

DOI: 10.4274/tjps.70288

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

Nöro-rejeneratör ajan olarak Seçici Rho Kinaz İnhibitörlerinin Kantitatif Yapı-Etki İlişkileri Analizi

Seema KESAR*, Sarvesh K. PALIWAL, Pooja MISHRA, Monika CHAUHAN

ABSTRACT

Introduction: Understanding 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 2D quantitative structure activity relationship (QSAR) model.

Materials and Methods: QSAR studies were executed on urea based scaffolds from anilines and benzylamines analogs, which were aligned for generation of chemometric based model. Multivariate statistical approaches have been applied includes linear and non-linear analysis such as multiple linear regression, partial least square and artificial neural network for the generation of model, and also an application of ADME studies was performed to ascertain the novelty and drug like properties of the intended molecules.

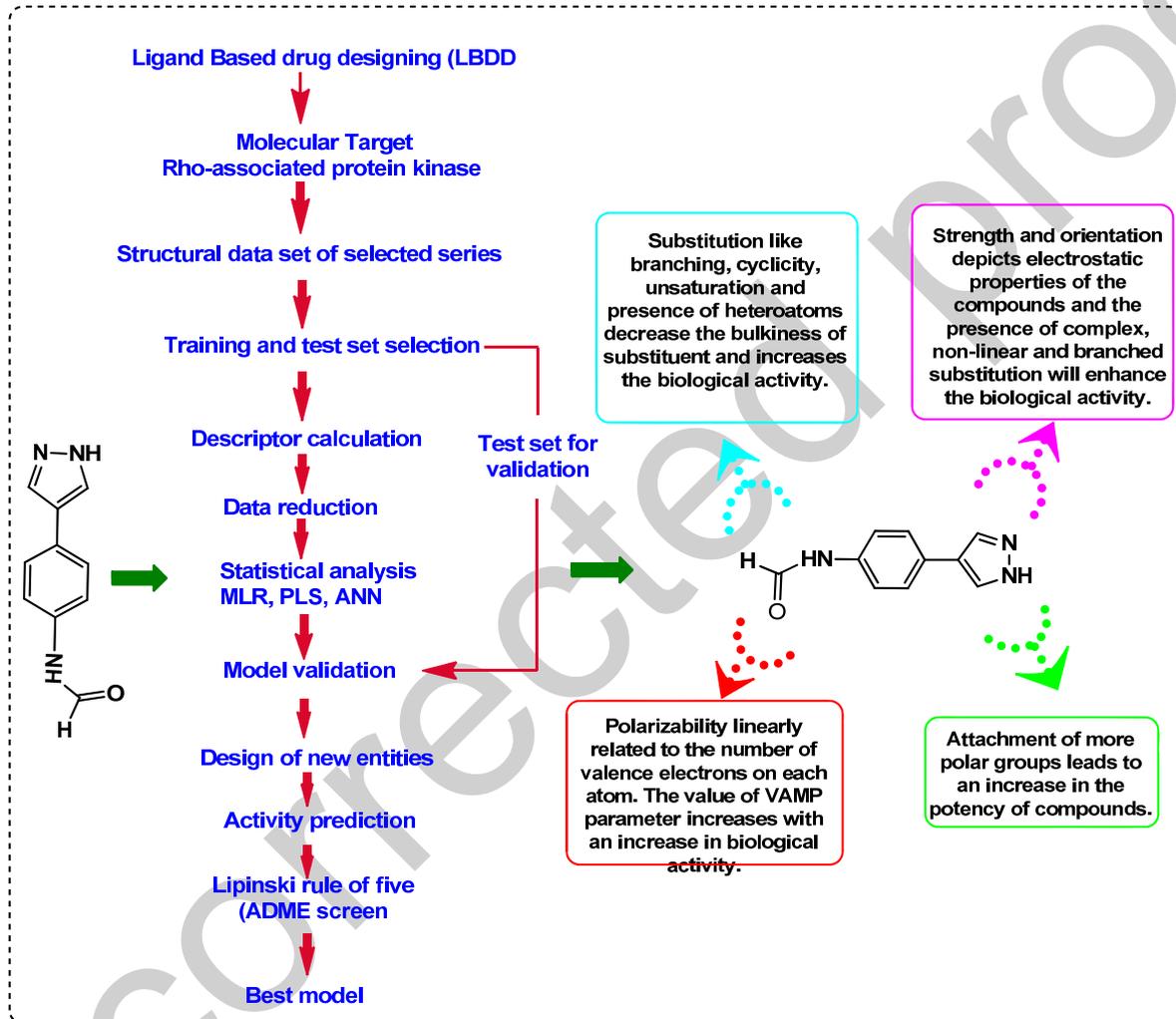
Results: Ligand based analysis have been implemented and having excellent statistical relevance such as S-value= 0.38, F-value= 48.41, $r = 0.95$, $r^2 = 0.91$ and $r^2_{cv} = 0.86$. Five illuminating variables, VAMP polarization YY component (whole molecule), VAMP dipole Y component (whole molecule), VAMP dipole Z component (whole molecule), Kier ChiV6 path index (whole molecule) and Moment of inertia 2 size (whole molecule) were found and own profound influence on the potency of the compounds.

Discussion and Conclusion: The values of standard statistical parameters reveal the predictive power and robustness of this model and also provided valuable insight to the significance of five descriptors. The acquired physicochemical properties (electronic, topological and steric) shows the important structural features required for activity against Rho kinase. After performing Lipinski's rule of five on urea based derivatives no molecule were violating the rule. So, these features can be effectively

employed for modeling and screening of active neurological agents as novel Rho-Kinase inhibitors.

Keywords: Quantitative structure activity relationship, Chemometric analysis, Lipinski's rule of five

Graphical Abstract



ÖZ

Giriş: Nörolojik kesimlerin tedavisinde Rho (serin / treonin) kinazların rolünü anlamak için, Rho enziminin güçlü kısıtlayıcılarını 2D niceliksel yapı aktivitesi ilişkisi (QSAR) modeliyle bulmak için girişimlerde bulunulmuştur.

Gereç ve Yöntemler: QSAR çalışmaları, kimyasal olarak modellenmiş modeli oluşturmak üzere hizalanan anilinler ve benzilamin analoglarından üre bazlı iskele üzerinde yürütülmüştür. Modelin oluşturulması için çoklu doğrusal regresyon, kısmi

en küçük kareler ve yapay sinir ağı gibi doğrusal ve doğrusal olmayan analizleri içeren çok değişkenli istatistiksel yaklaşımlar uygulanmış ve ayrıca Amaçlanan moleküllerin ilaç benzeri özelliklerinin saptanması için (in silico?) ADME çalışmaları uygulanmıştır.

Bulgular: Ligand temelli analizler uygulanmış ve S-değeri = 0.38, F-değeri = 48.41, $r = 0.95$, $r^2 = 0.91$ ve $r^2_{cv} = 0.86$ gibi mükemmel istatistiksel yöndedir. Beş aydınlatıcı değişken, VAMP polarizasyon YY bileşeni (tüm molekül), VAMP dipol Y bileşeni (tüm molekül), VAMP dipol Z bileşeni (tüm molekül), Kier ChiV6 yol indeksi (tüm molekül) ve atalet momenti 2 boyutu (tüm molekül) idi Bileşiklerin potensi üzerinde bulunmuş ve kendi üzerinde büyük etkisi vardır.

Tartışma ve Sonuç: Standart istatistiksel parametrelerin değerleri, bu modelin öngörücü gücü ve dayanıklılığını ortaya koymakta ve aynı zamanda beş tanımlayıcıya verilen önemi değerli bilgiler sağlamaktadır. Edinilen fizyolojik özellikler (elektronik, topolojik ve sterik), Rho kinaza karşı aktivite için gerekli olan önemli yapısal özellikleri gösterir. Lipinski'nin üre bazlı türevleri üzerindeki beş kuralını uyguladıktan sonra hiçbir kural bu kuralları ihlal etmedi. Dolayısıyla, bu özellikler, aktif nörolojik ajanların yeni Rho-Kinaz inhibitörleri olarak modellenmesi ve taranmasında etkili bir şekilde kullanılabilir.

Anahtar Kelimeler: Kantitatif yapı etki ilişkisi, Kemometrik analiz, Lipinski'nin beşler kuralı

INTRODUCTION

Inhibition of ROCK signalling pathway induces various neuronal functions such as activation of neurite outgrowth, axonal regeneration and pro-survival Akt. A recent outcome reveals the new site of ROCK as a new strategy to treat neurological disorders (NLDs). 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/ROK (Rho-kinase) associated to Protein kinaseA/G and C family of (serine-threonine kinases). RhoGTPases is a member of the Ras super-family of guanosine triphosphate hydrolase enzymes, purposes as molecular devices that can handle various signaling pathways by converting biochemically GDP-bound (inactive) to GTP-bound (active) state,¹⁻³ and the whole process is controlled by two main classes of proteins: GTPase-activating proteins (GAPs), and guanine nucleotide-exchange (GEFs) factors which increase intrinsic GTPase activity, and geared up the conversion of GDP to GTP respectively, which activates numerous of downstream effectors such as (Rho-kinase/ROCK)⁴.

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 announced the promising strategy for axon regeneration process.⁵ In the treatment of spinal cord injury, analysis of 30 preclinical studies reported that inhibition of ROCK is improves by 15% locomotor recovery.⁶ In the phase of 1/2 clinical trial, (Japan Primary Registries Network identifier UMIN000000825) designed to consider the safety and viability of the combination of (fasudil and olfactory mucosa autograft) into patients, but they don't divulge their results, and no results have been found effectual for the spinal cord injury. Thus, there is an urgent call for safe and affordable neuro-regenerative agents.

Optimization and development of one new drug contain a lot of efforts and nearly consume \$900 million of pharmaceutical or biotechnology companies. To conquer these unavoidable problems, applications of computer assisted drug design (CADD) or in-silico techniques been exclusively used to develop safe, cheap and biologically active new

chemical entities and establish their role of specificity in determining the biological activity of Rho-kinase, also gave prospect of possible molecular interactions of specific inhibitor which binds to active site of concerned (rho-kinase) enzyme. In sight of this, the current study relates to build a QSAR model by implementing diverse chemometric techniques to achieve improved inhibitory action towards Rho-kinase enzyme as neuro-regenerative agents.

2D QSAR or Hansch-Fujita

Quantitative Structure Activity Relationship (QSAR) analysis ascertains mathematical correlation among structural and/or physicochemical properties called descriptors and correlated to the experimentally measured (biological) activity. The most important recent developments in the field concur with a substantial increase in the size of experimental datasets available for the 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. Till the date no data (QSAR model) was reported for the treatment of (CNS) spinal cord injuries, and thus we rely

on our current model which has been able to provide important information for finding improved oral bioavailable new and safe molecules.

MATERIALS AND METHODS

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

The chemical structures of compounds⁷ were sketched using ChemDraw (8.0) and were imported in the TSAR3.3 sheet i.e., Tools for structure activity relationship (TSAR: assimilated analysis package). Once the structures were imported, the negative logarithm of the IC₅₀ is necessary as actual activities are often skewed and are measures the free energy of binding so it was introduced in another column of TSAR sheet. Prior to the descriptors calculation, the structures were subjected to the CORINA which is often used to create 3D structures of typically drug-like molecules.⁸ The geometry and energy were optimized in order to obtain the minimum potential energy conformer that is a measure of the stability of the conformer.⁹ Cosmic evaluates molecular energies by summing (bond-length, bond-angle, torsion-angle, Van der Waal's and coulombic) terms for all suitable sets of atom.

Preparation of data set and Data reduction

The main reason behind descriptor calculation is to decode the information, concerning physicochemical properties of each and every molecular structure liable for specific biological activity of the molecule. Up to 500 descriptors were computed in TSAR 3D sheet (inbuilt programming of calculating physicochemical properties) for a single molecule. In order to choose only relevant and significant set of descriptors, data reduction was carried out towards in direction to eliminate the prevalence of coincidental correlations as well as data redundancy. Firstly, descriptors with "0" values for each compound were deleted. Though, the data was reduced on the basis of correlation matrix developed between two descriptors. These correlation values indicate the height of co-linearity. Values are somewhat closer to 1 indicates the extent of better fitting of the regression model, If the inter-correlation values of the two variables is to be 0.5 or higher, then those descriptors were retained, but if they have below 0.5 then, those were eliminated from the data sheet. Finally,

the five descriptors i.e., Inertia Moment 2 size (WM), 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 & test set¹⁰

Series of 41 analogs were segregated into training (29 compounds) and test (11 compounds) set on the basis of diversity in the structures and the biological activity of the compounds. The compounds in training and test set were employed to develop and validate the predictability of the concerned model. Some of the compounds may behave as outliers, thus they have to be discarded during the process of model development.

Model Development and its Validation

1. Linear-Regression analysis

Linear-Regression analysis helps to establish the correlation between the independent and dependent variables ($\log 1/IC_{50}$) of the series. Construction of regression model was done using training set compounds and the significance of every model was determined on the basis of statistical

values¹¹ 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) should be high and s-value (standard error of estimate) ought to be minimum. Description of various statistical parameters was shown in (Table 1) Compounds of test set are for the prognostic ability of the proposed model.¹² For a predictive model, the value of correlation coefficient (r^2) of the test set should be greater than 0.6 and less than r^2 of the training set.

Partial least square analysis used to crosscheck the heftiness of the model generated by multiple least square regression analysis. Intend to validate the results obtained from MLR technique, the same data set was subjected and analysed for PLS analysis. The correlation coefficient r^2 , r^2_{cv} value of training and test set was evaluated to ascertain the quality of the developed PLS models.¹³

2. Non-Linear regression analysis (NNA)¹⁴

The artificial neural networks mimic the functioning of simulating learning process by neural network. It employs

interconnections of artificial neurons with the help of computational studies and capable to deal with non-linearity and irregular data. Neural network analysis was carried out on the same descriptors of MLR analysis. NN analysis involved three layers; input, hidden and output layer. The input layer worked by receiving data whereas the output layer generated the dependent variable. The hidden layer interconnects the two abovementioned layers. The final structural design of the generated NN model was (5-2-1).

3. Validation of model

- Cross-validation (r^2_{cv}): Its validation technique employs to appraise the reliability of the developed statistical models. Number of compounds was shuffled into test to training set and vice versa for the generation of the precised 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 compound) were performed to get the final model. Cross-validation test (r^2_{cv}) value should be greater than 0.60.¹⁵

- Activity prediction of test set compounds: In general, r^2 of test set

greater than 0.6 represents good prognostic ability of the model.¹⁶

4. Outliers

Outliers are the data points that are fitted far apart from the linear model as well as 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 reclined away from the regression line were deleted as outliers and they adversely affect the robustness of the prospective model.

RESULT AND DISCUSSION

Data sets of various analogs were employed for the present studies. Their chemical structures and biological activity (Log/IC₅₀) is presented in (Table 2). In order to assess major molecular factors that greatly affect the Rho-kinase inhibition of derivatives belonging to anilines and benzylamines three chemo-metric tools (MLR, PLS, and NN) were used to construct classical descriptor based (QSAR) models. Whole compounds were divided into (training and test set) and training set comprised of 29 compounds (excluding 1 outlier) and test set of 11 compounds.

1. Statistical analysis

1.1. MLR and PLS (Linear regression method)

Initially, more than 200 descriptors were calculate for regression analysis, due to large and redundant data, it showed very low value for r^2_{cv} (0.312), and implying inadequate internal predictability. There was a strong necessitates to build a reliable and informative set of descriptors having good correlation with the biological activity with no inter-correlation. After the deletion of undesirable set of descriptors, correlation matrix technique was employed and eventually gets five distinct physicochemical descriptors: 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 non-linear) 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 fabricating of the purposed model, one compound was shown as outlier and not to fit to either the training set or test sets also their residual value was

more than two orders of magnitude. Owing to this outlier (5b) with different prediction was omitting 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 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 output of MLR and PLS models are summarized in (Table 4). The statistical worthiness of the developed model was evaluated in terms of square of correlation coefficient value, where r^2 (MLR=0.913 and PLS=0.912) values explain 91% variance in the activity, indicates 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 the high prognostic ability of the model. Similarly in the PLS analysis, since 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 MLR and PLS models ($r^2_{cvMLR}= 0.862$ and $r^2_{cvPLS}=0.869$) were evaluated and both the models

have comparable r^2_{cv} value. F-test indicates the significance level of the model. The F-value of the final MLR model that is 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 less for better QSAR model generation; this is an approximation of how precisely the model will predict unknown 'Y' values. The value of 's' for best MLR model is 0.373 and for PLS 0.307. It signifies that a regression with 's' value of 0.3 should predict Y values with a standard error of 0.3 units. The final model has 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 depicting the strength of the model. To further assurance of 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 the five descriptors confirmed the significance of an individual descriptor in determining the importance of structural design in exhibiting Rho kinase inhibitory activity by a molecule. The generated model was headed to construe the structural dependency of the biological activity

demonstrated by the certain set of Rho kinase inhibitors, and generated following equation:

Equation 1:

$$\text{Original Data: } Y = -0.0012895674 \cdot X_1 + 0.092464715 \cdot X_2 + 0.22125481 \cdot X_3 + 0.34796265 \cdot X_4 + 1.5525457 \cdot X_5 - 3.9303896$$

$$\text{Standardized Data: } Y = -0.2757968 \cdot S_1 + 0.45911208 \cdot S_2 + 0.35567212 \cdot S_3 + 0.4776836 \cdot S_4 - 0.64006561 \cdot S_5 - 1.344354$$

Where X_1 = Inertia moment 2 size (whole molecule), X_2 = VAMP polarization YY (whole molecule), X_3 = VAMP dipole Y component (whole molecule), X_4 = VAMP dipole Z component (whole molecule), X_5 = Kier ChiV6 (path) index (whole molecule) and Y represents the biological activity.

PLS method is the addition to MLR approach which evaluates the results to get the assurance of the consequences which they show minimal variation and having comparable results. Statistically, PLS equation evaluated the unassailability of the developed model on the basis of statistical parameters i.e. r^2 .

Regression equation obtained by PLS method generated:

Equation 2:

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

1.2. Prediction of test set compounds

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

1.3. ANN (artificial non-linear regression analysis method)

ANN analyses were carried out with the same descriptors that were used for linear regression analysis in an endeavour to improve results obtained for the Rho-kinase inhibition. In the present study, inputs for the neural-network are the descriptors, while the output was the \log_1/IC_{50} values. In artificial neural network analysis, TSAR software automatically computes number of hidden neurons, as well as patterns of training and test. 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. NN with three hidden neurons and 30% of compounds exclude as test set was most successfully trained NN model for in vitro Rho-kinase inhibition data as compared to the other NN models. Each analysis was repeated several times so that test RMS fit and best RMS fit are 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 descriptors with the output value (\log_1/IC_{50}) clearly demonstrates that three descriptors VAMP polarization YY, VAMP dipole Y component, VAMP dipole Z component are positively correlated to output value (\log_1/IC_{50} value), while two descriptors Inertia Moment 2 size and Kier chiV6 path index are negatively correlated to (\log_1/IC_{50}) which is in agreement with the results of MLR analysis.

1.4. Comparison between linear and non-linear methods

Different statistical approaches were employed to quantify the predictive ability of generated models that to in terms of statistical fit (r^2). Both MLR and PLS (two linear methods) were studied and have comparable result.

The r^2 values for training set for QSAR analysis is $r^2_{MLR} = 0.91$, $r^2_{PLS} = 0.91$, and $r^2_{ANN} = 0.93$ and for test set the values are $r^2_{MLR} = 0.75$, $r^2_{PLS} = 0.75$, and $r^2_{ANN} = 0.73$. On the basis of predictive power of the model it is very clear that conventional MLR and PLS as well as ANN analysis has generated a highly robust and predictive model. The actual and predicted activity for the compounds of training and test set were obtained after MLR, PLS and NN analysis are shown in (Table 6) and their corresponding graph are shown in (Fig. 1, 2, 3). A summary of the performance of the different prediction models is given in (Table 7).

2. Mechanistic interpretation of the descriptors

The results of MLR, PLS and NN reveals 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 has been observed between Rho-kinase inhibitory activity and the five descriptors (Table 8). The dependence plots of descriptors with the output value (\log_1/IC_{50}) are shown in (Fig. 4)

and the Structure activity relationship derived from QSAR analysis is shown in Fig.5.

2.1. Kier ChiV6 path index

Out of five parameters, Kier ChiV6 path index (WM) shown to be an essential descriptor in defining the biological activity of urea based derivatives as evident from 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 ring and the amount of branching ring.¹⁷ In our study it has been observed that Kierchiv6 (path index) descriptor is negatively correlated with permeability, according to study the sixth order valence connectivity index (Kierchiv6) encodes structural complexity, such as size and heteroatom content of the rings. This complexity observed in least and most

active compound. As the size of the compound decreases, biological activity increases.

2.2. VAMP polarization YY

VAMP Polarization YY is a spatial descriptor which calculates the electronic properties of a compound and projects polarization towards YY planes. There is a direct relation in between polarizability and the number of valence electrons on every atom¹⁸. A positive correlation of VAMP polarization YY with the biological activity reveals the direct link of chemical reactivity index with the biological activity. The compounds 12a, 12b, 12c, 12e and 14c having high value of VAMP polarization YY are most active whereas compound 8e with low VAMP polarization YY are least active.

2.3. 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.¹⁹ A positive correlation of VAMP dipole Y with the biological activity reveals the direct link of chemical reactivity index with the biological activity.

2.4. *Moment of inertia 2 (size)*

Inertia Moment 2 size (WM) with reverence 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 behaviour with respect to the size of whole molecule is utmost important in the binding interaction with the receptor site as well as in imparting greater permeability. In present study the biological activity increases with a decrease in Moment of inertia 2 (size) of whole molecule. This phenomenon can be explained by taking the example of most active compounds which has less value of Moment of inertia 2 (size) whole molecule in comparison to the least active compound which has high value of Moment of inertia 2 (size) of whole molecule. Hence it can be concluded that by decreasing the Moment of inertia 2 (size) Rho kinase inhibitory activity can be increased.

2.5. *VAMP Dipole X Component*

Vamp is a (semi-empirical molecular orbital package) used to determine electrostatic-properties and perform

optimizing of structure such us total-energy, electronic-energy, nuclear repulsion-energy, accessible surface-area, atomic-charge, mean polarizability, heat of formation, HOMO and LUMO eigen-values, ionization potential, total-dipole, polarizability, and dipole components. The positive coefficient of this expression in proposed model elucidates that higher the value, better is the activity, and indicating that the biological activity with respect to Rho-kinase inhibition is directly dependent upon the chemical permanence of the compounds in biochemical systems. Some exciting facts reveals during the analysis of derived descriptors and their correlation with the structural designing of the molecules. Comparison of least active molecule with the most active compound of the selected series, we found that in least active compound (8e) is substituted with dimethylamine altered the shape and volume of the molecule which eventually changes optimal binding affinity of the molecule. However, in the most active compounds (12a, 12b, 12c, 14c and 12e) dimethylamine is 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 Rho-kinase inhibitory activity. Interestingly, both alkyl and oxygen-containing alkyl chain have analogous molecular mass and, thus, no further bulk was laden to the molecule. Attribution of alkyl or oxygen-containing alkyl substitutions positively affects electrostatic nature of the substituent, thus, additional atoms contributing to the energy resulting in enhanced biological activity, and also gave clear vision to the fact that hydroxyl or oxygen substitutions are being highly electronegative, which augments the overall polarizability of the molecule. Although, negative correlation of Kier Chiv6 (path) index at R₁ clearly unveiling the facts that reduces the bulkiness and volume at certain positions in the whole molecule leads to increase biological profile. Additionally, compounds 12a, 12b, 12c, 14c and 12e has lower molecular mass (378.52, 350.46, 362.47, 336.43 and 366.46) than compound 8e (407.57) also confirms the authentication of the descriptor, which leads to compression in the shape of the molecule, that allows it to conveniently enter into the binding site and align in such a way that it fits snugly with the walls of the active site.

ADME studies

“Rule of five” given by Lipinski, known as therapeutic relevance or property of the drug-likeness. It’s an empirical approach traditionally utilized for calculating drug like properties in a molecule which clearly postulates that molecules ought to have molecular weight less than 500, $\log P < 5$, hydrogen bond donors are less than 5, and the hydrogen bond acceptors are less than 10 exhibit excellent pharmacokinetics profile in terms of absorption or permeation through biological membrane.²⁰ This rule explains absorption, distribution, metabolism and excretion (ADME) 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 profile.

Study Limitations

The findings of the present QSAR analysis will be advantageous only for the modeling of potent Rho-Kinase Inhibitors active neurological agents. For the future aspects will try to screen multi-targeted novel rho kinase inhibitors.

CONCLUSION

2D QSAR study is performed to establish a structural and physicochemical relationship required for the inhibition of molecular target against neurological disease i.e. Rho kinase. The statistically significant model highlighted the significance of electronic, topological and steric descriptors. The authenticity of the projected model is checked by validation and cross-validation (r^2_{cv}) based on Leave-one-out (LOO) methodology. Over fitting of the models is checked by considering the difference between r^2 and r^2_{cv} . In a planned study, attempt has been made to understand the dependence of the biological activity on the structural design accountable for their specific Rho-kinase inhibition, the main problem is to target the CNS related drug is its ability to cross Blood Brain Barrier (BBB), and for crossing the blood brain barrier, there should be optimal log P, volume, shape, molecular mass and polarizability. The proposed model overtly point out towards the introduction of optimal bulk or charge distribution along with the shape and size of the molecules lead to determines the binding efficacy of the molecule to the receptor domain, which eventually increases

inhibitory profile of the selected molecules. VAMP polarization YY component (WM), VAMP dipole Y component (WM), and VAMP dipole Z component (WM) descriptors are positively correlated with activity while rest two Kier ChiV6 path index (WM), and Moment of inertia 2 (size) (WM) are negatively correlated descriptor and projected molecular structure information with reverence to a specific rotation axis or the rotational analogue to mass, groups that decreases Kier ChiV6 and inertia of moment at substitutions will increase the predictability of model. The above final model reveals the significance of selected descriptors and their correlation with the biological activity and provided substantial insights to intend new chemical scaffolds with improved selectivity outline. Designing of Rho kinase inhibitors through incorporating the appropriate essential features or descriptors and light up the hope to get better new molecules with enhanced inhibitory profile.

ACKNOWLEDGMENT

Authors are also thankful to Vice-chancellor, Banasthali University for providing necessary research facilities.

Uncorrected proof

REFERENCES

1. Aelst VL, Schorey CD. Rho GTPases and signaling networks. *Genes Dev.* 1997;11:2295–2322.
2. Burridge K, Wennerberg K. Rho and Rac take center stage. *Cell.* 2004;116:167-79.
3. Bishop AL, Hall A. Rho GTPases and their effector proteins. *Biochem J.* 2000;348:241-55.
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;6:167-80.
5. Fujita Y, Yamashita T. Axon growth inhibition by RhoA/ROCK in the central nervous system. *Front Neurosci.* 2014;8:1-12.
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;71:91-9
7. Lum, C, Kahe J and Kessler L. 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.
8. Dwivedi N, Mishra BN, Katoch VM. 2D-QSAR model development and analysis on variant groups of anti-tuberculosis drugs. *Bioinformation.* 2011;7:82–90.
9. Abraham MH, Whiting GS. Hydrogen bonding: XXI. Solvation parameters for alkylaromatic hydrocarbons from gas-liquid chromatographic data. *J. Chrom.* 1992;594:229-241.
10. 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.
11. Rose K, Hall LH. E-State Modeling of Fish Toxicity Independent of 3D Structure Information. SAR QSAR Environ Sci. 2003;14:113–129.
12. Rose K, LH, Kier LB. Modeling Blood-Brain Barrier Partitioning Using the Electrotopological State. *J. Chem. Inf. Comput. Sci.* 2002;42: 651–666.

13. Voelkel A. Structural descriptors in organic chemistry - new topological parameter based on electrotopological state of graph vertices. *Computers & Chemistry*, 1994, 18, 1-4.

14. Bottcher CJF. *Theory of Electric Polarization* (2nded). Amsterdam: Elsevier Press, 1973.

15. Hopfinger AJ. *Conformational Properties of Macromolecules*. New York: Academic Press.1973.

16. TSAR reference guide.

17. 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.

18. Hall LH, Kier LB. The molecular connectivity chi indices and kappa shape indices in structure-property modeling. In *Reviews of Computational Chemistry*. (K.B. Lipkowitz, and D.B. Boyd, eds). New York: VCH publishers. 1991;2:367-412.

19. Vadlamudi SM and Kulkarni MV. 3D-QSAR of Protein Tyrosine Phosphatase 1B Inhibitors by Genetic Function Approximation. *Internet Electron J. Mol Des*.2004; 3: 586-609.

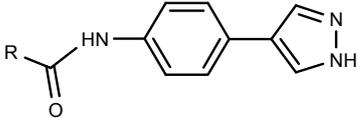
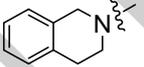
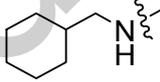
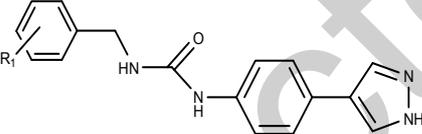
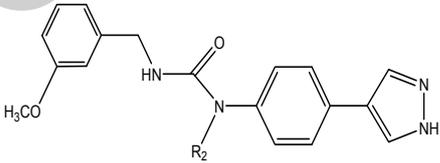
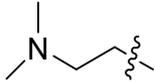
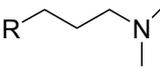
20. Lipinski CA, Lombardo F, Dominy DW, Feeney PJ. Experimental and Computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 1997; 23:3-25.

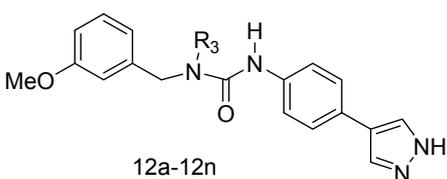
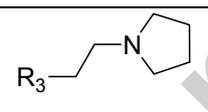
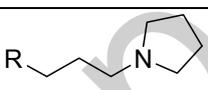
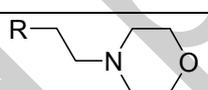
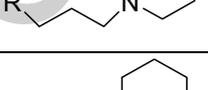
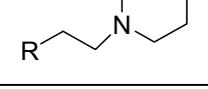
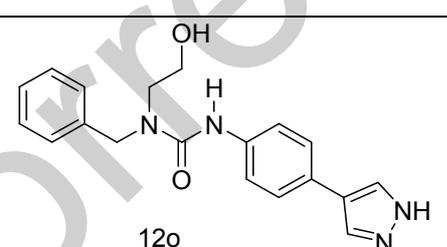
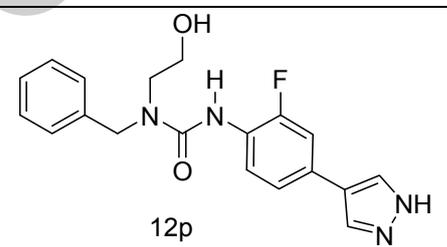
LIST OF TABLES

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.
SD (Standard Deviation)	Standard deviation (SD) 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 SD should be small but cannot have a value lower than the standard deviation of experimental data and the magnitude of SD 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 relatively more significant regression coefficients.

Table.2. Data set used for QSAR model development

Compd no.	Basic scaffold	R	ROCK-II IC ₅₀ (nM)
5a		PhNH	304
5b		PhCH ₂ NH	18
5c		PhCH ₂ CH ₂ NH	88
5d		PhCH ₂ CH ₂ CH ₂ NH	1017
5e		PhCH ₂ O	7
5f		PhCH ₂ CH ₂ O	55
5g		PhCH ₂ CH ₂	24
5h			280
5i			751
5j		3-OCH ₃	2
5k		2,3-di-OCH ₃	253
5l		2,4-di-OCH ₃	331
5m		2,5-di-OCH ₃	570
5n		2,6-di-OCH ₃	924
5o		3,4-di-OCH ₃	425
5p		3,5-di-OCH ₃	281
5q		3-F, 4-OCH ₃	357
8a			CH ₃
8b	CH ₂ CH ₂ OH		611
8c			2984
8e			3324
12a			CH ₃

12b	 <p>12a-12n</p>	CH ₂ CH ₃	1	
12c		Cyclopropyl	1	
12d		Isopropyl	3	
12e		CH ₂ CH ₂ OH	1	
12f		CH ₂ CH ₂ OCH ₃	17	
12h		CH ₂ CH ₂ N(CH ₃) ₂	1	
12i			3	
12j			5	
12k			13	
12l			4	
12m			5	
12n			4	
12o		 <p>12o</p>	CH ₂ CH ₂ OH	12
12p		 <p>12p</p>	CH ₂ CH ₂ OH	14

		R1	R2	
14a		OCH ₃	CH ₂ CH ₂ OH	2
14b		F	CH ₂ CH ₂ OH	2
14c		H	CH ₂ CH ₂ OH	2
14d		H	CH ₂ OH	1
14e		H	CH ₂ CH ₂ N(CH ₃) ₂	2

MLR Y = - 0.00128*X1 + 0.09246*X2 + 0.22125*X3 + 0.34796*X4 - 1.55254*X5 - 3.93038	0.95 5	0.91 3	0.86 2	0.38 7	48.4 1	Inertia Moment 2 size(w hole molec ule)	VAMP polariz ation YY (whol e molec ule)	VAMP dipole Y compo nent (whole molecu le)	VAMP dipole Z compo nent (whol e molec ule)	Kier ChiV6 path (whol e molec ule)
PLS Y = - 0.0013*X1 + 0.0914*X2 + 0.2306*X3 + 0.3482*X4 - 1.5262*X5 - 3.8318	0.94 3	0.91 2	0.86 9	0.30 7	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 , b_p are regression coefficients, b is the intercept, X_1 , X_2 ... 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 ... 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.

Table.5.The t-test values, Jackknife SE, and Covariance SE values of the descriptors used for regression analysis

Statistical parameter s Descriptors	Coefficient a	Jackknife SE b	Covariance SE c	t-value d	T – probability e
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

^aRepresents the regressions coefficient for each variable in the QSAR equations.

^b Represents the standard error estimation on each regression coefficient derived from a jack knife method on the final regression model.

^cRepresents an estimate of the standard error on each regression coefficient derived from covariance matrix.

^dMeasurement of the significance of each variable incorporated in the model.

^e Represents statistical significance for *t*-values.

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

Compound No.	Actual activity (Log1/IC ₅₀ values)	Predicted activity (Log1/IC ₅₀ values)		
		MLR	PLS	FFNN
Training set compounds				
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
Test set compounds				
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
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

Table.7. Correlation coefficients (r^2) between predicted and experimental values of the MLR, PLS, and NN models

r^2_{training}			r^2_{test}		
MLR	PLS	NN	MLR	PLS	NN
0.91	0.91	0.90	0.75	0.75	0.72

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

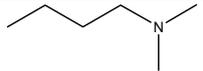
	Name of compound	Substitution at R ₃ (12a - 12h)	Biological activity LogIC ₅₀ (nM)	Inertia moment 2 size (WM)	VAMP Polarization YY (WM)	VAMP Dipole Y (WM)	VAMP Dipole Z (WM)	KierChiV6 (Path) (WM)	
Active Compounds	12a	CH ₃	1	1090.7	55.661	0.032	3.045	0	
	12b	CH ₂ CH ₃	1	1080.4	62.251	-3.086	-0.588	0	
	12c	Cyclopropyl	1	1099	63.302	-3.274	-0.314	0	
	12e	CH ₂ CH ₂ OH	1	1115.4	60.45	-2.487	0.492	0	
	12h	CH ₂ CH ₂ N(CH ₃) ₂	1	1216.1	54.60	-0.622	0.224	0	
	Substitution at R₁ R₄								
	14c	H	CH ₂ CH ₂ OH	1	1171.8	55.684	1.552	0.199	0
Inactive Compound	8e		3324	1537.5	43.504	-3.134	-4.05	0	

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

Name of the 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