World Journal of Biology Pharmacy and Health Sciences
eISSN: 2582-5542
Cross Ref DOI: 10.30574/wjbphs
Journal homepage: https://wjbphs.com/

(RESEARCH ARTICLE)

Check for updates

Employing CHEMBL Data with AAK1 and BIKE proteins to validate 3D-QSAR models

Shreya Goswami *

Department of Pharmacy, Erasmus Research Intern, University of Eastern Finland, Kuopio, Northern Savonia, Finland

World Journal of Biology Pharmacy and Health Sciences, 2025, 24(01), 554-569

Publication history: Received on 18 September 2025; revised on 25 October 2025; accepted on 27 October 2025

Article DOI: https://doi.org/10.30574/wjbphs.2025.24.1.0924

Abstract

AAK1 (Adaptor Associated Kinase 1) and BMP2-K (BMP 2 inducible kinase) are serine threonine human kinases that have been shown to perform endocytosis and are associated with notch signalling pathways. Their inhibition forms the basis of potential drug targets. Using tools from Schrodinger Maestro, protein structures from PDB and molecules from Chembl were considered for investigation. The common core structure based on the 3-acylaminoindazole ring, found from the literature study, was used for clustering and ligand alignment docking. Docking poses generated were looked for having proton acceptor and donor bonds with the hinge region of the kinase protein. A QSAR (Quantitative Structure Activity Relationship) model was built for those ligands having good docking poses. Internal validation of Chembl compounds was grouped into a training and test set (80:20) for 3D- Field Based QSAR, providing a promising model for the external test set of molecules with unknown activities.

Keywords: AAK1; BIKE; Virtual Screening; Molecular Modelling; Docking; Field Based QSAR; Inhibitors; OPLS 4; Comfa; Comsia

1. Introduction

This research focuses on kinases. Understanding their structure, function, molecular structure, and ligand-protein binding was critical. A thorough literature review revealed it. Cellular enzymes, kinases are vital. They transfer phosphoryl groups from high-energy ATP to targets. Target phosphorylation activates or inactivates. Sorrell et al. (2016) Kinases Structure overall: Two lobes compose the protein kinase core. The N-lobe beta loop has five strands and the C loop has alpha helices and strands. ATP binding pocket constituents include conserved amino acid residues in the hinge area between N and C loops that catalyse kinases. Glycine-rich N LOBE and ATP-binding loop. The conserved P loop amino acid sequence is GXGXXG. Sorrell et al. (2016).

These varied kinases make good pharmacological targets. Liu et al. (2009) connected BIKE to myopia, and Shi et al. (2014) linked AAK1 to ALS. Targeting BIKE may treat HIV, while AAK1 inhibitors enhance Hepatitis C virus entry. 2017 Bekerman et al. Finding drug targets, identifying them, connecting them, and assessing efficacy in people are all part of drug discovery. (2011) Hughes et al. An accurate system simulation may replace physical experimentation. This work models molecules using molecular docking, protein and ligand alignment, and field-based QSAR. Classical to AI modelling methods have altered. 2020 Maia et al. Generally, Data is used to assess and quantify ligand-receptor interactions.

Aligning the co-crystallised ligand with the PDB (Berman 2000) structure shows similarities and differences. This experiment uses Schrodinger Suite, a unique, cutting-edge chemical simulation tool with a simple user interface for glide docking and molecular dynamics modelling (Dixon et al. 2006). Glide sorts input ligand docking poses using Emodel scoring based on flexible energy optimisation on OPLS (Shivakumar et al. 2010). By Monte Carlo confirmation, Meng et

* Corresponding author: Shreya Goswami

al. (2011) sort docking postures. In the final treatment, QSAR models projected biological property changes to structural adjustments.

Validation is crucial for assessing the reliability, robustness, and confidence of QSAR model predictions for a dataset (Stanton 2003a). In other words, validation indicates the model's predictive power (Veerasamy et al., 2011). Internal and external validation techniques are used depending on the reason. Figure 7 shows this project's internal validation of CHEMBL chemical QSAR models. Molecular activity predictions are involved. As cross-validation divides n substances into calibration (training) and validation (test) subsets, internal validation is crucial. These data points are not used in calibration because the validation subset determines how well the model predicts incoming data. Model building uses the training/calibration subset. Internal validation is an effective method for assessing D model quality and fit. Used Internal Validation Methods. Least squares fitting is the most used internal model validation method. This validation method, similar to linear regression, compares anticipated and experimental activity using the R^2 (squared correlation coefficient). When calculating R2, the robust straight-line fit is superior since it gives less weight to data points far from the centre (over a given standard deviation from the model). Another method is to remove outliers (compounds from the training set) from the dataset to improve the QSAR model; however, it is only reliable if strict statistical standards are followed. The most common internal model validation method is least squares fitting (Stanton 2003 b). This validation method, similar to linear regression, compares anticipated and experimental activity using the R^2 (squared correlation coefficient). When calculating R2, the robust straight-line fit is superior since it gives less weight to data points far from the centre (over a given standard deviation from the model). Another method is to remove outliers (compounds from the training set) from the dataset to improve the QSAR model; however, it is only reliable if strict statistical standards are followed.

This study addresses questions regarding Kinases in two phases. The AAKI and BMP2K involved in NOTCH signalling pathway upregulation are involved in neurodegenerative diseases like Alzheimer's and many types of cancer.

The initial phase involves analysing the binding of candidate docking poses in relation to the hinge region of the co-crystallised ligand. If the poses demonstrate a suitable fit within the binding cavity, a QSAR model is anticipated for the corresponding ligands. achieving important statistical metrics. Candidates that exhibit strong binding capabilities produce effective poses and can be further analysed to develop a statistically significant model, thus facilitating the identification of novel drug targets. In the second phase, the model's internal validation demonstrates its capability to predict a set of compounds with unknown activities.

2. Material and methodology

The drug research framework encompasses target selection, high-throughput identification, lead optimisation, and subsequent pre-clinical and clinical evaluations. Computational modelling is essential for lead optimisation and hit identification.

2.1. Schrödinger-Maestro

The structural analysis was conducted using the Schrodinger suite, obtained through a license from the Molecular Modelling and Drug Design group led by Dr Antti Poso. The work was conducted on a computer in the modelling laboratory of the pharmacy department at the institute. The Schrodinger suite comprises a collection of tools utilising the Maestro graphical user interface. The project utilised Maestro as the platform for visualization and analysis. The tools included in the suite encompass a wide range of assessments, from basic visualization of chemical structures to molecular dynamics simulations. Schrödinger Release 2023-2: Maestro. Schrödinger, LLC, New York, NY, 2023.

2.2. Structure-Based Drug Designing

Molecular docking, a computational approach, has been the foundation of structure-based virtual screening for over a decade. SVS is crucial to medication development. Virtual screening involves docking an archive of chemicals to a protein (receptor or enzyme) binding cavity and generating docking poses based on rank. A part of the compounds is then tested for hits (Kireev 2016). Docking is detailed in detail below. SVS works by selecting a few chemicals from a big data collection.

Thus, these compounds can be tested for high-throughput screening (HTS) (Kontoyianni 2017).

2.3. Molecular docking

Guessing how two molecules will fit and interact. Fitting two puzzle pieces together yields the whole image. Docking studies use protein molecules and proteins or drugs as pieces. The goal is to discover how the protein and ligands fit together based on molecular structure geometries and atom chemical properties. This method finds drug discovery and optimisation candidate ligands. The suite's 3D-QSAR panel examined docking poses (Cappel et al. 2015). Schrodinger Suite, based on Maestro GUI, includes tools. The project uses maestro for analysis and visualisation. The suite includes too many assessment tools, from chemical structure visualisation to molecular dynamics simulations. Halgren (2009) used the suite's field-based QSAR panel to evaluate docking positions. Glide docking has three precision modes depending on the results and mode. This study generated docking poses of input ligands using Standard Precision (SP) mode with trustworthy results. It employs a hierarchical model to filter poses to discover receptor binding site locations. An exhaustive list of ligand torsions ranks compounds by positional degrees of freedom relative to the protein grid. Glide docking has three precision modes depending on the results and mode. This study generated docking poses of input ligands using Standard Precision (SP) mode with trustworthy results. It employs a hierarchical model to filter poses to discover receptor binding site locations. An exhaustive list of ligand torsions ranks compounds by positional degrees of freedom relative to the protein grid.

Based on the virtual drug screening design, the first phase methodology (Figure 5) was chosen. Discussion with the supervisor optimised it to meet rigorous outcomes faster. The procedure examined protein structure and inhibitors. Further steps include downloading and preparing protein PDB structures and using a receptor grid panel with ligand constraints to generate new PDB coordinates. After protein preparation, the receptor grid is retained for docking for reproducible and stable findings. AAK1 and BIKE-related chemicals were downloaded from CHEMBL (Mendez et al., 2019). Uniprot database (Bateman, 2019) has many structures with different bound ligands for each protein; therefore, they had to be prepared, docked, and compared.

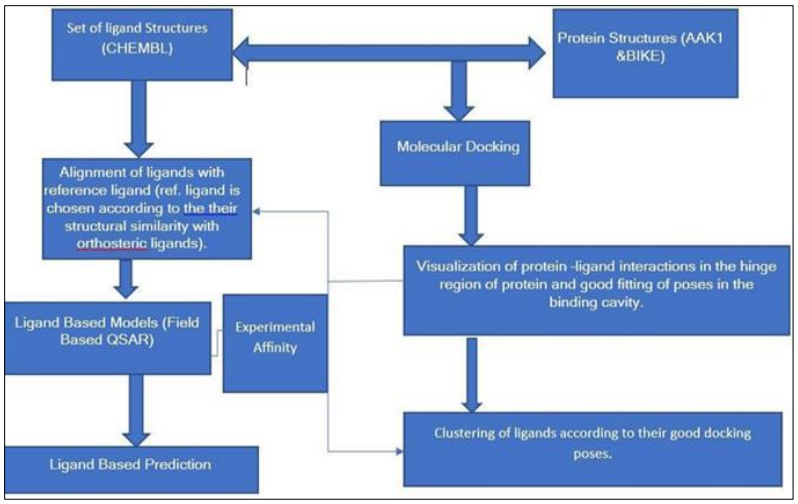


Figure 1 Structure-Based Drug Design research approach

The goal was to download the protein structure from the KLIFS database and use the Protein Preparation wizard to change PDB coordinates, remove superfluous water, and minimise hydrogen bonds for alignment and docking. SGC-AAK1 (CHEMBL ID: 4452939) (Wells et al. 2020) is aligned to certain PDB structures of AAK1 and BIKE to discover its key groups needed to bind efficiently to the protein inside the binding cavity. The 3-acylaminoindazole ring scaffold produces strong AAK1 and BIKE inhibitors (Bekerman et al. 2017).

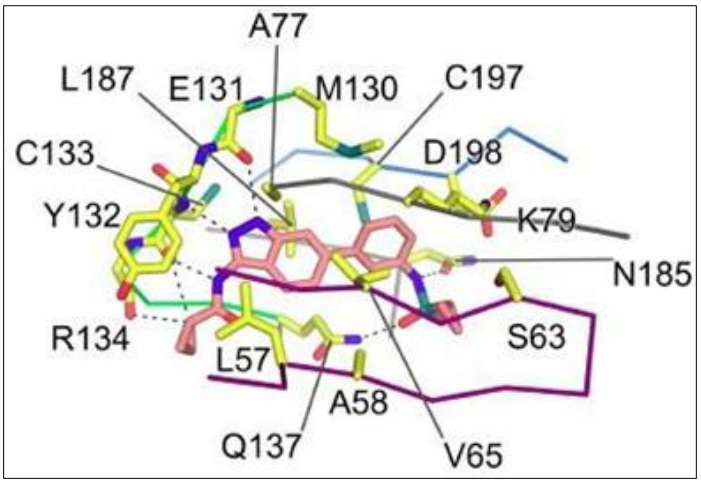


Figure 2 SGC-AAK1 probe bound to adjacent residues to BIKE ((Wells et al. 2020)

The idea was to first download the protein structure under study from the KLIFS database and use the Protein Preparation Wizard to prepare it (changing the PDB coordinates, removing excess water, and Performing Hydrogen bond minimisation) required for alignment and docking. The ligand SGC-AAK1(CHEMBL ID: 4452939) (Wells et al. 2020) is aligned to individual PDB structures of AAK1 and BIKE to determine the core groups of the ligand required to bind efficiently to the protein inside the binding cavity. The 3-acylaminoindazole ring scaffold has been a potent way to yield inhibitors both for AAK1 and BIKE (Bekerman et al. 2017)

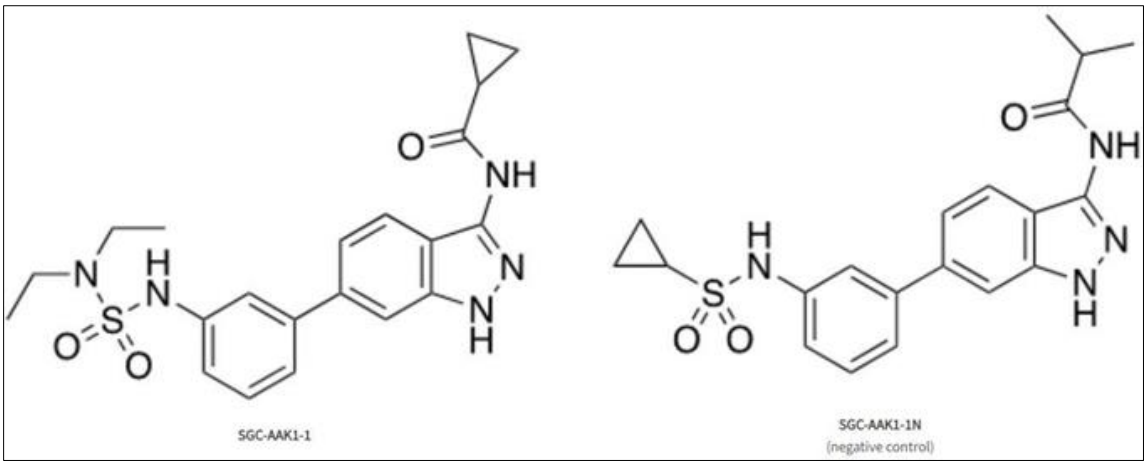


Figure 3 The compound SGC-AAK1-1 can effectively inhibit full-length ectopically produced AAK1- and BIKE-Nluc fusion proteins in a live cell NanoBRET experiment (Promega) (AAK1 IC50 = 230 nM; BIKE IC50 = 1.5 M). (Wells et al. 2020)

3. Protein and Ligand Structure Alignment Madhavi Sastry et al. (2013)

The aligned CHEMBL-prepared ligands were docked to each protein structure after being aligned using a ligand alignment tool.

Table 1 Different PDB IDs of AAK1 and BIKE with their intrinsic ligand and the reference ligands used for alignment

Serial NO.	PDBIDs	Resolution	Orthosteric Ligand	Reference ligand	Kinases
1	5l4q	1.97	LKBI	CHEMBL 516312	AAK1
2	5TE0	1.90	XIN	CHEMBL4452939	AAK1
3	4wsq	1.95	KSA	CHEMBL 4452939	AAK1
4	4W9W	1.72	YDJ	CHEMBL	BIKE
5	4W9X	2.14	Batricinib	CHEMBL	BIKE
6	513O	2.40	IDV	CHEMBL	BIKE
7	5I3R	2.40	IDK	CHEMBL	BIK
8	51KW	2.41	6BU	CHEMBL	BIKE

4. Protocol for Molecular Docking

The docking methodology creates novel docking routes for easy, reproducible results. Pose-generating co-crystallised ligand was processed using Ligprep. Docking each posture to the protein determined the correct match and interaction (emphasis on hinge region). When docking known inhibitors, these docking poses were employed as a reference.

Standard Precision (SP) Docking Implemented glide SP on the receptor grid before employing receptor-based constraints. Maestro posture viewer scanned docking poses. These poses were placed on the co-crystallised ligand to scan the conformers that matched the orthosteric ligand-protein interaction. Similar procedures were used for all CHEMBL ligands. The receptor grid was uploaded from the previously generated file, the ligands processed through ligprep were picked from the project table, the Standard Precision level was ticked, and constraint was selected as set in the receptor grid module in Glide. The output file limits posed to 10. The output files were saved by task day.

To ensure accurate assessment and reproducibility of results, controls were selected based on three acyl-aminoindazole rings and a literature study. SGC AAK1 was selected from the study and processed through ligprep and docking to the receptor grid.

5. Field-Based QSAR

The docking poses for each kinase were assessed based on their binding cavity fit and ligand PDB structure alignment. The assortment guided reversion to good pose-generating ligands. These groupings were put into the suite's field bases QSAR tool after visual examination. Twenty-five to thirty ligands are needed for QSAR. If there were fewer than twenty-five ligands, the structure similarity of PDB ligands was used to add more. The selected ligands were separated in the project table.

Field-based QSAR calculates field points like electrostatic fields of molecules in a rectangular grid in the training set using CoMFA and CoMSIA algorithms.

6. Comparative Molecular Field Analysis and CoMSIA

Both are computational methods for 3D QSAR modelling.

CoMFA analysis links the structural properties of a group of molecules to their biological function, such as target protein binding or medication efficacy. Below is the procedure.

Starting with a consistent and meaningful alignment of molecules in the dataset is the first stage. Many molecules are aligned to maximise structural similarity or superimposed depending on a shared substructure. This experiment used a 3-Acylaminoindazole Ring or reference ligands identical to the intrinsic ligand of proteins under study (see 2.4).

7. Grid Generation

A 3D grid is created around molecules after molecular alignment. Grids are three-dimensional lattices of uniformly spaced points. Grid points indicate where molecules' properties will be analysed.

8. Molecular field calculation

Each grid point computes various molecular properties or fields. Steric, van der Waals, electrostatic, and hydrophobic fields are used. The physicochemical environment molecules interact with in space is explained by these fields.

9. Partial Least Squares Regression (PLS)

Establishes a measurable link between molecular fields and biological activity. The PLS algorithm creates a mathematical model that links field fluctuations to activity to predict new molecular activity.

After PLS regression, CoMFA contour maps are created to identify regions of molecules that affect activity positively or negatively. These visualisations highlight activity-affecting structural features. Positive contours indicate locations where people or things improve the activity, whereas negative contours indicate areas where they inhibit it.

10. Model Validation and Interpretation

Identifies the most accurate forecasting power of the model, which may be evaluated on specific chemicals.

11. CoMSIA

considers molecular fields and similarity in QSAR modelling for a more complete analysis. Integrating the similarity field accounts for how structural similarity affects molecular activity (Klebe et al. 1994). This is beneficial for architecturally diverse and huge datasets or when considering new compound design using active scaffolds. This protocol is comparable to CoMFA.

CoMFA and CoMSIA use QSAR to link molecular structure and activity. Information like this can help create and optimise novel chemicals.

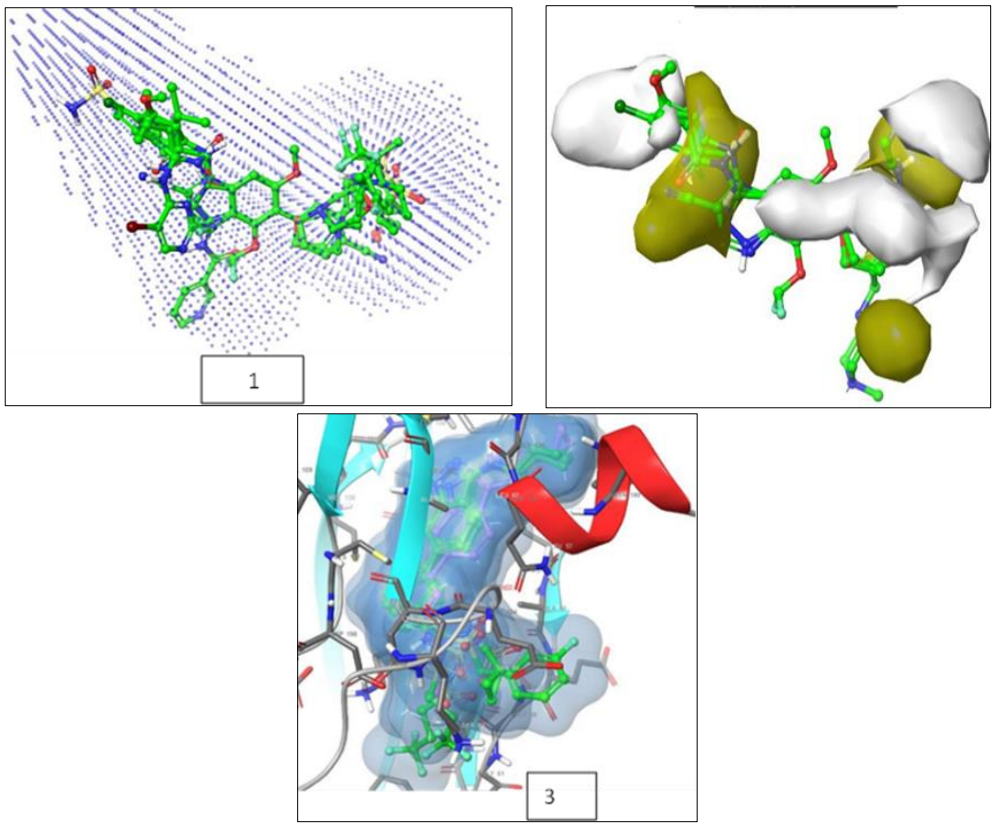


Figure 4 The electrostatic field contour map and Picture 2 showing the hydrophobic, steric interaction of QSAR test compounds of CHEMBL compounds related to BIKE (4w9x). Picture 3 showing how the docking poses fit in the binding cavity (blue shade) of BIKE (PDB ID: 5i3o)

OPLS (Lu et al. 2021 b) is a popular tool for computational chemistry, modelling, and molecular dynamics simulations. OPLS was created to explain atom-molecule interactions and energy. OPLS_4 improves accuracy and coverage. OPLS considers the following parameters: 1. Atomic Parameters: The force field assigns specific parameters to each atom in a molecule. Inappropriate dihedrals, partial charges, bond stretching, bond angle bending, van der Waals (Lennard-Jones) parameters, and bond stretching parameters are included. These variables control molecular atom and bond interactions.

12. Non-Bonded Interactions

OPLS considers Vander-walls and electrostatic interactions. Leonard Jonnes' Potential calculates atom size and partial charge. Each atom has a partial charge, describing electrostatic interactions.

Covalent interactions, including bond stretching, angle bending, and torsional rotations, are characterised by the OPLS force field. These interactions' characteristics are derived from data and quantum mechanical simulations; thus molecular geometries and bond energies may be precisely described.

This treatment allows for the investigation of solvation effects on molecular behavior and simulations of molecules in different solvent environments.

13. Application

Validated for widespread usage in molecular modelling, such as protein-ligand docking, molecular dynamics simulations, and structure-based drug design.

First phase of CHEMBL chemical docking (glide). Phase 2 procedure was followed (Figure 7).

In each case, project table ligand groups were chosen. The 80:20 random percentage ratio was used to sort the ligand into training and test sets. The following parameters were considered when developing the model. The 'field style' persisted.

Set Gaussian (recommended), 'maximum PLS factor' to 4, grid spacing to 1.0 Å, and other parameters to default. QSAR statistics of the training and test sets were then examined. Important statistical parameters (R^2, Q^2, stability, and RMSD) were used to establish the optimum activity prediction model. Each time the PLS graph (Predicted activity v/s Activity) incorporating the training and test set was examined, the gradual fitting of the points in the 45-degree line and outliers were noted.

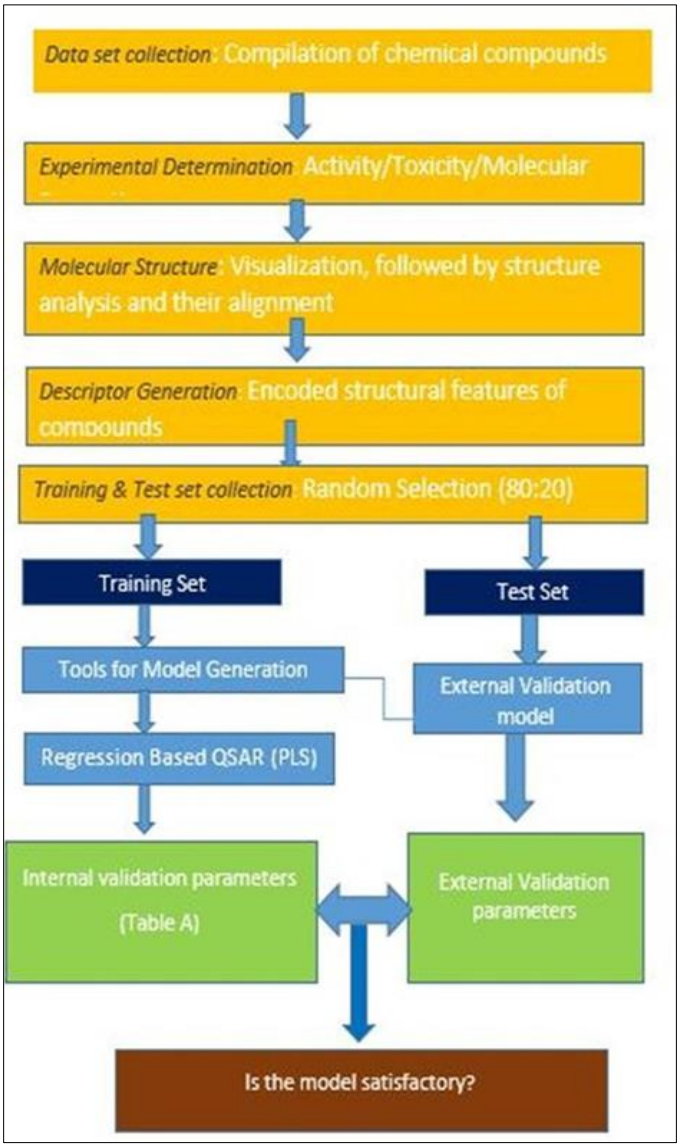


Figure 5 The methodology was followed up until internal validation for the production of 3D-QSAR models to Validation is crucial for assessing the reliability, robustness, and confidence of QSAR model predictions for a dataset (Stanton 2003 a). In other words, validation indicates model predictive power (Veerasamy et al. 2011). Internal and external validation techniques are used depending on the reason. Figure 7 shows this project's internal validation of CHEMBL chemical QSAR models. Molecular activity predictions are involved. As cross-validation divides n substances into calibration (training) and validation (test) subsets, internal validation is crucial. These data points are not used in calibration because the validation subset determines how well the model predicts incoming data. Model building uses the training/calibration subset. Internal validation is an excellent way to assess model quality and fit

Validation is crucial for assessing the reliability, robustness, and confidence of QSAR model predictions for a dataset (Stanton 2003 a). In other words, validation indicates model predictive power (Veerasamy et al. 2011). Internal and external validation techniques are used depending on the reason. Figure 7 shows this project's internal validation of CHEMBL chemical QSAR models. Molecular activity predictions are involved. As cross-validation divides n substances into calibration (training) and validation (test) subsets, internal validation is crucial. These data points are not used in calibration because the validation subset determines how well the model predicts incoming data. Model building uses the training/calibration subset. Internal validation is an excellent way to assess model quality and fit.

13.1.1. Used Internal Validation Methods

Least squares fitting is the most used internal model validation method. This validation method, similar to linear regression, compares anticipated and experimental activity using the R^2 (squared correlation coefficient). When calculating R2, the robust straight line fit is superior since it gives less weight to data points far from the centre (over a given standard deviation from the model). Another method is to remove outliers (compounds from the training set) from the dataset to improve the QSAR model; however, it is only reliable if strict statistical standards are followed. The most common internal model validation method is least squares fitting (Stanton 2003 b). This validation method, similar to linear regression, compares anticipated and experimental activity using the R^2 (squared correlation coefficient). When calculating R2, the robust straight line fit is superior since it gives less weight to data points far from the centre (over a given standard deviation from the model). Another method is to remove outliers (compounds from the training set) from the dataset to improve the QSAR model, however, it is only reliable if strict statistical standards are followed.

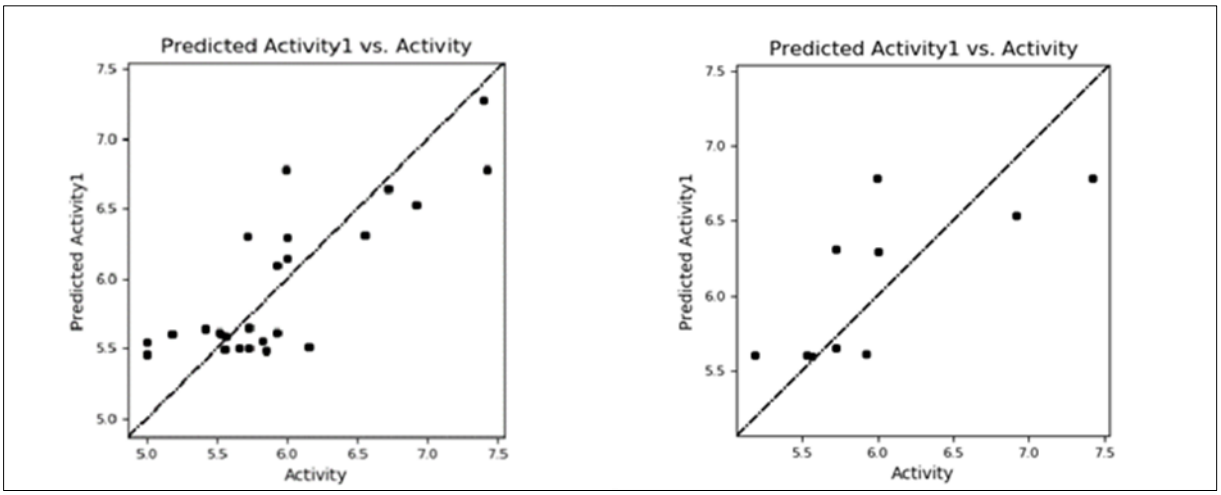


Figure 6 BIKE (4W9W) Ligand YDJ-: Partial least squares (PLS): PLS 1 figure (Training Set); PLS 2 figure (Test Set)

13.1.2. Cross-validation (cv)

A prominent method for internal validation of QSAR models. The regression is repeated on several data subsets by CV. The predicted values of each missing variable are calculated once (LOO) for each molecule to get the R value. CV often determines the model size needed to analyze a data set. A QSAR equation's R^2 is usually higher than its cross-validated value. It indicates an equation's predictive potential. CV's leave-one-out (LOO) method removes a molecule from the training set before developing and testing the model against each molecule. Upon completion, the mean of each Q^2 value is presented. A better training set of compounds (data points) for R calculation is employed to generate Q^2.

QSAR model validation usually employs LOO or LMO (Leave many out) CV approaches. Validated correlation coefficient R^2 represents the process's output. Model durability and predictive power are commonly measured by Q^2. Literature argues that high Q^2 (e.g., > 0.5) indicates the QSAR model's strong predictive power or is the strongest evidence. While Q^2 can indicate a model's predictive capability, it does not assess its ability to anticipate the behaviour of compounds not used in the QSAR model. This may limit these validation approaches. Some literature implies that developed models can forecast without validating exogenous molecules.

Serial No.	SD	R^2	R^2 CV	Scramble	Stability	F	P	RMSE	Q^2	Pearson	# Factor
1	0.343	0.700	0.45	0.22	0.89	93.6	5e-12	0.44	0.537	0.74	1
2	0.274	0.813	0.60	0.31	0.90	93.6	6.04e-15	0.42	0.57	0.77	2
3	0.242	0.8586	0.58	0.58	0.88	0.37	3.37e-16	0.39	0.62	0.79	3
4	0.237	0.867	0.86	0.42	0.866	60.4	1.03e-15	0.40	0.61	0.79	4

Figure 7 Example of a built model for BIKE (PDB ID-4W9W): the statistics show R^2 >R^CV and Q^2>0.5. Hence, the model shows potential predictive power. The parameters make a built model reliable, reproducible and robust

When a generated QSAR model satisfies the following criteria (the values listed below represent the minimum recommended thresholds for a significant QSAR model), it can be broadly applied in QSAR (for instance, PLS) investigations. Refer to Table 2 for the statistical metrics that are deemed acceptable for a qualified model.

PLS Data	Accepted Value
SD	SD of regression<SD of measured Activities
r^2	1>r^2>0.5
Stability	>0.6or =1
F	>0.6Large value of F
P	Small values indicate Confidence
Q^2	>0.5and(r^2-q^2)0.3

Figure 8 Demonstrating parameters to determine the acceptance of a built model

14. Findings and Examination

The results of the QSAR model, derived from data obtained from the CHEMBL database, concerning the proteins AAK1 and BIKE, are presented here.

Gathering Data: As previously noted, the dataset for BIKE comprised 29 compounds, while the dataset for AAK1 included 143 compounds along with their respective IC50 values sourced from CHEMBL. This is a collection of biologically active compounds that possess drug-like characteristics, and it is curated through manual editing. To facilitate the transformation of genomic data into effective new pharmaceuticals, it integrates chemical, activity, and genetic information (Mendez et al. 2019).

Collection of Training and Test Sets: The IC50 values underwent a negative logarithmic transformation to linearise the data, employing the equation [(pIC50 = -log10IC50)]. The dataset was partitioned into a training set and a test set (validation) in the ratio of [80:20]. The PLS graphs are included, illustrating the correlation between the ligand's potential to bind to the protein's binding cavities and their interaction with the hinge region. To evaluate the outcomes, it is essential to utilize these descriptors during the model construction process.

15. Results and Discussion

>The statistical metrics of the developed model for CHEMBL compounds are presented below

For AAK1, numerous models utilizing 143 CHEMBL compounds were developed through a trial-and-error approach to establish an effective predictive model, following the protocol outlined in Figure 7. The model generated for AAK1 that

demonstrates strong predictive capabilities is presented here, fulfilling the criteria for statistical significance. [R^2>2^CV; Q^2>0.5; Overall stability across PLS1-4 appears to be acceptable]

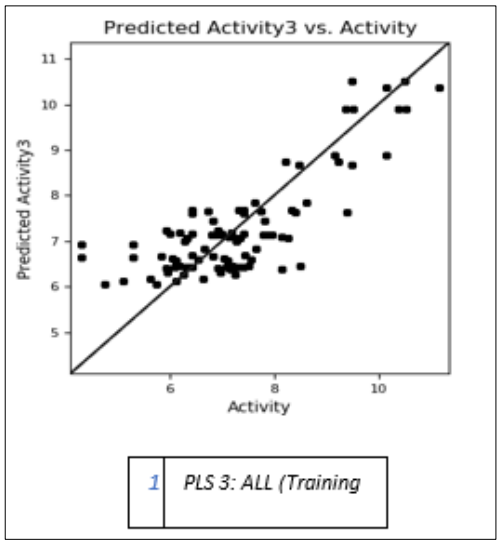
# Factor	Gauss_s	Gauss_e	Gauss_h	Gauss_a	Gauss_e
1	0.42	0.08	0.16	0.18	0.14
2	0.42	0.07	0.16	0.18	0.14
3	0.38	0.08	0.15	0.18	0.17
4	0.35	0.09	0.15	0.22	0.16

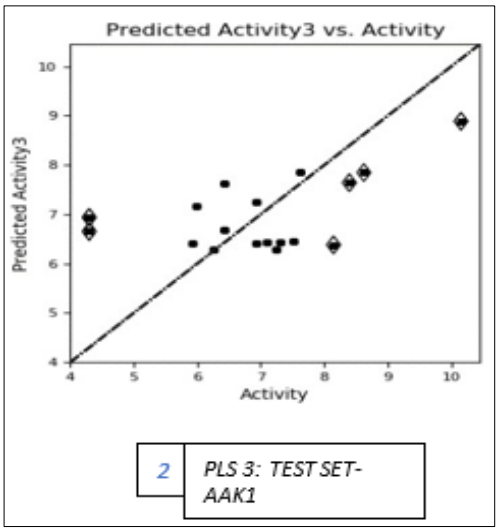
Figure 9 Field contributions for QSAR model ligands generated for BIKE (ligand IDV)

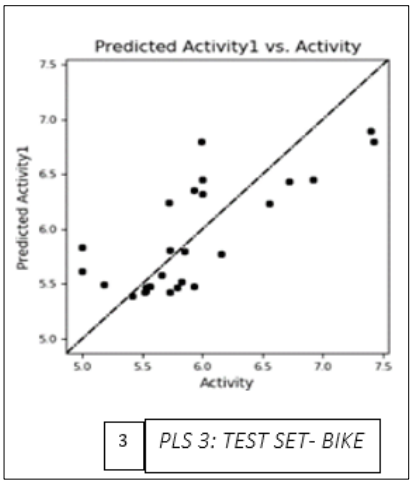
#Factors	Gauss_s	Gauss_e	Gauss_h	Gauss_d	Gauss_a
1	0.42	0.08	0.16	0.14	0.18
2	0.42	0.07	0.16	0.15	0.18
3	0.38	0.08	0.15	0.17	0.20
4	0.35	0.09	0.15	0.16	0.22

Figure 10 Field contributions for developed QSAR, AAK1 (ligand LKB)

Partial Least Figure (PLS) presented below







2

Figure 11 The relationship between experimental and predicted inhibitory activities (pIc50). To show the model's validity, the training and test sets predicted and experimental activity were plotted experimentally measured and predicted pIC50 values can be compared to show how well the Gaussian model predicts the activities of both training and test molecules

Table 2 Comparison of experimental versus predicted activities

Chembl Compounds	Activity	Predicted Activity 1	Predicted Activity 2	Predicted Activity 3	Predicted Activity 4
454376	4.3	6.5	6.6	6.9	6.8
4475067	4.3	6.7	6.7	6.6	5.8
450146	8.1	6.3	6.3	6.3	6.7
4439128	8.3	8.2	7.6	7.6	7.8
3102941	7.6	8.2	7.8	7.8	7.9
4857123	10.1	8.5	8.8	8.8	9.1

Table 2 showing comparison of experimental (pIC50) and predicted activities (pIC50) of test set compounds based on the model developed for AAK1 using 143 CHEMBL compounds, continued to next page.

Table 3 Comparison of experimental (pIC50) and predicted activities (pIC50) of test set compounds based on the model developed for BIKE using 29 CHEMBL compounds

Chembl Compounds	Activity	Predicted Activity 1	Predicted Activity 2	Predicted Activity 3	Predicted Activity 4
4561238	5.5	5.42	5.43	5.41	5.45
4470026	5.5	5.43	5.51	5.49	5.52
4473140	5.7	5.46	5.64	5.67	5.70
4593649	6.15	5.47	5.65	5.69	5.72
4097778	6.1	5.7	5.9	5.8	5.7
5079149	7.3	6.4	7.4	7.5	7.35
4542463	5.5	6.9	5.5	5.5	5.5
4531680	5.7	5.4	5.5	5.5	5.5
2205766	5.0	5.62	5.3	5.1	5.0

4872886	6.92	6.92	5.35	6.58	6.74

15.1. Gaussian Contour Maps

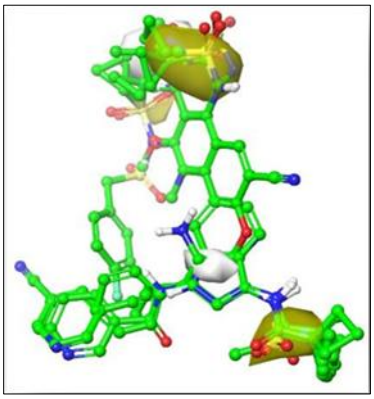


Figure 12 Example of the field contributions representing the electrostatic contribution of the QSAR model ligands for AAK1 (PDB ID: 5l4q). Yellow shade demonstrates the allowed regions for modification while the white shade depicts

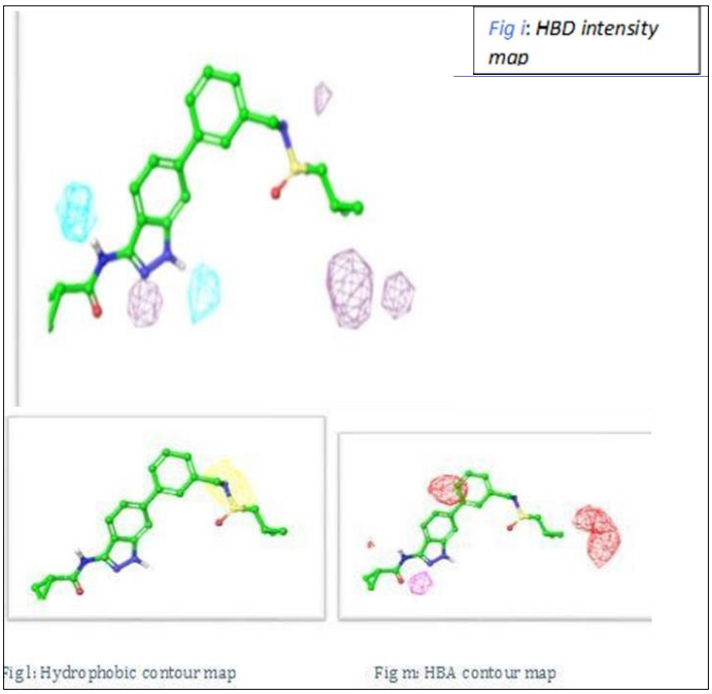


Figure 13 The relationship between experimental and predicted inhibitory activities (pIc50). To show the model's validity, the training and test sets' predicted and experimental activity were plotted experimentally measured and predicted pIC50 values can be compared to show how well the Gaussian model predicts the activities of both training and test molecules

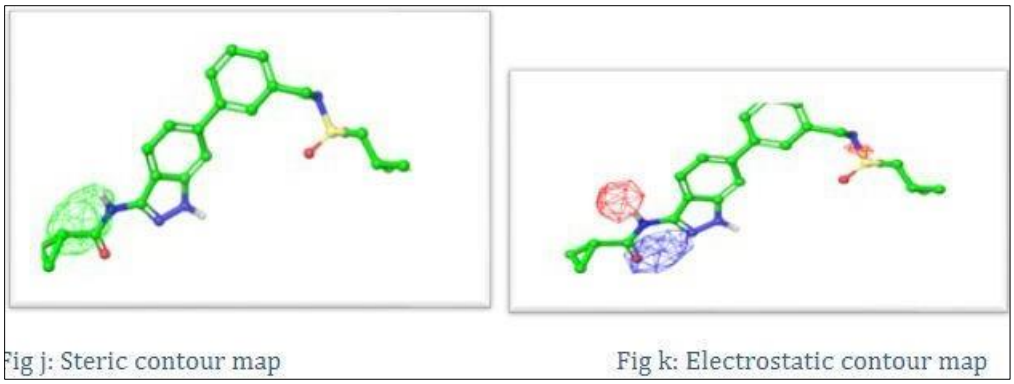


Figure 14 Visual representations of field contributions on QSAR model test ligands

Field contributions indicate that steric, hydrophobic, and HBA fields are stronger than electrostatic and HBD fields (Fig: i to k). The contour map shows CHEMBL 4516665, a chemical involved in AAK1 and BIKE activities. It is also a ligand alignment reference structure (Table 1). IUPAC name: N-[6-[3-(cyclopropylsulfonylamino) phenyl]-1H-indazol-3-yl] cyclopropanecarboxamide.

Gaussian contour map (Fig j) shows [1H-indazole-3-aryl] group (green contours) as favourable for steric interactions. In the electrostatic contour map, the [1H-indazole-3aryl] moiety had electropositive groups (Fig k). Yellow outlines in Fig. l indicate hydrophobic areas increasing activity.

The HBA contour map demonstrated unfavourable activity for the Hydrogen Bond Acceptor (HBA) groups in [3-(cyclopropyl sulfonyl amino) phenyl] and 1H indazole-aryl ring (red contours). HBD contour map (Fig n) shows purple contours for HBD group replacements that could increase activity and cyan contours for HBD group substitutions that are prohibited.

15.2. Validating Field-Based Gaussian QSAR Models

Predicted versus experimental activity was shown (5.2, Figure 13) to validate the model (Veerasamy et al. 2011). Comparing observed and predicted pIC50 values of 3-acylaminoindazole-basedigure inhibitors demonstrates that the Gaussian model predicts training and test set compound activities well (Table 2-3). The models predict CHEMBL compounds well and offer thorough and concise data on ligand chemical characteristics and antagonistic effects on AAK1 and BIKE based on statistical measures.

16. Discussion

This research uses receptor- and ligand-based methods to create models using CHEMBL substances with known IC50 activity. The compounds' docking poses show how a ligand fits in the protein's binding cavity, helping us virtually comprehend their similarity to the orthosteric ligand. We can profile similarities and filter a dataset of chemicals using molecular docking. After docking, good-pose CHEMBL molecules were matched to a reference ligand (Table 1). According to the literature, aligned ligands fit models among other factors. This effort aimed to sort kinase inhibitors by the 3-acylaminoindazole ring (Wells et al., 2020). Studies have targeted AAK1 and BIKE, initially targeting the kinase area (Couñago et al., 2017). Drug candidates can block the kinase by binding to the ATP phosphorylating region. Field-based QSAR models showed that steric, hydrophobic, and HBD fields outweighed electrostatic and HBA fields. Internally verified models allowed testing of unknown substances (external validation). Before QSAR model creation, aligning chemicals to their intrinsic structurally related compounds improved model predictability.

Active molecules binding to the hinge and ATP-binding area were studied. The chemicals interacted strongly in the active site. In vitro analysis will determine the process's validity, although the results are promising. Predictions that chemicals can be modified to increase biological activity can be tested.

17. Conclusion

In-silico drug designing for AAK1 and BIKE kinases provides a computationally efficient approach to develop inhibitors targeting proteins critically involved in diseases such as cancer and neural-muscular pain. CHEMBL compounds with experimentally determined IC50 values facilitated the creation of reliable 3D-field-based models, which can be used to

predict activities of novel compounds. These findings align with the literature and underscore the therapeutic potential of selectively inhibiting AAK1 and BIKE kinases to address related pathologies.

Compliance with ethical standards

Acknowledgments

Dr Antti Poso and Dr Tuomo Laitinen closely supervise this initiative. My sincerest thanks to the University of Verona for letting me do my thesis in Eastern Finland. I gained important drug discovery and molecular modelling skills. I am grateful to my supervisors and Erasmus personnel for their amazing help.

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Bekerman E, Neveu G, Shulla A et al.: Anticancer kinase inhibitors impair intracellular viral trafficking and exert broad-spectrum antiviral effects. J Clin Invest 127: 1338–1352, 2017 Berman HM: The Protein Data Bank. Nucleic Acids Res 28: 235–242, 2000

[2] Cappel D, Dixon SL, Sherman W, Duan J: Exploring conformational search protocols for ligand- based virtual screening and 3-D QSAR modeling. J Comput Aided Mol Des 29: 165–182, 2015 Conner SD, Schmid SL: Identification of an adaptor-associated kinase, AAK1, as a regulator of clathrin-mediated endocytosis. J Cell Biol 156: 921–929, 2002

[3] Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, Friesner RA: PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1.

[4] Methodology and preliminary results. J Comput Aided Mol Des 20: 647–671, 2006 Friesner RA, Murphy RB, Repasky MP et al.: Extra Precision Glide: Docking and Scoring

[5] Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes. J Med Chem 49: 6177–6196, 2006

[6] Greenwood JR, Calkins D, Sullivan AP, Shelley JC: Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. J Comput Aided Mol Des 24: 591–604, 2010

[7] Gupta-Rossi N, Ortica S, Meas-Yedid V et al.: The Adaptor-associated Kinase 1, AAK1, Is a Positive Regulator of the Notch Pathway. J Biol Chem 286: 18720–18730, 2011

[8] Halgren TA: Identifying and Characterizing Binding Sites and Assessing Druggability. J Chem Inf Model 49: 377–389, 2009

[9] Henderson DM, Conner SD: A Novel AAK1 Splice Variant Functions at Multiple Steps of the Endocytic Pathway. (Drubin, D. Ed.)Mol Biol Cell 18: 2698–2706, 2007

[10] Huang C, Ji C, Wang J: Current thoughts on cellular functions of numb-associated kinases. Mol Biol Rep 50: 4645–4652, 2023

[11] Hughes J, Rees S, Kalindjian S, Philpott K: Principles of early drug discovery: Principles of early drug discovery. Br J Pharmacol 162: 1239–1249, 2011

[12] Kanev GK, de Graaf C, Westerman BA, de Esch IJP, Kooistra AJ: KLIFS: an overhaul after the first 5 years of supporting kinase research. Nucleic Acids Res 49: D562–D569, 2021

[13] Kireev D: Structure-Based Virtual Screening of Commercially Available Compound Libraries. In: High Throughput Screen Methods Protoc. p. 65–76. Edited by Janzen WP, Springer New York, New York, NY 2016

[14] Klebe G, Abraham U, Mietzner T: Molecular Similarity Indices in a Comparative Analysis (CoMSIA) of Drug Molecules to Correlate and Predict Their Biological Activity. J Med Chem 37: 4130–4146, 1994

[15] Kontoyianni M: Docking and Virtual Screening in Drug Discovery. In: Proteomics Drug Discov Methods Protoc. p. 255–266. Edited by Lazar IM, Kontoyianni M, Lazar AC, Springer New York,

[16] New York, NY 2017 ligprep:

[17] Liu HP, Lin YJ, Lin WY et al.: A novel genetic variant of BMP2K contributes to high myopia. J Clin Lab Anal 23: 362–367, 2009

[18] Lu C, Wu C, Ghoreishi D et al.: a: OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. J Chem Theory Comput 17: 4291–4300, 2021

[19] Lu C, Wu C, Ghoreishi D et al.: b: OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. J Chem Theory Comput 17: 4291–4300, 2021

[20] Madhavi Sastry G, Adzhigirey M, Day T, Annabhimoju R, Sherman W: Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des 27: 221–234, 2013

[21] Maia EHB, Assis LC, De Oliveira TA, Da Silva AM, Taranto AG: Structure-Based Virtual Screening: From Classical to Artificial Intelligence. Front Chem 8: 343, 2020

[22] Mendez D, Gaulton A, Bento AP et al.: ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res 47: D930–D940, 2019

[23] Meng XY, Zhang HX, Mezei M, Cui M: Molecular Docking: A Powerful Approach for Structure- Based Drug Discovery. Curr Comput Aided-Drug Des 7: 146–157, 2011

[24] Neves BJ, Braga RC, Melo-Filho CC, Moreira-Filho JT, Muratov EN, Andrade CH: QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery. Front Pharmacol 9: 1275, 2018 Neveu G, Ziv-Av A, Barouch-Bentov R, Berkerman E, Mulholland J, Einav S: AP-2-Associated Protein Kinase 1 and Cyclin G-Associated Kinase Regulate Hepatitis C Virus Entry and Are Potential Drug Targets. (Ou, J. H. J. Ed.)J Virol 89: 4387–4404, 2015

[25] Ogasawara T, Kawaguchi H, Jinno S et al.: Bone Morphogenetic Protein 2-Induced Osteoblast Differentiation Requires Smad-Mediated Down-Regulation of Cdk6. Mol Cell Biol 24: 6560–6568, 2004

[26] Pu S, Schor S, Karim M et al.: BIKE regulates dengue virus infection and is a cellular target for broad-spectrum antivirals. Antiviral Res 184: 104966, 2020

[27] Shi B, Conner S, Liu J: Dysfunction of Endocytic Kinase AAK1 in ALS. Int J Mol Sci 15: 22918– 22932, 2014

[28] Shivakumar D, Williams J, Wu Y, Damm W, Shelley J, Sherman W: Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. J Chem Theory Comput 6: 1509–1519, 2010

[29] Sorensen EB, Conner SD: AAK1 Regulates Numb Function at an Early Step in Clathrin-Mediated Endocytosis. Traffic 9: 1791–1800, 2008

[30] Sorrell FJ, Szklarz M, Abdul Azeez KR, Elkins JM, Knapp S: Family-wide Structural Analysis of Human Numb-Associated Protein Kinases. Structure 24: 401–411, 2016

[31] Stanton DT: a: On the Physical Interpretation of QSAR Models. J Chem Inf Comput Sci 43: 1423– 1433, 2003

[32] Stanton DT: b: On the Physical Interpretation of QSAR Models. J Chem Inf Comput Sci 43: 1423– 1433, 2003

[33] The UniProt Consortium: UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47: D506–D515, 2019

[34] Veerasamy R, Rajak H, Jain A, Sivadasan S, Christapher PV, Agrawal R: Validation of QSAR Models - Strategies and Importance. Int J Drug Des Discov 2: 511–519, 2011

[35] Wells C, Couñago RM, Limas JC et al.: SGC-AAK1-1: A Chemical Probe Targeting AAK1 and BMP2K. ACS Med Chem Lett 11: 340–345, 2020

[36] Yung-Chi C, Prusoff WH: Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. Biochem Pharmacol 22: 3099–3108, 1973