



# Quantitative Structure–Activity Relationship Analysis of Selective Rho Kinase Inhibitors as Neuro-regenerator Agents

## Nöro-rejeneratör Ajan Olarak Seçici Rho Kinaz İnhibitörlerinin Kantitatif Yapı-Etkinlik İlişki Analizi

Seema KESAR\*, Sarvesh K. PALIWAL, Pooja MISHRA, Monika CHAUHAN  
Banasthali University, Faculty of Pharmacy, Department of Pharmacy, Rajasthan, India

### ABSTRACT

**Objectives:** To understand the role of Rho (serine/threonine) kinases in the treatment of neurological segments, attempts have been made to find potent inhibitors of Rho enzyme by a 2D quantitative structure–activity relationship (QSAR) model.

**Materials and Methods:** QSAR studies were executed on urea-based scaffolds from aniline and benzylamine analogues, which were aligned for generation of a chemometric-based model. Multivariate statistical approaches were applied including linear and nonlinear analysis such as multiple linear regression, partial least square and artificial neural network for the generation of model, and also an application of (*in silico*) absorption, distribution, metabolism, excretion studies was performed to ascertain the novelty and drug-like properties of the intended molecules.

**Results:** Ligand based analysis was implemented and showed excellent statistical relevance such as S value=0.38, F value=48.41,  $r=0.95$ ,  $r^2=0.91$ , and  $r_{cv}^2=0.86$ . Five illuminating variables, i.e., vesicle-associated membrane protein (VAMP) polarization YY component (whole molecule), VAMP dipole Y component (whole molecule), VAMP dipole Z component (whole molecule), Kier ChiV6 path index (whole molecule), and moment of inertia 2 size (whole molecule), were found and they have a profound influence on the potency of the compounds.

**Conclusion:** The values of standard statistical parameters reveal the predictive power and robustness of this model and also provide valuable insight into the significance of five descriptors. The acquired physicochemical properties (electronic, topological, and steric) show the important structural features required for activity against Rho kinase. After performing Lipinski's rule of five on urea-based derivatives no molecule was violating the rule. Therefore, these features can be effectively employed for the modeling and screening of active neurological agents as novel Rho kinase inhibitors.

**Key words:** Quantitative structure–activity relationship, chemometric analysis, Lipinski's rule of five

### ÖZ

**Amaç:** Rho (serin/treonin) kinazların nörolojik segmentlerin tedavisinde rolünü anlamak üzere, 2 boyutlu kantitatif yapı-aktivite ilişkisi (QSAR) modeli ile Rho enziminin potansiyel inhibitörlerini bulmak üzere girişimlerde bulunulmuştur.

**Gereç ve Yöntemler:** QSAR çalışmaları, kemometrik tabanlı bir modelin üretilmesi için anilin ve benzilamin analoglarından üre bazlı yapı iskeleleri üzerinde gerçekleştirildi. Modelin üretilmesi ve ayrıca (*in silico*) absorpsiyon, dağılım, metabolizma, eliminasyon çalışmalarının uygulaması, amaçlanan moleküllerin yenilik ve ilaç benzeri özelliklerinin kesin olarak anlaşılması için çoklu doğrusal regresyon, kısmi en küçük kare ve yapay sinir ağı gibi doğrusal ve doğrusal olmayan analizleri içeren çok değişkenli istatistiksel yaklaşımlar uygulandı.

**Bulgular:** Ligand bazlı analiz ile  $S=0.38$ ,  $F=48.41$ ,  $r=0.95$ ,  $r^2=0.91$  ve  $r_{cv}^2=0.86$  ile istatistiksel olarak mükemmel bir ilişki gösterdi. Beş aydınlatıcı değişkenin; vezikülle ilişkili zar proteini (VAMP) polarizasyonu YY bileşeni (bütün molekül), VAMP dipol Y bileşeni (bütün molekül), VAMP dipol Z bileşeni (bütün molekül), Kier ChiV6 yol indeksi (bütün molekül) ve atalet momenti 2 büyüklüğü (bütün molekül); bileşiklerin potensleri üzerinde önemli bir etkiye sahip oldukları bulundu.

**Sonuç:** Standart istatistiksel parametrelerin değerleri, bu modelin gücünü ve sağlamlığını ortaya koymakta ve aynı zamanda beş tanımlayıcıya ait değerli bilgiler sağlamaktadır. Elde edilen fizikokimyasal özellikler (elektronik, topolojik ve sterik), Rho kinaza karşı aktivite için gerekli olan önemli yapısal özellikleri göstermektedir. Üre bazlı türevler üzerinde Lipinski'nin beş kuralı uygulandığında bütün bileşikler bu kurala uymuştur. Bu nedenle, bu özellikler, yeni Rho kinaz inhibitörü aktif nörolojik ajanların modellenmesi ve taranmasında etkili bir şekilde kullanılabilir.

**Anahtar kelimeler:** Kantitatif yapı-etki ilişkisi, kemometrik analiz, Lipinski'nin beşer kuralı

\*Correspondence: E-mail: kesar.seema15@gmail.com, Phone: 08302916491 ORCID-ID: orcid.org/0000-0002-2307-9618

Received: 09.12.2017, Accepted: 25.01.2018

©Turk J Pharm Sci, Published by Galenos Publishing House.

## INTRODUCTION

Inhibition of the Rho-associated protein kinase (ROCK) signaling pathway induces various neuronal functions such as activation of neurite outgrowth, axonal regeneration, and pro-survival Akt. A recent outcome revealed the new site of ROCK as a new strategy to treat neurological disorders. One of the best-defined effectors of the small guanosine triphosphate (GTP) binding proteins of the Rho subfamily (RhoGTPases) is Rho-associated coiled-coil-containing protein kinase/ROCK/Rho kinase (ROK), associated with the protein kinase A/G and C family of serine-threonine kinases. Rho GTPases are members of the Ras superfamily of GTP hydrolase enzymes, used as molecular devices that can handle various signaling pathways by converting biochemically GDP-bound (inactive) to GTP-bound (active) state,<sup>1-3</sup> and the whole process is controlled by two main classes of proteins: GTPase-activating proteins and guanine nucleotide-exchange factors, which increase intrinsic GTPase activity, and gear up the conversion of GDP to GTP, respectively, which activates numerous downstream effectors such as (ROK/ROCK).<sup>4</sup> Activation of downstream Nogo receptor family members or chondroitin sulfate proteoglycan receptors by ROCK signaling stimulates axon growth inhibition; thus ROCK signaling pathway inhibition may be considered a promising strategy for the axon regeneration process.<sup>5</sup> In the treatment of spinal cord injuries, analysis of 30 preclinical studies reported that inhibition of ROCK improves locomotor recovery by 15%.<sup>6</sup> A phase 1 and 2 clinical trial (Japan Primary Registries Network identifier UMIN00000825) was designed to consider the safety and viability of the combination of fasudil and olfactory mucosa autograft in patients, but the results were not divulged, and no results have been found effectual for spinal cord injury. Thus, there is an urgent call for safe and affordable neuro-regenerative agents.

Optimization and development of a new drug involve a lot of effort and cost nearly \$900 million to pharmaceutical or biotechnology companies. To overcome these unavoidable problems, applications of computer assisted drug design or *in silico* techniques have been exclusively used to develop new safe, cheap, and biologically active chemical entities and establish their role of specificity in determining the biological activity of ROK, and also gave the prospect of possible molecular interactions of a specific inhibitor that binds to the active site of the concerned ROK enzyme. Thus, the current study aimed to build a Quantitative structure-activity relationship (QSAR) model by implementing diverse chemometric techniques to achieve improved inhibitory action towards ROK enzyme as a neuro-regenerative agent.

### 2D QSAR or Hansch-Fujita analysis

QSAR analysis ascertains mathematical correlations among structural and/or physicochemical properties called descriptors and correlated to the experimentally measured (biological) activity. The most important recent developments in the field involve a substantial increase in the size of experimental datasets available for analysis and an increased application of QSAR models as virtual screening tools to discover biologically

active molecules in chemical databases and/or virtual chemical libraries. To date no data (QSAR model) have been reported for the treatment of the central nervous system (CNS) spinal cord injuries, and thus we rely on our current model, which provided important information for finding improved new and safe oral bioavailable molecules.

## MATERIALS AND METHODS

### *Experimental data set (generation of 3D structure, charge calculation, and optimization)*

The chemical structures of compounds<sup>7</sup> were sketched using ChemDraw (8.0) and were imported in the Tools for structure-activity relationship (TSAR) 3.3 sheet, i.e. TSAR (TSAR: assimilated analysis package). Once the structures are imported, the negative logarithm of the IC<sub>50</sub> is necessary as actual activities are often skewed and are measures of the free energy of binding and so it was introduced in another column of the TSAR sheet. Prior to the descriptors' calculation, the structures were subjected to CORINA, which is often used to create 3D structures of typically drug-like molecules.<sup>8</sup> The geometry and energy were optimized in order to obtain the minimum potential energy conformer, which is a measure of the stability of the conformer.<sup>9</sup> Molecular energies are evaluated by summing (bond length, bond angle, torsion angle, Van der Waal's, and coulombic) terms for all suitable sets of atoms.

### *Preparation of the data set and data reduction*

The main reason behind descriptor calculation is to decode the information concerning the physicochemical properties of each and every molecular structure liable for specific biological activity of the molecule. Up to 500 descriptors were computed in the TSAR 3D sheet (inbuilt programming of calculating physicochemical properties) for a single molecule. In order to choose only relevant and significant sets of descriptors, data reduction was carried out to eliminate the prevalence of coincidental correlations as well as data redundancy. Firstly, descriptors with "0" values for each compound were deleted. The data were reduced on the basis of a correlation matrix developed between two descriptors. These correlation values indicate the height of co-linearity. Values somewhat close to 1 indicate the extent of better fitting of the regression model. If the intercorrelation values of the two variables were 0.5 or higher then those descriptors were retained, but if they were below 0.5 then they were eliminated from the data sheet. Finally, five descriptors, i.e., inertia moment 2 size (WM), vesicle-associated membrane protein (VAMP) polarization YY (WM), VAMP dipole Y component (WM), VAMP dipole Z component (WM), and Kier chiV6 path (WM), were found to be highly correlated with the actual activity.

### *Training and test set<sup>10</sup>*

Series of 41 analogues were segregated into training (29 compounds) and test (11 compounds) sets on the basis of diversity in the structures and the biological activity of the compounds. The compounds in the training and test sets

were employed to develop and validate the predictability of the concerned model. Some of the compounds may behave as outliers and thus they have to be discarded during the process of model development.

#### Model development and its validation

##### Linear regression analysis

Linear regression analysis helps to establish the correlation between the independent and dependent variables ( $\log 1/IC_{50}$ ) of the series. The construction of the regression model was done using training set compounds and the significance of every model was determined on the basis of statistical values<sup>11</sup> such as  $r^2$  value (correlation between dependent and independent variables), cross validation  $r^2_{cv}$  of the training set (which should be greater than 0.8),  $f$  value (degree of statistical confidence, which should be high) and  $s$ -value (standard error of estimate, which ought to be minimum). The various statistical parameters are shown in Table 1. Compounds of the test set are for the prognostic ability of the purposed model.<sup>12</sup> For a predictive model, the value of the correlation coefficient ( $r^2$ ) of the test set should be greater than 0.6 and less than the  $r^2$  of the training set.

Partial least square analysis was used to check the robustness of the model generated by the multiple least square regression analysis. To validate the results obtained from the multiple linear

regression (MLR) technique, the same data set was subjected to partial least squares (PLS) analysis. The correlation coefficient  $r^2$  and the  $r^2_{cv}$  value of the training and test sets were evaluated to ascertain the quality of the developed PLS models.<sup>13</sup>

##### Nonlinear regression analysis<sup>14</sup>

Artificial neural networks mimic the functioning of simulating the learning process by neural networks. They employ interconnections of artificial neurons with the help of computational studies and are capable of dealing with nonlinear and irregular data. Neural network analysis was carried out on the same descriptors of MLR analysis. The NNA involved three layers: input, hidden, and output layers. The input layer worked by receiving data, whereas the output layer generated the dependent variable. The hidden layer interconnected the two abovementioned layers. The final structural design of the generated NN model was (5-2-1).

##### Validation of the model

**Cross-validation ( $r^2_{cv}$ ):** This validation technique was employed to appraise the reliability of the developed statistical models. Number of compounds was shuffled from the test to the training set and vice versa for generation of the precise model. Leave-one-out methodology (when a molecule is removed from the training set and included in the test set and vice versa to predict the activity of a compound) was performed to get the final model. The cross-validation test ( $r^2_{cv}$ ) value should be greater than 0.60.<sup>15</sup>

**Activity prediction of test set compounds:** In general,  $r^2$  of the test set greater than 0.6 represents good prognostic ability of the model.<sup>16</sup>

##### Outliers

Outliers are the data points that are fitted far apart from the linear model and have some different mode of binding. An outlier of a QSAR model refers to a data point that falls outside the confidence interval of the regression line. The compounds with higher residual values that deviated from the regression line were deleted as outliers and they adversely affected the robustness of the prospective model.

## RESULTS AND DISCUSSION

Data sets of various analogues were employed for the present studies. Their chemical structures and biological activity ( $\log 1/IC_{50}$ ) are presented in Table 2. In order to assess the major molecular factors that greatly affect the ROK inhibition of derivatives belonging to anilines and benzylamines, three chemometric tools (MLR, PLS, and NN) were used to construct classical descriptor-based (QSAR) models. The compounds were divided into training and test sets and the training set comprised 29 compounds (excluding 1 outlier) and the test set 11.

##### Statistical analysis

##### MLR and PLS (linear regression method)

Initially, more than 200 descriptors were calculated for regression analysis. Due to the copious and redundant data,

**Table 1. Description of various statistical parameters**

Statistical parameters	Descriptions
$r$ (Correlation coefficient)	Correlation coefficient is a measure of quality of fit of the model. It constitutes the variance in the data and should be above 0.9.
$r^2$ (Coefficient of determination)	Coefficient of determination, also known as square of correlation coefficient, is a measure of the proportion of variability in the biological activity that can be explained by a linear relationship between independent variables and dependent variable and gives information about the goodness of fit of a model.
Standard deviation	Standard deviation is the typical amount by which an observation deviates from the regression line; it is an absolute measure of quality of fit of the model. The value of standard deviation should be small but cannot have a value lower than the standard deviation of experimental data and the magnitude of standard deviation may be attributed to some experimental error in the data as well as imperfection in the biological model.
F (Sequential Fisher test)	It is a measure of the level of statistical significance of the regression model. A higher value of F implies that a more significant correlation has been attained.
t (Test for statistical significance)	The t-test measures the statistical significance of the regression coefficients. The higher t-test values correspond to the more significant regression coefficients.

Table 2. Data set used for QSAR model development

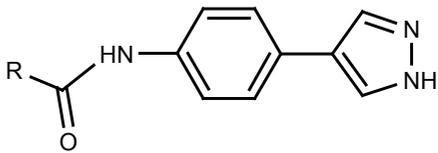
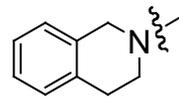
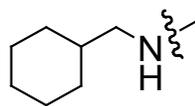
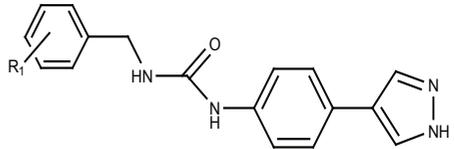
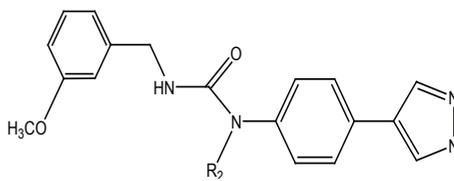
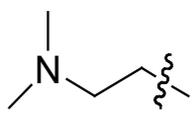
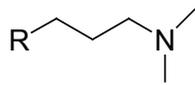
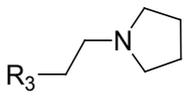
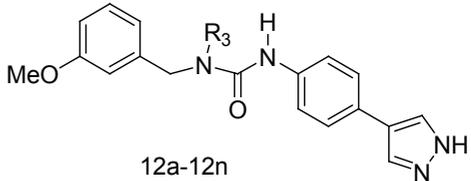
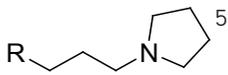
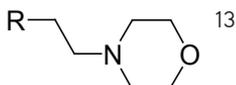
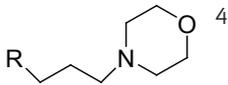
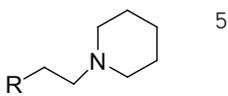
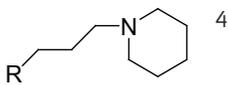
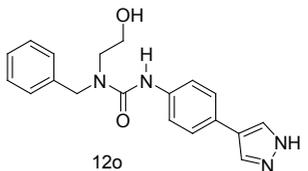
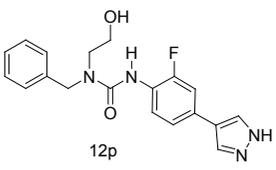
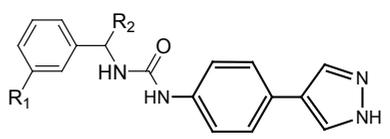
Compound no	Basic scaffold	R	ROCK-II IC <sub>50</sub> (nM)
5a		PhNH	304
5b		PhCH <sub>2</sub> NH	18
5c		PhCH <sub>2</sub> CH <sub>2</sub> NH	88
5d		PhCH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> NH	1017
5e		PhCH <sub>2</sub> O	7
5f		PhCH <sub>2</sub> CH <sub>2</sub> O	55
5g		PhCH <sub>2</sub> CH <sub>2</sub>	24
5h			280
5i		751	
5j		3-OCH <sub>3</sub>	2
5k		2,3-di-OCH <sub>3</sub>	253
5l		2,4-di-OCH <sub>3</sub>	331
5m		2,5-di-OCH <sub>3</sub>	570
5n		2,6-di-OCH <sub>3</sub>	924
5o		3,4-di-OCH <sub>3</sub>	425
5p		3,5-di-OCH <sub>3</sub>	281
5q		3-F, 4-OCH <sub>3</sub>	357
8a		CH <sub>3</sub>	87
8b		CH <sub>2</sub> CH <sub>2</sub> OH	611
8c		2984	
8e		3324	

Table 2. Continued

Compound no	Basic scaffold	R	ROCK-II IC <sub>50</sub> (nM)	
12a		CH <sub>3</sub>	1	
12b		CH <sub>2</sub> CH <sub>3</sub>	1	
12c		Cyclopropyl	1	
12d		Isopropyl	3	
12e		CH <sub>2</sub> CH <sub>2</sub> OH	1	
12f		CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	17	
12h		CH <sub>2</sub> CH <sub>2</sub> N(CH <sub>3</sub> ) <sub>2</sub>	1	
12i			3	
12j	 <p>12a-12n</p>		5	
12k			13	
12l			4	
12m			5	
12n			4	
12o	 <p>12o</p>	CH <sub>2</sub> CH <sub>2</sub> OH	12	
12p	 <p>12p</p>	CH <sub>2</sub> CH <sub>2</sub> OH	14	
		R <sub>1</sub>	R <sub>2</sub>	
14a		OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OH	2
14b		F	CH <sub>2</sub> CH <sub>2</sub> OH	2
14c		H	CH <sub>2</sub> CH <sub>2</sub> OH	2
14d		H	CH <sub>2</sub> OH	1
14e		H	CH <sub>2</sub> CH <sub>2</sub> N(CH <sub>3</sub> ) <sub>2</sub>	2

QSAR: 2D quantitative structure–activity relationship, ROCK: Rho-associated protein kinase

there was a very low value for  $r^2_{cv}$  (0.312), implying inadequate internal predictability. There was a strong necessity to build a reliable and informative set of descriptors having good correlation with the biological activity with no intercorrelation. After the deletion of the undesirable set of descriptors, the correlation matrix technique was employed and eventually five distinct physicochemical descriptors were obtained: inertia moment 2 size, VAMP polarization YY, VAMP dipole Y component, VAMP dipole Z component, and Kier chiV6 path. For QSAR analysis, the training set molecules were used to construct (linear and nonlinear) models so that a precise relationship could be established between molecules and biological activity and the molecules of test set served to check the robustness of the model. During the creation of the purposed model, one compound was shown as an outlier and not to fit either the training set or the test set; also the residual value was more than two orders of magnitude. Owing to this, outlier 5b with different prediction was omitted from the training set. The correlation matrixes of the descriptors are shown in Table 3. The selected model was assessed on the basis of different statistical values, such as regression coefficient ( $r$ ), coefficient of determination ( $r^2$ ), prognostic power of the model ( $r^2_{cv}$ ), standard deviation (SD), sequential Fisher's ratio (F), and test for statistical significance (t). The value of  $r^2$  ought to be  $>0.6$  and the value of  $r^2_{cv} >0.7$ . Statistical outputs of the MLR and PLS models are summarized in Table 4. The statistical worthiness of the developed model was evaluated in terms of square of the correlation coefficient, where  $r^2$  (MLR=0.913 and PLS=0.912) values explain 91% variance in activity, which indicates a measure of good fit by the regression equation. A small difference in (MLR)  $r^2$  (0.913) and  $r^2_{cv}$  (0.862) values implies high prognostic ability of the model. Similarly in the PLS analysis, there is a slight difference in  $r^2$  (0.912) and  $r^2_{cv}$  (0.869) values, which further ascertains the robustness of the model. The  $r^2_{cv}$  values for the MLR and PLS models ( $r^2_{cvMLR}=0.862$  and  $r^2_{cvPLS}=0.869$ ) were evaluated and both models had comparable  $r^2_{cv}$  values. The F-test indicates the significance level of the

model. The F-value of the final MLR model, 48.41, clearly shows the statistical significance of the derived model. The 's' value is the (standard) error of the regression model, and it should be low for better QSAR model generation; this is an approximation of how precisely the model will predict unknown 'Y' values. The value of 's' for the best MLR model is 0.373 and for PLS 0.307. It signifies that regression with an 's' value of 0.3 should predict Y values with a standard error of 0.3 units. The final model has the highest correlation coefficient (0.91), confirming the robustness of the model and also the model was cross validated and the  $r^2_{cv}$  value of 0.86 depicted the strength of the model. For further assurance of the significance of the descriptors, their t-test values, coefficient values, jackknife standard error (SE), and covariance SE values (Table 5) were evaluated. The values of all these parameters for all five descriptors confirmed the significance of an individual descriptor in determining the importance of structural design in exhibiting ROK inhibitory activity by a molecule. The generated model was used to understand the structural dependency of the biological activity demonstrated by the certain set of ROK inhibitors, and generated following equation 1:

*Original data*

$$Y = -0.0012895674 * X1 + 0.092464715 * X2 + 0.22125481 * X3 + 0.34796265 * X4 + 1.5525457 * X5 - 3.9303896$$

*Standardized data*

$$Y = -0.2757968 * S1 + 0.45911208 * S2 + 0.35567212 * S3 + 0.4776836 * S4 - 0.64006561 * S5 - 1.344354$$

Here X1= inertia moment 2 size (whole molecule), X2= VAMP polarization YY (whole molecule), X3= VAMP dipole Y component (whole molecule), X4= VAMP dipole Z component (whole molecule), X5= Kier ChiV6 (path) index (whole molecule), and Y represents the biological activity.

The PLS method is an addition to the MLR approach that evaluates the results to obtain assurance of the consequences when

**Table 3. Correlation matrix of the independent variables used in the final model demonstrating the degree of correlation**

Descriptors	Inertia moment 2 size (WM)	VAMP polarization YY (WM)	VAMP dipole Y component (WM)	VAMP dipole Z component (WM)	Kier ChiV6 path (WM)	Log1/IC <sub>50</sub>
Inertia moment 2 size (WM)	1	-0.0706	-0.1137	-0.3832	0.4636	-0.6950
VAMP polarization YY (WM)	-0.0706	1	-0.1074	-0.2828	0.0541	0.2270
VAMP dipole Y component (WM)	-0.1137	-0.1074	1	-0.0480	0.0266	0.2497
VAMP dipole Z component (WM)	-0.3832	-0.2828	-0.0480	1	-0.2635	0.5075
Kier ChiV6 path (WM)	0.4636	0.0541	0.0266	-0.2635	1	-0.7209
Log1/IC <sub>50</sub>	-0.6950	0.2270	0.2497	0.5075	-0.7209	1

they show minimal variation and have comparable results. Statistically, the PLS equation evaluated the robustness of the developed model on the basis of statistical parameters, i.e.,  $r^2$ . The regression equation obtained by the PLS method generated: *Equation 2*;

$$Y = -0.0013*X_1 + 0.0914*X_2 + 0.2306*X_3 + 0.3482*X_4 - 1.5262*X_5 - 3.8318$$

#### Prediction of test set compounds

The predictive ability of the intended model was validated by a set of eleven compounds. The difference between (actual and predicted) activity values of MLR and PLS analysis should be minimal for assessment of the quality of the developed models.

#### Artificial neural networks (artificial nonlinear regression analysis method)

Artificial neural networks (ANN) analyses were carried out with the same descriptors that were used for linear regression analysis in an endeavor to improve the results

obtained for ROK inhibition. In the present study, inputs for the neural network were the descriptors, while the output was the  $\log_1/IC_{50}$  values. In artificial neural network analysis, TSAR software automatically computes the number of hidden neurons, as well as patterns of the training and test sets. The number of neurons in the hidden layer and the number of rows in the training set are balanced to achieve the optimum predictive power for the neural network. The NN with three hidden neurons and 30% of compounds excluded as the test set was the most successfully trained NN model for *in vitro* ROK inhibition data as compared to the other NN models. Each analysis was repeated several times so that test RMS fit and best RMS fit were nearer to each other. Values of actual and predicted activity of the training/test set of compounds and their corresponding graph and also dependence plots of the descriptors with the output value ( $\log_1/IC_{50}$ ) clearly demonstrate that three descriptors, VAMP polarization YY, VAMP dipole Y component, and VAMP dipole Z component, are positively correlated to the output value ( $\log_1/IC_{50}$  value), while two descriptors, inertia moment 2 size and Kier chiV6

**Table 4. Equations, statistical values, and the descriptors used for the development of MLR and PLS models**

Equation Y=Biological Activity X=Descriptors	r	r <sup>2</sup>	r <sup>2</sup> <sub>cv</sub>	s value	F value	Descriptors				
						X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>
MLR Y = -0.00128*X1 + 0.09246*X2 + 0.22125*X3 + 0.34796*X4 - 1.55254*X5 - 3.93038	0.955	0.913	0.862	0.387	48.41	Inertia moment 2 size (whole molecule)	VAMP polarization YY (whole molecule)	VAMP dipole Y component (whole molecule)	VAMP dipole Z component (whole molecule)	Kier ChiV6 path (whole molecule)
PLS Y = -0.0013*X1 + 0.0914*X2 + 0.2306*X3 + 0.3482*X4 - 1.5262*X5 - 3.8318	0.943	0.912	0.869	0.307	46.0					

In linear regression analysis the dependence has the following linear form:

$$Y = b_1X_1 + b_2X_2 + \dots + b_pX_p + b,$$

where  $b_1, b_2, \dots, b_p$  are regression coefficients;  $b$  is the intercept;  $X_1, X_2, \dots, X_p$  are independent variables; and  $Y$  represents expected values of the dependent variable by the regression model. The regression coefficients  $b_1, b_2, \dots, b_p$  and the intercept are calculated by applying the method of least squares to give the smallest possible sum of squared differences between the true  $Y$  values of the dependent variable and the  $Y'$  values calculated by the regression model.

MLR: Multiple linear regression, PLS: Partial least squares, VAMP: Vesicle-associated membrane protein

**Table 5. The t-test values, jackknife SE, and covariance SE values of the descriptors used for regression analysis**

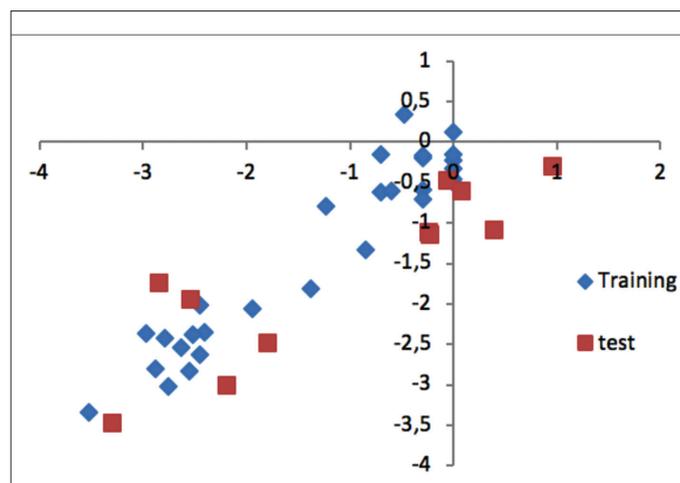
Statistical parameters	Descriptors	Coefficient <sup>a</sup>	Jackknife SE <sup>b</sup>	Covariance SE <sup>c</sup>	t-value <sup>d</sup>	T-probability <sup>e</sup>
Inertia moment 2 size (WM)		-0.0012896	0.00038428	0.00042338	-3.0459	0.0057359
VAMP polarization YY (WM)		0.092465	0.019823	0.015956	5.795	6.6441e-006
VAMP dipole Y component (WM)		0.22125	0.052874	0.046879	4.7197	9.3614e-005
VAMP dipole Z component (WM)		0.34796	0.065733	0.062124	5.6011	1.0638e-005
Kier ChiV6 path (WM)		-1.5525	0.23786	0.20289	-7.6521	9.1178e-008

<sup>a</sup>Represents the regressions coefficient for each variable in the QSAR equations, <sup>b</sup>Represents the standard error estimation on each regression coefficient derived from a jackknife method on the final regression model, <sup>c</sup>Represents an estimate of the standard error on each regression coefficient derived from the covariance matrix, <sup>d</sup>Measurement of the significance of each variable incorporated in the model, <sup>e</sup>Represents statistical significance for t-values, VAMP: Vesicle-associated membrane protein, QSAR: 2D quantitative structure-activity relationship, SE: Standard error

path index, are negatively correlated to  $\log 1/IC_{50}$ , which is in agreement with the results of MLR analysis.

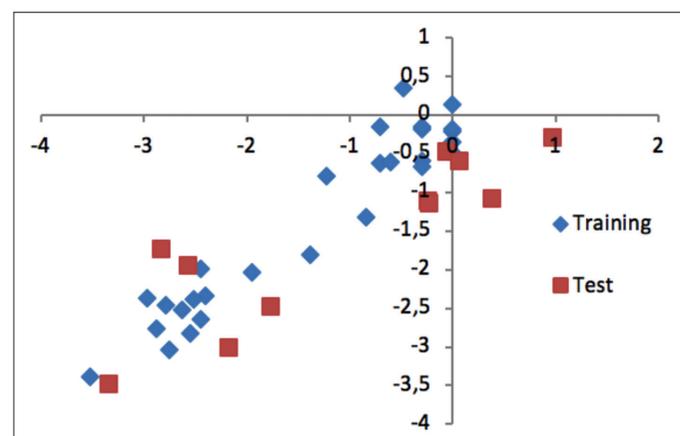
#### Comparison between linear and nonlinear methods

Different statistical approaches were employed to quantify the predictive ability of the generated models in terms of statistical fit ( $r^2$ ). Both MLR and PLS (two linear methods) were studied and had comparable results. The  $r^2$  values for the training set for QSAR analysis were  $r^2_{MLR}=0.91$ ,  $r^2_{PLS}=0.91$ , and  $r^2_{ANN}=0.93$  and for the test set the values were  $r^2_{MLR}=0.75$ ,  $r^2_{PLS}=0.75$ , and  $r^2_{ANN}=0.73$ . On the basis of the predictive power of the model it is very clear that conventional MLR and PLS as well as ANN analysis generated a highly robust and predictive model. The actual and predicted activity for the compounds of the training and test sets were obtained after MLR, PLS, and NNA and are shown in Table 6 and their corresponding graph are shown in Figures 1-3. A summary of the performance of the different prediction models is given in Table 7.



**Figure 1.** Graph plotted between actual and predicted activity of training and test set through MLR analysis

MLR: Multiple linear regression



**Figure 2.** Graph plotted between actual and predicted activity of training and test set through PLS analysis

PLS: Partial least squares

#### Mechanistic interpretation of the descriptors

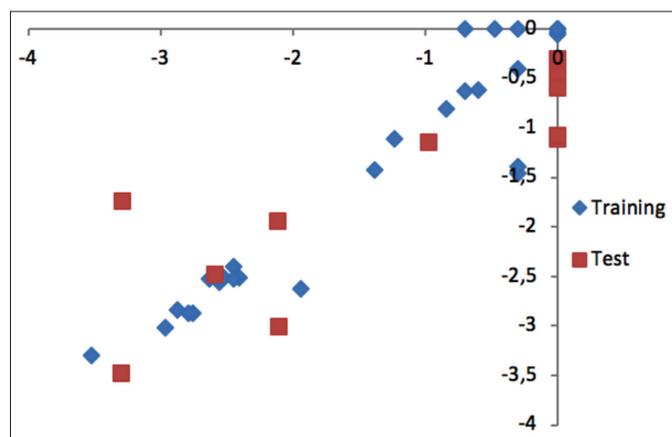
The results of MLR, PLS, and NN reveal the importance of inertia moment 2 size (WM), VAMP polarization YY (WM), VAMP dipole Y component (WM), VAMP dipole Z component (WM), and Kier chiV6 path (WM) and in fact a strong correlation was observed between ROK inhibitory activity and the five descriptors (Table 8). The dependence plots of the descriptors with the output value ( $\log 1/IC_{50}$ ) are shown in Figure 4 and the structure-activity relationship derived from QSAR analysis is shown in Figure 5.

#### Kier ChiV6 path index

Out of the five parameters, Kier ChiV6 path index (WM) was shown to be an essential descriptor in defining the biological activity of urea-based derivatives as evident from the correlation matrix and the t-value. Kier ChiV6 path index was initially defined by Randic and subsequently by Kier and Hall. It illustrates a number of series represented by "order" and "subgraph" type. By definition, a chi index is a calculation of a known type of subgraph such as path (P), cluster (C), path/cluster (PC), and ring (CH), weighted by a function of the delta values. The descriptor highlights different aspects of atom connectivity within a molecule. It also helps us to examine the substitution pattern in benzene rings and the amount of branching rings.<sup>17</sup> In our study it was observed that the KierchiV6 (path index) descriptor is negatively correlated with permeability; according to the study the sixth order valence connectivity index (KierchiV6) encodes structural complexity, such as the size and heteroatom content of the rings. This complexity is observed in the least and most active compound. As the size of the compound decreases, biological activity increases.

#### VAMP polarization YY

VAMP polarization YY is a spatial descriptor that calculates the electronic properties of a compound and projects polarization towards YY planes. There is a direct relation between polarizability and the number of valence electrons on every atom.<sup>18</sup> A positive correlation of VAMP polarization YY with the



**Figure 3.** Graph plotted between actual and predicted activity of training and test set through FFNN analysis

FFNN: Feed forward neural network

**Table 6. Actual activity versus predicted activity and corresponding residual for the training set of compound**

Compound no.	Actual activity (Log1/IC <sub>50</sub> values)	Predicted activity (Log1/IC <sub>50</sub> values)		
		MLR	PLS	FFNN
<b>Training set compounds</b>				
5c	-1.94	-2.05	-2.08	-2.62
5e	-0.84	-1.33	-1.37	-0.80
5g	-1.38	-1.80	-1.85	-1.42
5h	-2.44	-2.02	-2.10	-2.39
5i	-2.87	-2.79	-2.87	-2.84
5k	-2.40	-2.34	-2.34	-2.51
5l	-2.51	-2.38	-2.37	-2.51
5m	-2.75	-3.02	-3.02	-2.87
5n	-2.96	-2.37	-2.37	-3.02
5o	-2.62	-2.53	-2.49	-2.52
5p	-2.44	-2.63	-2.62	-2.60
5q	-2.55	-2.83	-2.82	-2.65
8b	-2.78	-2.43	-2.39	-2.74
8e	-3.52	-3.34	-3.30	-3.09
12a	0	-0.23	-0.29	-0.06
12b	0	-0.45	-0.45	-0.23
12c	0	-0.32	-0.32	-0.16
12d	-0.47	0.34	0.37	-0.04
12e	0	-0.15	0.13	-0.07
12f	-1.23	-0.79	-0.79	-0.09
12h	0	-0.50	-0.50	-0.09
12j	-0.69	-0.61	-0.64	-0.68
12l	-0.60	-0.61	-0.59	-0.79
12m	-0.69	-0.15	-0.14	-0.13
14a	-0.30	-0.58	-0.55	-0.07
14b	-0.30	-0.20	-0.22	-0.20
14d	-0.30	-0.17	-0.15	-0.06
14c	0	0.11	0.13	-0.05
14e	-0.30	-0.70	-0.80	-0.35
<b>Test set compounds</b>				
5a	-2.48	-2.28	-2.38	-0.60
5d	-3.00	-2.46	-2.15	-2.48
5f	-1.74	-2.54	-2.75	-2.68
8a	-1.93	-2.03	-1.85	-2.76
8c	-3.47	-3.09	-3.21	-3.09
12i	-0.47	-0.88	-0.93	-0.07

**Table 6. Continued**

Compound no.	Actual activity (Log1/IC <sub>50</sub> values)	Predicted activity (Log1/IC <sub>50</sub> values)		
		MLR	PLS	FFNN
12k	-1.11	-0.45	-0.81	-0.16
12n	-0.60	-0.93	-0.84	-0.10
12o	-1.07	-1.49	-1.04	-0.03
12p	-1.14	-0.84	-1.06	-0.20
5j	-0.30	-0.32	-0.29	-0.03

MLR: Multiple linear regression, PLS: Partial least squares, FFNN: Feed forward neural network

**Table 7. Correlation coefficients (r<sup>2</sup>) between predicted and experimental values of the MLR, PLS, and NN models**

r <sup>2</sup> <sub>training</sub>			r <sup>2</sup> <sub>test</sub>		
MLR	PLS	NN	MLR	PLS	NN
0.91	0.91	0.90	0.75	0.75	0.72

MLR: Multiple linear regression, PLS: Partial least squares, NN: Neural network

biological activity reveals a direct link between the chemical reactivity index and biological activity. The compounds 12a, 12b, 12c, 12e, and 14c, having high values of VAMP polarization YY, are the most active, whereas compound 8e, with low VAMP polarization YY, is the least active.

#### VAMP dipole Y component

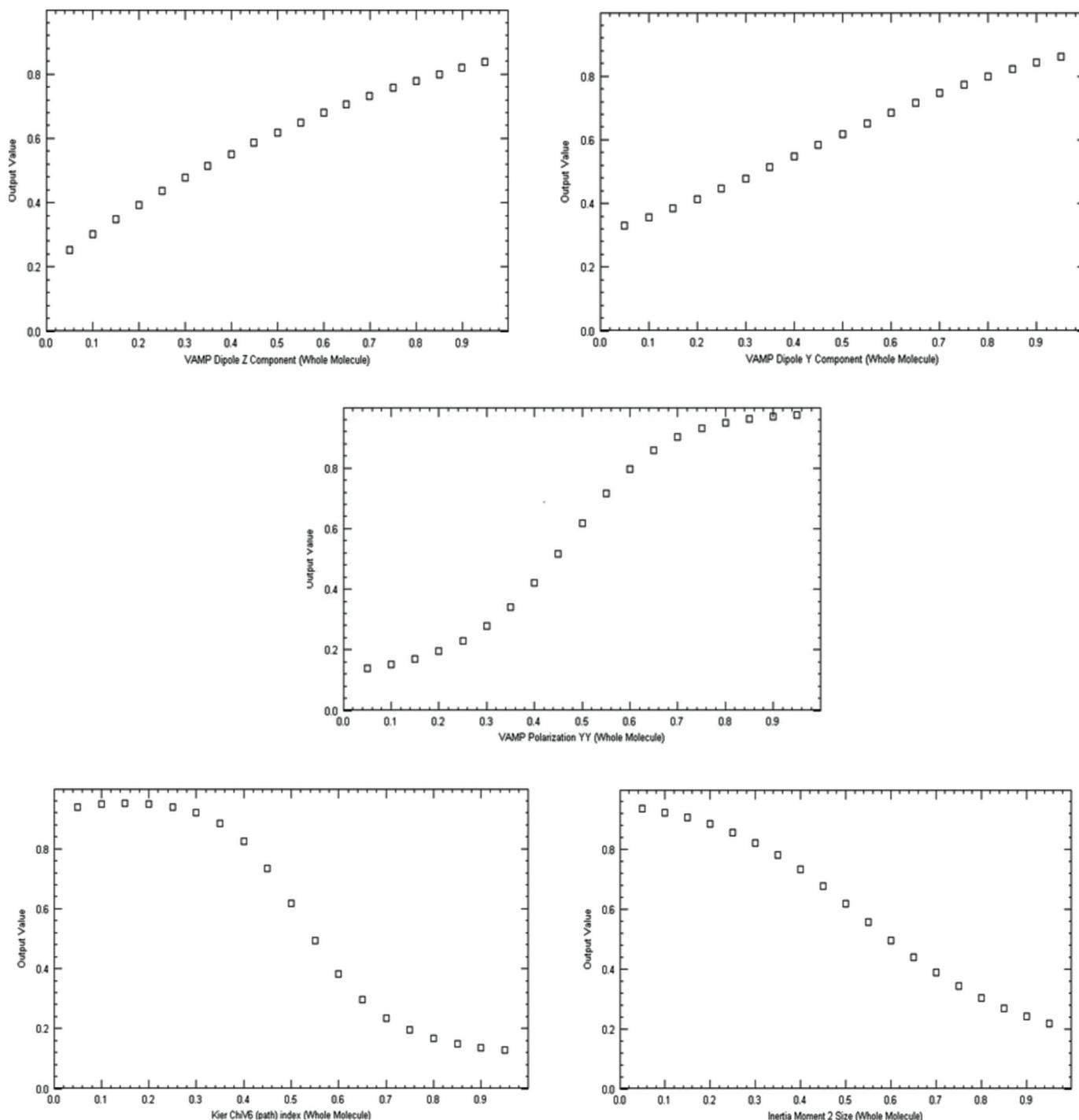
VAMP dipole Y component is an electronic parameter and is due to the degree of charge separation in a molecule. It describes the substituent point of attachment with the bond sited along the Y-axis.<sup>19</sup> A positive correlation between VAMP dipole Y and biological activity reveals a direct link between the chemical reactivity index and biological activity.

#### Moment of inertia 2 (size)

Inertia moment 2 size (WM) with reference to an axis is defined as the product of the mass times the distance from the axis squared. The higher positive correlation coefficient of inertia moment with permeability data and the high t-values (t-values define the statistical significance of a descriptor) suggest that the orientation behavior with respect to the size of whole molecule is of utmost importance in the binding interaction with the receptor site as well as in imparting greater permeability. In the present study the biological activity increases with a decrease in moment of inertia 2 (size) of the whole molecule. This phenomenon can be explained by taking the example of most active compound that has low value of moment of inertia 2 (size) of the whole molecule in comparison to the least active compound that has high value of moment of inertia 2 (size) of the whole molecule. Hence it can be concluded that by decreasing the moment of inertia 2 (size) ROK inhibitory activity can be increased.

#### VAMP dipole X component

Vamp is a semiempirical molecular orbital package used to determine electrostatic properties and perform optimizing



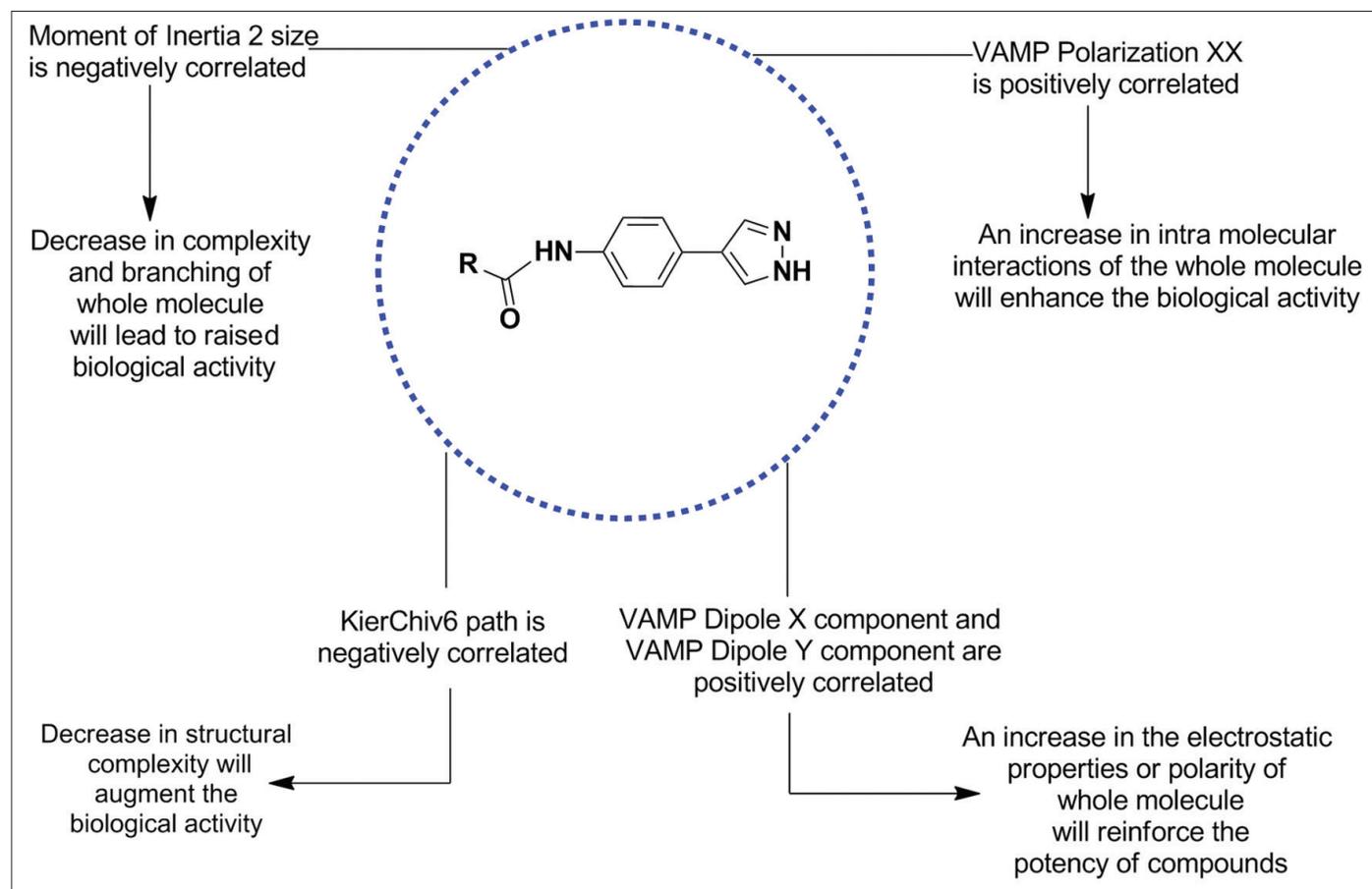
**Figure 4.** Dependency graph illustrating correlation between all five descriptors and actual activity data

of structure such as total energy, electronic energy, nuclear repulsion energy, accessible surface area, atomic charge, mean polarizability, heat of formation, highest occupied molecular orbital and lowest unoccupied molecular orbital eigenvalues, ionization potential, total dipole, polarizability, and dipole components. The positive coefficient of this expression in the proposed model elucidates that the higher the value, the better is the activity, and indicates that biological activity with respect to ROK inhibition is directly dependent upon the chemical

permanence of the compounds in biochemical systems. Some exciting facts were revealed during the analysis of the derived descriptors and their correlation with the structural design of the molecules. In the comparison of the least active molecule with the most active compound of the selected series, we found that when the least active compound (8e) was substituted with dimethylamine the shape and volume of the molecule were altered, which eventually changed the optimal binding affinity of the molecule. However, in the most

active compounds (12a, 12b, 12c, 14c, and 12e) dimethylamine was replaced with either simple alkyl or oxygen-containing alkyl substitutions, which gradually reduced the size and volume of the molecule, thus resulting in enhancement of the

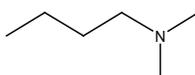
ROK inhibitory activity. Interestingly, both alkyl and oxygen-containing alkyl chains have analogous molecular mass and, thus, no further bulk was loaded onto the molecule. Attribution of alkyl or oxygen-containing alkyl substitutions



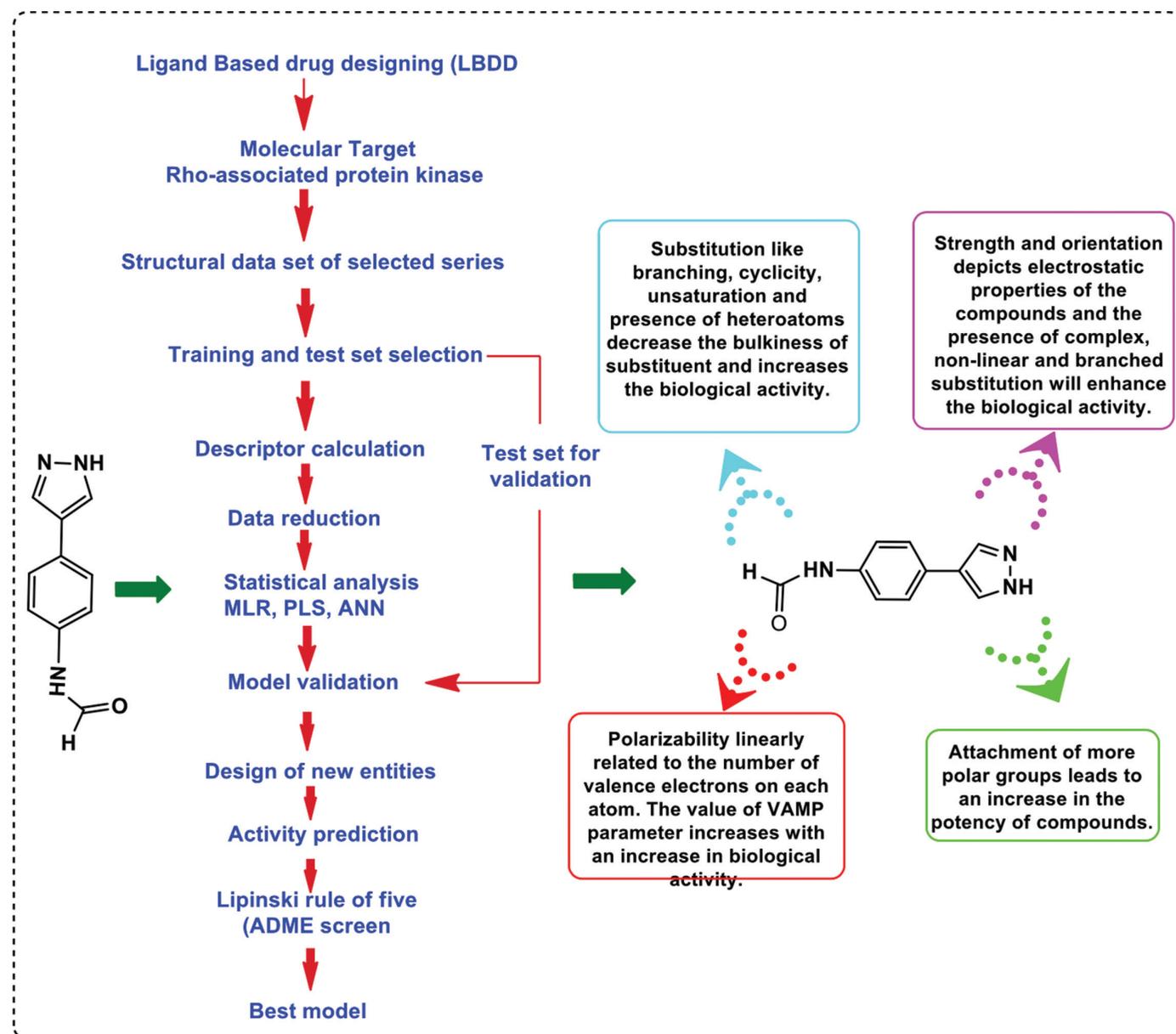
**Figure 5.** Systematic interpretation of the physicochemical descriptors used in the model, which briefly explained the structure–activity relationship

VAMP: Vesicle-associated membrane protein

**Table 8.** Correlation of biological activity of active and inactive molecules with all five descriptors

Name of compound	Substitution at R <sub>3</sub> (12a - 12h)	Biological activity LogIC <sub>50</sub> (nM)	Inertia moment 2 size (WM)	VAMP polarization YY (WM)	VAMP dipole Y (WM)	VAMP dipole Z (WM)	KierChi6 (path) (WM)
12a	CH <sub>3</sub>	1	1090.7	55.661	0.032	3.045	0
12b	CH <sub>2</sub> CH <sub>3</sub>	1	1080.4	62.251	-3.086	-0.588	0
12c	Cyclopropyl	1	1099	63.302	-3.274	-0.314	0
12e	CH <sub>2</sub> CH <sub>2</sub> OH	1	1115.4	60.45	-2.487	0.492	0
12h	CH <sub>2</sub> CH <sub>2</sub> N(CH <sub>3</sub> ) <sub>2</sub>	1	1216.1	54.60	-0.622	0.224	0
Substitution at R <sub>1</sub> , R <sub>4</sub>							
14c	H CH <sub>2</sub> CH <sub>2</sub> OH	1	1171.8	55.684	1.552	0.199	0
Inactive compound	8e 	3324	1537.5	43.504	-3.134	-4.05	0

VAMP: Vesicle-associated membrane protein



## Graphical Abstract

MLR: Multiple linear regression, PLS: Partial least squares, ANN: Artificial neural network, ADME: Association of destination management executives

positively affects the electrostatic nature of the substituent and thus additional atoms contributing to the energy, resulting in enhanced biological activity, and also showing clearly that hydroxyl or oxygen substitutions are highly electronegative, which augments the overall polarizability of the molecule. However, the negative correlation of Kier Chiv6 (path) index at  $R_1$  clearly reveals that reductions in the bulkiness and volume at certain positions in the whole molecule lead to an increased biological profile. Additionally, compounds 12a, 12b, 12c, 14c, and 12e have lower molecular mass (378.52, 350.46, 362.47, 336.43, and 366.46) than compound 8e (407.57), also confirming the authentication of the descriptor, which leads to compression in the shape of the molecule, allowing it to conveniently enter the binding site and align in such a way that it fits snugly with the walls of the active site.

*Absorption, distribution, metabolism, and excretion studies*

The 'rule of five', given by Lipinski, is known as the therapeutic relevance or property of drug-likeness. It is an empirical approach traditionally utilized for calculating drug-like properties in a molecule that clearly postulates that molecules with a molecular weight less than 500,  $\log p < 5$ , hydrogen bond donors less than 5, and hydrogen bond acceptors less than 10 exhibit an excellent pharmacokinetics profile in terms of absorption or permeation through the biological membrane.<sup>20</sup> This rule explains the absorption, distribution, metabolism, and excretion of bioactive compounds in a superior organism. Lipinski's rule of five was calculated for the particular series of compounds and no molecule was found to have violated the above stated set of rules (Table 9). This overtly indicates that all compounds showed adequate pharmacokinetic profiles.

**Table 9. Illustrating the values of various parameters constituting Lipinski's rule of five**

Compound	ADME weight (WM)	ADME H-bond acceptors (WM)	ADME H-bond donors (WM)	ADME log P (WM)	ADME violations (WM)
4	306.4	2	3	2.910	0
5	293.35	3	2	3.309	0
6	291.38	2	2	3.149	0
7	318.41	2	2	2.887	0
8	298.43	2	3	2.781	0
9	352.43	4	3	2.153	0
15	352.43	4	3	2.153	0
16	352.43	4	3	2.153	0
17	352.43	4	3	2.153	0
18	352.43	4	3	2.153	0
19	352.43	4	3	2.153	0
20	340.39	3	3	2.546	0
21	366.46	4	3	2.209	0
22	407.57	4	2	2.684	0
23	378.52	3	2	3.486	0
27	350.46	3	2	2.995	0
28	362.47	3	2	3.049	0
29	364.49	3	2	3.408	0
30	366.46	4	3	2.209	0
31	380.49	4	2	2.488	0
32	393.54	4	2	2.632	0
33	433.61	4	2	3.009	0
34	449.61	5	2	2.341	0
43	433.61	4	2	3.354	0
44	306.4	3	3	2.406	0
45	293.35	4	4	2.086	0
46	291.38	3	4	2.478	0

ADME: Association of destination management executives

### Study limitations

The findings of the present QSAR analysis will be advantageous only for the modeling of potent ROK inhibitors as active neurological agents. For the future aspects we will try to screen multitargeted novel ROK inhibitors.

### CONCLUSIONS

A 2D QSAR study was performed to establish a structural and physicochemical relationship required for the inhibition of a molecular target against a neurological disease, i.e., ROK. The

statistically significant model highlighted the significance of electronic, topological, and steric descriptors. The authenticity of the projected model was checked by validation and cross-validation ( $r^2_{cv}$ ) based on leave-one-out methodology. Overfitting of the models was checked by considering the difference between  $r^2$  and  $r^2_{cv}$ . In a planned study, an attempt was made to understand the dependence of biological activity on the structural design accountable for their specific ROK inhibition. The main problem regarding the CNS-related drug is its ability to cross the blood-brain barrier, and for crossing this barrier there should be optimal log P, volume, shape, molecular mass, and polarizability. The proposed model overtly points towards the introduction of optimal bulk or charge distribution along with the shape and size of the molecules to determine the binding efficacy of the molecule to the receptor domain, which eventually increases the inhibitory profiles of the selected molecules. VAMP polarization YY component (WM), VAMP dipole Y component (WM), and VAMP dipole Z component (WM) descriptors were positively correlated with activity while the other two, Kier ChiV6 path index (WM) and moment of inertia 2 (size) (WM), were negatively correlated descriptors, and projected molecular structure information with reference to a specific rotation axis or the rotational analogue to mass, and groups that decrease Kier ChiV6 and inertia of moment at substitutions will increase the predictability of the model. The above final model reveals the significance of selected descriptors and their correlation with biological activity and provided substantial insights to plan new chemical scaffolds with improved selectivity outlines. Design of ROK inhibitors incorporating the appropriate essential features or descriptors will increase the chances of getting new better molecules with enhanced inhibitory profiles.

### ACKNOWLEDGEMENT

The authors are thankful to the Vice-chancellor of Banasthali University for providing the necessary research facilities.

*Conflict of Interest:* No conflict of interest was declared by the authors.

### REFERENCES

1. Van Aelst L, D'Souza-Schorey C. Rho GTPases and signaling networks. *Genes Dev.* 1997;11:2295-2322.
2. Burridge K, Wennerberg K. Rho and Rac take center stage. *Cell.* 2004;116:167-179.
3. Bishop AL, Hall A. Rho GTPases and their effector proteins. *Biochem J.* 2000;348:241-255.
4. Rossman KL, Der CJ, Sondek J. GEF means go: turning on RHO GTPases with guanine nucleotide-exchange factors. *Nat Rev Mol Cell Biol.* 2005;6:167-180.
5. Fujita Y, Yamashita T. Axon growth inhibition by RhoA/ROCK in the central nervous system. *Front Neurosci.* 2014;8:338.
6. Watzlawick R, Sena ES, Dirnagl U, Brommer B, Kopp MA, Macleod MR, Howells DW, Schwab JM. Effect and reporting bias of RhoA/ROCK-

- blockade intervention on locomotor recovery after spinal cord injury: a systematic review and meta-analysis. *JAMA Neurol.* 2014;71:91-99.
- Lum C, Kahl J, Kessler L, Kucharski J, Lundström J, Miller S, Nakanishi H, Pei Y, Pryor K, Roberts E, Sebo L, Sullivan R, Urban J, Wang Z. 2,5-Diaminopyrimidines and 3,5-disubstituted azapurines as inhibitors of glycogen synthase kinase-3 (GSK-3). *Bioorg Med Chem Lett.* 2008;18:3578-3581.
  - Dwivedi N, Mishra BN, Katoch VM. 2D-QSAR model development and analysis on variant groups of anti-tuberculosis drugs. *Bioinformation.* 2011;7:82-90.
  - Abraham MH, Whiting GS. Hydrogen bonding: XXI. Solvation parameters for alkylaromatic hydrocarbons from gas-liquid chromatographic data. *J Chrom.* 1992;594:229-241.
  - Wu W, Walczak B, Massart DL, Heuerding S, Erni F, Last IR, Prebble KA. Artificial neural networks in classification of NIR spectral data: design of the training set. *Chemometrics Intell Lab Syst.* 1996;33:35-46.
  - Rose K, Hall LH. E-State Modeling of Fish Toxicity Independent of 3D Structure Information. *SAR QSAR Environ Res.* 2003;14:113-129.
  - Rose K, Hall LH, Kier LB. Modeling Blood-Brain Barrier Partitioning Using the Electrotopological State. *J Chem Inf Comput Sci.* 2002;42:651-666.
  - Voelkel A. Structural descriptors in organic chemistry - new topological parameter based on electrotopological state of graph vertices. *Computers & Chemistry.* 1994;18:1-4.
  - Bottcher CJF. *Theory of Electric Polarization.* (2nd ed). Amsterdam; Elsevier Press; 1973.
  - Hopfinger AJ. *Conformational Properties of Macromolecules.* New York; Academic Press; 1973.
  - TSAR reference guide.
  - Rinaldi D, Rivail JL. Molecular polarisability and dielectric effect of medium in the liquid phase: Theoretical study of the water molecule and its dimers. *Theor Chim Acta.* 1973;32:57-70.
  - Hall LH, Kier LB. The molecular connectivity chi indices and kappa shape indices in structure-property modeling. In: Lipkowitz KB, Boyd DB, eds. *Reviews of Computational Chemistry.* New York; VCH publishers; 1991;2:367-412.
  - Vadlamudi SM, Kulkarni MV. 3D-QSAR of Protein Tyrosine Phosphatase 1B Inhibitors by Genetic Function Approximation. *Internet Electron J Mol Des.* 2004;3:586-609.
  - Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and Computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev.* 1997;23:3-26.