Science Letters ISSN 2345-5463

SL

Research article
2020 | Volume 8 | Issue 1 | Pages 28-35

ARTICLE INFO

Received
December 02, 2019
Revised
February 19, 2020
Accepted
February 22, 2020
Published
April 10, 2020

***Corresponding Author**
Babatunde Samuel Obadawo
E-mail
Sobadawo@gmail.com
otunbasamtz009@gmail.com

Keywords
Enterovirus
Applicability
Domain
Y-Randomization
Williams plot
QSAR

How to Cite
Obadawo BS, Oyeneyin OE, Anifowose MM, Fagbohungbe KH, Amoko JS. QSAR Evaluation of C-8-Tert-Butyl Substituted 4-Aryl-6,7,8,9-tetrahydrobenzo[4,5]thieno[3,2-e] [1,2,4]triazolo [4,3-a] pyrimidin-5(4H)-one Derivatives as Potent Anti-enterovirus Agents. Sci Lett 2020; 8(1):28-35

Open Access

QSAR Evaluation of C-8-Tert-Butyl Substituted 4-Aryl-6,7,8,9-tetrahydro benzo[4,5]thieno[3,2-e] [1,2,4]triazolo [4,3-a]pyrimidin-5(4H)-one Derivatives as Potent Anti-enterovirus Agents

Babatunde Samuel Obadawo[1,2]*, Oluwatoba Emmanuel Oyeneyin[2], Mayowa Monday Anifowose[2], Kehinde Henry Fagbohungbe[3], Justinah Solayide Amoko[4]

[1] Department of Chemistry, Ahmadu Bello University, Zaria, Kaduna State, Nigeria
[2] Theoretical and Computational Chemistry Unit, Department of Chemical Sciences, Adekunle Ajasin University, Akungba-Akoko, Ondo, State, Nigeria
[3] Department of Chemistry, University of Ibadan, Oyo State, Nigeria
[4] Department of Chemistry, Adeyemi College of Education, Ondo, Ondo State, Nigeria

Abstract

A study was performed on twenty compounds of C-8-tert-butyl substituted 4-aryl-6,7,8,9-tetrahydrobenzo[4,5]thieno[3,2-e][1,2,4]triazolo [4,3-a] pyrimidin-5(4H)-one derivatives to formulate a mathematical linear equation that can be used to estimate the activity of new compounds as anti-enterovirus inhibitors. The genetic function algorithm technique was employed to generate six different models for the six different studied enteroviruses. The models were established to have coefficient of determination (R^2), cross-validation coefficient (Q^2_{cv}), coefficient of determination for Y-randomization (cR^2_p) to be in cordial agreement with the recommended values. Furthermore, the built models were externally validated to have R^2_{test} to be 0.7565, 0.7399, 0.9353, 0.9084, 0.8631, and 0.7768 for Cox B1, Cox B3, PV3, HRV 14, HRV 21 and HRV71, respectively, which ascertained the prognostic power of the model. The applicability domain evaluation revealed that there were no outliers and influencing compounds in the built models since the warning limit ($h^* = 1.07$) was greater than the leverage values of the compounds. Due to the dependability, validity and stability of the built model, C-8-tert-butyl substituted 4-aryl-6,7,8,9-tetrahydrobenzo[4,5]thieno[3,2-e][1,2,4]triazolo [4,3-a]pyrimidin-5(4H)-one can be improved as a potent enterovirus inhibitor.

SCAN ME

Introduction

Viral Infections are life periling issues that lead to abrupt death in humans. *Picornaviridae*, one of the prominent families of the virus has been regarded as the most significant viral pathogen discovered and capable of causing harm to humans and animals and the intestinal transmission route account for their nomenclature [1]. *Enteroviruses* are one of the family of *Picornaviridae* with the ssRNA genome. Examples of viruses that belong to this genus include Coxsackievirus A, Coxsackievirus B, Poliovirus, enterovirus and Human Rhinovirus. Also, terrible diseases have been affiliated with this viral family, such as respiratory illness [2], myocarditis [3], Chronic Obstructive Pulmonary Disease (COPD) [4], asthma [5], etc.

Some compounds have been synthesized to inhibit enterovirus strains. Benserazide was found to be the first inhibitor with allosteric properties against Coxsackievirus B3 [6]. Non-peptide compounds, which include a series of novel heteroaromatic esters were synthesized and proved to function as anti-human rhinovirus [7]. Despite the high inhibition potencies of some Rupintrivir analogs, they proved to be unsuccessful in a test carried out with infected human Rhabdomyosarcoma cell lines [8]. This inactivity unfolded discoveries for the design of novel anti-enterovirus inhibitors. Recently, Quantitative Structure-Activity Relationship (QSAR) utilized computational software in building a mathematical equation, which helped to design novel compounds with potent activity [9]. This technique aims at correlating the descriptors with molecular properties of the compound (biological activities) such as inhibition concentration. Therefore, this work aimed to formulate a mathematical model that can be used to evaluate the activity of new compounds as anti-enteroviruses through QSAR method using Genetic Function Approximation Multiple Linear R egression (GFA-MLR) method.

Materials and Methods

Data collection

Antiviral activities of a series twenty compounds of 4-aryl-6,7,8,9-tetrahydrobenzo [4,5]thieno[3,2-e] triazolo[4,3-a]pyrimidin-5(4H)-one derivatives [10] were selected out of twenty-two compounds synthesized for each of the enteroviruses (Cox B1, Cox B3, PV3, HRV 14, HRV 21, HRV 71) from the literature. The *in vivo* antiviral activities of these compounds were given in EC_{50} (µM) and all were converted to their corresponding negative logarithmic scale and labeled as pEC_{50} values (*i.e.*, -log EC_{50} = pEC_{50}) in order to make the activities agree to an array of values and also fit into the normal distribution curve. The structures of the compounds are presented in Table 1 and Fig. 1.

Table 1 Structures of C-8-tert-butyl substituted 4-aryl-6,7,8, 9tetrahydrobenzo[4,5]thieno[3,2-e][1,2,4]triazolo[4,3-a] pyrimidin-5(4H)-one derivatives.

S/No	R^1	R^2	R^3
1	H	H	H
2	6-Methyl	H	H
3	8-Methyl	H	H
4	8-Propyl	H	H
5	8-(tert-Butyl)	H	H
6	8-Phenyl	H	H
7	8,8-Dimethyl	H	H
8	7, 7, 9, 9-Tetramethyl	H	H
9	8-(tert-Butyl)	2'-Me	H
10	8-(tert-Butyl)	4'-Me	H
11	8-(tert-Butyl)	2'-Cl	H
12	8-(tert-Butyl)	3'-Cl	H
13	8-(tert-Butyl)	4'-Cl	H
14	8-(tert-Butyl)	3'-Br	H
15	8-(tert-Butyl)	4'-Br	H
16	8-(tert-Butyl)	4-OMe	H
17	8-(tert-Butyl)	4'-CF_3	H
18	8-(tert-Butyl)	4'-(p-Me-Ph)	H
19	8-(tert-Butyl)	H	Me
20	8-(tert-Butyl)	H	Ph
21	8-(tert-Butyl)	H	NH_2
22	8-(tert-Butyl)	H	SH

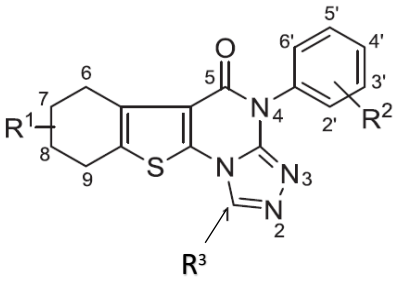

Fig. 1 Structures of C-8-tert-butyl substituted 4-aryl-6,7,8,9 tetrahydrobenzo[4,5]thieno[3,2-e][1,2,4]triazolo[4,3-a] pyrimidin-5(4H)-one derivatives. The R values correspond to the R values presented in Table 1.

Equilibrium geometry optimization at ground state

The 2D structure of the compounds presented in Table 1 and Fig. 1 were drawn using ChemDraw Ultra 12.0 and then saved as cdx file. The cdx file format of the compounds was exported to Spartan 14 version 1.1.4 software for optimization using Molecular Mechanics with the MMFF followed by

Density Functional Theory (DFT) employing Becke's three-parameter Lee-Yang-Parr hybrid functional (B3LYP) with the 6-31G* basis set. [11, 12]. After the geometry optimization process was done, the optimized structures initially saved as spartan sdf file.

Molecular descriptor calculation

The optimized compounds initially saved as sdf file were subjected to PaDEL-Descriptor software version 2.20 in order to calculate the 1D, 2D and 3D descriptors of the compounds. After removing salt, detecting tautomer and retaining the file name as a molecule name, the result was saved as Microsoft Excel Comma Separated value (csv) file.

Data pre-treatment and data division

In order to reduce collinearity, filter descriptors with redundant data and highly correlated data, the descriptors were subjected to pre-treatment using data pre-treatment software obtained from the Drug Theoretical and Cheminformatics Laboratory (DTC Lab). The pre-treated data were made to pass through the data division software obtained from the Drug Theoretical and Cheminformatics Laboratory (DTC Lab). This is to divide the pre-treated data into training and test sets using by employing Kennard and Stone's algorithm [13]. The permutation algorithm divided the data into 70% (14 compounds) of the total data set, which was considered as a training set and used to build the model. Around 30% of the total data sets (6 compounds) were considered as a test set and also used to validate the built model externally.

Model building

The compounds used as a training set were subjected to Material studio 2017 software employing the Genetic Function Approximation (GFA) method to generate a valid model with the biological activities (pEC_{50}) as the dependent variable and the physiochemical properties (descriptors) as the independent variables.

Internal validation of the model

The generated models with a scaled LOF smoothness parameter of 0.5 were appraised using the Friedman formula (LOF), which measured the fitness score of the model [14]. LOF is defined as follows:

$$LOF = \frac{SEE}{\left(1-\frac{C+dp}{M}\right)^2} \quad \text{Eq. 1}$$

Where SEE is the standard error of estimation, p is the total number of descriptors in the model, d is a userdefined smoothing parameter, c is the number of terms in the model, and M is the number compound in the training set [15].

SEE is the standard error of estimation which equals the standard deviation of the model and a model is said to be good when it has lower SEE value. SEE is given as:

$$SEE = \sqrt{\frac{\left(Y_{exp}-Y_{pred}\right)^2}{N-P-1}} \quad \text{Eq. 2}$$

The correlation coefficient (R^2) is the most frequently used internal assessments for the QSAR model. The closer the value of R^2 to 1.0, the better the model generated. R^2 is expressed as follows:

$$R^2 = 1 - \frac{\Sigma\left(Y_{exp}-Y_{pred}\right)^2}{\Sigma\left(Y_{exp}-Y_{training}\right)^2} \quad \text{Eq. 3}$$

Where $Y_{training}$, Y_{exp} and Y_{pred} are the mean experimental activity, experimental activity and the predicted activity in the training set, respectively.

R^2 value varies directly with the increase in the number of descriptors, thus, R^2 is not reliable to measure the stability of the model. Therefore, R^2 is adjusted in order to have a reliable and stable model. The adjusted R^2 is defined as follows:

$$R^2_{adj} = \frac{R^2-P(n-1)}{n-p+1} \quad \text{Eq. 4}$$

where p and n are the number of descriptors in the model and the number of compounds that made up the training set.

The strength of the QSAR model to predict the activity of a new compound was determined using a cross-validation test. The cross-validation coefficient (Q^2_{cv}) is defined as follows:

$$(Q^2_{cv}) = 1 - \left\{\frac{\Sigma\left(Y_{pred}-Y_{exp}\right)^2}{\Sigma\left(Y_{exp}-Y_{training}\right)^2}\right\} \quad \text{Eq. 5}$$

Where $Y_{training}$, Y_{exp} and Y_{pred} are the mean experimental activity, experimental activity and the predicted activity in the training set, respectively.

External validation of the model

External validation of the developed model was assessed by the value R^2_{test} value. The R^2_{test} value is the most commonly used parameter to validate a built model despite other parameters because once the R^2_{test} value is considered satisfied, the remaining parameters will also be satisfied. Also, the closer the value of R^2_{test} to 1.0, the better the stability the model generated. This stability will account for the

Table 2 Generally accepted values for the validation parameters of a built QSAR model.

Validation parameter	Definition	Recommended
R^2	Coefficient of determination	≥0.6
$P_{(95\%)}$	Confidence interval at 95% confidence level	<0.05
Q^2_{cv}	Cross-validation coefficient	≥0.5
R^2 -Q^2_{cv}	Difference between R^2 and Q^2_{cv}	<0.3
$N_{(ext\ \&\ test\ set)}$	Minimum number of the external test set	≥5
cR^2_p	Coefficient of determination for Y-randomization	≥0.5

reliability of the model in predicting the activity of a new compound. The R^2_{test} is defined by as follows:

$$R^2 = 1 - \frac{\Sigma\left(Y_{pred_{test}} - Y_{exp_{test}}\right)^2}{\Sigma\left(Y_{pred_{test}} - \bar{Y}_{training}\right)^2} \quad \text{Eq. 6}$$

where $Y_{pred_{test}}$ and $Y_{exp_{test}}$ are the predicted and experimental activity test set. While $\bar{Y}_{training}$ is mean values of experimental activity of the training set.

Y-randomization test

The Y-Randomization test method is another way to validate the QSAR model. To guarantee that the model is built strong and not deduced by chance, the test was performed on the training set data [16]. For the built QSAR model to robust and reliable, the model is expected to have a low R^2 and Q^2 values fo r several trials. Coefficient of determination cR^2_p for Y-randomization is another parameter calculated whichshould be greater than 0.5 for passing this test.

$$cR^2_p = R \times [R^2 - (R_r)^2]^2 \quad \text{Eq. 7}$$

Where cR^2_p is the coefficient of determination for Y-randomization, R is the coefficient of determination for Y-randomization and R_r is average 'R' of the random model.

Evaluation of the applicability domain of the model

This is another statistical method of validating a built QSAR model. Evaluation of the applicability domain of the QSAR model is a vital step in establishing the at the model is good to make predictions within the chemical space for which it was built [16]. The leveraged approach was applied in analysing the applicability domain of the QSAR models [17] and it is defined for the *ith* compound as follows:

$$hi = X_i(X^TX)^{-1}X_i^T \quad \text{Eq. 8}$$

where X_i is training compounds matrix of i. X is the m k descriptor matrix of the training set compound and X^T is the transpose matrix of X used to build the model. As a prediction tool, the warning leverage (h^*) is the benchmark of normal values for Xoutliers and is defined as follows:

$$h^* = 3\frac{(k+1)}{n} \quad \text{Eq. 9}$$

where n and k are the descriptors and the training set compounds, respectively.

To interpret the pertinence of the model within the chemical space, the standardized residual activity is plotted against the leverage value calculated for the test compounds (Williams plot). Compounds with leverage value less than the warning leverage value (h^*), *i.e.*, $hi < h^*$ and the standardized residual values within the 2 standard deviation unit (±2) are considered to be within the chemical space and accepted as Y outlier. In a similar manner, compounds whose leverage values are greater than the warning leverage value, *i.e.*, $hi > h^*$, are considered unreliable because they are justified to be extrapolated by the plot [18].

Variance inflation factor (VIF)

To check for outliers and multi-collinearity among the descriptors contained in the model, Variance Inflation Factor (VIF) was calculated and it is defined as follows:

$$VIF = \frac{1}{1-R^2} \quad \text{Eq. 10}$$

Where R^2 is the correlation coefficient of the multip le regression between the variables within the model. If VIF value is less than 10, it indicates that there is no multicollinearity among the descriptors. Therefore, the model is accepted. But if the VIF value is greater than 10, it connotes that the model contains multicollinearity among the descriptors and such model is unacceptable due to its instability [19].

Quality assurance of the model

The validation parameters are employed to measure the strength, dependability and predictive ability of a built QSAR model. Therefore, Table 2 gives the g eneral minimum requirement values for both interna l and external validation parameters for the assessm ent of a QSAR model [17].

Table 3 List of descriptors used to build the QSAR model and their dimension.

S/No	Descriptor	Description	Dimension
1	Di	D total accessibility index / weighted by relative first ionization potential	3D
2	E3e	3rd component accessibility directional WHIM index / weighted by relative Sanderson electronegativities	3D
3	ETA_Shape_Y	Extended Topochemical Atom Shape index Y	2D
4	GATS3i	Geary autocorrelation - lag 3 / weighted by first ionization potential	2D
5	GATS5m	Geary autocorrelation - lag 5 / weighted by mass	2D
6	GATS6m	Geary autocorrelation - lag 6 / weighted by mass	2D
7	GATS8e	Geary autocorrelation - lag 8 / weighted by Sanderson electronegativities	2D
8	MATS4e	Moran autocorrelation - lag 4 / weighted by Sanderson electronegativities	2D
9	MATS5e	Moran autocorrelation - lag 5 / weighted by Sanderson electronegativities	2D
10	maxssCH2	Maximum atom-type E-State: -CH2-	2D
11	RDF130m	Radial distribution function - 130 / weighted by relative mass	3D
12	RDF35m	Radial distribution function - 035 / weighted by relative mass	3D
13	RDF40v	Radial distribution function - 040 / weighted by relative van der Waals volumes	3D
14	RNCS	Relative negative charge surface area -- most negative surface area	3D
15	SCH-7	Simple chain, order 7	2D
16	SdsCH	Sum of atom-type E-State: = CH-	2D
17	SRW9	Self-returning walk count of order 9 (ln(1+x)	2D
18	TDB10u	D topological distance-based autocorrelation - lag 10 / unweighted	3D
19	TDB6i	3D topological distance-based autocorrelation - lag 6 / weighted by first ionization potential	3D
20	TDB8u	3D topological distance-based autocorrelation - lag 8 / unweighted	3D
21	VR3_Dzp	Logarithmic Randic-like eigenvector-based index from the topological distance matrix	2D

Results and Discussion

Genetic Function Approximation (GFA) method of BIOVIA Material studio 2017 software was used to generate six (6) different models for each of the ent eroviruses (Cox B1, CoxB3, PV3, HRV 14, HRV 2 1, HRV 71). The validation parameters for the six m odels were presented below.

Cox B1

pEC_{50} = -1.256715385 * GATS5m - 3.315298541* SRW9 - 0.098057700 * RDF40v + 3.386573966 * Di + 27.389746577

Friedman LOF = 0.01147100, R-squared = 0.98150100, Adjusted R-squared = 0.97327900, Cross validated R-squared = 0.95239700, Significance-of-regression F-value = 119.37926600

Cox B3

pEC_{50} = 0.279906959 * SdsCH + 1.132256876 * T DB8u + 0.037094666 * RNCS + 0.066184728 * RDF35m - 5.795236555

Friedman LOF= 0.00692800, R-squared = 0.97872700, Adjusted R-squared = 0.96927300, Cross validated R-squared = 0.94597100, Significance-of-regression F-value = 103.51914700

PV3

pEC_{50} = 3.232848542 * MATS4e + 4.307559759 * MATS5e + 1.285914796 * GATS6m - 11.432011982 * SRW9 + 89.223229486

Friedman LOF = 0.04074600, R-squared = 0.98421900, Adjusted R-squared = 0.97720600, Cross validated R-squared = 0.95801600, Significance-of-regression F-value = 140.33039000

HRV14

pEC_{50} = 4.091076076 * GATS3i + 0.074748474 * VR3_Dzp - 9.031427748 * SRW9 - 0.145939506 * RDF130m + 67.670904769

Friedman LOF = 0.02651700, R-squared = 0.98196000, Adjusted R-squared = 0.97394200, Cross validated R-squared = 0.96827800, Significance-of-regression F-value = 122.47414300

HRV21

pEC_{50} = 1.469759027 * GATS8e - 4.668639602 * SCH-7 - 12.812092700 * ETA_Shape_Y + 5.704024913 * E3e + 10.561949183

Friedman LOF = 0.09142800, R-squared = 0.94353500, Adjusted R-squared = 0.91843900, Cross validated R-squared = 0.83036300, Significance-of-regression F-value = 37.59765400

HRV 71

pEC_{50} = 6.214871348 * maxssCH2 + 2.152251641 * TDB10u - 0.025676301 * TDB6i + 6.799169097 * Di - 8.715210762

Friedman LOF = 0.04011200, R-squared = 0.98246300, Adjusted R-squared = 0.97466800, Cross validated R-squared = 0.95980600, Significance-of-regression F-value = 126.04730500

Table 4 Descriptive statistics of the inhibition data.

Parameters	COX B1	Predicted	COX B3	Predicted	PV3	Predicted
Mean	4.530097	4.656111	4.429669	4.463166	4.776802	4.731882
Standard Error	0.068554	0.063617	0.05804	0.045521	0.130576	0.142502
Median	4.443793	4.670678	4.412259	4.465973	4.630303	4.509993
Standard Deviation	0.306585	0.284503	0.259562	0.203578	0.583954	0.637286
Sample Variance	0.093994	0.080942	0.067373	0.041444	0.341003	0.406134
Kurtosis	-0.74617	-1.27127	0.873758	0.367582	-1.09984	-0.98481
Skewness	0.327229	0.075504	0.562233	0.451531	0.434039	0.446511
Range	1.053057	0.871961	1.035269	0.83386	1.735182	2.103453
Minimum	4	4.235166	4	4.056224	4	3.66754
Maximum	5.053057	5.107127	5.035269	4.890084	5.735182	5.770993
Sum	90.60194	93.12222	88.59338	89.26331	95.53604	94.63764
Count	20	20	20	20	20	20

Table 5 Descriptive statistics of the inhibition data.

Parameters	HRV 14	Predicted	HRV21	Predicted	HRV 71	Predicted
Mean	5.091626	5.264658	4.694677	4.691709	4.939632	4.97716
Standard Error	0.12261	0.102978	0.096715	0.100005	0.129439	0.117507
Median	5.057999	5.401715	4.705584	4.677891	5.006704	5.086849
Standard Deviation	0.548329	0.460531	0.432521	0.447236	0.578868	0.525509
Sample Variance	0.300664	0.212088	0.187074	0.20002	0.335088	0.27616
Kurtosis	-0.40803	-1.40831	0.600609	0.219571	-0.90195	-0.87385
Skewness	-0.52487	-0.43013	0.550211	-0.05611	-0.21416	-0.41135
Range	1.769551	1.344256	1.619789	1.73497	1.853872	1.798604
Minimum	4	4.496024	4	3.857881	4	4.022685
Maximum	5.769551	5.84028	5.619789	5.592851	5.853872	5.821289
Sum	101.8325	105.2932	93.89353	93.83418	98.79264	99.5432
Count	20	20	20	20	20	20

The descriptors and their corresponding dimensions used in building the model are reported in Table 3. From the table, both 2D and 3D descriptors played a vital role in predicting the activity of a new molecule that can inhibit enterovirus species. The negative coefficient of the descriptors in the models inferred that the pEC_{50} of the compounds that fall between the warning limit of William's plot decreases as the value of the descriptor increases. Inversely, the positive coefficient of the descriptors in the models inferred that the pEC_{50} of the compounds that fall between the warning limit of William's plot increases as the value of the descriptor decreases. This implies that to design a potent compound with high pEC_{50} value, the negative coefficient of the descriptor will have to be reduced and the positive coefficient of the descriptors will have to be increased.

Pearson's correlation analysis and the variance inflation factor of the descriptors are presented in Table S1 to Table S6. The results show that there is no correlation between most of the descriptors used in building the models and just one descriptor shows strong correlation (VR3_Dzp / SRW9 (0.733338) in Table S4. Also, the variance inflation factor for the descriptors used in constructing the model ranges from (1.064199 - 4.811878), which is less than the recommended value of 10. This ascertains that the models are acceptable since there is no multicollinearity among the descriptors used in the models.

The descriptive statistics of the inhibition data for both experimental activity and predicted activity shown in Table 4 and Table 5 reveals that the means for experimental activity and the predicted activity for each of the enteroviruses are (4.530097 / 4.656111, 4.429669 / 4.463166, 4.776802/ 4.731882, 5.091626 / 5.264658, 4.694677 / 4.691709 and 4.939632 / 4.97716) for COX B1, COX B3, PV3, HRV 14, HRV 21 and HRV 71, respectively. This inferred that the equations used to build the models have great influence in predicting new compounds with higher pEC_{50}. Other parameters such as median, standard deviation, variance and range among others also confirmed that there is no difference between the experimental activity and the activity predicted by the models.

The results for Y-randomization analysis are presented in Table S7-S12. The low values of R^2, Q^2 for several iterations and cRp^2 generated for the six models (0.795698 (COX B1), 0.85344 (COX B3), 0.815391 (PV3), 0.878643 (HRV14), 0.784737 (HRV21) and 0.811355 (HRV71) were all greater

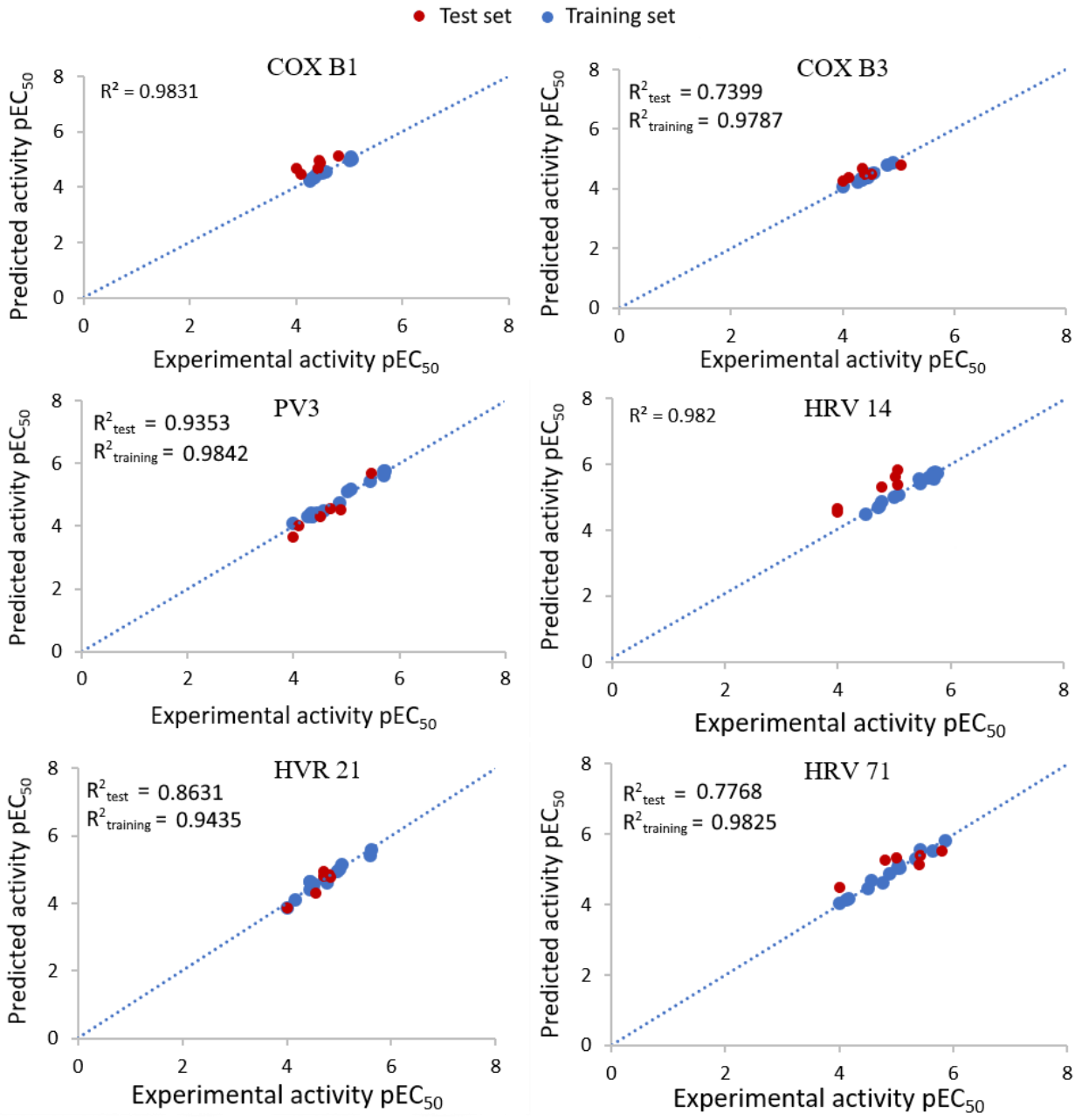

Fig. 2 Plots of predicted activity against the experimental activity of training set and the test set of six different enteroviruses, Cox B1, Cox B3, PV3, HRV 14, HRV 21 and HRV71, respectively.

than 0.5, which ascertain that the models built were not by chance and it is reliable to predict the activity of a new molecule. The coefficient of determination R^2 for both the training set and test was reported in Fig. 2. The high value of R^2 shown on the plot confirmed that the model can successfully predict the activity of a new compound due to the correlation of the experimental activity with the predicted activity.

The randomness of the activities on both negative and positive sides of the y-axis shown on the scatter plot between standardized residual activity and the experimental activity confirmed that the built model is free from systematic error (Fig. S1-S6). To discover outliers and influential compounds in the built model, the standardized residual activity for the entire data set was plotted against the leverages. The results are reported in Fig. S7-S12. The Williams plot

confirmed that there is no outlier and influential compound in the built models.

Conclusions

The quest for designing anti-enterovirus drug has led to the QSAR study of C-8-tert-butyl substituted 4-aryl-6,7,8,9 tetrahydrobenzo[4, 5]thieno[3,2-e] [1,2,4] triazolo [4,3-a]pyrimidin-5(4H)-one derivatives. The genetic function approximation multilinear regression (GFA-MLR) method was employed to construct a model that can predict the activity of new compounds. The built models were validated internally using the training set and extern ally using the test set. The validation parameters such as R^2, Q^2_{cv}, cR^2_p, VIF and others were found to be in cordial agreement with the recommended standard for an acceptable QSAR model. The applicability domain analysis reveals that there is no outlier and influencing compounds among the compounds used in building the model. So, the aim of the study has been accomplished after constructing a mathematical linear equation that can be used to design C-8-tert-butyl substituted 4-aryl-6,7,8,9tetrahydrobenzo[4, 5] thieno[3,2-e][1,2,4]triazolo[4,3-a]pyrimidin-5(4H)-one compounds with potent inhibitory ability against enteroviruses.

Conflict of interest

The authors have no conflict of interest.

References

[1] Ryu WS. Molecular virology of human pathogenic viruses, academic press, Korea, 2016; pp. 153-164.

[2] Royston LA, Tapparel C. Rhinoviruses and respiratory Enteroviruses: not as simple as ABC. Viruses 2016; 8(1):16.

[3] Chuang YY, Huang YC. Enteroviral infection in neonates. J Microbiol Immunol Infect 2019; 52(6):851-857.

[4] Biancardi E, Fennell M, Rawlinson W, Thomas PS. Viruses are frequently present as the infecting agent in acute exacerbations of chronic obstructive pulmonary disease in patients presenting to hospital. Intern Med J 2016; 46(10):1160–1165.

[5] Wang YC, Tsai CS, Yang YH, Huang KY, Hsieh WC, Kuo TY, et al. Association between enterovirus infection and asthma in children: a 16-year nationwide population-based cohort study. Pediatr Infect Dis J 2018; 37(9):844-849.

[6] Bo-Kyoung K, Joong-Heui C, Pyeonghwa J, Youngjin L, Jia JL, Kyoung RP, et al. Benserazide, the first allosteric inhibitor of *coxsackievirus* B3 3C protease. FEBS Lett 2015; 589:1795–1801.

[7] Hsyu PH, Pithavala YK, Gersten M, Penning CA, Kerr BA. The antiviral compound enviroxime targets. Antimicrob Agents Chemother 2002; 46:392–397.

[8] Tan YW, Ang MJY, Lau QY, Poulsen A, Ng FM, Then SW, et al. Antiviral activities of peptide-based covalent inhibitors of the Enterovirus 71 3C protease. Scient Rep 2016; 6:33663

[9] Obadawo BS, Oyeneyin OE, Anifowose MM, Fagbohungbe KH, Amoko JS. QSAR modeling of novel substituted 4-Phenylisoquinolinones as potent BET bromodomain (BRD4-BD1) inhibitors. Biomed Lett 2019; 5(2):69-78.

[10] Bishyajit KB, Yashwardhan RM, Neul H, Do-Hyun K, Jin-Soo S, Hae-Soo K, et al. Enterovirus inhibitory activity of C-8-tert-butyl substituted 4-aryl- 6,7,8,9-tetrahydrobenzo [4,5] thieno [3,2-e] [1,2,4] triazolo [4,3-a] pyrimidin-5(4H)-ones. Bioorg Med Chem Lett 2017; 27:3582–3585.

[11] Becke AD. Becke's three parameter hybrid method using the LYP correlation functional. J Chem Phys 1993; 98:5648–5652.

[12] Lee C, Yang W, Parr RG. 1988. Development of the Colle-Salvetti correlation energy formula into a functional of the electron density. Phy Rev B 1988; 37:785.

[13] Kennard RW, Stone LA. Computer aided design of experiments. Technometrics 1969; 11:137–148.

[14] Friedman JH. Multivariate adaptive regression splines. Annal Statist 1991; 19:1–67.

[15] Khaled KF. Modeling corrosion inhibition of iron in acid medium by genetic function approximation method: A QSAR model. Corrosion Sci 2011; 53:3457 –3465.

[16] Tropsha A, Gramatica P, Gombar VK. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. Mol Informat 2003; 22:69–77.

[17] Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP, Agrawal RK. Validation of QSAR models-strategies and importance. Int J Drug Des Discov 2011; 3, 511–519.

[18] Gramatica P, Giani E., Papa E., Statistical external validation and consensus modelling. A QSPR case study for KOC prediction. J Mol Graph Model 2007; 25:755-766.