to AutoDock Vina score (Guterres, & Im 2020). In this thesis, although only two protein targets are evaluated, we reported up to 17 and 35% of improvement in the classification task compared to XP scoring for PKACA and AURKA, respectively. Nonetheless, despite having promising results, in order to achieve a stronger validation of the developed ML models herein, further analysis should be performed in a wider array of protein targets belonging to different families aiming to ensure a robust performance.

Among the conclusions of a comprehensive review on ML scoring functions, it is highlighted the fact ML scoring functions outperform a wide range of classical scoring functions when predicting binding affinities (Ain et al. 2015). For instance, ML scoring functions build upon 200 ligand-based molecular descriptors achieved up to 0.82 pearson correlation towards binding affinity prediction in the PDBbind 2016 core sets (Fergus et al. 2019).

In this thesis, the performance of random forest and neural network models during the testing in terms of R^2 lies within 0.2-0.65 (Figures 5 and 9) and for RMSE is distributed in the 0.7-0.55 range (Figures 4 and 10). This suggests a good agreement with the experimental pKi of the training set ligands obtained from the large ChEMBL database.

In the line with the previous, the increasing amount of chemical and biological data publicly available, together with the growing evidence that ML regression provides an accurate prediction of binding affinities and inhibition profiles of prospective drug candidates, fuels the existing optimism of improving the performance of ML regression models as long as the amount of

available data continues growing (Ain et al. 2015; H. Li et al. 2015). Nonetheless, the existing experimental error in the publicly available data, the lack of standardized pharmacological measurements and the tendency of ML models to over fit the training data establish a barrier to be surpassed for prospective ML models (Shepperd et al. 2019; Cao, & L. Li 2014).

Regarding over fitting, the implemented ML models performance was comparable between training and testing sets. Nonetheless, the fact that the violin plots displayed a wider distribution of the regression metrics compared to the testing, might be related k-fold splitting into folds with different proportions of active and decoy molecules.

Additionally, it is important to consider that most of the features used in the ML models built in this thesis were derived from the results from molecular docking (per residue interaction scores and XP descriptors) and molecular features obtained from RDKit. The latter, not only facilitates the possibility of reproducing the results here reported for third parties, but also, opens the door to implementing ML regression as an scoring post-processing alternative without loosing docking high efficiency and capacity for enriching chemical libraries. This would be advantageous in processes of hit finding and hit-to-lead in drug discovery projects.

Finally, XP docking poor performance in identifying active and inactives based on their score is mainly due to the simplifications in the method, introduced to make the process highly efficient, as stated by Pantsar et al. Docking scoring failure when accurately predicting ligand binding energies has been assessed in several studies and it is widely acknowledged (Ramírez, & Caballero 2016; Ramírez, & Caballero 2018; Pagadala et al. 2017).

However, the molecular docking tool is already providing both industry and scientist the possibility to virtually screen up to millions of compounds in early stages of the drug discovery pipeline. Also, it has the virtue of predicting the preferred geometries of a ligand and its moieties inside a given binding site. Fortunately, the emerging popularity of ML scoring functions together with the computationally expensive free energy calculation methods (i.e. thermodynamic integration and free energy perturbations) provide feasible alternatives to compensate docking inaccuracy, specially at the late stages of the drug discovery process.

Acknowledgment

Thanks to CSC for the computational resources, support and access to both Taito and Puhti clusters.

Thanks for to all my supervisors which have answered all my doubts, aid me when in trouble and helped me writing this thesis. Thanks to Tatu Pantsar for helping me with AURKA and Amirhossein Saktheman for the knowledge-transfer.

Thanks to my mom.

References

Abadi, M. et al. (2016). "TensorFlow: A system for large-scale machine learning". In: *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 265–283. URL: https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf.

Ain, N. Q. et al. (2015). Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 5.6, 405–424.

Akamine, P. et al. (2003). Dynamic features of cAMP-dependent protein kinase revealed by apoenzyme crystal structure. *Journal of Molecular Biology* 327.1, 159–171.

Aliagas-Martin, I. et al. (2009). A Class of 2, 4-Bisanilinopyrimidine Aurora A Inhibitors with Unusually High Selectivity against Aurora B†. *Journal of Medicinal Chemistry* 52.10, 3300–3307.

Bayliss, R., S. G. Burgess, and P. J. McIntyre (2017). Switching Aurora-A kinase on and off at an allosteric site. *The FEBS Journal* 284.18, 2947–2954.

Bayliss, R., T. Sardon, et al. (2003). Structural Basis of Aurora-A Activation by TPX2 at the Mitotic Spindle. *Molecular Cell* 12.4, 851–862.

Bento, A. P. et al. (2013). The ChEMBL bioactivity database: an update. *Nucleic Acids Research* 42.D1, D1083–D1090.

Berman, H. M. (Jan. 2000). The Protein Data Bank. *Nucleic Acids Research* 28.1, 235–242.

Cao, Y. and L. Li (2014). Improved protein-ligand binding affinity prediction by using a curvature-dependent surface-area model. *Bioinformatics* 30.12, 1674–1680.

Carnegie, G., C. Means, and J. Scott (2009). A-kinase anchoring proteins: From protein complexes to physiology and disease. *IUBMB Life* 61.4, 394–406.

Cicenas, J. et al. (2018). Kinases and cancer. *Cancers* 10.3, 1–7.

Cohen, P. (2002). Protein kinases - The major drug targets of the twenty-first century? *Nature Reviews Drug Discovery* 1.4, 309–315.

Dalton, G. D. and W. L. Dewey (2006). Protein kinase inhibitor peptide (PKI): A family of endogenous neuropeptides that modulate neuronal cAMP-dependent protein kinase function. *Neuropeptides* 40.1, 23–34.

Drews, J. (2000). Drug Discovery : A Historical Perspective. *Science* 287.March, 1960–1965.

Enyedy, I. J. and W. J. Egan (2008). Can we use docking and scoring for hit-to-lead optimization ? *J Comput Aided Mol Des.* January, 161–168.

Fergus, B., D. Charlotte M, and M. M. Garrett (Aug. 2019). Learning from the ligand: using ligand-based features to improve binding affinity prediction. *Bioinformatics*. Ed. by A. Elofsson.

Fraser, T. and A. Crum Brown (1868). On the Connection between Chemical Constitution and Physiological Action; with special reference to the Physiological Action of the Salts of the Ammonium Bases derived from Strychnia, Brucia, Thebaia, Codeia, Morphia, and Nicotia. *J Anat Physiol.* 2.2, 224–242.

Friesner, R., J. Banks, et al. (2004). Glide : A New Approach for Rapid , Accurate Docking and Scoring . *Journal of Medicinal Chemistry*, 1739–1749.

Friesner, R., R. B. Murphy, et al. (2006). Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein - Ligand Complexes. *Journal of Medicinal Chemistry*, 6177–6196.

Fu, J. et al. (2007). Roles of aurora kinases in mitosis and tumorigenesis. *Molecular Cancer Research* 5.1, 1–10.

Gaulton, A. et al. (Sept. 2011). ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Research* 40.D1, D1100–D1107.

Gong, C. X. et al. (2006). Dysregulation of protein phosphorylation/dephosphorylation in Alzheimer's disease: A therapeutic target. *Journal of Biomedicine and Biotechnology* 2006, 1–11.

Gustafson, W. et al. (Sept. 2014). Drugging MYCN through an Allosteric Transition in Aurora Kinase A. *Cancer Cell* 26.3, 414–427.

Guterres, H. and W. Im (Apr. 2020). Improving Protein-Ligand Docking Results with High-Throughput Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling*. URL: https://doi.org/10.1021/acs.jcim.0c00057.

Hansch, C. et al. (Apr. 1962). Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients. *Nature* 194.4824, 178–180.

Johnson, D. A. et al. (2001). Dynamics of cAMP-dependent protein kinase. *Chemical Reviews* 101.8, 2243–2270.

Jones, A. W. (2011). Early drug discovery and the rise of pharmaceutical chemistry. *Drug testing and analysis* April, 337–344.

Kannan, N. et al. (2008). The hallmark of AGC kinase functional divergence is its C-terminal tail, a cis-acting regulatory module. *PNAS - USA* 105.26, 9130.

Kooistra, A. J. et al. (2016). KLIFS: A structural kinase-ligand interaction database. *Nucleic Acids Research* 44.D1, D365–D371.

Kornev, A., S. Taylor, and L. Ten Eyck (2008). A helix scaffold for the assembly of active protein kinases. *Proceedings of the National Academy of Sciences* 105.38, 14377–14382.

Kundu, I., G. Paul, and R. Banerjee (2018). A machine learning approach towards the prediction of protein-ligand binding affinity based on fundamental molecular properties. *RSC Advances* 8.22, 12127–12137.

Kuntz, I. et al. (1982). A Geometric Approach to Macromolecule-Ligand Interactions. *Journal of Molecular Biology*.

Lauber, B. et al. (2016). Addressing the Glycine-Rich Loop of Protein Kinases by a Multi-Facetted Interaction Network: Inhibition of PKA and a PKB Mimic. *Chemistry A European Journal* 22.1, 211–221.

Li, H. et al. (2015). Improving AutoDock Vina Using Random Forest: The Growing Accuracy of Binding Affinity Prediction by the Effective Exploitation of Larger Data Sets. *Molecular Informatics* 34.2-3, 115–126.

Lietava, J. (1992). Medicinal plants in a Middle Paleolithic grave Shanidar IV? *Journal of Ethnopharmacology* 35, 263–266.

Lukasiewicz, K. and W. Lingle (2009). Aurora A, centrosome structure, and the centrosome cycle. *Environmental and Molecular Mutagenesis* 50.8, 602–619.

Lusted, L. B. (1971). Signal Detectability and Medical Decision-Making. *Science* 171.3977, 1217–1219.

Modi, V. and R. Dunbrack (2019). Defining a new nomenclature for the structures of active and inactive kinases. *Proceedings of the National Academy of Sciences* 116.14, 6818–6827.

Narayana, N. et al. (1997). A binary complex of the catalytic subunit of cAMP-dependent protein kinase and adenosine further defines conformational flexibility. *Structure* 5.7, 921–935.

Nikonova, A. S. et al. (2013). Aurora A kinase (AURKA) in normal and pathological cell division. *Cellular and Molecular Life Sciences* 70.4, 661–687.

O'Boyle, N. et al. (2011). Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, 1–14.

Pagadala, N. S., K. Syed, and J. Tuszynski (2017). Software for molecular docking: a review. *Biophysical Reviews* 9.2, 91–102.

Pantsar, T. and A. Poso (2018). Binding Affinity via Docking : Fact and Fiction. *Molecules*, 1–11.

Pedregosa, F. et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 28252830.

Pinzi, L. and G. Rastelli (2019). Molecular docking: Shifting paradigms in drug discovery. *International Journal of Molecular Sciences* 20.18.

Ramírez, D. and J. Caballero (2016). Is it reliable to use common molecular docking methods for comparing the binding affinities of enantiomer pairs for their protein target? *International Journal of Molecular Sciences* 17.4, 1–15.

– (2018). Is It Reliable to Take the Molecular Docking Top Scoring Position as the Best Solution without Considering Available Structural Data? *Molecules* 23.5, 1–17.

Rannels, S. and J. Corbin (1980). Studies of functional domains of the regulatory subunit from cAMP-dependent protein kinase isozyme I. *Journal of Cyclic Nucleotide Research* 6.3, 201–15.

RCSB PDB (2012). *cAMP-dependent Protein Kinase (PKA)*. [Online; accessed December, 2019]. URL: `https://pdb101.rcsb.org/motm/152`.

Reddy, M. et al. (2014). Free Energy Calculations to Estimate Ligand-Binding Affinities in Structure-Based Drug Design. *Current Pharmaceutical Design* 20.20, 3323–3337.

Ringheim, G. E. and S. S. Taylor (1990). Effects of cAMP-binding site mutations on intradomain cross-communication in the regulatory subunit of cAMP-dependent protein kinase I. *Journal of Biological Chemistry* 265.32.

Sastry, M. et al. (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design* 27.3, 221–234.

Schmidt, B. et al. (2008). A natural history of botanical therapeutics. *Metabolism* 57.

Schneider, P. et al. (2019). Rethinking drug design in the artificial intelligence era. *Nature Reviews Drug Discovery*.

Shepperd, M. et al. (2019). "The Prevalence of Errors in Machine Learning Experiments". In: *International Conference on Intelligent Data Engineering and Automated Learning*, 102–109. eprint: `1909.04436`.

Sim, A. and J. Scott (1999). Targeting of PKA, PKC and protein phosphatases to cellular microdomains. *Cell Calcium* 26.5, 209–217.

Swets, J. A. (1973). The relative operating characteristic in psychology. *Science* 182.4116, 990–1000.

Szilveszter, K. P., T. Németh, and A. Mócsai (2019). Tyrosine kinases in autoimmune and inflammatory skin diseases. *Frontiers in Immunology* 10.AUG, 1–21.

Taylor, S. S., R. Ilouz, et al. (2012). Assembly of allosteric macromolecular switches: Lessons from PKA. *Nature Reviews Molecular Cell Biology* 13.10, 646–658.

Taylor, S. S. and A. P. Kornev (2011). Protein kinases: evolution of dynamic regulatory proteins. *Trends in Biochemical Sciences* 36.2, 65–77.

Tenreiro, S., K. Eckermann, and T. F. Outeiro (2014). Protein phosphorylation in neurodegeneration: Friend or foe? *Frontiers in Molecular Neuroscience* 7, 1–30.

Terasawa, K. et al. (2006). Cdc37 Interacts with the Glycine-Rich Loop of Hsp90 Client Kinases. *Molecular and Cellular Biology* 26.9, 3378–3389.

Treiber, D. K. and N. P. Shah (2013). Ins and outs of kinase DFG motifs. *Chemistry and Biology* 20.6, 745–746.

Tsuchiya, Y. et al. (2020). Covalent Aurora A regulation by the metabolic integrator coenzyme A. *Redox Biology* 28, 101318.

Turnham, R. E. and J. D. Scott (2016). Protein kinase A catalytic subunit isoform PRKACA; History, function and physiology. *Gene* 577.2, 101–108.

Van Drie, J. H. (2007). Computer-aided drug design : the next 20 years. *Journal of Computer-Aided Molecular Design* November, 591–601.

Walker, C. et al. (2019). Cushing's syndrome driver mutation disrupts protein kinase A allosteric network, altering both regulation and substrate specificity. *Science Advances* 5.8, 1–12.

Zheng, J. et al. (1993). Crystal Structure of the Catalytic Subunit of cAMP-Dependent Protein Kinase Complexed with MgATP and Peptide Inhibitor. *Biochemistry* 32.9, 2154–2161.

Zorba, A. et al. (2014). Molecular mechanism of Aurora A kinase autophosphorylation and its allosteric activation by TPX2. *eLife* 3.