PENERBIT PRESS
UNIVERSITI TEKNOLOGI MARA
eISSN: 2785-924X
2021 September, Volume 1, Issue 1

JOURNAL OF SMART SCIENCE AND TECHNOLOGY

Editorial Board

Journal of Smart Science and Technology

PENERBIT PRESS UNIVERSITI TEKNOLOGI MARA

2021 September Volume 1 Issue 1

eISSN: : 2785-924x

Predicting AAK1/GAK Dual-Target Inhibitor against SARS-CoV-2 Viral Entry into Host Cells: An *in silico* Approach

Xavier Chee Wezen[1*], Clement Sim Jun Wen[1], Lilian Siaw Yung Ping[1], Yeong Kah Ho[1], Kong Hao Qing[1], Christopher Ha[1], Hwang Siaw San[1]

[1]*School of Chemical Engineering and Science, Swinburne University of Technology* (*Sarawak Campus*)*, 93350 Kuching, Sarawak*, *Malaysia*

Received: 22-06-2021;
Accepted: 27-07-2021;
Published: 30-09-2021

*Correspondence
Email: xchee@swinbune.edu.my
(Xavier Chee Wezen)

Abstract

Clathrin-mediated endocytosis (CME) is a normal biological process where cellular contents are transported into the cells. However, this process is often hijacked by different viruses to enter host cells and cause infections. Recently, two proteins that regulate CME – AAK1 and GAK – have been proposed as potential therapeutic targets for designing broad-spectrum antiviral drugs. In this work, we curated two compound datasets containing 83 AAK1 inhibitors and 196 GAK inhibitors each. Subsequently, machine learning methods, namely Random Forest, Elastic Net and Sequential Minimal Optimization, were used to construct Quantitative Structure Activity Relationship (QSAR) models to predict small molecule inhibitors of AAK1 and GAK. To ensure predictivity, these models were evaluated by using Leave-One-Out (LOO) cross validation and with an external test set. In all cases, our QSAR models achieved a q^2_{LOO} in range of 0.64 to 0.84 (Root Mean Squared Error; RMSE = 0.41 to 0.52) and a q^2_{ext} in range of 0.57 to 0.92 (RMSE = 0.36 to 0.61). Besides, our QSAR models were evaluated by using additional QSAR performance metrics and y-randomization test. Finally, by using a concensus scoring approach, nine chemical compounds from the Drugbank compound library were predicted as AAK1/GAK dual-target inhibitors. The electrostatic potential maps for the nine compounds were generated and compared against two known dual-target inhibitors, sunitinib and baricitinib. Our work provides the rationale to validate these nine compounds experimentally against the protein targets AAK1 and GAK.

Keywords

QSAR models; Machine learning; AAK1; GAK; Dual-target inhibitors; Viral entry

1 Introduction

The clathrin-mediated endocytosis (CME) is a key process where cargo molecules or proteins are trafficked from the cell surface to the interior by clathrin-coated vesicles (CCVs)[1]. The CME process is vital to physiological processes such as nutrient uptake, internalization of receptors, signal transduction regulation and synaptic vesicle recycling[2]. The process starts with the nucleation of clathrin-coated pits followed by the recruitment of heterotetrameric protein complexes known as adaptor proteins (APs; Figure 1a)[2]. There are four known adaptors proteins (AP1-4), of which AP2 is most commonly found in CCVs[3].

AP2 is a stable complex formed by four adaptins termed as α, β, σ2 and μ2 subunits. These four subunits give rise to a core domain joined to two appendage domains by polypeptide linkers[4]. The core domain would bind to the cargo while the appendages would interact with other accessory proteins to initiate CME[5]. Several studies showed that the phosphorylation of the Threonine-156 residue in AP2 is critical for endocytosis to occur. Olusanya et al.[6] showed that transferrin (receptor in iron transport) internalization requires the phosphorylation of the μ2 subunit[6]. Additionally, another study by Ricotta *et al.* showed that μ2 phosphorylation leads to a stronger binding affinity to tyrosine- or di-leucine based sorting motifs on membrane proteins[7]. Clearly, the phosphorylation of the μ2 subunit of AP-2 is important for cargo recruitment and vesicle assembly (Figure 1b)[8,9,10]. This phosphorylation of the μ2 subunit is regulated by two serine/threonine kinases, namely the AP-2-associated Protein Kinase 1 (AAK1) and the Cyclin-G-associated Kinase (GAK)[7,11].

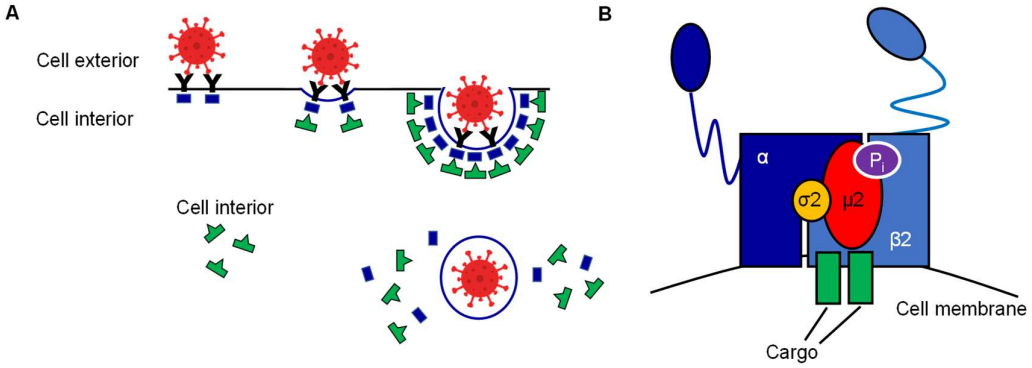


Figure 1. Clathrin-mediated endocytosis. A. Viral attachment on cell surface receptor leading to assembly of CCV and endocytosis. B. Phosphorylated AP2 complex binding to cargo molecules or proteins. Phosphate group abbreviated as Pi

Recently, AAK1 and GAK have been proposed as potential protein targets for the design of broad-spectrum antiviral drugs[12]. This is because a majority of viruses hijacks the same CME process to gain entry to host cells. Viral entry often starts with the attachment of viruses to membrane receptors followed by endocytosis[13]. For example, the Hepatitis C virus (HCV) contains a conserved tyrosine-based sorting motif (YXXΦ) that is recognized by μ2 subunit of AP2 followed by internalization via CME. This same YXXΦ motif exists for many other viruses, including the Human Immunodeficiency Virus (HIV) and the severe acute respiratory syndrome-Coronavirus-2 (SARS-CoV-2)[14,15].

As proof-of-concept for the druggability of AAK1 and GAK, Neveu

et al.[16] measured the protein-inhibitor dissociation constant (*Kd*) and showed that erlotinib binds to GAK (*Kd* = 3.16 nM) and sunitinib to AAK1 (*Kd* = 11.22 nM). Furthermore, both sunitinib and erlotinib significantly reduced viral infections when incubated with mammalian cells[16]. Subsequent studies showed that sunitinib/erlotinib combination protected mammalian cells against Flaviviridae (HCV, Dengue virus, West Nile virus, Zika virus), Filoviridae (Ebola virus, Marburg virus), Togaviridae (Chikugunya virus), Arenaviridae (June virus), Retroviridae (HIV), Paramyxoviridae (Respiratory Syncytial virus), Rhadoviridae (Rabies virus) and Coronaviridase (SARS-CoV, MERS-CoV and SARS-CoV-2)[12,17,18,19]. Although the data is encouraging, one study showed that inhibition of both AAK1 and GAK is essential for therapeutic efficacy[12]. In both Dengue and Ebola murine model studies, treatment with erlotinib alone did not alter survival of infected mice, whereas sunitinib alone offered partial protection[12]. The greatest protection was observed when both erlotinib and sunitinib were administered together[12]. However, recent studies showed that sunitinib is an AAK1/GAK dual-target inhibitor itself[19]. This might explain why sunitinib alone was able to confer partial protection against SARS-CoV, MERS-CoV and SARS-CoV-2.20.

To date, only four AAK1/GAK dual-target inhibitors were identified, namely sunitinib, AZD7762, an isothiazolo[5,4-b]pyridine-based compound and baricitinib[12,19]. Interestingly, baricitinib is a dual-target inhibitor predicted by the London-based artificial intelligence platform BenevolentAI. Barticitinib was granted an Emergency Use Authorization by the US Food and Drug Administration (FDA) and is currently in Phase 3 clinical trial. With AAK1/GAK dual-target inhibitors showing promises, we set out to predict novel GAK/AAK1 dual-target inhibitors.

Our strategy involves using machine learning (ML) algorithms to construct Quantitative-Structure Activity Relationship (QSAR) models that could predict AAK1 and GAK inhibitors. Machine learning methods are widely deployed to discover new inhibitors against various drug targets. For example, Bayesian machine learning models were used successfully to repurpose ruboxistaurin (an investigational drug for diabetic retinopathy) for targeting an Alzheimer's disease-related protein target, Glycogen Synthase Kinase 3β[21]. Additionally, a Support Vector Machine (SVM) model identified a dual-target inhibitor against the cancer-associated proteins Fibroblast Growth Factor Receptor 4 and the Epidermal Growth Factor Receptor[22]. Other recent successes of machine learning application in drug discovery include discovering inhibitors for Janus Kinase 2[23], Indoleamine 2,3-dioxygenase[24], RNA Polymerase of Hepatitis C virus25 and 3CL-Proteinase of SARS-CoV-2[26]. In-depth reviews of the techniques and application of machine learning in drug discovery have been covered elsewhere[25,27,28].

Given the potential of machine learning applications in drug discovery, we would deploy both QSAR models for AAK1 and GAK in our work to screen the Drugbank compound library simultaneously with the intent of discovering AAK1/GAK dual-target inhibitors.

2 Methods

2.1 Construction of Compound Datasets

Structural information of chemical compounds could be represented by the Simplified Molecular Input Line Entry system (SMILES). The SMILES of chemical compounds that were tested against the proteins AAK1 and GAK were retrieved from the bioassay database ChEMBL[29]. Additionally, other inhibitors reported in literature were curated, converted to SMILES and pooled together to form the AAK1 and GAK compound datasets. Next, we filtered for compounds that were annotated with protein-ligand dissociation constant (*Kd*) against the protein AAK1 or GAK. A *Kd* value is an

experimentally-determined measurement that characterises the binding affinity of a compound to a protein target[30]. A compound that binds tightly to its protein target would exhibit a low *Kd* value. To ease numerical handling, all *Kd* values are expressed as their inverse logarithmic values, *pKd* where:

$$pKd = -\log Kd$$

Therefore, a chemical compound with a high *pKd* value indicates high binding affinity to proteins. In total, the AAK1 and GAK compound datasets consisted of 83 and 195 compounds, respectively.

2.2 Featurization and Train/Test Splitting

The AAK1 and GAK compound datasets were subsequently featurized using the Extended Connectivity Fingerprint 4 (ECFP4)[31] by using the open-source chemoinformatics software, RDKit[32] in Python. The ECFP4 is a circular topological fingerprint that could be rapidly calculated and could capture essential structural information such as compound substructure and stereochemical information[31]. We chose ECFP4 for featurization as several benchmarking studies showed that the performance of ECFP4 is among the best for virtual screening[33]. After featurization, the *pKd* values were appended to the dataset as class attribute. The featurized compound datasets for AAK1 and GAK were then split into training and test sets with a 90:10 split in a stratified fashion. The purpose of the stratified splitting is to ensure that both the training and test sets have the similar distribution of *pKd* values. After this train-test splitting process, we produced four datasets: (a) AAK1 training set (74 inhibitors), (b) AAK1 external test set (9 inhibitors), (c) GAK training set (175 inhibitors) and (d) GAK external test set (20 inhibitors).

2.3 Feature Selection

Feature selection is an important pre-processing step to remove highly correlated or irrelevant fingerprints to prevent model overfitting[34]. We performed this step by using the Attribute Evaluator function available on the Waikato Environment for Knowledge Analysis (WEKA), a platform with a collection of machine learning algorithms[35]. The CfsSubsetEval function using the Best First search method was used with its default setting to identify important chemical fingerprints. The CfsSubsetEval function evaluates the worth of a subset of features by considering the individual predictive ability of each feature along with the degree of redundancy between them[36]. In essence, the CfsSubsetEval function would select a subset of ECFP4 fingerprints that correlates with the *pKd* values and yet, unrelated to each other[37].

2.4 Chemical Space and Applicability Domain Analysis

PCA is a dimensionality reduction method that allows for the visualization of multi-dimensional datasets on a two- or three-dimensional plots[38]. In the context of structural diversity, PCA could be used to visualize similarities of a collection of compounds based on their structural and physicochemical properties[38,39]. Due to its usefulness in displaying and revealing structural diversity in a convenient graphical format, PCA plots are often used to (a) analyse chemical space and (b) define QSAR applicability domain (AD). As it is impossible for a single QSAR model to be applicable for all chemical compounds, the predicted response (*pKd* values in this work) for a modelled compound is only valid if the modelled compound falls within the AD of the model[40]. To construct the PCA plots, features such as molecular weight, total polar surface area, number of rotatable bonds, number of hydrogen bond donors and acceptors as well as the solubility (LogP) for compounds in the datasets were calculated. These features were then combined and transformed into a set of principal components (PCs). The PCA plots are then visualized with only the first and second PCs. These steps are automated by using the Platform for Unified Molecular Analysis (PUMA)[41] version 1.0.

2.5 Construction of QSAR Models

The QSAR models were constructed by using machine learning models available in WEKA. For both AAK1 and GAK, we tested the algorithms Gaussian Processes (GP), Elastic Net (EN), Support Vector Machine (SVM), Sequential Minimal Optimization (SMO), k-Instance Based Learner (IBK), K* Based Learner (K*) and Random Forest (RF) to identify which algorithms have high predictivity. The performance of these models was evaluated by using 10-fold cross validation (CV) and Leave-One-Out (LOO) CV performance. Between them all, the top three algorithms were GP, EN and SMO. The hyperparameters of these selected algorithms were further tuned to enhance performance. For the EN model for AAK1, the alpha and lambda sequence values were set to 0.0032 and 40. Meanwhile, the EN model for GAK has an alpha value of 0.008 and lambda sequence value of 100. For SMO, the complexity parameter (c value) was set to 0.5 and 0.1 for AAK1 and GAK models, respectively. Unless stated otherwise, all other hyperparameters were set to the default WEKA values.

2.6 Evaluation of QSAR Model Performance

The fitness and the robustness of the QSAR models developed in this study were assessed by using the statistical parameters, namely the coefficient of determination, r^2 (q^2 for internal or external validation methods), mean absolute error (MAE) and root mean square error (RMSE).

2.6.1. Coefficient of Determination (r^2 or q^2).

The r^2 or q^2 is a measure of fit to determine how well the regression lines of QSAR models fit the experimentally determined *pKd* values. An r^2 or q^2 of 1.0 indicates a perfect fit. They are calculated as follows:

$$r^2 = \frac{1 - \text{SSE}}{\text{SST}} \quad (1)$$

where SSE = sum of squared errors and SST = total sum of squares

Conventionally, a $q^2 > 0.5$ obtained from a model validation procedure is indicative of good model performance[42].

2.6.2. Mean Absolute Error (MAE)

The MAE represents the difference between experimentally determined and the predicted *pKd* values of the compounds.

$$\text{MAE} = \frac{\sum_{i=1}^{N} |y_i - \hat{y}_i|}{N} \quad (2)$$

where N = total number of compounds and $|y_i - \hat{y}_i|$ = absolute difference between the experimentally determined and the predicted *pKd* values.

2.6.3. Root Mean Squared Error (RMSE)

Similar to MAE, RMSE is another indicator to measure the variation between the experimentally determined and the predicted *pKd* values. However, as the errors are squared during the calculation, the RMSE metric gives a higher penalty to large errors.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}} \quad (3)$$

Conventionally, an RMSE < 0.5 obtained from a model validation procedure indicates good model predictivity[43].

2.6.4. Additional Criteria of Predictive QSAR models

Although a high q^2 and a low RMSE are important indications of a good QSAR model, other metrics could be used to indicate high predictive power[44]. To ensure model robustness, Golbraikh and Trophsha[44] suggested using an external test set and formulated several statistical criteria to evaluate QSAR models based on the external test set:

a. $q^2_{\text{ext}} > 0.6$
b. $(|q^2_{\text{ext}} - q_o{}^2_{\text{ext}}|)/q^2_{\text{ext}}$ (henceforth referred as *D* value)
c. $0.85 < k < 1.15$

where $q_o^2_{ext}$ and q^2_{ext} are correlations of determination of the external test set with or without passing through the origin and *k* is the slope of the regression line passing through the origin. QSAR models that satisfy these criteria would be considered as predictive [42,44,45].

2.6.5. Y-randomization Test

The purpose of the y-randomization test is to quantify chance correlation. Chance correlation is the situation where a handful of descriptors fits the data reasonably well without having any true connection to the response variable[46]. To conduct a y-randomization test, the response variable i.e. *pKd* values were randomly permutated while keeping the independent variables i.e. chemical fingerprints untouched[46]. This process generated Y-scrambled datasets with no meaningful connections between the chemical fingerprints and the *pKd* values. Hence, any QSAR models generated using the Y-scrambled datasets should fail when assessed using any form of validation. Each of the QSAR models in this work was tested 10 times with 10 Y-scrambled datasets.

2.7 Calculation of Electrostatic Potential Surface

Complementary electrostatic potentials of the chemical compounds to their target proteins are essential for molecular recognition and binding. To calculate the electrostatic potential maps (EPMs) of chemical compounds, we used the Rapid Overlay of Chemical Structures (ROCS)[47] software by OpenEye. To perform this step, a multi-conformer file consisting of the predicted chemical compounds were generated by the OMEGA[47] software. The multi-conformer file was then parsed to the ROCS software to match the compound conformers to the chemical structure of baricitinib or sunitinib using a shape-based superposition method. The EPMs were then visualized by using VIDA[48].

3 Methods

3.1 Construction of Training and Test Sets for AAK1 and GAK

In a QSAR modelling process, machine learning algorithms would be evaluated using an external test set to determine the accuracy and the generalizability of the predictive models. For the evaluation to be valid, the external test set must be representative of the training set i.e. captures most of the information from the training set[49]. To make sure that the training-test set pairs were congruent, we used a stratified splitting strategy to make sure the set pairs for each protein target had similar *pKd* value distributions (Figure 2A and Figure 2B). Additionally, we used PCA and showed that the compounds in the training-test set pairs occupy similar chemical space (Figure 2C and Figure 2D). The first and second principal compounds of PCA captured 81.3% of the chemical diversity in the AAK1 training-test sets and 79.4% of that for GAK.

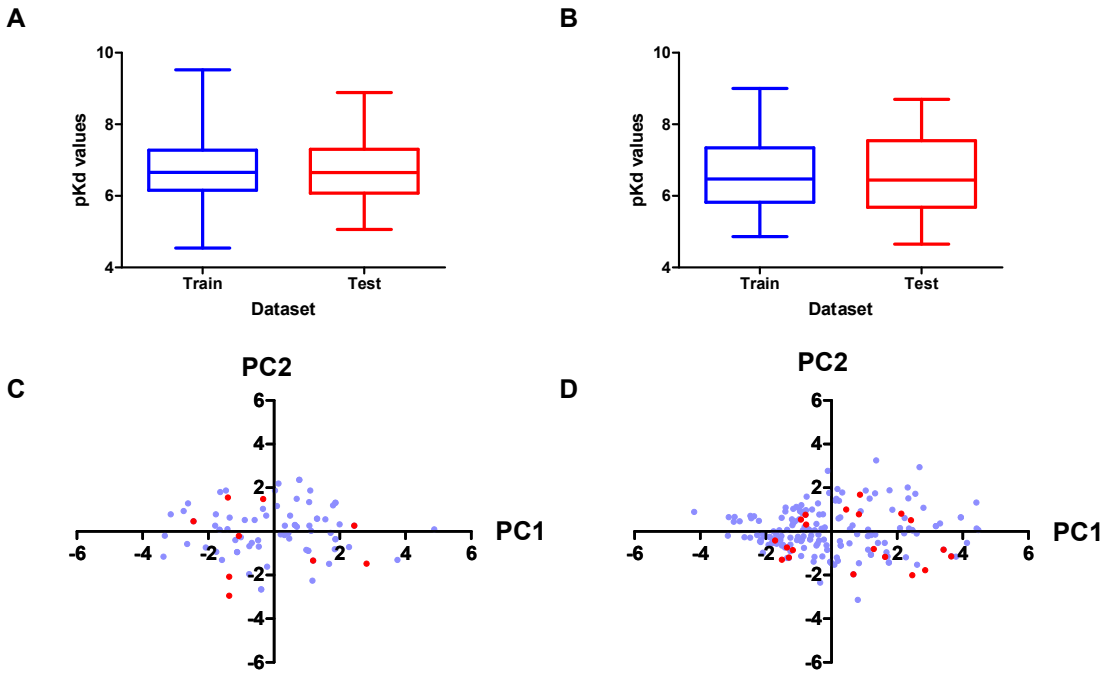


Figure 2. Similarity of AAK1 and GAK Training and Test Sets. Distribution of *pKd* values in the training and test sets for A. AAK1 and B. GAK. Chemical space coverage represented by PCA plots for C. AAK1 and D. GAK. Blue dots represent compounds from training sets; red dots represent compounds from test sets

Having constructed congruent training-test set pairs for both AAK1 and GAK, we featurized the datasets by using ECFP4 chemical fingerprints. For every compound, the ECFP4 algorithm generated a 1024-bit string to represent the chemical structure. While ECFP4 is known to require higher dimensional representations i.e. many chemical fingerprints to perform well, redundant or collinear chemical fingerprints would result in model overfitting[50]. To avoid this, we used the CfsSubsetEval function in WEKA to identify a subset of chemical fingerprints that correlated well with the *pKd* values. The result from this process was an (*N*, 110) array for AAK1 datasets and an (*N*, 75) array for GAK datasets, where *N* refers to the number of compounds in each dataset. The corresponding *pKd* values are stored in a separate (*N*, 1) array. Details of the featurized datasets used for subsequent QSAR modelling are presented in Table 1.

Table 1. Details of featurized datasets for AAK1 and GAK QSAR modelling

Descriptions	AAK1	GAK
Number of compounds in training set	74	175
Number of compounds in external test set	9	20
Number of features	110 chemical fingerprints	75 chemical fingerprints

3.2 Construction of Training and Test Sets for AAK1 and GAK

With the training-test set pairs for both AAK1 and GAK, we proceeded to identify ML algorithms that could predict the *pKd* values of chemical compound based on chemical fingerprints. We trained the ML algorithms using the training sets and evaluated their performance using 10-fold

CV and LOO CV. In both cases, three machine learning models (EN, GP and SMO) were identified as the best performers (Supplementary Table S1 and Supplementary Table S2). Next, we further tuned the hyperparameters of the ML algorithms to generate the best possible QSAR models. This yielded three QSAR models each for AAK1 (AAK1-GP, AAK1-EN and AAK1-SMO) and GAK (GAK-GP, GAK-EN and GAK-SMO).

The robustness of these models was assessed by both internal and external validation procedures. For internal validation, LOO CV procedure was adopted. First, we noted that all the QSAR models had $q^2_{LOO} > 0.5$ and an RMSE of ≈ 0.5 (Table 2). Although ideally, the QSAR models should have an RMSE of < 0.5, the RMSE of the QSAR models for GAK were still acceptable. Secondly, it was also noteworthy that the QSAR models for AAK1 outperform (refer to QSAR models) those for GAK.

We also assessed the predictivity of the QSAR models by using the external test sets. Similarly, the QSAR models for AAK1 fared better than those of GAK. Although all six QSAR models had $q^2_{ext} > 0.5$, the QSAR models for AAK1 had q^2_{ext} of 0.88 or higher (Table 2). The RMSE for all six models were also close to 0.5 (Table 2). However, the *D* values for the QSAR models of GAK were high (Table 2). This was because the q^2_{ext} of these GAK QSAR models were skewed by the mis-prediction of four compounds across all models. The *pKd* values of one compound were consistently over-estimated while the remaining three compounds were consistently underestimated (Supplementary Figure S1). However, the *D* values improved dramatically for the models GAK-GP, GAK-EN and GAK-SMO if the four outliers were removed (Table 2).

Table 2. Performance of QSAR models for AAK and GAK when assessed using internal and external validation methods

	Internal validation		External validation			
Models	q^2_{LOO}	RMSE	q^2_{ext}	RMSE	*D*	***k***
AAK1-GP	0.80	0.48	0.88	0.54	0.13	0.98
AAK1-EN	0.81	0.44	0.90	0.41	0.11	0.99
AAK1-SMO	0.84	0.41	0.92	0.36	0.09	0.99
GAK-GP	0.64	0.54	0.61	0.58	0.64*	0.99
GAK-EN	0.65	0.52	0.57	0.61	0.75**	0.99
GAK-SMO	0.67	0.52	0.67	0.54	0.49***	0.99

*Value after outlier removal = 0.18
**Value after outlier removal = 0.23
***Value after outlier removal = 0.08

From the validation procedures above, only the AAK1-SMO QSAR model fulfilled all six stringent criteria for QSAR model predictivity while all other models had near misses with three or less criterion. However, we noted that all of the QSAR models were still beneficial in approximating the *pKd* values of chemical compounds against AAK1 and GAK. Our reasoning was based on (a) our QSAR models were still able to produce predicted *pKd* values that could rank chemical compounds correctly (Figure 3A-E) and (b) the residuals of the experimental vs predicted *pKd* values were generally low (Supplementary Figure S1). The error of our models often came from the underestimation of *pKd* values for compounds with experimental *pKd* values > 8 (Supplementary Figure S1). We posit that this was because a majority of the compounds in our training sets had *pKd* values in the range of 6.0-7.5 (Figure 2A and Figure 2B).

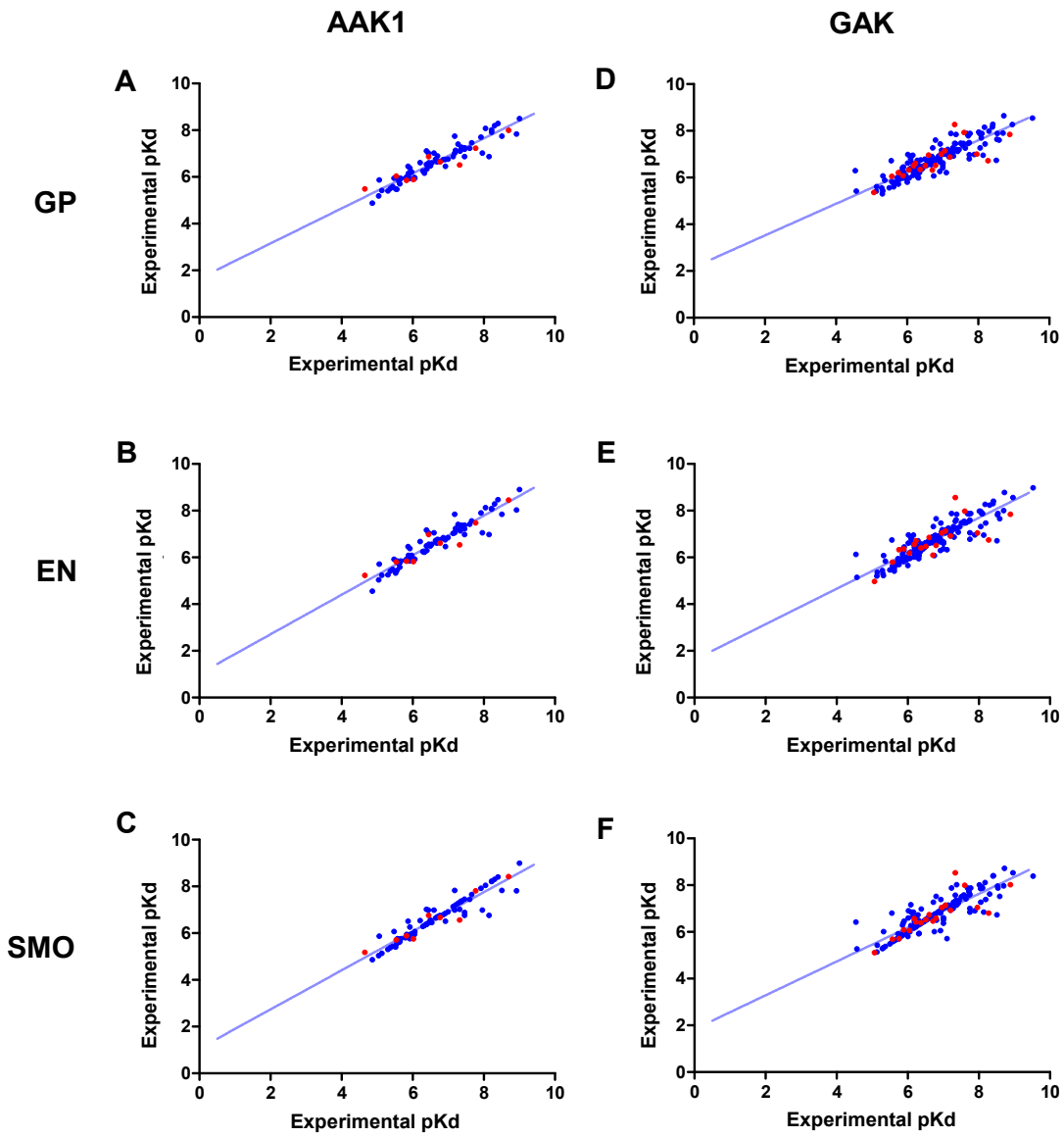


Figure 3. Correlation plot of experimental *pKd* vs predicted *pKd*. These plots are for the QSAR models A. AAK1-GP, B. AAK1-EN, C. AAK1-SMO, D. GAK-GP, E. GAK-EN and F. GAK-SMO. Blue dots represent training set data; red dots represent test set data; regression line shown in pale blue

3.3 Y-randomization Test

Next, we conducted Y-randomization test to rule out chance correlation in our QSAR models. In all cases, the QSAR models Y-randomized datasets did not have predictive power when assessed using LOO CV (Figure 4A-B). Only the models constructed from the original AAK1 and GAK datasets yielded high r^2 and q^2. This indicated that the correlation of the predicted *pKd* values and the experimental *pKd* values from our QSAR models are not due to chance.

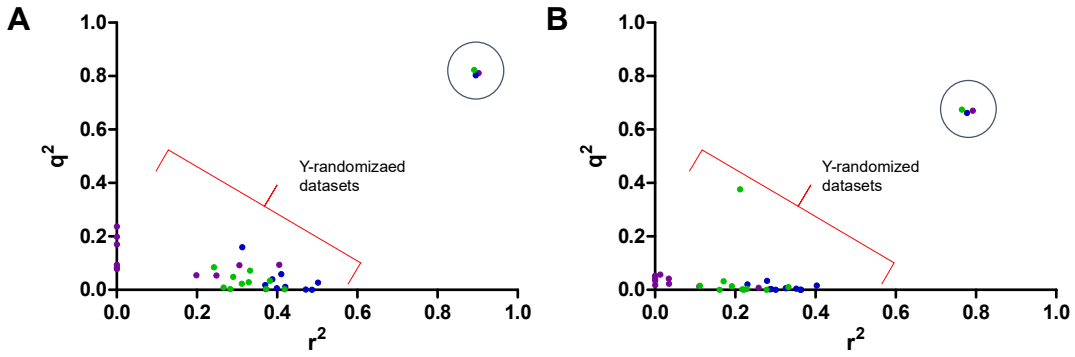


Figure 4. Performance of QSAR models in Y-randomization test. Y randomization plots for A. AAK1 and B. GAK. Purple dots for datasets trained using EN; Blue dots for datasets trained using GP; green dots for datasets trained using SMO. Original dataset highlighted in blue circle

3.4 Model Deployment

Drug repositioning refers to the strategy of discovering new bioactivities of existing drugs beyond their original intended purpose[51]. This strategy is attractive because known drugs often have determined pharmacological and safety profile; these information would help in accelerating their development into new therapeutics[51]. As part of our drug repositioning effort, we deployed our QSAR models on the Drugbank compound library[52] consisting of investigational, experimental and FDA-approved drugs. To identify chemical compounds that could potentially exhibit high *pKd* values against the protein targets AAK1 and GAK, we decided to take a concensus scoring approach (Figure 5A). Concensus scoring could improve model accuracy as the predicted *pKd* values of the different QSAR models are averaged (henceforth referred to as concensus *pKd* values)[53,54]. In our strategy, a compound would be considered as a potential AAK1/GAK dual-target inhibitor if its concensus *pKd* values > 8 for both AAK1 and GAK. In total, 252 compounds and 361 compounds fulfilled this criterion (Figure 5B-C). Among them, 19 compounds had concensus *pKd* values > 8 against both AAK1 and GAK inhibitors (Figure 5B). Next, we conducted AD analysis by means of PCA. Only compounds that fall within the chemical space defined by the training set were considered for further analysis. From this, only nine of the potential dual inhibitors fall within the AD of the QSAR models (Figure 5D-E). The identities of these nine compounds are given in Table 3 and their chemical structures are presented in Supplementary Figure S2.

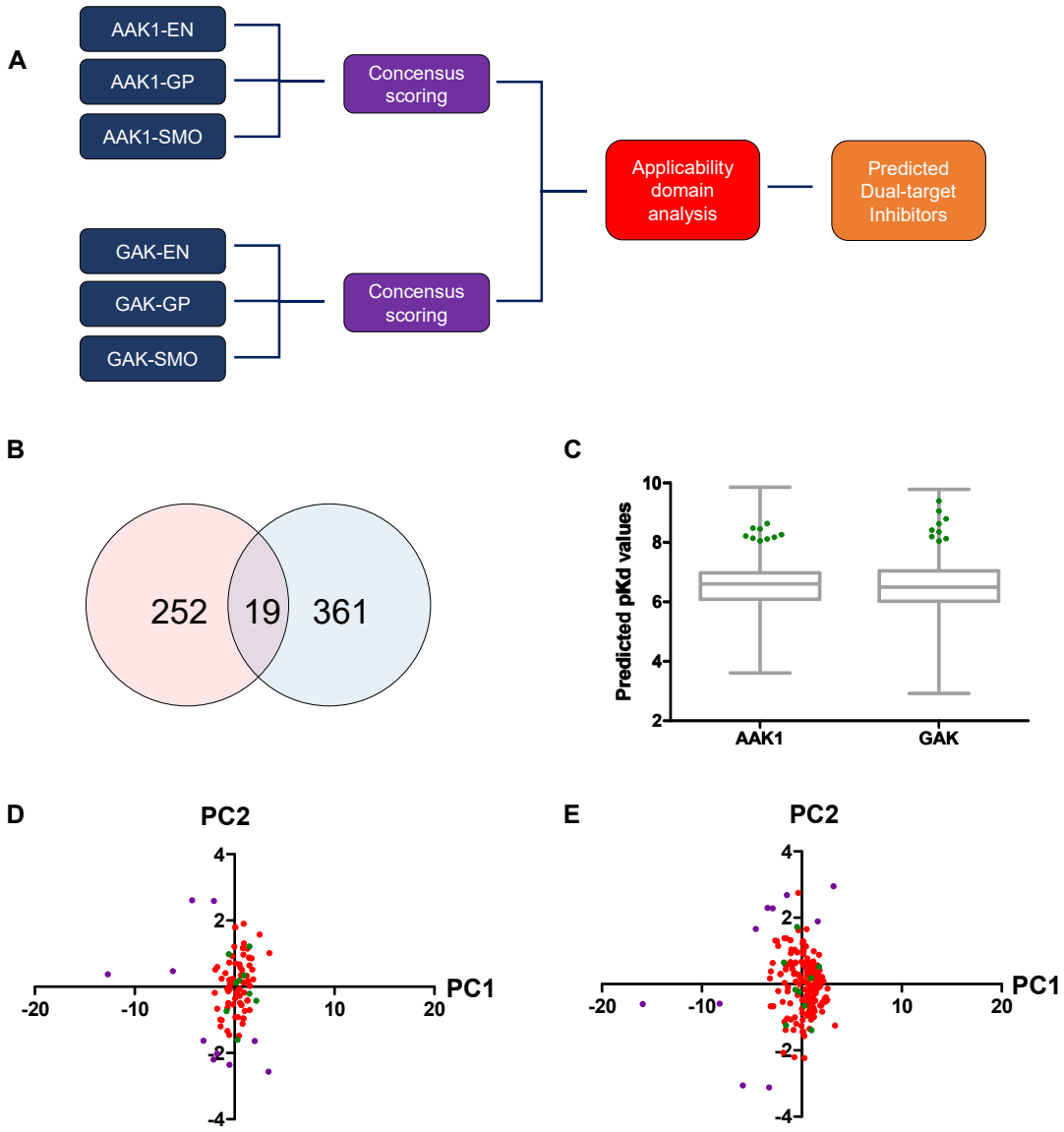


Figure 5 Deployment of QSAR models. A. General strategy for QSAR model deployment using a concensus scoring method. B. Venn diagram showing number of compounds from Drugbank with concensus *pKd* values > 8. Red circle represents AAK1 inhibitors; blue circle represents GAK inhibitors. C. Distribution of concensus *pKd* values obtained from AAK1 and GAK QSAR models. The selected nine compounds shown in green. D. PCA plots for AD analysis for AAK1, and E. GAK. Red dots for compounds in training sets (that define the AD). Green dots for selected Drugbank compounds that fall within the AD. Purple dots for selected Drugbank compounds that fall outside of AD

Table 3. Description of Drugbank compounds that were predicted as AAK1/GAK dual-target inhibitors

	Drugbank ID	Common name	Description
1	DB12675	PF-4995274	Investigational drug against serotonin 4 receptor (5HT-4) for treatment of depression
2	DB12137	GSK-256066	Investigational drug against phosphodiesterase 4B (PDE4B) for treatment of asthma and chronic obstructive pulmonary disease (COPD)
3	DB12066	Orteronel	Investigational drug against Cytochrome P450 for treatment of prostate cancer
4	DB00802	Alfentanil	FDA-approved drug against opioid μ-receptor; used as anesthesia and analgesic
5	DB08219	-	Experimental drug against Cyclin-dependent kinase 2 (CDK2) for treatment against cancer
6	DB12949	PF-03382792	Investigational drug against serotonin 4 receptor (5HT-4)
7	DB04704	-	Experimental drug that binds to RAR-related orphan receptor gamma (RORγ)
8	DB13307	Proscillaridin	Experimental drug against topoisomerase I and II for treatment against cancer
9	DB13185	Oxabolone cipionate	Prodrug of oxabolone, an anabolic-androgenic steroid; used as a performance enhancing drug

3.5 Electrostatic Potential Map Comparison

Chemical compounds often complement their protein target in shape and electrostatics[55]. This implies that chemical compounds with similar shape and electrostatic properties may bind to the same receptor[55]. This principle has been used to identify small molecule inhibitors similar to natural substrates or known inhibitors by screening for compounds with similar shape, volume and electrostatics[55,56,57]. Therefore, we hypothesize that our nine predicted compounds should exhibit similar EPMs as baricitinib and sunitinib. Based on the virtual inspection of the electrostatic features, the EPMs of DB12066, DB12137, DB04704, DB13185 and DB13307 were similar to baricitinib. The EPMs of the compounds showed a balance of positively-charged and negatively-charged regions (Figure 6) with hydrophobic patches (absence of positive or negative electrostatic field). Meanwhile, the EPMs of DB12675, DB12949, DB08219 and DB00802 were similar to that of sunitinib (Figure 6) with the positive electrostatics field featured dominantly.

4 Discussion

To date, the COVID-19 pandemic has affected 178 million people and caused 3.86 million fatalities globally. Despite the availability of vaccines, COVID-19 is likely to be endemic with occasional regional outbreaks globally[58,59]. Besides, rapid environmental changes due to anthropogenic activities are poised to cause future COVID-19-like pandemics and other zoonotic diseases[60]. As such, there is an urgent need to develop broad-spectrum antivirals that may work against future viral pandemics of unknown origin[61]. Current anti-viral drugs often target specific viral proteins; therefore, they are termed as direct-acting antiviral agents (DAAs). However, another strategy to combat viral infections is by designing drugs that target host proteins that viruses often exploit in viral entry and replication, such as AAK1 or GAK. Such drugs are known as host-directed antiviral agents (HDAs). Besides the fact that HDAs are likely to be broad-spectrum antiviral drugs, one study showed that HDAs could provide high genetic barrier to the development of drug resistance[61].

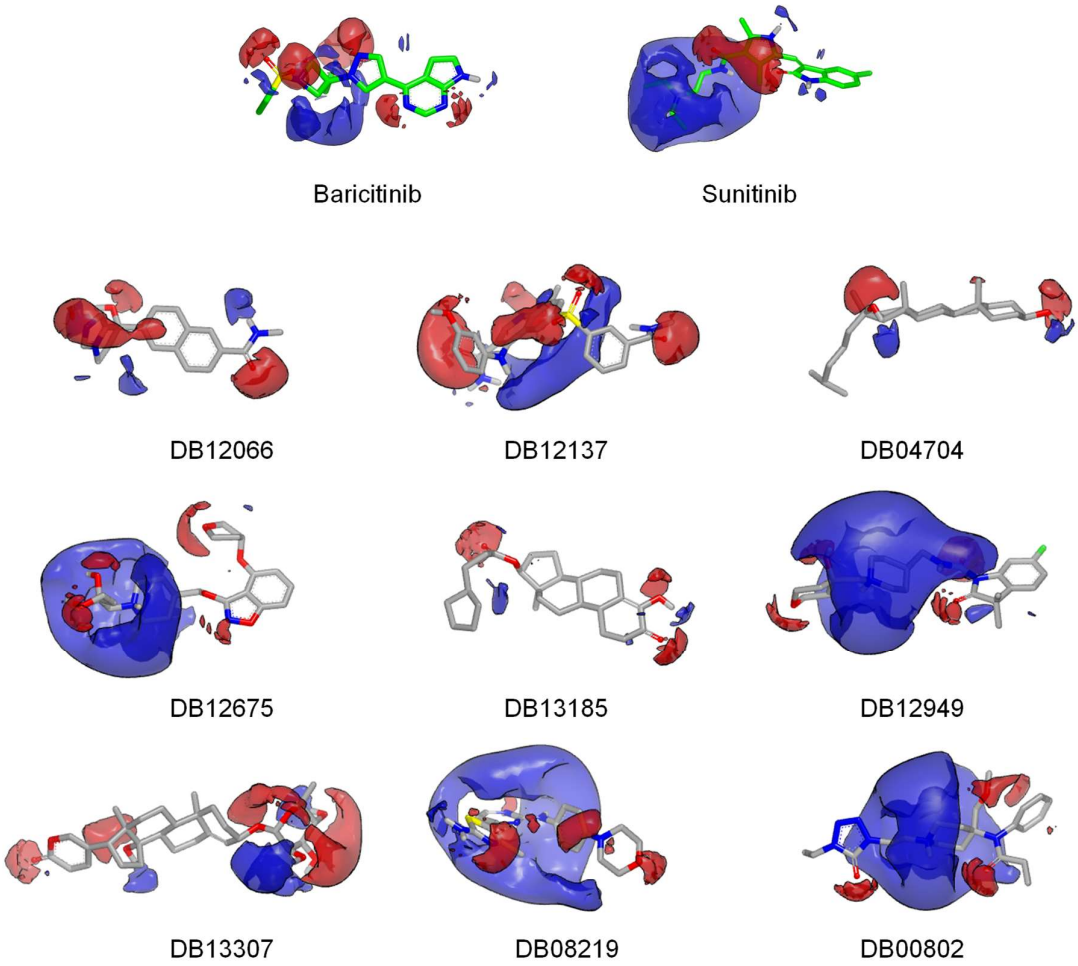


Figure 6. Electrostatics Potential Maps of baricitinib, sunitinib and the predicted dual-target inhibitors. Positive electrostatic field is denoted as blue patches; negative electrostic field is denoted as red patches; hydrophobic field is denoted as absence of both blue or red patches

Therefore, designing HDA drugs that target host proteins such as AAK1 and GAK could be advantageous in future-proofing against pandemics that could occur in years to come.

In this work, we have used machine learning method/approach to construct QSAR models to discover AAK1/GAK dual-target inhibitors. This is in line with the study showing that an erlotinib/sunitinib combination treatment that targets both AAK1 and GAK exhibits higher therapeutic efficacy than targeting either protein alone[12]. We are also interested to design dual-target inhibitors because such drugs could reduce therapeutic doses, exhibit less side effects and reduce the risk of the emergence of drug resistance[62].

First, we constructed three ECFP4-based QSAR models for each of the protein targets AAK1 and GAK. Although our QSAR models performed modestly in predicting the absolute *pKd* values, they were able to correctly rank compounds in the external test set and exhibited good q^2 values when tested with both internal and external validation methods. Next, we used a concensus scoring approach to identify chemical compounds that were predicted to have high concensus *pKd* values for both AAK1 and GAK. In total, 19 chemical compounds from Drugbank were predicted to have *pKd* values > 8 against both target proteins. However, PCA plots showed that only nine compounds fell within the AD of our QSAR models. Out of these nine compounds, two compounds were

identified as potential COVID-19 therapeutics by other labs. Compound GSK-256066 is a phosphodiesterase 4B inhibitor that can inhibit the Main Protease protein of SARS-CoV-2[63]. Likewise, Proscillaridin is predicted to be an inhibitor of the SARS-CoV-2 non-structural protein 14, an N7-methyltransferase[64]. More significantly, Proscillaridin was found to inhibit the post-attachment step of HBV before viral RNA replication through an unknown mechanism[65]. Given that HBV enters primary hepatocytes through CME[66], it is therefore plausible that Proscillaridin could have inhibited AAK1 and/or GAK to prevent entry of HBV into human cells. To further bolster our confidence on our predicted compounds, we calculated the EPMs of two AAK1/GAK dual-target inhibitors (baricitinib and sunitinib) and compared them with our predicted compounds. Five of our compounds showed similar electrostatics to baricitinib while the other four are similar to sunitinib. Therefore, it is plausible that these compounds could mimic baricitinib or sunitinib in interacting with AAK1 and GAK.

To the best of our knowledge, our work represents the first QSAR models constructed using machine learning methods for AAK1 and GAK. One field-based QSAR model for GAK inhibitors was reported by Asquith et al. in 2019[67]. Their QSAR model was constructed based on a series of derivatives bearing the quinoline scaffold. Similar to their findings, our EPMs for our proposed inhibitors have either (a) strong hydrophobic fields or (b) with strong positive electrostatics. Based on their WaterMap analysis, Asquith et al.[67] posit that hydrophobic fields are needed by GAK inhibitors to displace a high-energy water in the inhibitor binding site[67].

5 Conclusion

Targeting AAK1 and GAK is a valid strategy in designing broad-spectrum antivirals against known and future viruses. In our work, we constructed two QSAR models that were able to predict the *pKd* values of AAK1 and GAK inhibitors. However, our QSAR models would benefit from the incorporation of data from AAK1 or GAK inhibitors with higher binding affinities in future. Nonetheless, our models were still able to rank and identify chemical compounds that could potentially be developed as AAK1 or GAK inhibitors. With the QSAR models, we screened the Drugbank compound library and identified nine compounds that performed well in both QSAR models. Hence, from this work, we provided *in silico* justification and sets the foundation for future experimental exploration of these nine compounds as *bona fide* AAK1/GAK dual-target inhibitors capable of inhibiting viral entry into host cells.

Conflict of Interest

The authors declare that they have no conflict of interest.

Appendices or Supplementary Material

Supplementary Table S1.
Performance of ML algorithms in predicting *pKd* values of AAK1 inhibitors when assessed using internal validation methods.

Supplementary Table S2.
Performance of ML algorithms in predicting *pKd* values of GAK inhibitors when assessed using internal validation methods.

Supplementary Figure S1.
Residual plots between experimental and predicted pIC50 values for QSAR models.

Supplementary Figure S2.
Chemical structures of the predicted AAK1/GAK dual-target inhibitors. The Drugbank ID and their predicted *pKd* values against GAK and AAK1 are provided.

Acknowledgment

We would like to acknowledge the Swinburne Research Internship Grant 2020 for funding the studentship of Clement Sim. We also extend our gratitude to OpenEye for providing a free Academic License to use the software ROCS and EON.

References

1. Kaksonen, M., & Roux, A. (2018). Mechanisms of clathrin-mediated endocytosis. *Nature Reviews Molecular Cell Biology, 19*(5), 313–326.
2. McMahon, H. T., & Boucrot, E. (2011). Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nature Reviews Molecular Cell Biology,12*(8), 517–533.
3. Blondeau, F., Ritter, B., Allaire, P. D., Wasiak, S., Girard, M., Hussain, N. K., Angers, A., Legendre-Guillemin, V., Roy, L., Boismenu, D., Kearney, R.E., Bell, A.W., Bergeron, J.J.M., & McPherson, P. S. (2004). Tandem MS analysis of brain clathrin-coated vesicles reveals their critical involvement in synaptic vesicle recycling. *Proceedings of the National Academy of Sciences*, *101*(11), 3833–3838.
4. Pearse, B. M., Smith, C. J., & Owen, D. J. (2000). Clathrin coat construction in endocytosis. *Current Opinion in Structural Biology*, *10*(2), 220–228.
5. Smythe, E. (2002). Regulating the clathrin-coated vesicle cycle by AP2 subunit phosphorylation. *Trends in Cell Biology*, *12*(8), 352-354.
6. Olusanya, O., Andrews, P. D., Swedlow, J. R., & Smythe, E. (2001). Phosphorylation of threonine 156 of the μ2 subunit of the AP2 complex is essential for endocytosis *in vitro* and *in vivo*. *Current Biology, 11*(11), 896–900.
7. Ricotta, D., Conner, S. D., Schmid, S. L., Von Figura, K., & Höning, S. (2002). Phosphorylation of the AP2 μ subunit by AAK1 mediates high affinity binding to membrane protein sorting signals. *The Journal of Cell Biology*, *156*(5), 791–795.
8. Ghosh, P., & Kornfeld, S. (2003). AP-1 binding to sorting signals and release from clathrin-coated vesicles is regulated by phosphorylation. *The Journal of Cell Biology*, *160*(5), 699–708.
9. Conner, S. D., & Schmid, S. L. (2003). Differential requirements for AP-2 in clathrin-mediated endocytosis. *The Journal of Cell Biology*, *162*(5), 773-780.
10. Umeda, A., Meyerholz, A., & Ungewickell, E. (2000). Identification of the universal cofactor (auxilin 2) in clathrin coat dissociation. *European Journal of Cell Biology*, *79*(5), 336-342.
11. Korolchuk, V. I., & Banting, G. (2002).CK2 and GAK/auxilin2 are major protein kinases in clathrin-coated vesicles. *Traffic*, *3*(6), 428-439.
12. Bekerman, E., Neveu, G., Shulla, A., Brannan, J., Pu, S. Y., Wang, S., Xiao, F., Barouch-Bentov, R., Bakken, R.R., Mateo, R., Govero, J., Nagamine, C.M., Diamond, M.S., De Jonghe, S., Herdewijn, P., Dye, J.M., Randall, G., & Einav, S. (2017). Anticancer kinase inhibitors impair intracellular viral trafficking and exert broad-spectrum antiviral effects. *The Journal of Clinical Investigation*, *127*(4), 1338–1352.
13. Robinson, M., Schor, S., Barouch-Bentov, R., & Einav, S. (2018). Viral journeys on the intracellular highways. *Cellular and Molecular Life Sciences*, *75*(20), 3693–3714.
14. Majumdar, P., & Niyogi, S. (2020). ORF3a mutation associated higher mortality rate in SARS-CoV-2 infection. *Epidemiology and Infection*, *148*, e262, 1–6.
15. Bhakta, S. J., Shang, L., Prince, J. L., Claiborne, D. T., & Hunter, E. (2011). Mutagenesis of tyrosine and di-leucine motifs in the HIV-1 envelope cytoplasmic domain results in a loss of Env-mediated fusion and infectivity. *Retrovirology*, *8*(1), 1-17.
16. Neveu, G., Barouch-Bentov, R., Ziv-Av, A., Gerber, D., Jacob, Y., & Einav, S. (2012). Identification and targeting of an interaction between a tyrosine motif within Hepatitis C Virus core protein and AP2M1 essential for viral assembly. *PLoS Pathogens*, *8*(8), e1002845.
17. Dorosky, D., Prugar, L. I., Pu, S., O'Brien, C., Bakken, R., De Jonghe, S., Herdewijn, P., Brannan, J., Dye, J. M., & Einav, S. (2018). AAK1 and GAK inhibitors demonstrate activity against Filoviruses. *Journal of Immunology*, *200*(1 Supplement), 50-57.
18. Wang, C., Wang, J., Shuai, L., Ma, X., Zhang, H., Liu, R., Chen, W., Wang, X., Ge, J., Wen, Z., & Bu, Z. (2020). The serine/threonine kinase AP2-associated kinase 1 plays an important role in rabies virus entry. *Viruses*, *12*(1), 45.
19. Stebbing, J., Krishnan, V., de Bono, S., Ottaviani, S., Casalini, G., Richardson, P. J., ... & Sacco Baricitinib Study Group. (2020). Mechanism of baricitinib supports artificial intelligence-predicted testing in COVID-19 patients. *EMBO Molecular Medicine*, *12*(8), e12697.
20. Wang, P. G., Tang, D. J., Hua, Z., Wang, Z., & An, J. (2020). Sunitinib reduces the infection of SARS-CoV, MERS-CoV and SARS-CoV-2 partially by inhibiting AP2M1 phosphorylation. *Cell Discovery*, *6*(1), 1-5.
21. Vignaux, P. A., Minerali, E., Foil, D. H., Puhl, A. C., & Ekins, S. (2020). Machine learning for discovery of GSK3β inhibitors. *ACS Omega*, *5*(41), 26551-26561.
22. Chen, X., Xie, W., Yang, Y., Hua, Y., Xing, G., Liang, L., Deng, C., Wang, Y., Fan, Y., Liu, H., Lu, T., Chen, Y., & Zhang, Y. (2020). Discovery of dual FGFR4 and EGFR inhibitors by machine learning and biological evaluation. *Journal of Chemical Information and Modeling*, *60*(10), 4640-4652.
23. Yang, M., Tao, B., Chen, C., Jia, W., Sun, S., Zhang, T., & Wang, X. (2019). Machine learning models based on molecular fingerprints and an extreme gradient boosting method lead to the discovery of JAK2 inhibitors. *Journal of Chemical Information and Modeling*, *59*(12), 5002-5012.
24. Zhang, H., Liu, W., Liu, Z., Ju, Y., Xu, M., Zhang, Y., Wu, X., Gu, Q., Wang, Z. and Xu, J. (2018). Discovery of indoleamine 2,3-dioxygenase inhibitors using machine learning based virtual screening. *Medicinal Chemistry Communications*, *9*(6), 937-945.
25. Weidlich, I. E., Filippov, I. V., Brown, J., Kaushik-Basu, N., Krishnan, R., Nicklaus, M. C., & Thorpe, I. F. (2013). Inhibitors for the hepatitis C virus RNA polymerase explored by SAR with advanced machine learning methods. *Bioorganic & Medicinal Chemistry*, *21*(11), 3127-3137.

26. Xu, Z., Yang, L., Zhang, X., Zhang, Q., Yang, Z., Liu, Y., Wei, S., & Liu, W. (2020). Discovery of potential flavonoid inhibitors against COVID-19 3CL proteinase based on virtual screening strategy. *Frontiers in Molecular Biosciences, 7,* 556481.
27. Stephenson, N., Shane, E., Chase, J., Rowland, J., Ries, D., Justice, N., Zhang, J., Chan, L., & Cao, R. (2019). Survey of machine learning techniques in drug discovery. *Current Drug Metabolism, 20*(3)*,* 185-193.
28. Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., & Zhao, S. (2019). Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery, 18*(6)*,* 463-477.
29. Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., Mutowo, P., Atkinson, F., Bellis, L. J., Cibrián-Uhalte, E. and Davies, M. (2017). The ChEMBL database in 2017. *Nucleic Acids Research, 45*(D1), D945-D954.
30. Dias, R., & Kolaczkowski, B. (2017). Improving the accuracy of high-throughput protein-protein affinity prediction may require better training data. *BMC Bioinformatics, 18*(5)*,* 7-18.
31. Rogers, D., & Hahn, M. (2010). Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling, 50*(5), 742-754.
32. Bento, A. P., Hersey, A., Félix, E., Landrum, G., Gaulton, A., Atkinson, F., Bellis, L.J., Veij, M.D., & Leach, A. R. (2020). An open source chemical structure curation pipeline using RDKit. *Journal of Cheminformatics, 12*(1), 1-16.
33. O'Boyle, N. M., & Sayle, R. A. (2016). Comparing structural fingerprints using a literature-based similarity benchmark. *Journal of Cheminformatics, 8*(1), 1-14.
34. Chauhan, J. S., Dhanda, S. K., Singla, D., Agarwal, S. M., Raghava, G. P., & Open Source Drug Discovery Consortium. (2014). QSAR-based models for designing quinazoline/imidazothiazoles/ pyrazolopyrimidines based inhibitors against wild and mutant EGFR. *PLoS One, 9*(7), e101079.
35. Frank, E., Hall, M., Trigg, L., Holmes, G., & Witten, I. H. (2004). Data mining in bioinformatics using Weka. *Bioinformatics, 20*(15), 2479-2479.
36. Hall, M. A. (1999). Class CfsSubsetEval (weka-dev 3.9.5 API). Retrieved from https://weka.sourceforge.io/doc.dev/weka/attributeSelection/CfsSubsetEval.html. Accessed 30 April 2021.
37. Hall, M. A. (1999). Correlation-based feature selection for machine learning. Ph.D. thesis*,* The University of Waikato.
38. Wenderski, T. A., Stratton, C. F., Bauer, R. A., Kopp, F., & Tan, D. S. (2015). Principal component analysis as a tool for library design: A case study investigating natural products, brand-name drugs, natural product-like libraries, and drug-like libraries. *Methods in Molecular Biology, 1263,* 225-242.
39. Suvannang, N., Preeyanon, L., Malik, A. A., Schaduangrat, N., Shoombuatong, W., Worachartcheewan, A., Tantimongcolwat, T., & Nantasenamat, C. (2018). Probing the origin of estrogen receptor alpha inhibition: Via large-scale QSAR study. *RSC Advances, 8*(21), 11344-11356.
40. Roy, K., Kar, S., & Das, R. N. (2015). Validation of QSAR models. In *Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment* (pp.231–286). Academic Press.
41. Gonzalez-Medina, M., & Medina-Franco, J. L. (2017). Platform for Unified Molecular Analysis: PUMA. *Journal of Chemical Information and Modeling, 57*(8), 1735-1740.
42. Tropsha, A. (2010). Best practices for QSAR model development, validation, and exploitation. *Molecular Informatics, 29*(6-7), 476-488.
43. Veerasamy, R., Rajak, H., Jain, A., Sivadasan, S., Varghese, C. P., & Agrawal, R. K. (2011). Validation of QSAR models - strategies and importance. *International Journal of Drug Design and Discovery, 2*(3)*,* 511-519.
44. Golbraikh, A., & Tropsha, A. (2002). Beware of q2!. *Journal of Molecular Graphics and Modelling*, *20*(4), 269-276.
45. Alexander, D. L., Tropsha, A., & Winkler, D. A. (2015). Beware of R 2: Simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models. *Journal of Chemical Information and Modeling, 55*(7), 1316-1322.
46. Rücker, C., Rücker, G., & Meringer, M. (2007). y-randomization and its variants in QSPR/QSAR. *Journal of Chemical Information and Modeling, 47*(6), 2345-2357.
47. Hawkins, P. C., Skillman, A. G., Warren, G. L., Ellingson, B. A., & Stahl, M. T. (2010). Conformer generation with OMEGA: Algorithm and validation using high quality structures from the protein databank and cambridge structural database. *Journal of Chemical Information and Modeling, 50*(4), 572-584.
48. OpenEye Scientific. (2021). Visualization & Communication of Modeling Results. Retrieved from https://www.eyesopen.com/vida.
49. Borovicka, T., Jirina Jr, M., Kordik, P., & Jirina, M. (2012). Selecting representative data sets. In *Advances in data mining knowledge discovery and applications* (pp. 43-70). InTech, Rijeka, Croatia.
50. Probst, D., & Reymond, J. L. (2018). A probabilistic molecular fingerprint for big data settings. *Journal of Cheminformatics, 10*. 66.
51. Low, Z. Y., Farouk, I. A., & Lal, S. K. (2020). Drug repositioning: New approaches and future prospects for life-debilitating diseases and the COVID-19 pandemic outbreak. *Viruses, 12*(9), E1058.
52. Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., Pon, A., Knox, C., & Wilson, M. (2018). DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Research, 46*(D1), D1074-D1082.
53. Khan, K., Benfenati, E., & Roy, K. (2019). Consensus QSAR modeling of toxicity of pharmaceuticals to different aquatic organisms: Ranking and prioritization of the DrugBank database compounds. *Ecotoxicology and Environmental Safety, 168,* 287-297.

54. Valsecchi, C., Grisoni, F., Consonni, V., & Ballabio, D. (2020). Consensus versus individual QSARs in classification: Comparison on a large-scale case study. *Journal of Chemical Information and Modeling, 60*(3), 1215-1223.
55. Voet, A., Berenger, F., & Zhang, K. Y. (2013). Electrostatic similarities between protein and small molecule ligands facilitate the design of protein-protein interaction inhibitors. *PLoS One, 8*(10), e75762.
56. Naylor, E., Arredouani, A., Vasudevan, S. R., Lewis, A. M., Parkesh, R., Mizote, A., Rosen, D., Thomas, J.M., Izumi, M., Ganesan, A., Galione, A., & Churchill, G. C. (2009). Identification of a chemical probe for NAADP by virtual screening. *Nature Chemical Biology, 5*(4), 220-226.
57. Boström, J., Grant, J. A., Fjellström, O., Thelin, A., & Gustafsson, D. (2013). Potent fibrinolysis inhibitor discovered by shape and electrostatic complementarity to the drug tranexamic acid. *Journal of Medicinal Chemistry, 56*(8), 3273-3280.
58. Phillips, N. (2021). The coronavirus is here to stay - here's what that means. *Nature, 590*(7846), 382-384.
59. Torjesen, I. (2021). COVID-19 will become endemic but with decreased potency over time, scientists believe. *British Medical Journal*, *372*, n494.
60. Shaghayegh, G., & Gorji, A. (2021). COVID-19 pandemic: The possible influence of the long-term ignorance about climate change. *Environmental Science and Pollution Research International, 28*(13), 15575-15579.
61. Chitalia, V. C., & Munawar, A. H. (2020). A painful lesson from the COVID-19 pandemic: The need for broad-spectrum, host-directed antivirals. *Journal of Translational Medicine, 18*(1), 1-6.
62. Liu, T., Wan, Y., Xiao, Y., Xia, C., & Duan, G. (2020). Dual-target inhibitors based on HDACs: Novel antitumor agents for cancer therapy. *Journal of Medicinal Chemistry, 63*(17), 8977-9002.
63. Shitrit, A., Zaidman, D., Kalid, O., Bloch, I., Doron, D., Yarnizky, T., Buch, I., Idan Segev, I., Ben Zeev, E., Segev, E., & Kobiler, O. (2020). Conserved interactions required for inhibition of the main protease of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *Scientific Reports, 10*(1), 1-11.
64. Xu, C., Ke, Z., Liu, C., Wang, Z., Liu, D., Zhang, L., Wang, J., He, W., Xu, Z., Li, Y., Yang, Y., Huang, Z., Lv, P., Wang, X., Han, D., Li, Y., Qiao, N., & Liu, B. (2020). Systemic in silico screening in drug discovery for coronavirus disease (COVID-19) with an online interactive web server. *Journal of Chemical Information and modeling, 60*(12), 5735-5745.
65. Okuyama-Dobashi, K., Kasai, H., Tanaka, T., Yamashita, A., Yasumoto, J., Chen, W., ... & Moriishi, K. (2015). Hepatitis B virus efficiently infects non-adherent hepatoma cells via human sodium taurocholate cotransporting polypeptide. *Scientific Reports, 5*(1), 1-14.
66. Huang, H. C., Chen, C. C., Chang, W. C., Tao, M. H., & Huang, C. (2012). Entry of hepatitis B virus into immortalized human primary hepatocytes by clathrin-dependent endocytosis. *Journal of Virology, 86*(17), 9443-9453.
67. Asquith, C.R.M., Laitinen, T., Bennett, J. M., Wells, C. I., Elkins, J. M., Zuercher, W. J., Tizzard, G.J., & Poso, A. (2020). Design and analysis of the 4-anilinoquin(az)oline kinase inhibition profiles of GAK/SLK/STK10 using quantitative structure-activity relationships. *ChemMedChem, 15*(1), 26-49.

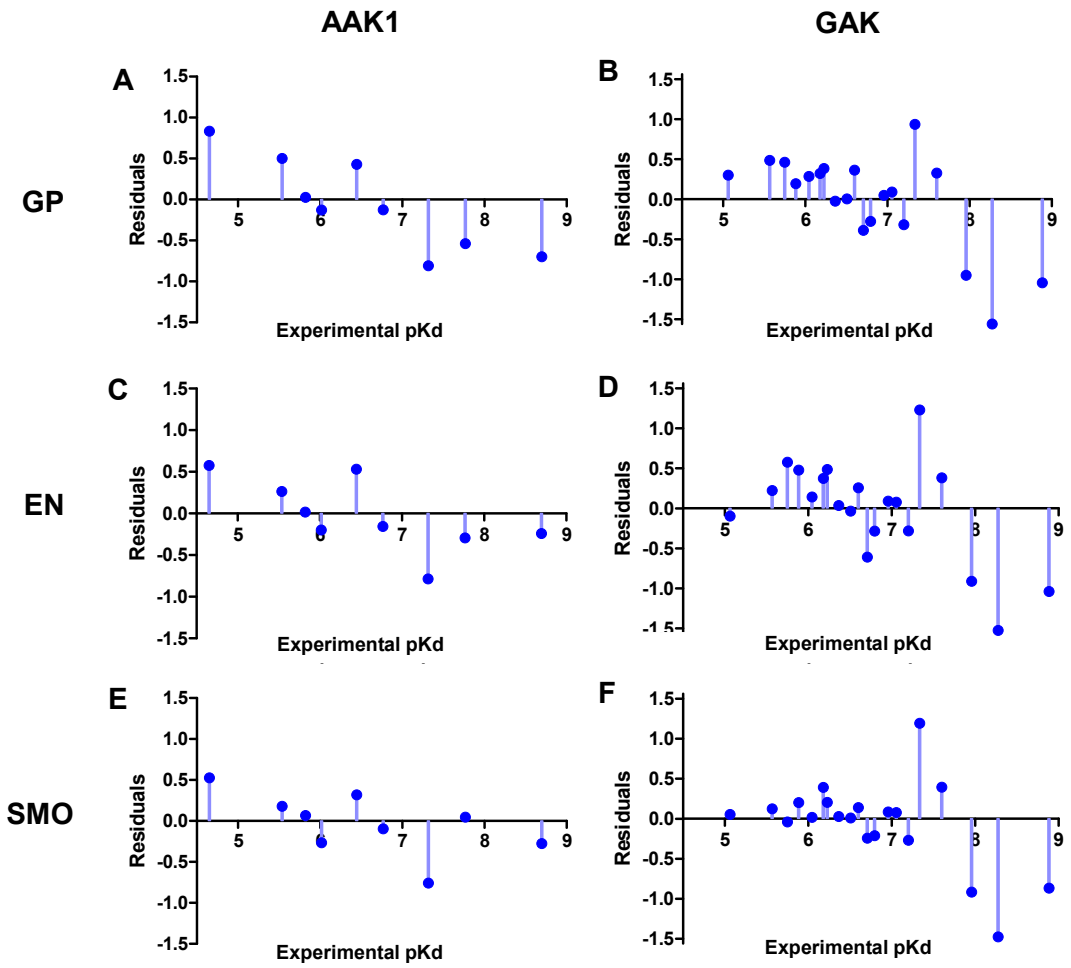
Supplementary Information

Supplementary Table S1. Performance of ML algorithms in predicting *pKd* values of AAK1 inhibitors when assessed using internal validation methods

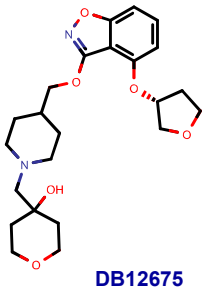
Performance of Machine Learning Algorithms							
	GP	**EN**	**SVM**	**SMO**	**IBK**	**K***	**RF**
Regression Model Statistics							
r^2	0.90	0.90	0.74	0.89	0.94	0.93	0.91
MAE	0.26	0.25	0.79	0.18	0.12	0.16	0.27
RMSE	0.36	0.35	0.95	0.34	0.26	0.28	0.34
10-fold Cross Validation Performance							
q^2_{10CV}	0.80	0.81	0.44	0.82	0.35	0.42	0.52
MAE	0.38	0.37	0.82	0.33	0.67	0.64	0.59
RMSE	0.49	0.48	0.99	0.44	0.84	0.78	0.73
LOO-Cross Validation Performance							
q^2_{LOO}	0.80	0.80	0.52	0.83	0.35	0.41	0.53
MAE	0.37	0.37	0.81	0.29	0.66	0.64	0.58
RMSE	0.48	0.48	0.98	0.42	0.83	0.79	0.72

Supplementary Table S2. Performance of ML algorithms in predicting *pKd* values of GAK inhibitors when assessed using internal validation methods

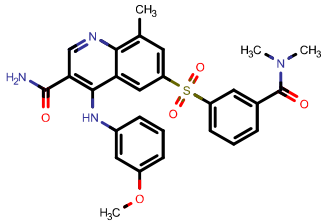
Performance of Machine Learning Algorithms							
	GP	**EN**	**SVM**	**SMO**	**IBK**	**K***	**RF**
Regression Model Statistics							
r^2	0.78	0.74	0.47	0.77	0.85	0.84	0.81
MAE	0.28	0.34	0.57	0.22	0.16	0.19	0.28
RMSE	0.43	0.48	0.75	0.43	0.35	0.36	0.41
10-fold Cross Validation Performance							
q^2_{10CV}	0.66	0.59	0.32	0.66	0.10	0.14	0.22
MAE	0.36	0.42	0.61	0.35	0.65	0.63	0.57
RMSE	0.52	0.58	0.79	0.52	0.94	0.87	0.79
LOO-Cross Validation Performance							
q^2_{LOO}	0.64	0.58	0.33	0.64	0.10	0.15	0.22
MAE	0.37	0.43	0.60	0.36	0.63	0.62	0.56
RMSE	0.54	0.59	0.78	0.54	0.94	0.86	0.79



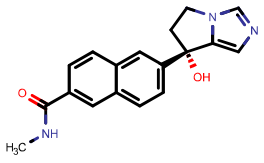
Supplementary Figure S1. Residual plots between experimental and predicted pIC50 values for QSAR models



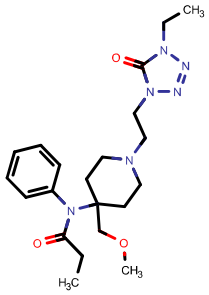
DB12675
AAK1 pKd: 8.65
GAK pKd: 9.07



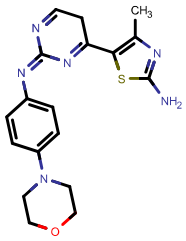
DB12137
AAK1 pKd: 8.16
GAK pKd: 9.41



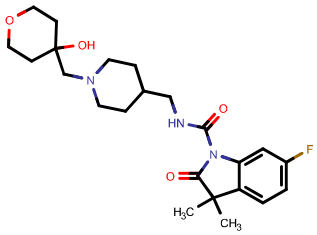
DB12066
AAK1 pKd: 8.47
GAK pKd: 8.43



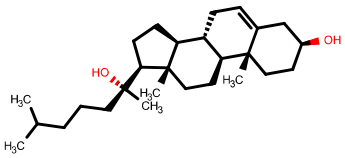
DB00802
AAK1 pKd: 8.28
GAK pKd: 8.06



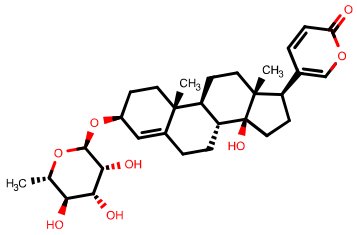
DB08219
AAK1 pKd: 8.50
GAK pKd: 8.81



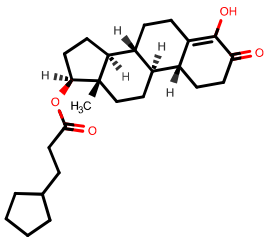
DB12949
AAK1 pKd: 8.07
GAK pKd: 8.64



DB04704
AAK1 pKd: 8.13
GAK pKd: 8.14



DB13307
AAK1 pKd: 8.23
GAK pKd: 8.37



DB13185
AAK1 pKd: 8.19
GAK pKd: 8.21

Supplementary Figure S2. Chemical structures of the predicted AAK1/GAK dual-target inhibitors. The Drugbank ID and their predicted *pKd* values against GAK and AAK1 are provided