

SAR and QSAR in Environmental Research

Quantitative structure–activity relationship study on BTK inhibitors by modified multivariate adaptive regression spline and CoMSIA methods

A. Xu^a, Y. Zhang^a, T. Ran^a, H. Liu^a, S. Lu^a, J. Xu^a, X. Xiong^a, Y. Jiang^a, T. Lu^{a,b} and Y. Chen^{a*}

^aLaboratory of Molecular Design and Drug Discovery, School of Basic Science, China Pharmaceutical University, Nanjing, Jiangsu, P.R. China; ^bState Key Laboratory of Natural Medicines, China Pharmaceutical University, Nanjing, Jiangsu, P.R. China

(Received 20 December 2014; in final form 2 March 2015)

Bruton's tyrosine kinase (BTK) plays a crucial role in B-cell activation and development, and has emerged as a new molecular target for the treatment of autoimmune diseases and B-cell malignancies. In this study, two- and three-dimensional quantitative structure–activity relationship (2D and 3D-QSAR) analyses were performed on a series of pyridine and pyrimidine-based BTK inhibitors by means of genetic algorithm optimized multivariate adaptive regression spline (GA-MARS) and comparative molecular similarity index analysis (CoMSIA) methods. Here, we propose a modified MARS algorithm to develop 2D-QSAR models. The top ranked models showed satisfactory statistical results (2D-QSAR: $Q^2 = 0.884$, $r^2 = 0.929$, $r^2_{\text{pred}} = 0.878$; 3D-QSAR: $q^2 = 0.616$, $r^2 = 0.987$, $r^2_{\text{pred}} = 0.905$). Key descriptors selected by 2D-QSAR were in good agreement with the conclusions of 3D-QSAR, and the 3D-CoMSIA contour maps facilitated interpretation of the structure–activity relationship. A new molecular database was generated by molecular fragment replacement (MFR) and further evaluated with GA-MARS and CoMSIA prediction. Twenty-five pyridine and pyrimidine derivatives as novel potential BTK inhibitors were finally selected for further study. These results also demonstrated that our method can be a very efficient tool for the discovery of novel potent BTK inhibitors.

Keywords: Bruton's tyrosine kinase; multivariate adaptive regression spline; genetic algorithms; CoMSIA; SAR

1. Introduction

Bruton's tyrosine kinase (BTK) is a non-receptor tyrosine kinase of the Tec kinase family only expressed in B cells [1]. Studies demonstrated that BTK is intimately involved in multiple signal transduction pathways regulating activation, proliferation, survival and differentiation of B-lineage lymphoid cells [2–6]. BTK serves as an important character for the function of osteoclasts, monocytes and mast cells, and its deficiency in humans is concerned with reduction in tumour necrosis factor- α (TNF- α) production by activated monocytes [7]. Consequently, BTK has emerged as a new molecular target for the treatment of autoimmune, allergic disorders and inflammatory diseases, including rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), asthma, idiopathic thrombocytopenic purpura (ITP) and other related disorders [8–11].

*Corresponding author. Email: ydchen@cpu.edu.cn

Since BTK was discovered as a drug target in the 1990s, a variety of BTK inhibitors have been developed [12] such as dasatinib [13], LFM-A13 [14], CGI-1746 [15], RN-486 [16,17], GDC-0834 [18] and ibrutinib [19]. However, most inhibitors failed in clinical trials and only ibrutinib was approved, in November 2013, to treat patients with mantle cell lymphoma [3]. Thus, there is still an unmet need to design promising and potent BTK inhibitors as lead candidates for the drug development.

Nowadays, quantitative structure–activity relationship (QSAR) methods are widely applied to predict the activities of newly designed drugs and to identify key chemical features essential for inhibitory activities [20]. Therefore, QSAR could provide structural insights to optimize a certain property and then design the molecules containing desired features [21]. Two-dimensional (2D) QSAR is a well-known computational method widely used in drug design, which has been successfully applied to the prediction and design of kinase inhibitors [22], such as inhibitors of multidrug transporter P-glycoprotein [20], human β -secretase (BACE-1) [23] and HIV-1 non-nucleoside reverse transcriptase [24].

A number of factors including the variable selection may affect the quality of 2D-QSAR model. The most commonly used variable selection methods are forward selection [25], backward elimination [25,26], stepwise regression [27], genetic algorithms [28,29] and so on. Recently, multivariate adaptive regression splines (MARS) technique was developed as a tool for QSAR study of SARS inhibitors [30], which was firstly proposed by Friedman [31]. Basically, it divides the whole space of each independent variable into various sub-regions and then defines a different mathematical equation for each area [32], which makes each sub-region of independent variable respond separately to the system. Thus the MARS approach is different from other regression methods and useful for modelling non-linear and complicated systems [33,34]. Generally, only a few molecular descriptors are essential to develop a MARS model. To select the most important descriptors and decrease the over-fitting risk simultaneously, an in-house program was coded in our laboratory using a genetic algorithm (GA) [35], which could help to establish a robust 2D-QSAR model. In addition, three-dimensional (3D) QSAR can take into account the structure-based information and lead to high quality contour maps to establish a correlation between chemical structures and biological activities [36]. In particular, comparative molecular similarity indices analysis (CoMSIA) is able to handle the data better than other methods due to its lower sensitivity to alignment and smoother Gaussian function-based potentials [36].

In the present work, 2D- and 3D-QSAR models were developed on pyridine and pyrimidine derivatives which have been reported as potent BTK inhibitors. A GA optimized MARS method (GA-MARS), combined with analysis of variance (ANOVA), was applied to select important descriptors for the 2D-QSAR model [17,37]. The modified model not only maintains high capacity in handling large data sets and modelling nonlinearities [34], but also increases the numerical stability of the model construction process and attains better robustness. The 2D-QSAR model was validated by Y-randomization, 10-fold cross-validation, external validation and calculation of the applicability domain. In addition, a CoMSIA model was established to interpret the impact of physicochemical and structural parameters of compounds on inhibitory activity by obtaining the contour maps. As a useful tool to accurately predict activities of newly designed inhibitors, the combination of GA-MARS based 2D-QSAR and CoMSIA provided insightful information on structural modifications for more potent BTK inhibitors.

2. Materials and methods

2.1 Data set preparation

The chemical structures and biological activities of 68 pyridine and pyrimidine-based novel BTK inhibitors (Table 1) were collected from the literature [37]. The inhibitory activities IC_{50} values were converted into negative logarithmic scale pIC_{50} [$-\log IC_{50}$ (M)]. The pIC_{50} values were then used as the dependent variables in subsequent GA-MARS and CoMSIA analysis. Considering the structural diversity and inhibitory activity gradients, we randomly divided the 68 compounds into a training set and a test set in a ratio of 5:1.

2.2 2D-QSAR method

2.2.1 Preprocessing descriptors

Various physiochemical properties were selected as descriptors used for ligand-based QSAR model constructions. A total of 316 molecular descriptors were calculated using MOE (Molecular Operating Environment), v.2009.10. To improve the robustness of the model, the initial 316 features were preprocessed to reduce redundancy. Firstly, descriptors with 80% zero values out of the total training compounds were eliminated. Secondly, descriptors with very small standard deviation values ($<0.5\%$) were removed. Thirdly, pairs of descriptors with a correlation coefficient >0.90 were classified as inter-correlated, and the one with greater F -score [Equation (1)] in each correlated pair was kept. Finally, a total of 173 molecular descriptors (see Table S1 in supplementary material which is available via the multimedia link on the online article webpage) were selected after the preprocessing, and their values were scaled to a range of 0 to 1.

$$F_i = \frac{(\bar{x}_i^+ - \bar{x}_i)^2 + (\bar{x}_i^- - \bar{x}_i)^2}{\frac{1}{n_+ - 1} \sum_{k=1}^{n_+} (x_{k,i}^+ - \bar{x}_i^+)^2 + \frac{1}{n_- - 1} \sum_{k=1}^{n_-} (x_{k,i}^- - \bar{x}_i^-)^2} \quad (1)$$

where $\bar{x}_i$ is the mean of the entire data set in feature i , $\bar{x}_i^+$ is the mean of positive data set in feature i , $\bar{x}_i^-$ is the mean of negative data set in feature i , $x_{k,i}^+$ is the value of the positive type sample k in feature i , and $x_{k,i}^-$ is the value of the negative type sample k in feature i .

2.2.2 GA-MARS

An in-house developed program GA-MARS was used to select key descriptors and attain better robustness. GA initializes a population of solutions randomly and then improves the fitness through iterations of mutation, crossover and selection. Each feature subset is referred to as a chromosome. The value of the feature subset is 0 or 1, where 0 represents the corresponding descriptor is abandoned and 1 indicates to keep the descriptor. To reduce over-fitting risk, the training set was randomly divided into a T1 set and a T2 set at a ratio of 4:1 in the GA-MARS model. The T2 set was considered as an internal validation set to evaluate the model, which was trained by the T1 set. The internal validation parameter r^2_{T2} (non-cross-validated correlation coefficient for the T2 set) of the MARS model was used as the fitness of GA and feature subset was selected simultaneously by GA [38,39] so that the overall MARS prediction accuracy was improved during the training process. Parameters of GA were set as follows: (i) the generation number was set to 1000; (ii) the crossover rate was 0.8; (iii) the mutation rate was 0.1; and (iv) the initial population was 100.

Table 1. Chemical structures of BTK inhibitors with their experimental and predicted activities.

Compd.	Substituents							pIC_{50exp}	pIC_{50pred}	
	R_1	R_2	R_3	R_4	R_5	X	A		GA-MARS	CoMSIA

1		CH ₃	CH ₃	H			N	N	5.97	6.16	5.95
2		H	H	H			N	N	6.92	6.78	6.95
3		CH ₃	CH ₃	H			N	N	6.98	7.09	7.08
4		CH ₃	CH ₃	H			N	N	6.49	6.47	6.65
5		CH ₃	CH ₃	H			N	N	6.90	7.13	6.83
6	-CONHCH ₃	CH ₃	CH ₃	H			N	N	7.00	6.93	7.18
7		CH ₃	CH ₃	H			N	N	6.96	6.93	6.97
8		CH ₃	H	H			N	N	6.92	6.78	6.79
9		CH ₃	CH ₃	H			N	N	7.35	7.17	7.30
10	-CONHCH ₂ CH ₃	CH ₃	CH ₃	H			N	N	6.91	7.45	7.12
11		CH ₃	H	H			N	C	5.77	5.80	5.85
12		CH ₃	H	H			N	C	6.81	6.31	6.85
13		CH ₃	H	H			N	C	6.09	6.05	6.01
14		CH ₃	H	H			N	C	4.79	4.92	4.72
15		CH ₃	H	H			N	C	6.69	6.43	6.61
16		CH ₃	H	H			N	C	4.47	4.38	4.84
17		CH ₃	H	H			N	N	6.68	6.77	6.65
18		CH ₃	H	H			N	C	5.63	5.71	5.49

(Continued)

Table 1. (Continued).

Compd.	Substituents							pIC_{50exp}	pIC_{50pred}	
	R_1	R_2	R_3	R_4	R_5	X	A		GA-MARS	CoMSIA
19		CH ₃	H	H		N	C	4.21	4.41	4.18
20		CH ₃	CH ₃	H		N	N	6.91	6.90	6.77
21		CH ₃	CH ₃	H		N	N	4.81	4.80	4.88
22	-CONHCH ₃	CH ₃	CH ₃	H		C	N	6.88	6.96	6.86
23		CH ₃	-CH ₂ OH	H		C	N	7.51	7.80	7.43
24	-CONHCH ₃	CH ₃	CH ₃	H		C	N	7.40	6.81	7.28
25	-CONHCH ₃	CH ₃	CH ₃	H		C	—	6.95	6.65	7.03
26	-CONHCH ₃	CH ₃	CH ₃	H		C	N	7.02	6.91	6.85
27	-CONHCH ₃	CH ₃	CH ₃	H		C	N	5.64	6.16	5.71
28		CH ₃	CH ₃	H		C	N	6.89	7.01	6.77
29		CH ₃	CH ₃	H		C	N	7.42	7.25	7.42
30	-CONHCH ₃	CH ₃	CH ₃	H		C	N	6.51	6.48	6.52
31	-CONHCH ₃	CH ₃	CH ₃	H		C	N	6.64	7.10	6.62
32	-CONHCH ₃	CH ₃	H	H		C	C	5.85	5.62	5.77
33	-CONHCH ₃	CH ₃	H	H		C	C	5.36	5.36	5.18
34	-CONHCH ₃	CH ₃	H	H		C	C	4.83	5.09	4.80
35	-CONHCH ₃	CH ₃	H	H		C	C	5.81	5.91	5.98
36	-CONHCH ₃	-CH ₃	H	H		C	C	5.46	5.47	5.43
37	-CONHCH ₃	-CH ₃	H	H		C	C	4.26	4.17	4.24
38	-CONHCH ₂ CH ₃	-CH ₃	-CH ₃	H		C	N	7.89	7.16	7.76
39		-CH ₃	-CH ₂ OH	H		C	N	7.70	7.74	7.77

(Continued)

Table 1. (Continued).

Compd.	Substituents							pIC_{50exp}	pIC_{50pred}	
	R_1	R_2	R_3	R_4	R_5	X	A		GA-MARS	CoMSIA
40		-CH ₃	-CH ₃	H		C	N	6.51	6.94	6.58
41		-CH ₃	-CH ₂ OH	H		C	N	7.82	7.73	7.92
42		-CH ₃	-CH ₂ OH	H		C	N	7.62	7.62	7.51
43		-CH ₃	-CH ₂ OH	H		C	N	7.12	7.16	7.30
44	-CONHCH ₂ CF ₃	-CH ₃	-CH ₃	H		C	N	6.06	6.05	6.06
45		-CH ₃	-CH ₂ OH	H		C	N	7.30	7.22	7.35
46		-CH ₃	-CH ₂ OH	H		C	N	6.67	6.95	6.58
47		-CH ₃	-CH ₂ OH	H		C	N	7.37	7.08	7.32
48		-CH ₃	-CH ₂ OH	H		C	N	7.27	6.94	7.26
49		-CH ₃	-CH ₂ OH	H		C	N	7.22	7.44	7.29
50		-CH ₃	-CH ₂ OH	H		C	N	7.34	7.39	7.25
51		-CH ₃	-CH ₃	H		C	N	5.91	6.04	6.00
52		-CH ₃	-CH ₂ OH	H		C	N	6.53	6.44	6.52
53		-CH ₃	-CH ₂ OH	H		C	N	7.62	7.68	7.52
54		-CH ₃	-CH ₂ OH	H		C	N	7.70	7.51	7.55
55		-CH ₃	-CH ₂ OH	H		C	N	7.05	7.46	7.15
56		-CH ₃	-CH ₂ OH	H		C	N	7.96	8.30	8.07
57		-CH ₃	-CH ₂ OH	H		C	—	8.40	7.84	8.46
58*		-CH ₃	-CH ₃	H		N	N	7.29	6.84	7.24
59*		-CH ₃	-CH ₃	H		N	N	7.03	7.01	6.54

(Continued)

Table 1. (Continued).

Compd.	Substituents							pIC_{50exp}	pIC_{50pred}	
	R_1	R_2	R_3	R_4	R_5	X	A		GA-MARS	CoMSIA
60*		-CH ₃	H	H		N	C	5.66	6.18	5.15
61*		-CH ₃	H	H		N	C	6.17	6.46	6.20
62*		-CH ₃	-CH ₃	H		N	N	6.57	6.57	6.09
63*		-CH ₃	-CH ₂ OH	H		C	N	7.17	6.97	7.01
64*		-CH ₃	-CH ₂ OH	H		C	N	7.3	7.24	7.34
65*		-CH ₃	-CH ₂ OH	H		C	N	7.15	6.92	7.35
66*		-CH ₃	-CH ₂ OH	H		C	N	7.4	6.95	7.50
67*		-CH ₃	-CH ₂ OH	H		C	N	6.8	6.80	6.91
68*		-CH ₃	-CH ₂ OH	H		C	N	7.47	7.18	7.46

*Compounds in test set.

2.2.3 2D-QSAR model validation

The model was further validated using Y-randomization, 10-fold cross-validation, external validation and applicability domain calculation. Y-randomization was used to ensure that there is no random correlation within the data matrix [40] and the 10-fold cross-validation of training set was used to test the robustness of the model. External validation was carried out to test the model's prediction ability, while the applicability domain calculation was applied to investigate the possible presence of outliers in the whole data set.

2.3 3D-QSAR method

The CoMSIA method is used to predict the biological activities of a set of aligned compounds as well as to supply a visual guide in three-dimensions about the location of property fields crucial for activity [41]. These property fields are based on similarity indices of aligned compounds and are calculated by a partial least squares (PLS) analysis, which is a generalization of regression.

2.3.1 Molecular modelling and alignment

The CoMSIA models were built using the same data set as the GA-MARS model, i.e. 57 compounds as the training set for developing model and 11 compounds as the test set for

external validation (Table 1). All compounds were optimized using the LigPrep program from Schrödinger 2009. Since molecular alignment is crucial for the quality of CoMSIA models, we adopted docking-based alignments by Glide to search active conformation in this study. The crystal structure (PDB ID: 3GEN) of the kinase domain of BTK bound to a pyrrolopyrimidine-containing compound was taken as the receptor structure and prepared with the *Protein Preparation Wizard workflow*. All the water molecules were removed and the binding region was defined by an enclosing box centred on the centroid of the crystal ligand and in similar size to the ligand. The standard precision (SP) mode was applied for the docking and scoring. All other parameters were kept in default. The best pose based on the Glide score and the protein–ligand interaction was obtained [42]. CoMSIA studies were then performed based on the molecular alignment as described above so as to build a robust 3D-QSAR model. The aligned molecules were placed in a 3D grid box of 2.0 Å such that the entire set was included in the simulation box. The CoMSIA descriptors, steric, electrostatic, hydrophobic, hydrogen bond donor (HBD) and hydrogen bond acceptor (HBA) fields with a distance-dependent dielectric at each grid point were calculated using the default settings in SYBYL.

2.3.2 Training and internal validation

Leave-one-out cross-validation (LOOCV) was performed to gain the optimum number of components leading to the highest cross-validated coefficient q^2 [Equation (2)], which is necessary but not sufficient to prove predictive ability and the robustness of models. After that, non-cross-validation was carried out to acquire the final PLS regression models with F -ratio and the explained variance r^2 .

$$q^2 = 1 - \frac{\sum_y (Y_{\text{predict}} - Y_{\text{exp}})^2}{\sum_y (Y_{\text{exp}} - Y_{\text{mean}})^2} \quad (2)$$

where Y_{exp} is the experimentally obtained pIC_{50} of the BTK inhibitor, Y_{predict} is the predicted pIC_{50} and Y_{mean} is the mean predicted pIC_{50} .

2.3.3 External validation

The quality of the external prediction was evaluated by the standard deviation of error prediction (r^2_{pred}), according to Equation (3):

$$r^2_{\text{pred}} = 1 - \frac{\sum_y (Y_{\text{predict}(\text{test})} - Y_{\text{exp}(\text{test})})^2}{\sum_y (Y_{\text{exp}(\text{test})} - Y_{\text{mean}(\text{training})})^2} \quad (3)$$

where $Y_{\text{predict}(\text{test})}$ and $Y_{\text{exp}(\text{test})}$ indicate the predicted and experimental pIC_{50} of the test set compounds, respectively, and $Y_{\text{mean}(\text{training})}$ indicates the mean predicted pIC_{50} value of the training set. The activity values of the inhibitors in the external test set (11 compounds) were predicted utilizing the same CoMSIA calculation parameters as in the model's generation. The non-cross-validated analyses were applied to display the coefficient contour maps. Finally, CoMSIA results were graphically presented by field contour maps, where the coefficients were generated using the StDev*Coefffield type [43]. The favoured and disfavoured level was fixed at 80% and 20%, respectively.

2.4 Molecular fragment replacement (MFR)

As an integrated tool for scaffold replacement and fragment linking and growing, MFR aims only at replacing a fragment of the original molecule, maintaining the other structural elements that are known or suspected to have a critical role in ligand binding; this makes it distinct from the other general virtual screening [44,45]. MOE's scaffold replacement module was applied to generate the fragment database for further screening. The compounds selected as leads were modified by removing some groups of their structures. Based on the suggestions gained from 3D-QSAR models, two pharmacophore models were established respectively for R₁ and R₅ substituents to screen the fragment database as a 3D SMARTS filter. As molecular weight, topological polar surface area (TPSA) and SlogP are used for finding drug-like molecules [46]; these three molecular properties of the resultant molecules were estimated during the search phase to avoid undesirable investigating linkers. Finally, we selected *Add group to ligand mode* and extended a compound at one or more sites to find suitable fragments for constructing potential leads.

3. Results and discussion

3.1 2D-QSAR analysis

3.1.1 GA-MARS model generation and validation

The training set molecules were used to build the GA-MARS model, following the workflow of molecular descriptor selection, spline ratios, knot points and optimum number of basic functions. Initially, 316 molecular descriptors were calculated for each compound, which covered physical, topological, geometrical and electronic properties. After the preprocessing, 173 descriptors were retained (see Table S1 in the supplementary material which is available via the multimedia link on the online article webpage) for building the GA-MARS model. The basic functions and coefficients in the developed MARS model are shown in Table 2. Every basic function in the MARS model is an equation of linear regression for a definite region of an independent variable, which describes the relationship between a molecular descriptor and the activity of the compound. The overall model is a weighted sum of basic functions and used to predict the activity of BTK inhibitors. ANOVA decomposition of the MARS model (Table 3) was used to estimate the importance of each descriptor. The standard of deviation of the function indicates its relative importance to the overall model. The generalized

Table 2. Definition of the basic functions (BF) and their coefficients using the GA-MARS technique.

BF. no.	Definition
1	5.19[E _{tor} -0.52]+[SlogP_VSA1-0.23]+
2	4.15[E _{tor} -0.52]+[0.23-SlogP_VSA1]-
3	0.38[PM3_HOMO-0.58]+
4	9.89[E _{tor} -0.52]+[SMR_VSA5-0.31+
5	0.47[0.30-SlogP_VSA7]-
6	2.34[logP(o/w)-0.37]+[0.61-MNDO_dipole]-
7	7.01[SlogP_VSA7-0.30]+[0.34-Q_VSA_FPOL]-
8	0.51[PM3_HOMO-0.58]+[0.25-a _{acc}]-
9	3.97[PM3_HOMO-0.58]+[0.26-a _{acc}]-
10	0.40[0.29-BCUT_SLOGP_1]

Table 3. ANOVA decomposition of the GA-MARS model.

ANOVA function	STD ^a	GCV ^b	Descriptor(s)
1	0.09	0.050	BCUT_SLOGP_1
2	0.04	0.017	PM3_HOMO
3	0.07	0.038	SlogP_VSA7
4	0.05	0.024	a_acc,PM3_HOMO
5	0.15	0.085	E_tor,SlogP_VSA1
6	0.19	0.141	E_tor,SMR_VSA5
7	0.05	0.020	logP(o/w),MNDO_dipole
8	0.19	0.150	Q_VSA_FPOL,SlogP_VSA7

^aStandard deviation of the ANOVA function.^bGeneralized cross-validation error values.

cross-validation error values (GCV) provide another indication of the significance of the corresponding ANOVA function. The GCV is a score for a model with all of the basic functions corresponding to that particular ANOVA function removed, which can be utilized to evaluate whether this function could contribute significantly to the model.

Some parameters were shown to be crucial for the activity (Table 3), including SlogP_VSA7 (sub-divided surface areas), SlogP_VSA1, SMR_VSA5, a_acc (number of hydrogen bond acceptor atoms), logP(o/w) (log of the octanol–water partition coefficient), E_tor (torsion potential energy), MNDO_dipole (MOPAC descriptors, the dipole moment calculated using the MNDO Hamiltonian), PM3_HOMO (MOPAC descriptors, the energy of the highest occupied molecular orbital calculated using the PM3 Hamiltonian) and Q_VSA_FPOL (fractional polar van der Waals surface area). Although the same term such as (E_tor -0.52) occurred multiple times in different fitting coefficients, it was not contradictory. For example, the fitting coefficient of basic function 1 was 5.19 only when E_tor >0.52 and SlogP_VSA1 >0.23; the fitting coefficient of basic function 2 was 4.15 only when E_tor >0.52 and SlogP_VSA1 <0.23; and the fitting coefficient of basic function 4 was 9.89 only when E_tor >0.52 and SMR_VSA5 >0.31. According to the GCV score, the most significant factors were E_tor, Q_VSA_FPOL and SlogP_VSA7; this was consistent with the fact that molecular activity was closely related to the ‘U’ shaped active conformation (Figure 5) and hydrophobic interaction.

The final GA-MARS model was validated through Y-randomization test and 10-fold cross-validation (Table 4). In the Y-randomization test, all values of r^2_{pred} were <0, indicating there was no possibility of correlation or structural dependency of the training set. In terms of 10-fold cross-validation, the GA-MARS model yielded a Q^2 of 0.884 and r^2 of 0.929 on the

Table 4. Validation of GA-MARS model.

Model	Q^2 ^a	r^2 ^b	r^2_{pred} ^c	$RMSE_{\text{test}}$ ^d	Y-randomization($n = 30$)
GA-MARS	0.884	0.929	0.878	0.292	<0

^a10-fold cross-validated correlation coefficient.^bNon cross-validated correlation efficient for the train set.^cPredicted correlation coefficient for the test set.^dRoot mean square error (RMSE) for the test set.

training set and a r^2_{pred} of 0.878 on the test set (Figure 1A). High values of Q^2 and r^2_{pred} indicated that our 2D-QSAR model had good robustness and predictive power.

3.1.2 Applicability domain

The Williams graph (Figure 2) was used to investigate the possible presence of outliers in the whole data set, in which the standardized residuals and the leverage value (h) are plotted [32]. If the leverage value (h) of a compound is higher than the critical value (h^*), the prediction of the compound can be considered as unreliable [47]. The cutoff line (high leverage value) for this investigation was calculated as $h^* = 0.353$. Besides, a value of $\pm 3\delta$ for standardized residual is widely used as a cutoff value for accepting the predicted value due to its 99% coverage of normally distributed data. As illustrated in Figure 1, all the molecules were within the applicability domain, and therefore the established model was reliable and could be used to predict the activity of new pyridine and pyrimidine inhibitors.

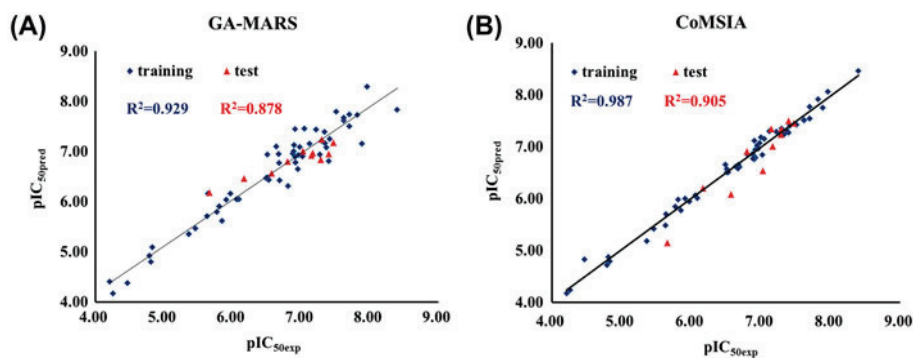


Figure 1. Correlation between the predicted and experimental activities of the training and test set compounds: (A) GA-MARS model and (B) CoMSIA model.

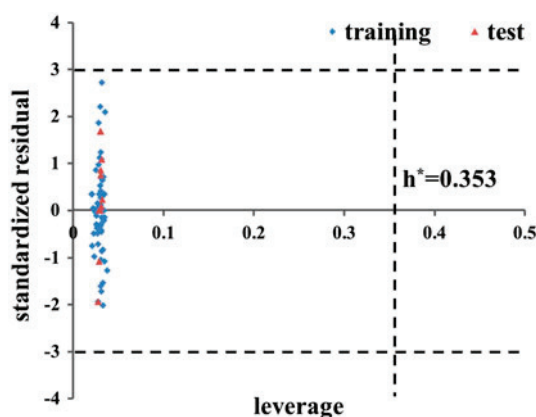


Figure 2. Williams plot of GA-MARS model for the training and test sets.

3.2 3D-QSAR analysis

3.2.1 Molecular alignment

The CoMSIA model was developed and evaluated with the same training and test sets used in 2D-QSAR analysis. Molecular alignment is the most sensitive factor affecting the quality of 3D-QSAR models. The common structure-based alignment has been widely applied, while an alignment based on the active conformations is supposed to be more reasonable. In this study, the conformation of crystal structure (PDB ID: 3GEN) was well reproduced by the best docking conformation [root mean square deviation (RMSD) = 0.4577] and thus docking conformation-based alignment was adopted for building the CoMSIA model. All the training and test set compounds were docked into the active site of BTK and the best pose for each compound was used for molecular alignment (see Figure S1 in supplementary material which is available via the multimedia link on the online article webpage). The alignment based on docking conformations also contributed to understanding the contour maps of the CoMSIA models [44].

3.2.2 CoMSIA analysis

After trying various combinations of the different fields, the best CoMSIA model was developed based on five fields (steric, electrostatic, hydrophobic, hydrogen bonding donor and hydrogen bonding acceptor). For the CoMSIA model, the LOO cross-validated PLS analysis of this model resulted in a q^2 of 0.616 using six principal components. The non-cross-validated PLS analysis yielded a higher r^2 of 0.987 with a very low standard error of estimate (SEE) 0.115 (Table 5). The contributions to the variance of the steric, electrostatic, hydrophobic, hydrogen-bonding donor and hydrogen-bonding acceptor fields were 8.9%, 14.5%, 29.2%, 25.5% and 21.9%, respectively. Moreover, the CoMSIA model demonstrated good predictive ability with a r^2_{pred} of 0.905 against the test set (Figure 2B). Experimental and predicted activities for 68 compounds are shown in Table 1.

Table 5. Statistical parameters of the best CoMSIA model.

Statistical parameter	CoMSIA
q^{2a}	0.616
ONC ^b	6
SEE ^c	0.115
r^{2d}	0.987
F^e	657.303
$r^2_{\text{pred}}{}^f$	0.905
Fraction of field contributions	
Steric	0.089
Electrostatic	0.145
Hydrophobic	0.292
Hydrogen bond donor	0.255
Hydrogen bond acceptor	0.219

^aLeave-one-out(LOO) cross-validated correlation coefficient.

^bONC, optimum number of components.

^cSEE, standard error of estimate.

^dNon cross-validated correlation efficient.

^e F -test value.

^fPredicted correlation coefficient for the test set.

3.2.3 CoMSIA contour maps

To visualize the information contained in the 3D-QSAR models, contour maps for CoMSIA were generated, which provided a comprehensive understanding of key structural requirements responsible for the biological activity. Five CoMSIA contour fields based on PLS analysis are shown in Figure 3. The steric contour maps gave information about the spatial volume of substituted groups on different positions. There were two green and three red contour maps located in the active site, while green means bulky groups were favoured and red means bulky groups were disfavoured.

The green contour map appeared on the position of R_3 group, indicating that bulky groups in this region were beneficial for improving activity. For example, when a methyl was added to the site of R_3 of Compound 8 (Table 1), inhibitory activity was improved slightly as seen in Compound 7. For electrostatic contour maps, increasing positive charges would lead to enhance the inhibitory activity in blue regions while the presence of negative charges was favoured and may result in higher activity in red regions.

The red contour map was found to cover the nitrogen atom of pyridazine, indicating negative electrostatic groups here may be beneficial for improving activity. Therefore, it was found in the scaffold of compounds where there was a hydroxyl in this position. In a hydrophobic field, yellow and red contours were highlighted areas where the yellow contours represented the presence of hydrophobic substituents would be favourable for better activity, while red contours indicated regions where hydrophobic groups decreased the biological activity. The yellow region located on the R_5 group can be explained by comparing Compounds 15 and 16 where increasing the hydrophobicity by adding more alkyl groups resulted in improvement of biological activities ($6.69 > 4.47$).

Hydrogen bonds were also considered to be very important for the activity of ligands. The hydrogen bond donor (HBD) and acceptor (HBA) contour maps were provided by the CoMSIA model to facilitate structural optimization. For the donor contour maps, the HBD groups located in the cyan regions were favourable for enhancing activity. The cyan contour map clearly explained that why Compounds 9, 24, 31, 41 and 56, which had -NH- in site A, showed a higher activity than Compounds 12, 13, 16, 19, 36 and 37) without such a substituent. While in the HBA contour maps, the HBA groups located in the magenta regions were suitable for the enhancement of inhibitory activity. The magenta contour maps indicated that R_1 with hydrogen bond acceptor was crucial to the activity. This also conformed well with almost all positive compounds where R_1 was a hydrogen bond acceptor.

In short, all the cases mentioned above indicated that the CoMSIA model had good predictive ability.

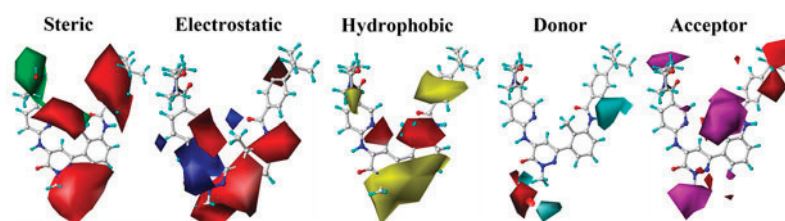


Figure 3. Green, blue, yellow, cyan and magenta indicate regions favourable for BTK inhibitor activity, whereas red indicates contours not favourable for activity. Compound 7 is placed as a reference.

3.3 Connection of GA-MARS and CoMSIA models

Some parameters chosen by GA-MARS are significantly correlated with the inhibition activity, including Q_VSA_FPOL, PM3_HOMO, MNDO_dipole, logP(o/w), a_acc, SlogP_VSA7, SlogP_VSA1, SMR_VSA5 and BCUT_SLOGP_1. The first three descriptors are electrical parameters, the fourth is a hydrophobic parameter, the fifth is the number of hydrogen bond acceptors and the rest are steric parameters. This was consistent with the results of the CoMSIA model. GCV score showed that E_tor, Q_VSA_FPOL and SlogP_VSA7 were the most significant descriptors. Meanwhile, the CoMSIA contour plots highlighted the importance of electrostatic and hydrophobic field interactions, accounting for 14.5% and 29.2% contribution of the model, respectively.

All of these results proved that the electrical and hydrophobic properties were closely related to the biological activity of pyridine and pyrimidine inhibitors. In addition, both the GA-MARS model and CoMSIA model indicated that hydrogen bond acceptor was important for the activity. Table 2 shows that increasing a_acc and decreasing PM3_HOMO can improve the activity. Similarly, the CoMSIA model demonstrated that the contribution to the variance of HBA field was 21.9%.

In this paper, we studied the structure–activity relationship (SAR) from different aspects. A GA-MARS model was established to undertake quantitative analysis and CoMSIA models was developed to produce more intuitive guidance for the design of BTK inhibitors.

3.4 SAR summary and novel BTK inhibitors generated

SAR information on different R substituents is summarized in Figure 4. In the binding pocket, bulky volume and hydrogen bond interactions were preferable for R₁ substituents. Apart from the existing hydrogen bond formed between the nitrogen atom of pyridine and carbonyl oxygen atom of ASN484, one more hydrogen bond between the carbonyl oxygen atom of Leu408 and R₁ may be favourable for further activity improvement. The linker is a key part for ligands to

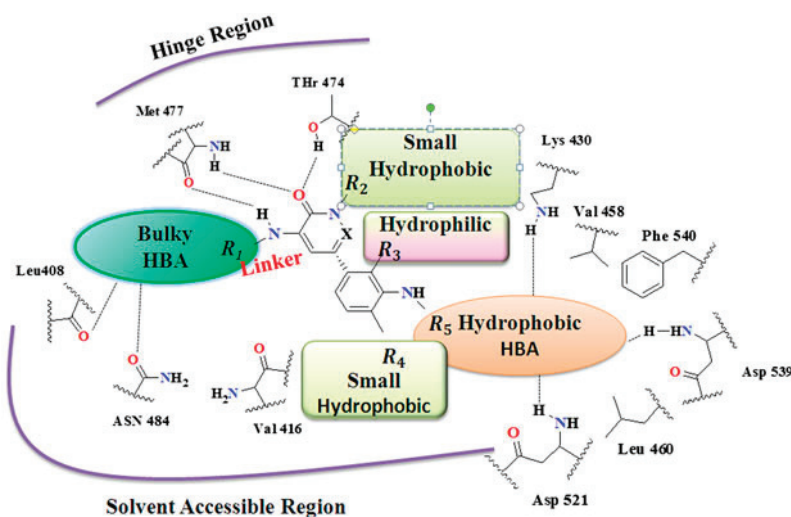


Figure 4. SAR summary in the presence of scaffold. HBA: hydrogen bond acceptor; HBD: hydrogen bond donor.

keep the U-shaped conformation (Figure 5B) and -NH- was suggested as be the most suitable substituent. The GA-MARS model demonstrated that E_{tor} was the most significant descriptor, which was closely related to the U-shaped conformation. For the scaffold, the hydrogen bond with Met477 or Thr474 was considered one of the most important protein–ligand interactions. As R_3 was extended to the solvent accessible region, hydrophilic substituents were favoured. Small volume and hydrophobicity were desirable for R_2 and R_4 substituents. As R_5 was located in the narrow but deep channel, bulky and hydrophobic substituents inserted into the channel lengthwise were favoured. The R_5 substituent with one hydrogen bond acceptor was preferred to interact with Lys430, Asp539 or Asp521. All these effects could be observed from the results of the 3D-QSAR and GA-MARS models.

To maintain the essential interactions and binding affinity of the hinge region, only R_1 and R_5 substituents were modified. Applying the MFR method, the following experiments were carried out on the modification of Compound 55, which was more representative due to their structural similarities and mapped well with the 3D-QSAR models. First, the R_1 substituent of Compound 55 was replaced by a set of fragments. Then, according to the SAR summary mentioned above, a two-point pharmacophore (Figure 5A) model was established based on Compound 55. The pharmacophore consisted of an aromatic centre and a hydrogen acceptor, serving as a query for screening the compounds generated by MFR.

To further validate our pharmacophore's accuracy, the pocket surface of BTK was mapped automatically (Figure 5B). It conformed well to the two pharmacophoric features where the acceptor query was just within the hydrogen bond interaction regions (red) and the aromatic centre was in a hydrophobic pocket (blue). In addition, the obtained molecules were initially constrained to ensure a molecular weight <500 Da, TPSA in the range of 40 to 140, and SlogP in the range of -4 to 8 [46].

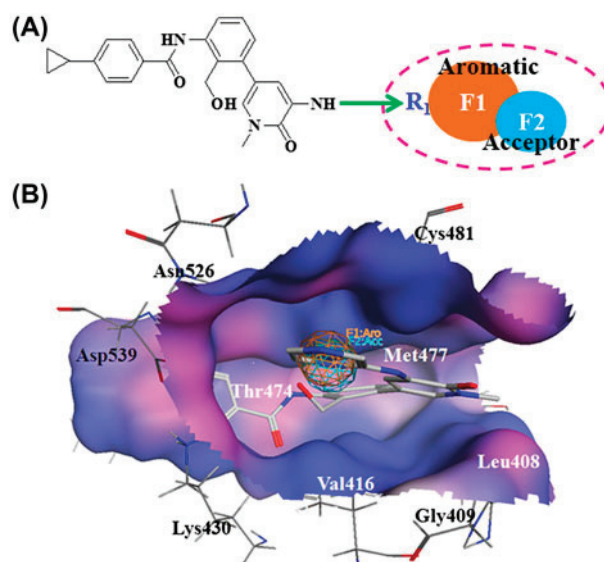


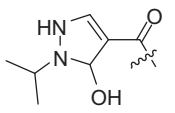
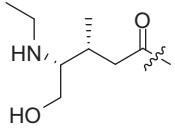
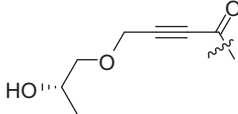
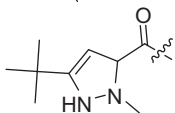
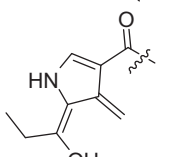
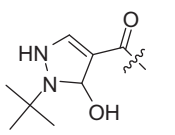
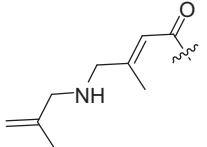
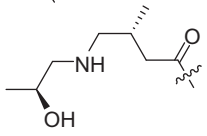
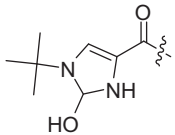
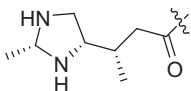
Figure 5. Two-point pharmacophore generated by Compound 55 and coordinate BTK crystal complex (PDB ID: 3GEN): (A) 2D structure of the pharmacophore; and (B) 3D structure of the pharmacophore in the active pocket. Blue represents hydrophobic region, red represents hydrogen region and white represents mild polar region.

Table 6. Chemical structures of newly designed BTK inhibitors with respective predicted activities.

No.	Substituents		Predicted pIC_{50}	
	R_1	R_5	GA-MARS	CoMSIA
N1			9.00	7.21
N2			8.82	7.00
N3			7.58	7.13
N4			7.05	7.04
N5			8.42	7.34
N6			8.30	7.48
*N7			8.30	7.67
N8			8.10	7.10
N9			7.91	7.38
N10			7.87	7.36

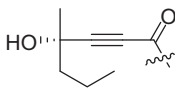
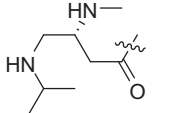
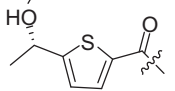
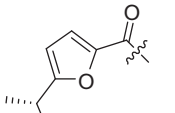
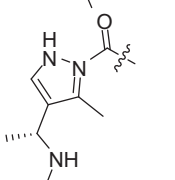
(Continued)

Table 6. (Continued).

No.	Substituents		Predicted pIC_{50}	
	R_1	R_5	GA-MARS	CoMSIA
N11			7.83	7.29
*N12			7.82	7.59
N13			7.82	7.14
N14			7.80	7.25
N15			7.74	7.45
N16			7.67	7.23
N17			7.65	7.42
N18			7.62	7.34
N29			7.58	7.18
N20			7.51	7.18

(Continued)

Table 6. (Continued).

No.	Substituents		Predicted pIC_{50}	
	R_1	R_5	GA-MARS	CoMSIA
N21			7.50	7.07
N22			7.46	7.20
N23			7.45	7.01
N24			7.40	7.12
N25			7.35	7.09

*High activity compounds with pIC_{50} values >7.50 in both predicted models.

The pharmacophore was used to search a set of 785,007 fragments extracted from a combination of ZINC fragment library and vendor catalogues. Under the condition of the two-point pharmacophore constraint, a subset of 6582 compounds was generated. The obtained compounds were further predicted by GA-MARS and CoMSIA models, and the 12 compounds with $pIC_{50} > 7.00$ were selected. Compounds were considered as candidates for further study when the docking conformations satisfied the binding model and the key hydrogen bonds formed with crucial residues in the active site of BTK kinase including Thr474, Met477, Leu408 and Val416 [48]. Finally, four novel potent BTK inhibitors were designed and predicted (Table 6).

Similarly, the R_5 substituent of Compound 55 was replaced by a set of fragments. A three-point pharmacophore (Figure 6A) model was developed based on Compound 55. The pharmacophore model consisted of a hydrogen acceptor, an aromatic centre and a hydrophobic feature, serving as a query for screening the compounds generated by MFR. Three pharmacophoric features were well proved by the pocket surface of BTK (Figure 6(B) where the acceptor query was just within the hydrogen bond interaction regions (red), and the aromatic centre and hydrophobic point were embedded in a hydrophobic pocket (blue). Eventually, 21 novel potent BTK inhibitors (Table 6) were designed and predicted using the same workflow. The trend of the value predicted by GA-MARS model was consistent with the CoMSIA model.

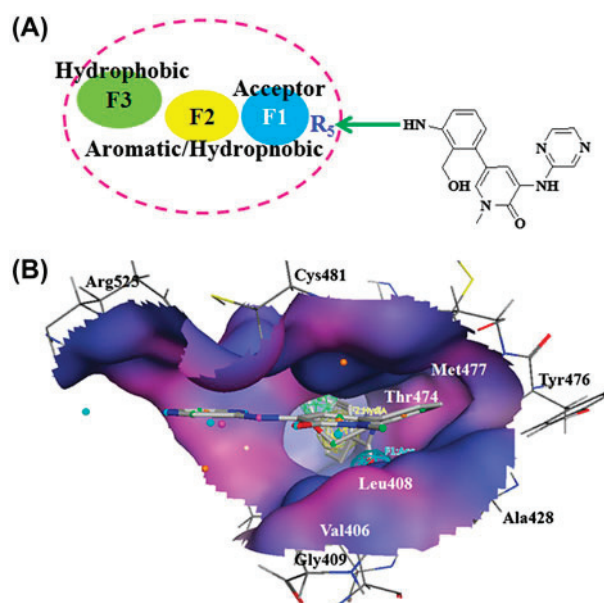


Figure 6. Three-point pharmacophore generated by Compound 55 and coordinate BTK crystal complex (PDB ID: 3GEN): (A) 2D structure of the pharmacophore and (B) 3D structure of the pharmacophore within activity pocket. Blue represents hydrophobic region, red represents hydrogen region and white represents mild polar region.

4. Conclusions

GA-MARS and CoMSIA models were developed and validated for a series of BTK inhibitors to study the quantitative structure–activity relationship. The GA-MARS technique was successfully applied for selecting significant molecular descriptors and predicting the inhibitory activities of the pyridine and pyrimidine derivatives. The MARS algorithm optimized by GA increased the numerical stability of the model construction process and attained better robustness. The GA-MARS model obtained the high r^2_{pred} and the low $\text{RMSE}_{\text{test}}$, which proved that it had perfect predictive power. Additionally, the CoMSIA model described in this investigation also performed well by getting high q^2 and r^2_{pred} values. The CoMSIA contour maps highlighted the importance of hydrophobic and hydrogen bonding donor field interactions. Both the GA-MARS and CoMSIA model showed that the electrical and hydrophobic properties were closely related to the biological activity of pyridine and pyrimidine-based BTK inhibitors. Finally, a new molecular database generated by MFR was evaluated with GA-MARS and CoMSIA prediction, and 25 pyridine and pyrimidine-based novel potential BTK inhibitors were obtained. Generally, GA-MARS and CoMSIA model could serve together as a useful tool package to help medicinal chemists design novel BTK inhibitors.

Acknowledgement

This work was financially supported by the Qinglan Project of Jiangsu Province, a project funded by the Priority Academic Programme Development of Jiangsu Higher Education Institutions and the Jiangsu Province Postgraduate Innovative Research Foundation (KYLX_0637).

References

- [1] M.Y. Choi and T.J. Kipps, *Inhibitors of B-cell receptor signaling for patients with B-cell malignancies*, Cancer J. 18 (2012), pp. 404–410.
- [2] R.W. Hendriks, S. Yuvaraj, and L.P. Kil, *Targeting Bruton's tyrosine kinase in B cell malignancies*, Nat. Rev. Cancer 14 (2014), pp. 219–232.
- [3] X. Li, Y. Zuo, G. Tang, Y. Wang, Y. Zhou, X. Wang, T. Guo, M. Xia, N. Ding, and Z. Pan, *Discovery of a series of 2,5-diaminopyrimidine covalent irreversible inhibitors of Bruton's tyrosine kinase with in vivo antitumor activity*, J. Med. Chem. 57 (2014), pp. 5112–5128.
- [4] P. Sideras and C.I.E. Smith, *Molecular and cellular aspects of X-linked agammaglobulinemia*, Adv. Immunol. 59 (1995), pp. 135–224.
- [5] R.C. Rickert, *New insights into pre-BCR and BCR signalling with relevance to B cell malignancies*, Nat. Rev. Immunol. 13 (2013), pp. 578–591.
- [6] R.E. Davis, V.N. Ngo, G. Lenz, P. Tolar, R.M. Young, P.B. Romesser, H. Kohlhammer, L. Lamy, H. Zhao, Y. Yang, W. Xu, A.L. Shaffer, G. Wright, W. Xiao, J. Powell, J.-K. Jiang, C.J. Thomas, A. Rosenwald, G. Ott, H.K. Muller-Hermelink, R.D. Gascoyne, J.M. Connors, N.A. Johnson, L.M. Rimsza, E. Campo, E.S. Jaffe, W.H. Wilson, J. Delabie, E.B. Smeland, R.I. Fisher, R.M. Braziel, R.R. Tubbs, J.R. Cook, D.D. Weisenburger, W.C. Chan, S.K. Pierce, and L.M. Staudt, *Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma*, Nature 463 (2010), pp. 88–92.
- [7] T.U. Marron, M. Martinez-Gallo, J.E. Yu, and C. Cunningham-Rundles, *Toll-like receptor 4-, 7-, and 8-activated myeloid cells from patients with X-linked agammaglobulinemia produce enhanced inflammatory cytokines*, J. Allergy Clin. Immunol. 129 (2012), pp. 184–190.e184.
- [8] J.A. Whang and B.Y. Chang, *Bruton's tyrosine kinase inhibitors for the treatment of rheumatoid arthritis*, Drug Discov. Today 19 (2014), pp. 1200–1204.
- [9] J. Hutcheson, K. Vanarsa, A. Bashmakov, S. Grewal, D. Sajitharan, B.Y. Chang, J.J. Buggy, X.J. Zhou, Y. Du, A.B. Satterthwaite, and C. Mohan, *Modulating proximal cell signaling by targeting Btk ameliorates humoral autoimmunity and end-organ disease in murine lupus*, Arthritis Res. Ther. 14 (2012), p. R243.
- [10] L. Vargas, A. Hamasy, B.F. Nore, and C.I. Smith, *Inhibitors of BTK and ITK: State of the new drugs for cancer, autoimmunity and inflammatory diseases*, Scand. J. Immunol. 78 (2013), pp. 130–139.
- [11] O.J. D'Cruz and F.M. Uckun, *Novel Bruton's tyrosine kinase inhibitors currently in development*, Onco. Targets Ther. 6 (2013), pp. 161–176.
- [12] Z. Pan, *Bruton's tyrosine kinase as a drug discovery target*, Drug News Perspect. 21 (2008), pp. 357–362.
- [13] O. Hantschel, U. Rix, U. Schmidt, T. Bürckstümmer, M. Kneidinger, G. Schütze, J. Colinge, K.L. Bennett, W. Ellmeier, P. Valent, and G. Superti-Furga, *The Btk tyrosine kinase is a major target of the Bcr-Abl inhibitor dasatinib*, Proc. Natl. Acad. Sci. U. S. A. 104 (2007), pp. 13283–13288.
- [14] F.M. Uckun, H.E. Tibbles, and A.O. Vassilev, *Bruton's tyrosine kinase as a new therapeutic target*, Anticancer Agents Med. Chem. 7 (2007), pp. 624–632.
- [15] J.A. Di Paolo, T. Huang, M. Balazs, J. Barbosa, K.H. Barck, B.J. Bravo, R.A.D. Carano, J. Darrow, D.R. Davies, L.E. DeForge, L. Diehl, R. Ferrando, S.L. Gallion, A.M. Giannetti, P. Gribbling, V. Hurez, S.G. Hymowitz, R. Jones, J.E. Kropf, W.P. Lee, P.M. Maciejewski, S.A. Mitchell, H. Rong, B.L. Staker, J.A. Whitney, S. Yeh, W.B. Young, C. Yu, J. Zhang, K. Reif, and K.S. Currie, *Specific Btk inhibition suppresses B cell- and myeloid cell-mediated arthritis*, Nat. Chem. Biol. 7 (2011), pp. 41–50.
- [16] Y. Lou, X. Han, A. Kuglstatter, R.K. Kondru, Z.K. Sweeney, M. Soth, J. McIntosh, R. Litman, J. Suh, B. Kocer, D. Davis, J. Park, S. Frauchiger, N. Dewdney, H. Zecic, J.P. Taygerly, K. Sarma, J. Hong, R.J. Hill, T. Gabriel, D.M. Goldstein, and T.D. Owens, *Structure-based drug design of RN486, a potent and selective Bruton's tyrosine kinase (BTK) inhibitor, for the treatment of rheumatoid arthritis*, J. Med. Chem. 58 (2015), pp. 512–516.

- [17] D. Xu, Y. Kim, J. Postelnek, M.D. Vu, D.Q. Hu, C. Liao, M. Bradshaw, J. Hsu, J. Zhang, A. Pashine, D. Srinivasan, J. Woods, A. Levin, A. O'Mahony, T.D. Owens, Y. Lou, R.J. Hill, S. Narula, J. DeMartino, and J.S. Fine, *RN486, a selective Bruton's tyrosine kinase inhibitor; abrogates immune hypersensitivity responses and arthritis in rodents*, J. Pharmacol. Exp. Ther. 341 (2012), pp. 90–103.
- [18] L. Liu, J.S. Halladay, Y. Shin, S. Wong, M. Coraggio, H. La, M. Baumgardner, H. Le, S. Gopaul, J. Boggs, P. Kuebler, J.C. Davis Jr, X.C. Liao, J.W. Lubach, A. Deese, C.G. Sowell, K.S. Currie, W.B. Young, S.C. Khojasteh, C.E. Hop, and H. Wong, *Significant species difference in amide hydrolysis of GDC-0834, a novel potent and selective Bruton's tyrosine kinase inhibitor*, Drug Metab. Dispos. 39 (2011), pp. 1840–1849.
- [19] S. Parmar, K. Patel, and J. Pinilla-Ibarz, *Ibrutinib (imbruvica): A novel targeted therapy for chronic lymphocytic leukemia*, P. T. 39 (2014), pp. 483–519.
- [20] I. Jabeen, P. Wetwitayaklung, P. Chiba, M. Pastor, and G.F. Ecker, *2D- and 3D-QSAR studies of a series of benzopyranes and benzopyrano[3,4b][1,4]-oxazines as inhibitors of the multidrug transporter P-glycoprotein*, J. Comput. Aided Mol. Des. 27 (2013), pp. 161–171.
- [21] J. Polanski, A. Bak, R. Gieleciak, and T. Magdziarz, *Modeling robust QSAR*, J. Chem. Inf. Model. 46 (2006), pp. 2310–2318.
- [22] S. Sharma, B. Bagchi, S. Mukhopadhyay, and A.K. Bothra, *2D QSAR studies of several potent aminopyridine, anilinoypyrimidine and pyridine carboxamide-based JNK inhibitors*, Indian J. Pharm. Sci. 73 (2011), pp. 165–170.
- [23] D. Santos Cruz and M. Santos Castilho, *2D QSAR studies on series of human beta-secretase (BACE-1) inhibitors*, Med. Chem. 10 (2014), pp. 162–173.
- [24] S.V. Jain, M. Ghate, K.S. Bhadoriya, S.B. Bari, A. Chaudhari, and J.S. Borse, *2D, 3D-QSAR and docking studies of 1,2,3-thiadiazole thioacetanilides analogues as potent HIV-1 non-nucleoside reverse transcriptase inhibitors*, Org. Med. Chem. Lett. 2 (2012), p. 22.
- [25] M. Goodarzi, B. Dejaegher, and Y.V. Heyden, *Feature selection methods in QSAR studies*, J. AOAC Int. 95 (2012), pp. 636–651.
- [26] D. Dunkler, M. Plischke, K. Leffondré, and G. Heinze, *Augmented backward elimination: A pragmatic and purposeful way to develop statistical models*, PloS One 9 (2014), p. e113677.
- [27] R.D. Cramer, J.D. Bunce, D.E. Patterson, and I.E. Frank, *Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies*, Quant. Struct.-Act. Relat. 7 (1988), pp. 18–25.
- [28] S.P. Niculescu, *Artificial neural networks and genetic algorithms in QSAR*, J. Mol. Struc.-Theochem 622 (2003), pp. 71–83.
- [29] R. Leardi, *Genetic algorithms in chemometrics and chemistry: A review*, J. Chemom. 15 (2001), pp. 559–569.
- [30] M. Jalali-Heravi and M. Asadollahi-Baboli, *Quantitative structure–activity relationship study of serotonin (5-HT₇) receptor inhibitors using modified ant colony algorithm and adaptive neuro-fuzzy interference system (ANFIS)*, Eur. J. Med. Chem. 44 (2009), pp. 1463–1470.
- [31] J.H. Friedman, *Multivariate adaptive regression splines*, Ann. Statist. (1991), pp. 1–67.
- [32] M. Asadollahi-Baboli, *Docking and receptor-based QSAR approaches for modeling of CETP inhibitors*, Med. Chem. Res. 23 (2013), pp. 1162–1169.
- [33] R. Menon, G. Bhat, G.R. Saade, and H. Spratt, *Multivariate adaptive regression splines analysis to predict biomarkers of spontaneous preterm birth*, Acta Obstet. Gyn. Scan. 93 (2014), pp. 382–391.
- [34] J.B. Ghasemi and E. Zolfonoun, *Application of principal component analysis–multivariate adaptive regression splines for the simultaneous spectrofluorimetric determination of dialkyltins in micellar media*, Spectrochim. Acta A 115 (2013), pp. 357–363.
- [35] X.L. Zhu, H.Y. Cai, Z.J. Xu, Y. Wang, H.Y. Wang, A. Zhang, and W.L. Zhu, *Classification of 5-HT_{1A} receptor agonists and antagonists using GA-SVM method*, Acta. Pharmacol. Sin. 32 (2011), pp. 1424–1430.

- [36] J.K. Buolamwini and H. Assefa, *CoMFA and CoMSIA 3D QSAR and docking studies on conformationally-restrained cinnamoyl HIV-1 integrase inhibitors: Exploration of a binding mode at the active site*, J. Med. Chem. 45 (2002), pp. 841–852.
- [37] N.J. Dewdney, R.K. Kondru, Y. Lou, M. Soth, and T. Gabriel, *BTK protein kinase inhibitors*, US Patent 20090105209 A1 (2009).
- [38] K.-H. Kim, A. Maderna, M.E. Schnute, M. Hegen, S. Mohan, J. Miyashiro, L. Lin, E. Li, S. Keegan, J. Lussier, C. Wrocklage, C.L. Nickerson-Nutter, A.J. Wittwer, H. Soutter, N. Caspers, S. Han, R. Kurumbail, K. Dunussi-Joannopoulos, J. Douhan Iii, and A. Wissner, *Imidazo[1,5-a]quinoxalines as irreversible BTK inhibitors for the treatment of rheumatoid arthritis*, Bioorg. Med. Chem. Lett. 21 (2011), pp. 6258–6263.
- [39] A.S. Reddy, S. Kumar, and R. Garg, *Hybrid-genetic algorithm based descriptor optimization and QSAR models for predicting the biological activity of Tipranavir analogs for HIV protease inhibition*, J. Mol. Graph. Model. 28 (2010), pp. 852–862.
- [40] P. Izadiyan, M.H. Fatemi, and M. Izadiyan, *Elicitation of the most important structural properties of ionic liquids affecting ecotoxicity in limnic green algae; a QSAR approach*, Ecotoxicol. Environ. Saf. 87 (2013), pp. 42–48.
- [41] J. Perez-Villanueva, J.L. Medina-Franco, T.R. Caulfield, A. Hernandez-Campos, F. Hernandez-Luis, L. Yopez-Mulia, and R. Castillo, *Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) of some benzimidazole derivatives with trichomonocidal activity*, Eur. J. Med. Chem. 46 (2011), pp. 3499–3508.
- [42] H. Yuan, J. Zhuang, S. Hu, H. Li, J. Xu, Y. Hu, X. Xiong, Y. Chen, and T. Lu, *Molecular modeling of exquisitely selective c-Met inhibitors through 3D-QSAR and molecular dynamics simulations*, J. Chem. Inf. Model. 54 (2014), pp. 2544–2554.
- [43] S. Pirhadi and J.B. Ghasemi, *3D-QSAR analysis of human immunodeficiency virus entry-1 inhibitors by CoMFA and CoMSIA*, Eur. J. Med. Chem. 45 (2010), pp. 4897–4903.
- [44] Y. Zhang, H. Liu, Y. Jiao, H. Yuan, F. Wang, S. Lu, S. Yao, Z. Ke, W. Tai, Y. Jiang, Y. Chen, and T. Lu, *De novo design of N-(pyridin-4-ylmethyl)aniline derivatives as KDR inhibitors: 3D-QSAR, molecular fragment replacement, protein-ligand interaction fingerprint, and ADMET prediction*, Mol. Divers. 16 (2012), pp. 787–802.
- [45] Z. Ke, T. Lu, H. Liu, H. Yuan, T. Ran, Y. Zhang, S. Yao, X. Xiong, J. Xu, A. Xu, and Y. Chen, *3D-QSAR and molecular fragment replacement study on diaminopyrimidine and pyrrolotriazine ALK inhibitors*, J. Mol. Struct. 2014 (1067), pp. 127–137.
- [46] P. Ertl, B. Rohde, and P. Selzer, *Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties*, J. Med. Chem. 43 (2000), pp. 3714–3717.
- [47] E. Pourbasheer, R. Aalizadeh, S. Shokouhi Tabar, M.R. Ganjali, and J. Shadmanesh, *2D and 3D quantitative structure–activity relationship study of hepatitis C virus NS5B polymerase inhibitors by comparative molecular field analysis and comparative molecular similarity indices analysis methods*, J. Chem. Inf. Model. 54 (2014), pp. 2902–2914.
- [48] H.L. Wan, Z.R. Wang, L.L. Li, C. Cheng, P. Ji, J.J. Liu, H. Zhang, J. Zou, and S.Y. Yang, *Discovery of novel Bruton's tyrosine kinase inhibitors using a hybrid protocol of virtual screening approaches based on SVM model, pharmacophore and molecular docking*, Chem. Biol. Drug Des. 80 (2012), pp. 366–373.

Structure–Activity Relationship Study of QL47: A Broad-Spectrum Antiviral Agent

Yanke Liang,^{†,‡,||} Melissanne de Wispelaere,^{§,||} Margot Carocci,^{§,||} Qingsong Liu,^{†,‡,||} Jinhua Wang,^{†,‡} Priscilla L. Yang,^{*,§} and Nathanael S. Gray^{*,†,‡,||}

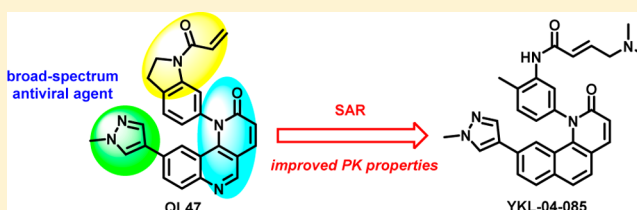
[†]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, United States

Departments of [‡]Biological Chemistry & Molecular Pharmacology and [§]Microbiology and Immunobiology, Harvard Medical School, Boston, Massachusetts 02115, United States

S Supporting Information

ABSTRACT: Here we report the structure–activity relationship (SAR) investigations of QL-XII-47 (QL47), a compound that possesses broad-spectrum antiviral activity against dengue virus and other RNA viruses. A medicinal chemistry campaign initiated from QL47, a previously reported covalent BTK inhibitor, to derive YKL-04-085, which is devoid of any kinase activity when screened against a panel of 468 kinases and with improved pharmacokinetic properties. Both QL47 and YKL-04-085 are potent inhibitors of viral translation and exhibit cellular antiviral activity at 35-fold lower concentrations relative to inhibition of host-cell proliferation.

KEYWORDS: Dengue virus, broad-spectrum antiviral, QL47, structure–activity relationship



Dengue virus (DENV) is a significant human pathogen for which no effective drugs or vaccines currently exist. DENV is a plus-stranded RNA virus and a member of the *Flaviviridae* family. The four DENV serotypes (DENV-1, -2, -3, -4) are defined by their immunoreactivity and differ from one another by 25–40% at the amino acid level.¹ Due to the large number of people at risk for infection, DENV is the most widespread mosquito-borne viral disease affecting humans today. An estimated 2.5 billion people live in areas at risk for epidemic transmission, and an estimated 390 million DENV infections occur annually, of which 96 million (67–136) manifest clinically.² Infection with DENV is responsible for disease ranging from dengue fever to the much more severe and life-threatening dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS) that are characterized by vascular leakage. An estimated 500,000 cases of DHF/DSS occur annually and are associated with 2.5% fatality although fatality rates for DHF/DSS can exceed 20% if untreated.³ Since both the number of DENV infections and the proportion of infections resulting in DHF have escalated in recent years, the absence of anti-DENV therapeutics or an anti-DENV vaccine is a growing public health concern.

Classical antiviral approaches involve development of so-called “direct-acting” antivirals that target essential, virally encoded proteins, which in the case of DENV would include the viral polymerase and protease.⁴ Despite tremendous success in developing antivirals that target the analogous enzymes of HIV and HCV,^{5–8} to-date efforts to develop antivirals against the dengue polymerase or protease have not yielded promising clinical candidates. We pursued a complementary approach to identify compounds that target host-factors that are essential for viral replication. The potential advantage of this approach is that viral resistance might emerge more slowly because there is not a

direct route to viral resistance arising from mutations that block drug-binding to a viral target. A second advantage is that host-directed antivirals might possess broader spectrum antiviral activity if they target processes commonly utilized by different viruses.⁹

Previously we have implemented cellular phenotypic screens of DENV infection that resulted in the identification of inhibitors of host signaling and lipid metabolism.^{10,11} We extended this screening to a large collection of cysteine-reactive, covalent kinase inhibitors that have the potential to target a subset of the approximately 150 kinases that possess a cysteine near the ATP-binding site of the kinase. From this screen,¹² we identified a tricyclic quinoline compound, QL-XII-47 (QL47), previously identified as a potent and covalent inhibitor of BTK and other Tec-family kinases.¹³ Here we describe a medicinal chemistry campaign focused on the goal of understanding the structural features required for antiviral activity of this scaffold relative to inhibition of host-cell kinases. Surprisingly, we discovered compounds such as YKL-04-085, which lack the critical quinoline nitrogen required for hydrogen bonding to the kinase hinge and are devoid of any kinase activity, that retain the potent antiviral activity. We demonstrate that this series of compounds inhibit DENV infection at concentrations that are not cytotoxic and also inhibit the replication of other RNA viruses in cell culture.

To identify new host-targeted inhibitors of DENV, we used two different approaches to screen small molecule libraries. Since a

Received: January 9, 2017

Accepted: February 3, 2017

Published: February 3, 2017

detailed account of the two screens is described elsewhere,¹² we present here a qualitative description as background. A first screen was based on the utilization of DENV reporter viral particles (RVP), which are composed of DENV2 E and prM/M proteins and contain a nucleocapsid formed by DENV2 C and a subgenomic replicon of the closely related West Nile virus (WNV), another Flavivirus.¹⁴ In a second screen, we employed a previously characterized HEK-293T derived stable cell line encoding a GFP tagged DENV2 replicon, the T-REx-293-DGZ cell line.¹⁴ In both screens, the expression levels of the reporter gene encoded by either the RVP-packaged WNV replicon or the DENV2 replicon were measured after treatment with a library of small molecules.

The library of small molecules screened was a collection of acrylamide-modified, ATP-site directed heterocycles designed to target cysteine residues located in or around the kinase ATP-site. This library contains experimentally validated covalent inhibitors of EGFR,¹⁵ BTK,¹⁶ BMX,¹⁷ BLK,¹⁶ TEC,¹⁶ JAK3,¹⁸ FGFR1,2,3,4,¹⁹ JNK1,2,3,²⁰ CDK7,²⁰ CDK12,²¹ and CDK13²¹ as well as potential inhibitors of a subset of over 150 kinases that possess a potentially reactive cysteine.²² In addition, proteomic studies have demonstrated that these compounds can target other non-kinase targets either due to the presence of a hyper-reactive cysteine or due to the presence of a suitable binding pocket and an adjacent cysteine residue.²³

The screen resulted in the identification of a series of small molecules belonging to a class of tricyclic quinoline inhibitors including QL47 (Figure 1A),¹² a previously reported inhibitor of

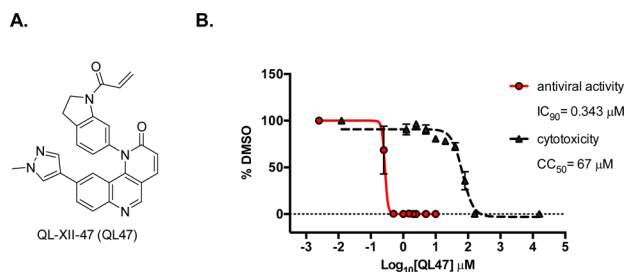


Figure 1. (A) Chemical structure of QL47. (B) Huh7 cells were infected with DENV2 then treated with a range of QL47 concentrations. Cell viability (black) and viral yield (red) were measured 24 hours later. CC_{50} and IC_{90} values—respectively, corresponding to the concentrations of QL47 that lead to 50% cytotoxicity and to a 1- $\log_{10}$ unit reduction in the number of infectious progeny virions produced—were determined by non-linear regression. Representative data (mean $\pm$ standard deviation of experimental duplicates) out of $n \geq 2$ repeats are shown.

BTK, and other TEC-family kinases.¹³ Among the screening hits, QL47 exhibited the most potent antiviral activity, with IC_{90} value against DENV2 of 0.343 μ M, and a desirable window between potency and cytotoxicity (Figure 1B). In order to determine whether BTK/BMX kinases are the pharmacologically relevant targets for viral replication, we further evaluated other known BTK/BMX kinase inhibitors, CGI-1746,²⁴ Ibrutinib,¹⁶ and BMX-IN-1.²⁵ However, none of these inhibitors showed comparable antiviral activity (Figure 2), indicating BTK/BMX are not likely to be the pharmacologically relevant targets of QL47. To further investigate the potent antiviral activity of QL47, we examined the inhibition of other unrelated viruses and observed that QL47 exhibited broad-spectrum antiviral activity against other Flaviviruses, such as West Nile virus (WNV), as well as against Enteroviruses, Vesiculoviruses, Pneumoviruses (Figure 3), among others. The potent inhibition of a variety of viruses by

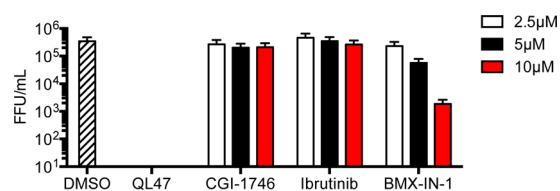


Figure 2. Huh7 cells were infected with DENV2 then treated with each of the indicated small molecules. The antiviral activity of each small molecule against DENV2 was measured by titration of the infectious particles that had been secreted to culture supernatants at 24h post-infection. A representative experiment out of $n = 2$ repeats is shown. Representative data (mean $\pm$ standard deviation of experimental duplicates) out of $n = 2$ repeats are shown. FFU, focus forming unit.

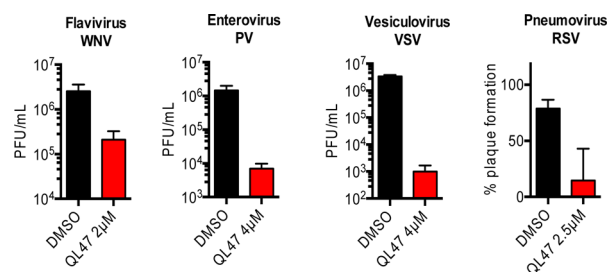


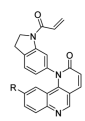
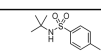
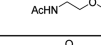
Figure 3. Huh7 cells were infected with West Nile virus (WNV), then treated with 2 μ M of QL47 for 24h. Vero cells were infected with poliovirus (PV) or vesicular stomatitis virus (VSV), then treated with 4 μ M of QL47 for 12h and 6h, respectively. The infectious virus released to the supernatants were quantified by PFA. For respiratory syncytial virus (RSV), Vero cells were treated with 2.5 μ M of QL47 post-infection. The number of plaque formed at 7 days post-infection were counted and the results were plotted relative to DMSO-treated samples. Representative data (mean $\pm$ standard deviation of experimental duplicates) out of $n \geq 2$ repeats are shown. PFU, plaque forming unit.

QL47 indicates that the molecular target of QL47 might be a host cell factor commonly utilized in the infectious cycle of these viruses. Despite the risk of side effects, targeting host cell proteins could be used as a complementary approach to direct-targeting of essential viral proteins.

Encouraged by the broad spectrum antiviral activity of QL47, we sought to elucidate the structure–activity relationships required to impart antiviral activity and to improve the poor pharmacokinetic properties of QL47 (Table S1). QL47 was originally developed as a covalent BTK inhibitor using structure-based design starting from the highly potent mTOR inhibitor, Torin2.^{26,27} Torin2 has been cocrystallized with mTOR, which revealed how this scaffold binds to the ATP-binding cleft (Figure S1).²⁸ These tricyclic quinolines use the quinoline N nitrogen to form an essential hydrogen bond to the kinase hinge segment, while the two side chains extend toward the back and the front of the ATP binding site. The acrylamide arm of QL47 extends toward Cys481 of BTK located at the entrance of the ATP-binding site.¹³

We first examined whether QL47 acrylamide moiety, which is required for covalent inhibition of BTK, is essential for the antiviral activity. We used QL47R in which the acrylamide moiety was replaced with a nonreactive propyl amide functional group (Figure 4A). QL47R did not inhibit DENV2 at concentrations below 10 μ M (Figure 4B), suggesting the acrylamide moiety is essential for antiviral activity. Consistent with the interpretation that QL47 covalently modifies its target, we also found that the full anti-DENV activity was recapitulated when cells were pretreated with QL47 for 6 h followed by extensive washing prior to DENV infection.¹²

Table 1. Structure–Activity Relationship of QL47 and Quinolinyln Analogues

Com- pound	R	Antiviral activity ΔLog_{10} FFU (DMSO/ compound)		Luciferase activity: cellular translation of DENV2(GVD) (% DMSO)		CC ₅₀ (μM)	Micro- somal stability $T_{1/2}$ (min)
		2 μM	10 μM	2 μM	10 μM		
	QL47						
	1	0	1.9	81	100	13	2.3
	2	0.8	4	44	75	75	6.4
	3	0.8	4	29	76	77	6.5
	4	1.0	4	40	34	> 160	4.1
	5	2.7	4	82	95	> 160	1.3
	6	2.8	4	32	94	46	2.7
	7	0.1	1.2	0	23	103	9
	8	0.1	0	0	2	44	5
	9	0	0.4	23	24	107	7.9
	10	4	4	97	99	105	4.5
	11	1.1	4	39	90	99	nd ^a
	12	2.0	4	4	93	83	nd
	13	0	0	3	19	>160	nd
	14	0.1	0.8	49	53	68	2
	15	2.5	4	94	99	79	2
	16	2.0	4	77	98	126	2

^and = not determined.

ring systems were used (compounds 7–9), the stability was moderately improved. However, these compounds again completely lost viral inhibition even at 10 μM , suggesting a conjugated ring system is required to achieve antiviral activity.

Next, we evaluated the effects of the substituents at the position *ortho* to the acrylamide (Table 1, compounds 10–13). Compound 10 demonstrated comparable potency to QL47, presumably due to the intramolecular hydrogen bonding of the *ortho*-fluoro group, which forms a pseudo-five-membered ring. Compound 10 also showed potent *in cellulo* translation inhibition and antiviral activity, but the half-life of the compound was only marginally improved to 4.5 min. The *ortho*-ether analogue 11 exhibited reduced antiviral activity and translation inhibition. The benzamide analogue 12 showed moderate activity at 2 μM , while the benzoic acid analogue 13 did not inhibit the virus at all, which might be due to the poor cellular permeability.

The unsuccessful effort to improve microsomal stability by varying methylpyrazole and acrylamide moieties prompted us to examine the tricyclic quinolinyln core of QL47 (Table 1, compounds 14–16). Interestingly, compounds containing a six-membered urea moiety, such as 14, only showed minimal antiviral activity at 10 μM , suggesting that this region of the molecule is a

critical determinant for antiviral activity. We further evaluated five-membered methylurea analogues 15 and 16. Both displayed increased antiviral activity and inhibition of DENV(GVD) reporter replicon translation, probably due to the restoration of some conjugated features of the tricyclic core. However, none of these compounds exhibited substantially improved metabolic stability.

The SAR indicated that an extended conjugated system from methylpyrazole to acrylamide is required to preserve antiviral activity. Combining the hypothesis that the kinase targets of QL47, BTK/BMX, are not likely responsible for the effects on viral replication, we were interested to assess naphthyl analogues without the critical kinase hinge-interacting nitrogen atom. The antiviral activity, inhibition of reporter replicon translation, and MLM stability for naphthyl analogues are shown in Table 2. As

Table 2. Structure–Activity Relationship of Naphthyl Analogues

Com- pound	R	Antiviral activity ΔLog_{10} FFU (DMSO/ compound)		Luciferase activity: cellular translation of DENV2(GVD) (% DMSO)		CC ₅₀ (μM)	Micro- somal stability $T_{1/2}$ (min)
		2 μM	10 μM	2 μM	10 μM		
	17	4	4	76	95	>160	4
	18	2.0	4	39	79	29	4
	19	1.9	4	82	94	75	9
	20	0.5	1.0	49	60	>160	nd ^a
	21	0.5	0.1	77	88	>160	nd
	22	3.3	4	17	66	21	6.3
	YKL-04-085	4	4	65	95	12	15.6

^and = not determined.

expected, compounds 17–19 exhibited potent viral inhibition at 10 μM , suggesting that the molecular target of these compounds might not be a protein kinase. Among these compounds, 17, with a methylpyrazole substituent, demonstrated potency comparable to QL47, although its MLM half-life was only marginally improved to 4 min. We also tested hydrogenated analogues 20 and 21, which exhibited no antiviral activity, further confirming the necessity of a conjugated ring system to achieve viral inhibition. Intriguingly, when a dimethylamine tail was installed at the acrylamide site, the antiviral activity of compound 22 and YKL-04-085 was greatly improved at 2 μM . More promisingly, the MLM half-life of compound YKL-04-085 was significantly improved to 15.6 min.

To further characterize compound YKL-04-085, it was screened against a panel of 468 kinases and disease relevant mutant variants using an active site-directed competition binding assay at 10 μM drug concentration (Figure 5A). Only a handful of kinases, such as PIMs and DDRs, exhibited more than 80% binding with YKL-04-085. We further determined the IC₅₀ values for these kinases, as well as BTK/BMX (Figure 5B), and none of them showed significant inhibition by YKL-04-085, which confirmed YKL-04-085 is devoid of any kinase activity due to the lack of the critical hinge-contacting nitrogen. The anti-DENV2 IC₉₀ value of YKL-04-085 was determined to be 0.555

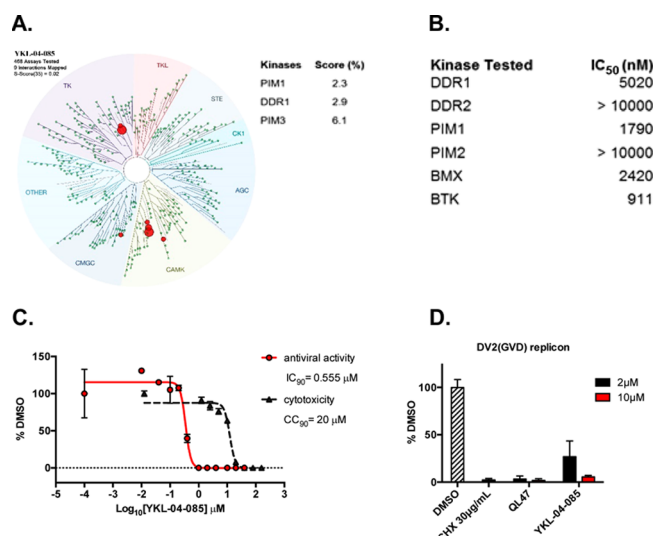


Figure 5. (A) DiscoverX KINOMEScan selectivity profile of YKL-04-085. Compound was screened at 10 μ M concentration against 468 kinases and mutants. Scores for primary screen hits were reported as a percent of the DMSO control (% control). The lower the score, the lower the K_d is likely to be. Scores for kinase targets, which showed more than 80% binding, are shown in the table. (B) Enzymatic IC_{50} s of YKL-04-085. IC_{50} s against DDR1/2 were obtained with LanthaScreen binding assays, while IC_{50} s against PIM1/2 and BTK/BMX were obtained with Z'-Lyte activity assays. (C) Huh7 cells were infected with DENV2 and then treated with a range of YKL-04-085 concentrations. Cell viability (black) and viral yield (red) were measured 24 hours later. CC_{90} and IC_{90} values—respectively, corresponding to the concentrations of QL47 that lead to 50% cytotoxicity and to a 1- $\log_{10}$ unit reduction in the number of infectious progeny virions produced—were determined by non-linear regression. Representative data (mean $\pm$ standard deviation of experimental duplicates) out of $n \geq 2$ repeats are shown. (D) Huh7 cells were transfected with the in vitro-transcribed DENV2(GVD) reporter replicon RNA, and then immediately treated with the indicated compounds. Intracellular firefly luciferase activity was quantified at 6h post-transfection as a measure of successful RNA translation. Reporter activity was measured as a function of the luciferase activity obtained with the DMSO treated samples. Representative data (mean $\pm$ standard deviation of experimental duplicates) out of $n = 3$ repeats are shown.

μ M, and the inhibitor exhibited more than 35-fold window between viral inhibition and cytotoxicity (Figure 5C). In addition, YKL-04-085 demonstrated good inhibition of translation of the DENV2(GVD) reporter replicon *in cellulo* (Figure 5D). Encouraged by the potent inhibition of DENV2 and improved metabolic stability of YKL-04-085, we then evaluated the mouse PK properties to determine whether it is a suitable candidate for future *in vivo* efficacy studies (Table 3). Although the compound displayed high clearance via intravenous (IV) route and low bioavailability, intraperitoneal (IP) administration gave acceptable plasma drug exposure. Furthermore, a 40 mg/kg once daily IP dosing regime was both well-tolerated and sufficient to maintain plasma concentrations of YKL-04-085 above the IC_{90} value measured *in cellulo* (Figure S2).

Table 3. In Vivo Pharmacokinetic Properties of YKL-04-085

route	dose (mg/kg)	$T_{1/2}$ (h)	T_{max} (h)	C_{max} (ng/mL)	AUC_{last} (ng·h/mL)	AUC_{inf} (ng·h/mL)	CL (mL/min/kg)	V_{ss} (L/kg)	F (%)
i.v.	2	1.07		461.71	161.33	167.28	199.26	9.58	
p.o.	10	1.53	1.00	40.41	95.40	117.89			12
i.p.	10		0.25	786.78	1112.65	1213.28			
i.p.	50		0.08	4505.75	8222.94	8554.01			

In conclusion, we report the identification, characterization, and preliminary structure–activity relationships (SAR) for a novel class of covalent inhibitors of DENV. Preliminary mechanistic studies of QL47 suggest that its antiviral activity is correlated with effects on the steady-state abundance of viral proteins, likely due to an effect on translation of the viral RNA genome. The correlation of anti-DENV activity with effects on translation of the DENV2(GVD) reporter replicon in the SAR studies reported here provide further support that effects on translation of the viral genome are a key contributor to the antiviral activity of these compounds. Nevertheless, there are several outliers that fall outside this correlation, indicating that additional mechanisms may contribute to the overall antiviral activity of these compounds. While the precise molecular target of this series of compounds is currently under investigation, KINOMEScan of QL47 and YKL-04-085 suggests that protein kinases are unlikely to be targets. Lead optimization yielded compound YKL-04-085, which might be an appropriate candidate for further *in vivo* studies to test efficacy of this class of compounds. Further optimization of the lead compound, target identification, and *in vivo* efficacy studies are the subject of ongoing research.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsmmedchemlett.7b00008.

Supplemental figures, tables, materials and methods, experimental procedures, and compound characterization (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: nathanael_gray@dfci.harvard.edu.

*E-mail: priscilla_yang@hms.harvard.edu.

ORCID

Qingsong Liu: 0000-0002-7829-2547

Nathanael S. Gray: 0000-0001-5354-7403

Author Contributions

^{||}These authors contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We gratefully acknowledge Dr. Tinghu Zhang for the synthesis of relevant tool compounds and helpful discussions; and Dr. Michael Cameron at the Scripps Research Institute for conducting *in vitro* mouse microsome study and *in vivo* tolerability study. This work was funded by the Linde program in chemical biology.

■ REFERENCES

- (1) Kuno, G.; Chang, G.-J. J.; Tsuchiya, K. R.; Karabatsos, N.; Cropp, C. B. Phylogeny of the Genus Flavivirus. *J. Virol.* **1998**, *72*, 73–83.

- (2) Bhatt, S.; Gething, P. W.; Brady, O. J.; Messina, J. P.; Farlow, A. W.; Moyes, C. L.; Drake, J. M.; Brownstein, J. S.; Hoen, A. G.; Sankoh, O.; Myers, M. F.; George, D. B.; Jaenisch, T.; Wint, G. R. W.; Simmons, C. P.; Scott, T. W.; Farrar, J. J.; Hay, S. I. The global distribution and burden of dengue. *Nature* **2013**, *496*, 504–507.
- (3) World Health Organization. Dengue and dengue haemorrhagic fever. <http://www.who.int/mediacentre/factsheets/fs117/en/>.
- (4) Lim, S. P.; Wang, Q.-Y.; Noble, C. G.; Chen, Y.-L.; Dong, H.; Zou, B.; Yokokawa, F.; Nilar, S.; Smith, P.; Beer, D.; Lescar, J.; Shi, P.-Y. Ten years of dengue drug discovery: Progress and prospects. *Antiviral Res.* **2013**, *100*, 500–519.
- (5) Kiser, J. J.; Flexner, C. Direct-Acting Antiviral Agents for Hepatitis C Virus Infection. *Annu. Rev. Pharmacol. Toxicol.* **2013**, *53*, 427–449.
- (6) Gerber, L.; Welzel, T. M.; Zeuzem, S. New therapeutic strategies in HCV: polymerase inhibitors. *Liver Int.* **2013**, *33*, 85–92.
- (7) Anderson, J.; Schiffer, C.; Lee, S.-K.; Swanstrom, R. Viral Protease Inhibitors. In *Antiviral Strategies*; Kräusslich, H.-G., Bartenschlager, R., Eds.; Springer: Berlin Heidelberg, 2009; Vol. 189, pp 85–110.
- (8) Deeks, S. G.; Smith, M.; Holodniy, M.; Kahn, J. O. HIV-1 protease inhibitors: A review for clinicians. *JAMA, J. Am. Med. Assoc.* **1997**, *277*, 145–153.
- (9) Tan, S.-L.; Ganji, G.; Paepfer, B.; Proll, S.; Katze, M. G. Systems biology and the host response to viral infection. *Nat. Biotechnol.* **2007**, *25*, 1383–1389.
- (10) Chu, J. J. H.; Yang, P. L. c-Src protein kinase inhibitors block assembly and maturation of dengue virus. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 3520–3525.
- (11) Carocci, M.; Hinshaw, S. M.; Rodgers, M. A.; Villareal, V. A.; Burri, D. J.; Pilankatta, R.; Maharaj, N. P.; Gack, M. U.; Stavale, E. J.; Warfield, K. L.; Yang, P. L. The Bioactive Lipid 4-Hydroxyphenyl Retinamide Inhibits Flavivirus Replication. *Antimicrob. Agents Chemother.* **2015**, *59*, 85–95.
- (12) de Wispelaere, M.; Carocci, M.; Liang, Y.; Liu, Q.; Sun, E.; Vetter, M. L.; Wang, J.; Gray, N. S.; Yang, P. L. Discovery of host-targeted covalent inhibitors of dengue virus. *Antiviral Res.* **2017**, *139*, 171–179.
- (13) Wu, H.; Wang, W.; Liu, F.; Weisberg, E. L.; Tian, B.; Chen, Y.; Li, B.; Wang, A.; Wang, B.; Zhao, Z.; McMillin, D. W.; Hu, C.; Li, H.; Wang, J.; Liang, Y.; Buhrlage, S. J.; Liang, J.; Liu, J.; Yang, G.; Brown, J. R.; Treon, S. P.; Mitsiades, C. S.; Griffin, J. D.; Liu, Q.; Gray, N. S. Discovery of a Potent, Covalent BTK Inhibitor for B-Cell Lymphoma. *ACS Chem. Biol.* **2014**, *9*, 1086–1091.
- (14) Ansarah-Sobrinho, C.; Nelson, S.; Jost, C. A.; Whitehead, S. S.; Pierson, T. C. Temperature-dependent production of pseudoinfectious dengue reporter virus particles by complementation. *Virology* **2008**, *381*, 67–74.
- (15) Zhou, W.; Ercan, D.; Chen, L.; Yun, C.-H.; Li, D.; Capelletti, M.; Cortot, A. B.; Chirieac, L.; Jacob, R. E.; Padera, R.; Engen, J. R.; Wong, K.-K.; Eck, M. J.; Gray, N. S.; Janne, P. A. Novel mutant-selective EGFR kinase inhibitors against EGFR T790M. *Nature* **2009**, *462*, 1070–1074.
- (16) Liu, F.; Zhang, X.; Weisberg, E.; Chen, S.; Hur, W.; Wu, H.; Zhao, Z.; Wang, W.; Mao, M.; Cai, C.; Simon, N. I.; Sanda, T.; Wang, J.; Look, A. T.; Griffin, J. D.; Balk, S. P.; Liu, Q.; Gray, N. S. Discovery of a Selective Irreversible BMX Inhibitor for Prostate Cancer. *ACS Chem. Biol.* **2013**, *8*, 1423–1428.
- (17) Hur, W.; Velentza, A.; Kim, S.; Flatauer, L.; Jiang, X.; Valente, D.; Mason, D. E.; Suzuki, M.; Larson, B.; Zhang, J.; Zagorska, A.; DiDonato, M.; Nagle, A.; Warmuth, M.; Balk, S. P.; Peters, E. C.; Gray, N. S. Clinical stage EGFR inhibitors irreversibly alkylate Bmx kinase. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5916–5919.
- (18) Honigberg, L. A.; Smith, A. M.; Sirisawad, M.; Verner, E.; Loury, D.; Chang, B.; Li, S.; Pan, Z.; Thamm, D. H.; Miller, R. A.; Buggy, J. J. The Bruton tyrosine kinase inhibitor PCI-32765 blocks B-cell activation and is efficacious in models of autoimmune disease and B-cell malignancy. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 13075–13080.
- (19) Zhou, W.; Hur, W.; McDermott, U.; Dutt, A.; Xian, W.; Ficarro, S. B.; Zhang, J.; Sharma, S. V.; Brugge, J.; Meyerson, M.; Settleman, J.; Gray, N. S. A Structure-Guided Approach to Creating Covalent FGFR Inhibitors. *Chem. Biol.* **2010**, *17*, 285–295.
- (20) Kwiatkowski, N.; Zhang, T.; Rahl, P. B.; Abraham, B. J.; Reddy, J.; Ficarro, S. B.; Dastur, A.; Amzallag, A.; Ramaswamy, S.; Tesar, B.; Jenkins, C. E.; Hannett, N. M.; McMillin, D.; Sanda, T.; Sim, T.; Kim, N. D.; Look, T.; Mitsiades, C. S.; Weng, A. P.; Brown, J. R.; Benes, C. H.; Marto, J. A.; Young, R. A.; Gray, N. S. Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature* **2014**, *511*, 616–620.
- (21) Zhang, T.; Kwiatkowski, N.; Olson, C. M.; Dixon-Clarke, S. E.; Abraham, B. J.; Greifenberg, A. K.; Ficarro, S. B.; Elkins, J. M.; Liang, Y.; Hannett, N. M.; Manz, T.; Hao, M.; Bartkowiak, B.; Greenleaf, A. L.; Marto, J. A.; Geyer, M.; Bullock, A. N.; Young, R. A.; Gray, N. S. Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors. *Nat. Chem. Biol.* **2016**, *12*, 876–884.
- (22) Liu, Q.; Sabnis, Y.; Zhao, Z.; Zhang, T.; Buhrlage; Sara, J.; Jones, Lyn, H.; Gray; Nathanael, S. Developing Irreversible Inhibitors of the Protein Kinase Cysteine. *Chem. Biol.* **2013**, *20*, 146–159.
- (23) Lanning, B. R.; Whitby, L. R.; Dix, M. M.; Douhan, J.; Gilbert, A. M.; Hett, E. C.; Johnson, T. O.; Joslyn, C.; Kath, J. C.; Niessen, S.; Roberts, L. R.; Schnute, M. E.; Wang, C.; Hulse, J. J.; Wei, B.; Whiteley, L. O.; Hayward, M. M.; Cravatt, B. F. A road map to evaluate the proteome-wide selectivity of covalent kinase inhibitors. *Nat. Chem. Biol.* **2014**, *10*, 760–767.
- (24) Di Paolo, J. A.; Huang, T.; Balazs, M.; Barbosa, J.; Barck, K. H.; Bravo, B. J.; Carano, R. A. D.; Darrow, J.; Davies, D. R.; DeForge, L. E.; Diehl, L.; Ferrando, R.; Gallion, S. L.; Giannetti, A. M.; Gribbling, P.; Hurez, V.; Hymowitz, S. G.; Jones, R.; Kropf, J. E.; Lee, W. P.; Maciejewski, P. M.; Mitchell, S. A.; Rong, H.; Staker, B. L.; Whitney, J. A.; Yeh, S.; Young, W. B.; Yu, C.; Zhang, J.; Reif, K.; Currie, K. S. Specific Btk inhibition suppresses B cell- and myeloid cell-mediated arthritis. *Nat. Chem. Biol.* **2011**, *7*, 41–50.
- (25) Tan, L.; Akahane, K.; McNally, R.; Reyskens, K. M. S. E.; Ficarro, S. B.; Liu, S.; Herter-Sprie, G. S.; Koyama, S.; Pattison, M. J.; Labella, K.; Johannessen, L.; Akbay, E. A.; Wong, K.-K.; Frank, D. A.; Marto, J. A.; Look, T. A.; Arthur, J. S. C.; Eck, M. J.; Gray, N. S. Development of Selective Covalent Janus Kinase 3 Inhibitors. *J. Med. Chem.* **2015**, *58*, 6589–6606.
- (26) Liu, Q.; Wang, J.; Kang, S. A.; Thoreen, C. C.; Hur, W.; Ahmed, T.; Sabatini, D. M.; Gray, N. S. Discovery of 9-(6-Aminopyridin-3-yl)-1-(3-(trifluoromethyl)phenyl)benzo[h][1,6]naphthyridin-2(1H)-one (Torin2) as a Potent, Selective, and Orally Available Mammalian Target of Rapamycin (mTOR) Inhibitor for Treatment of Cancer. *J. Med. Chem.* **2011**, *54*, 1473–1480.
- (27) Liu, Q.; Xu, C.; Kirubakaran, S.; Zhang, X.; Hur, W.; Liu, Y.; Kwiatkowski, N. P.; Wang, J.; Westover, K. D.; Gao, P.; Ercan, D.; Niepel, M.; Thoreen, C. C.; Kang, S. A.; Patricelli, M. P.; Wang, Y.; Tupper, T.; Altabel, A.; Kawamura, H.; Held, K. D.; Chou, D. M.; Elledge, S. J.; Janne, P. A.; Wong, K.-K.; Sabatini, D. M.; Gray, N. S. Characterization of Torin2, an ATP-Competitive Inhibitor of mTOR, ATM, and ATR. *Cancer Res.* **2013**, *73*, 2574–2586.
- (28) Yang, H.; Rudge, D. G.; Koos, J. D.; Vaidialingam, B.; Yang, H. J.; Pavletich, N. P. mTOR kinase structure, mechanism and regulation. *Nature* **2013**, *497*, 217–223.
- (29) Liu, Q.; Chang, J. W.; Wang, J.; Kang, S. A.; Thoreen, C. C.; Markhard, A.; Hur, W.; Zhang, J.; Sim, T.; Sabatini, D. M.; Gray, N. S. Discovery of 1-(4-(4-Propionylpiperazin-1-yl)-3-(trifluoromethyl)phenyl)-9-(quinolin-3-yl)benzo[h][1,6]naphthyridin-2(1H)-one as a Highly Potent, Selective Mammalian Target of Rapamycin (mTOR) Inhibitor for the Treatment of Cancer. *J. Med. Chem.* **2010**, *53*, 7146–7155.
- (30) Stowasser, F.; Baenziger, M.; Garad, S. D. Preparation of salts and crystalline forms of 2-methyl-2-[4-(3-methyl-2-oxo-8-quinolin-3-yl-2,3-dihydro-imidazo[4,5-c]quinolin-1-yl)-phenyl]-propionitrile and its use as a drug. WO 2008/064093, May 29, 2008.
- (31) de Wispelaere, M.; LaCroix, A. J.; Yang, P. L. The Small Molecules AZD0530 and Dasatinib Inhibit Dengue Virus RNA Replication via Fyn Kinase. *J. Virol.* **2013**, *87*, 7367–7381.