

QSAR Model Development using MLR, PLS and NN Approach to Elucidate the Physicochemical Properties Responsible for Neurologically Important JNK3 Inhibitory Activity.

Ashima Nagpal*, Sarvesh Kumar Paliwal

Department of Pharmacy, Banasthali Vidyapith, Banasthali, Rajasthan, India.

Abstract

Aim: In order to decipher the information encoded by the molecular structure of the compounds, a classical physicochemical descriptors based QSAR study was performed on a data set of tri-substituted derivatives reported to be selective c-Jun-N-terminal kinase 3 (JNK3) inhibitor.

Materials and methods: A battery of statistical methods have been applied in the present study which include linear methods of analysis such as multiple linear regression (MLR), partial least square PLS and non-linear approach like artificial neural networks (ANN). The developed models were further subjected to validation using various statistical tools and methods which evidently confirmed their high predictability and precision.

Results: The predictive power and robustness of this model was further ascertained through certain statistical parameters and the model was found to be of excellent statistical relevance as depicted by the value of the standard statistical parameters such as s value: 0.22, F-value: 89.24, r: 0.97, r^2 : 0.94, r^2CV : 0.88. The generated model provided valuable insight to the relevance of four descriptors, Molecular surface area (whole molecule), verloop B2 (subs 2), Verloop B4 (subs 3) and KierChiv 4 (subs 3), and thus implied that certain changes in the substitution pattern can bring about dramatic increase in the JNK3 inhibitory activity.

Conclusion: The developed model did not only explain the dependence of bioactivity on the structures of the molecules but also suggested the changes that can be incorporated and applied to design novel molecules with enhanced inhibitory activity profile against JNK3 enzyme.

Key Words: JNK (c-Jun N-terminal Kinase), QSAR (Quantitative Structure Activity Relationship), MLR, PLS, NN (Neural Network).

INTRODUCTION

Mitogen-activated protein kinases (MAPKs) are conventionally known to modulate embryogenesis, differentiation, proliferation and apoptosis in mammalian cells. The mammalian cells are known to exhibit the presence of greater than a dozen MAPK genes. The genes known to comprise MAPK family are: 1) p38 code four isoforms p38R, p38 β , p38 γ , and p38 δ 2) the extracellular-signal regulated kinases code ERK1, ERK2, ERK3/4, ERK5, and ERK7/8 and 3) c-Jun N-terminal kinase code three genes JNK1, JNK2 and JNK3. All these sub-members of MAPK family exhibit distinct functions and involvement in diverse regulatory pathways. The families of these three proteins form the core of MAPK cascades. Despite having not so complex structure of this pathway, these enzymes have ability to respond to quite a large number of extracellular stimuli that gives intricately specific cellular outcomes. The strength of the responses is governed by the kinetics of their activation or inactivation, the availability of substrates, their localization to subcellular regions, and the presence of the complexes for enzymes to act. The cascade activity is modulated through the messages given, to the scaffolding accessory proteins, by the primary kinases.

JNK1/SAPKb, JNK2/SAPKa and JNK3/SAPKg are the three genes known to encode JNK family. These three genes manifest high extent of similarity, approximately 85%, in the amino acid sequence of the binding domain. Three genes, in total, make 10 spliced isoforms,

constituting JNK subfamily^[1,2], depending on the cleavage at different sites and thus weigh from 46 to 55 kDa^[3]. Out of these three genes, JNK1 and JNK2 are expressed in a ubiquitous manner whereas JNK3 is primarily expressed in the brain and at somewhat lower levels in the heart and testes^[2,5-6]. JNKs are known to be activated upon exposure to an extracellular stimuli such as stress caused by U.V irradiation, cytokines, and number of mitogens^[7-11]. The JNKs act by phosphorylating the trans-activation domain present at the amino terminal of transcription factor c-Jun on the specific sites which in turn capacitate c-Jun to trigger the transcription process of some specific kind of genes^[12-14]. In addition to c-Jun, certain other transcriptional factors that make contribution to Apolipoprotein-1 activity are also phosphorylated and stimulated by the JNKs. The intense curiosity in the area of neurodegeneration has led researchers to unfold many facts underlying the causes of multiple sclerosis, Alzheimer's disease, Parkinson's disease etc. The findings such as restricted concentration of JNK3 in brain has made it a putative target to unravel the factors that are responsible for the development and progression of various neurological disorders and thereby, has provided an opportunity to discover effective therapy to prevent or cure the aforementioned brain disorders. A plethora of evidence to support the notion of JNK3 been involved in neurological disorders is available in literature. For an instance, JNK3 gene knockout mice models exhibited a discernible resistance to seizures induced by Kainic acid clearly

indicating the potential involvement of JNK3 in modulation of seizures^[15]. Experiments done on JNK2, JNK3 and double knockout JNK2/JNK3 mice models showed resistance to neurodegeneration (caused by deficit in motor functions) induced by MPTP (1- methyl-4-phenyl-1,2,3,4-tetrahydropyridine) and exhibited profound improvement in motor activity when compared to MPTP lesioned mice (wild type). These studies further assured the crucial role of JNK3 in causing wide array of consequences that usually surface during Parkinson's disease, a leading neurological disorder^[16]. Additionally, sections of brain of patients with Alzheimer's disease, upon post-mortem, showed anomaly in distribution as well as activity of JNKs in diverse subcellular structures of specific regions targeted by Alzheimer's disease^[17, 18]. Also, JNK3 null mice showed profound decrease in cell death in cortical neurons, which is mediated by JNK3 via c-jun activation and increase in Fas ligand expression, when induced by β -amyloid^[19]. JNK3, via phosphorylation at Thr668^[20], directly regulate APP and is also found to phosphorylate Tau in-vitro^[21]. According to a recent study, JNK also facilitates TNF-R mediated regulation of γ -secretase^[22]. In addition to this, the neuropathological effect of Alzheimer's A β 42 mediated through Toll \rightarrow NF κ B pathway results via apoptosis modulated by JNK^[23]. All these findings clearly suggests the crucial role of JNK3 in the development as well as progression of Alzheimer's disease.

The major problem is the inefficiency of the JNK3 inhibitors, tested so far, in terms of selectivity. The inhibitor, if not selective, may pose serious toxicity and adverse effects complications, when used for treatment of chronic diseases. These undesirable outcomes can be accounted to the complete enzyme inhibition rather than its specific isoform. As different isoforms of JNK play diverse roles through modulation of a number of distinct pathways, this problem is very typical while selecting a JNK inhibitor for a certain disease. The high extent of similarity, commonly called as homology, in the binding pocket of the kinases makes the designing of selective and potent inhibitors a very difficult task^[1]. But this challenge and involvement of kinases in modulating diverse array of functions makes it an interesting target to explore. In view of strengths of in-silico techniques in drug design and role of specificity in determining the biological activity of JNKs, we have attempted to implement diverse chemometric techniques to develop models which could aid design of specific JNK3 inhibitors. The reported study has also helped to determine the possible molecular interactions which takes place when a specific inhibitor binds to the active site of JNK3 and the significant changes that can be made in substitution pattern on the molecule in order to optimize the reported inhibitors and to achieve selectivity in their inhibitory action towards JNK3 enzyme.

MATERIALS AND METHODS

sketching of preliminary chemical structures and data set preparation

The sketching of chemical structures of all the compounds, selected from the reported series of tri-substituted thiophenes^[24] (Table 4 and Table 5), as JNK3 inhibitors,

was performed using Chemdraw 4.0 followed by cleaning of their geometries using clean structure option of chemdraw and the structures of all the compounds were imported to the TSAR worksheet. TSAR, which is an integrated package, analyse the types of interactions of QSARs. IC₅₀ values of the compounds which were determined experimentally and reported in the series were used and before importing into TSAR work sheet, changed into negative logarithm. The negative log value of the experimentally determined IC₅₀ values of the compounds were used for QSAR model development. Tools for Structure Activity Relationship (TSAR) used in the present study is an integrated analysis package often used in pharmaceutical and toxicological research. TSAR generates a mathematical model through the structural features of a molecule^[25-28], and the derived output provide deeper insights into the relationship between the biological activity and chemical structure of the molecule.

defining substituents and three-dimensional optimized structure building.

The substitution pattern, which was found to exhibit least ambiguity and thus provided clear and explicit image for better understanding of the structural architecture of the molecule, was selected as a template. The three substituents, were defined around a common nucleus (R₁, R₂, and R₃) (Figure 1) through the application of "define substituents" option present in the TSAR worksheet's toolbar (version 3.3; Accelrys Inc., Oxford, England). Since some properties of the molecules are well represented through their 3D orientation, all the structures were converted 3D molecular structures of high quality was achieved using Corina^[28]. CORINA program automatically generate molecule's atomic co-ordinates in three dimensional form through information obtained by connection table or linear string^[29]. Energy of all the molecules were calculated using Cosmic force field. Cosmic option in TSAR assess the total energy of a molecule by adding up van der Waals, coulombic, bond length, bond angle and torsion angle terms for each suitable set of atoms. The valence electrons from the molecular atoms are included in these calculations. The calculations were terminated after the energy gradient became smaller than 1×10^{-5} and 1×10^{-10} kcal/mol, respectively^[30].

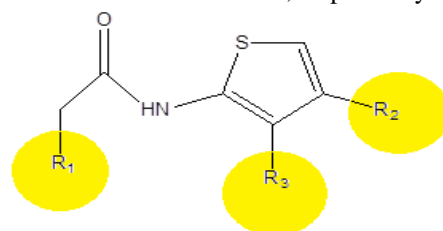


Figure 1: Depicting substitution pattern around thiophene nucleus.

Calculation of descriptors and data reduction

The main aim of descriptor calculation is to decode the information, about the physicochemical properties, encoded by the features or groups present in each molecular

structure, which are accountable for the specific biological activity of the molecule. TSAR typically has capacity to make calculations for up to 500 descriptors of different types, such as geometrical, electrostatic, constitutional and topological, obtained from the structure of the whole molecule as well as its defined substituents. Since the number of descriptors calculated are very high, it is mandatory to carry out data reduction in order to eliminate the incidence of chance correlations as well as data redundancy. The Correlation matrix was employed for data reduction as well as to identify the major descriptors, based on the criteria of minimum inter-correlation, to be included in final model. The degree or an extent of linear correlation between the two independent variables can be determined with the help of a Correlation coefficient^[31]. If the values of coefficient, describing the inter-correlation of the two variables, was detected to be 0.5 or higher, in that case, the descriptor possessing a greater correlation values with biological activity was retained whereas the others exhibiting the lower values were removed from the data sheet. This step was carried out because when the values of correlation coefficient, which are basically the measure of the extent of fitting of the regression model, are somewhat closer to 1 indicates the model to be of a better fit. The descriptor remaining after pairwise correlation analysis were further sorted by employing backward elimination method. The descriptors with smaller t-value were removed from the data set^[32].

Data set preparation and statistical analysis

Finally, a set of four independent descriptors, manifesting good correlation with biological activity and no correlation with each other, were left. The selected descriptors including, Molecular surface area (whole molecule), Verloop B2 (Subst. 2), Verloop B4 (Subst. 3) and Kier Chiv4 (path/cluster) index (Subst. 1) were used for model development. The compounds of the series were segregated into the training set and the test set. Training set consisted of two-third of the molecules of the series and rest were included in the test set. The training set containing 25 compounds (**Table 4**) was employed for building the QSAR models in order to obtain accurate relationship between the structural features and the biological activity. Whereas, the test set containing 7 compounds (Table 5) was used as a tool for validating the predictability of the developed model. MLR, PLS and NN were used as statistical tools to develop quantitative models with an aim to elucidate important structural features of the compounds and their role in determining the biological activity^[33].

The variables with least inter-correlation were used in the MLR technique^[34]. The cross-validation step was performed by employing leave-one-out (LOO) method in which removal of one compound at a time was done and the model obtained from the rest of the data set was used to make the calculations. The evaluation of statistical significance of the regression equations was performed on the basis of the values of the standard error of estimate (s)^[35], Fischer's ratio (F)^[36] and conventional regression coefficient (r^2)^[37]. With an aim to validate the results obtained from MLR technique, the same data set was subjected to PLS analysis and the results were analysed.

ANN was used to estimate the conformity of the obtained descriptors with the biological activity and hence, the usefulness of the developed model. ANN in the present study is comprised of two hidden neurones, and an output layer represented by the experimentally determined logIC50 values of biological activity. The final architecture of the generated ANN model was (4-2-1).

Cross-validation techniques

In order to determine the precision with which a model has been developed, it must undergo a number of validation steps before it can be declared to be a reliable one^[38]. Cross-validation is a conventional yet popular technique utilized to evaluate the reliability of the developed statistical models. According to this technique, a diverse number of data sets, by shuffling the number of compounds in the training set and the test set, are created. This method is typically known as "leave-one-out" (when a single molecule is removed from the training set and included in the test set and vice versa) and "leave-some-out" (when a number of compounds are removed from the training set and included in the test set and vice versa). For each data set, an input-output model was generated and its predictability was evaluated using the test set compounds.

RESULTS

In order to develop the 2D QSAR model, nearly 200 classical descriptors, including geometric, electronic, topological, constitutive, etc., were generated for the selected series of compounds, listed in **Table 4** and Table 5, using TSAR software version 3.3. The initial model, including all the 200 descriptors upon regression analysis, showed very low value for r^2_{cv} (0.52), due to large and redundant data, implying inadequate internal predictability. Therefore, to build a reliable and informative set of descriptors, with no inter correlation but having good correlation with the biological activity, there was a strong need to abridge the data to get better clarity about the impact of physical properties, due to distinct substitutions around a common moiety, on the JNK3 inhibitory activity of the molecules. The correlation matrix technique was employed to delete the undesirable descriptors. Eventually four physicochemical parameters or descriptors were obtained: Molecular surface area (whole molecule), verloop B2 (subs 2), Verloop B4 (subs 3) and KierChiv 4 (subs 3) which were further evaluated to study the impact of diverse substituents on JNK3 inhibitory activity. Out of these four descriptors Verloop B2 (subs 2), Verloop B4 (subs 3) and KierChiv 4 (subs 3) manifested very high correlation values of 0.79 and 0.73 and 0.76 respectively with the biological activity. The Molecular surface area (whole molecule) showed relatively lower, correlation values than the other three descriptors as depicted in the correlation matrix in Table 1. Additionally, no inter-correlation was found among these descriptors, further supporting the fact that a good fit can be obtained when these four descriptors will be used in unification. To further ascertain the significance of the descriptors used to build the model, their t-test values, coefficient values, jackknife standard error (SE), and covariance SE values (Table 2) were evaluated. The adequate values of all these parameters for all the four

descriptors confirmed the importance of an individual descriptor in determining the importance of structural architecture in exhibiting JNK3 inhibitory activity by a molecule. After gaining insight into the structural pattern of the molecules, exhibiting JNK3 inhibitory activity, and confirming the statistical soundness as well as relevance of the developed model, certain statistical tools were employed to further analyse the data to ascertain its validity as well as its reliability.

After meticulous reduction of data, the compounds were divided into the training set and the test set to evaluate the predictive power of the 2D QSAR models generated using MLR, PLS and ANN. The activity of the training set compounds, used to build the model, was validated using the test set compounds that were not included to build the model. Similar methods of evaluation, as that used for training set compounds, were used for the test set compounds. For a QSAR model, to be in concordance with the standards of the statistical parameters, must possess a minimum of 0.80 value for r^2 [30]. The high r^2 value for Equation 1 is 0.94 which is excellent statistically and indicates the soundness of the generated model. The model was further analysed for the determination of the outliers. But this model did not show the inclusion of any outlier. The actual and predicted values of the descriptors were close enough, which simply indicates that the extent of prediction using the developed QSAR equation was very close to the observed values (**Table 4**). The quality of the best 2D QSAR model was described by the value of certain statistical parameters (**Table 3**) and includes: $r = 0.97$, $r^2 = 0.94$, $r^2CV = 0.88$, $s = 0.22$, $f = 89.29$. Where, r measures the extent of fitting of the developed model. Nearer is its

value to 1, the better is the quality of fit of the generated model. The statistical parameter r^2 describes the percent data present in a specific equation. The value of r^2 for this model was found out to be 0.94 which represents 94% variance in the biological activity of the used data. The predictability of the developed model was further assessed by its r^2CV (cross validated r^2). Its value for the model 0.88 clearly indicates the excellent predictability of the generated regression equation. The value obtained for the standard deviation parameter 's' was 0.22 and its low value affirms that the error in the regression is minimal for this developed model. The F-test is basically defined as the ratio of variance manifested by the developed model and variance that arise due to the regression error. The higher is its value, more statistically sound a model is. Here, for the generated model, its value 89.24 further supports the notion of the statistical significance of the model. The developed model was further used to interpret the structural dependency of the biological activity exhibited by the selected set of JNK3 inhibitors. The model consisted of 25 compounds in the training set and 7 compounds in the test set and generated the following equation:

Equation 1

Original Data: $Y = -0.0031632297 * X1 - 1.5969231 * X2 - 1.5210433 * X3 + 2.240274 * X4 + 3.7058432$

Standardized Data: $Y = -0.12631048 * S1 - 0.411349 * S2 - 0.36352766 * S3 + 0.30546683 * S4 - 2.0672114$

Where $X1$ = Molecular surface area (whole molecule), $X2$ = Verloop B2 (subs 2), $X3$ = Verloop B4 (subs 3), $X4$ = KierChiv 4 (subs 3) and Y represents the biological activity.

Table 1: Representing the correlation matrix depicting the relationship of the derived descriptors with the biological activity.

	$-\log IC_{50}$	Molecular surface area (Whole molecule)	Verloop B2 (R2)	Verloop B4 (R3)	Kier Chiv4(path/cluster) index (R1)
$-\log IC_{50}$	1	0.18633	-0.79616	-0.73012	0.76163
Molecular surface area (Whole molecule)	0.18633	1	-0.22954	-0.43716	0.31488
Verloop B2 (R2)	-0.79616	-0.22954	1	0.46612	-0.62156
Verloop B4 (R3)	-0.73012	-0.43716	0.46612	1	0.43442
Kier Chiv4(path/cluster) index (R1)	0.76163	0.31488	-0.62156	0.43442	1

Table 2: Representing the relevance of descriptors entered in the selected model.

Descriptors	Coefficient ^a	Jackknife ^b	Covariance SE ^c	t-value ^d	t-probability ^e
Molecular surface area (Whole molecule)	-0.0031632	0.0010174	0.0013164	-2.4029	0.026088
Verloop B2 (R2)	-1.5969	0.47948	0.27511	-5.806	1.1148e-005
Verloop B4 (R3)	-1.521	0.39171	0.21269	-7.1514	6.2991e-007
Kier Chiv4(path/cluster) index (R1)	2.2403	0.46315	0.49884	4.491	0.00022343

^a Represents the regressions coefficient for each variable in the QSAR equations.

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

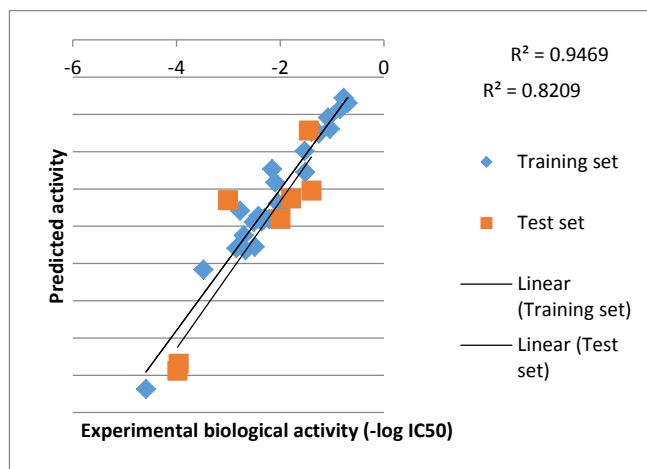
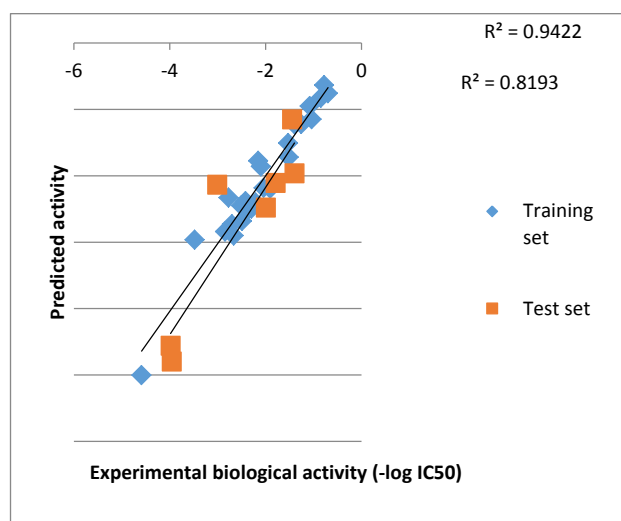
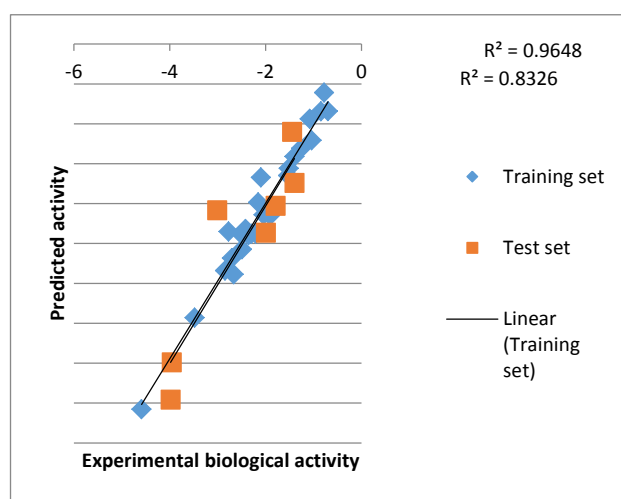
^c Represents an estimate of the standard error on each regression coefficient derived from covariance matrix.

^d Is the Measure of the significance of each variable included in the final model.

^e Represents statistical significance for t values.

Table 3: statistical parameters of the models generated.

	Test set compounds	r	r ²	r ² CV	s value	f value
Final Model	5,6,17,19,22,33,43	0.97	0.94	0.88	0.22	89.24

**Figure 2: MLR graph showing plot between the observed vs predicted activity.****Figure 3: PLS graph showing plot between the observed vs predicted activity.****Figure 4: ANN graph showing plot between the observed vs predicted activity.**

In addition to MLR approach, PLS method was used to obtain the regression equation. This step was performed to further validate the results derived from MLR and to ascertain that the results obtained from both the methods are comparable i.e. they show minimal variation.

The output that equation from PLS gave, was also evaluated on the basis of statistical parameters such as r² to assure the soundness of the generated model in terms of statistics.

Regression equation obtained by PLS method generated:

Equation 2

$$Y = -0.0040894211 * X_1 - 1.4459748 * X_2 - 1.8684793 * X_3 + 2.5747774 * X_4 + 4.1907516$$

The closeness in the values of the observed and predicted activity of the training and the test set of the molecules (Table 4 and Table 5) clearly indicates the relevance of the equation obtained through MLR and PLS methods for statistical evaluation. Figure 2 and Figure 3 depicting the graphs between the observed activity versus predicted activity for the training and test set compounds obtained from MLR and PLS method respectively.

To achieve the goal of this study, that is, to design a validated 2D QSAR model manifesting reliability and robustness, Artificial neural network (ANN) method was employed to determine the stability of the 2D QSAR model generated using MLR and PLS approach. The observed activity versus predicted activity graph of ANN as depicted in Figure 4 further corroborates the developed model to be of excellent statistical quality. The dependency plots (Figure 5, Figure 6, Figure 7 and Figure 8) obtained from NN (Neural Network) eventually made the prediction of the dependence of biological activity on the structural architecture even more reliable.

Equation 1 and Equation 2 depicts that Molecular Surface Area (Whole Molecule), Verloop B2 (subs 2), Verloop B4 (subs 3) and Kier Chiv 4 (Subst. 1) are the major independent physicochemical parameters comprising the model and manifesting excellent correlation with the biological activity. The negative coefficients of the Molecular surface area (whole molecule), Verloop B2 (Subst. 2), Verloop B4 (Subst. 3) indicates that decreasing the values of these descriptors may lead to augmented biological activity, whereas the positive coefficient of the KierChiv 4 (Subst. 3) depicts that increasing its value can improve the JNK3 inhibitory activity Figure 9.

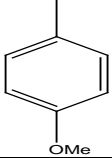
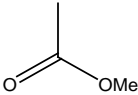
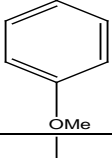
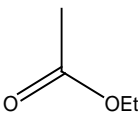
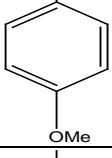
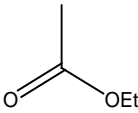
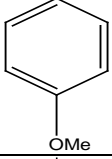
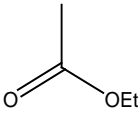
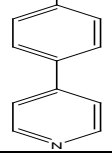
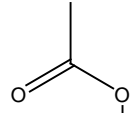
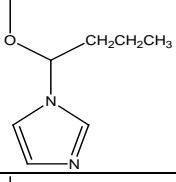
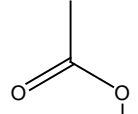
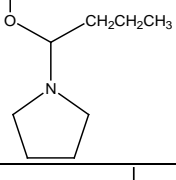
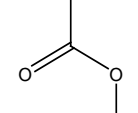
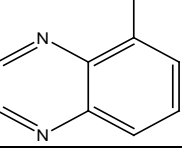
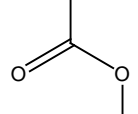
Lipinski's rule of five

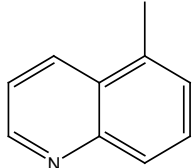
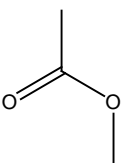
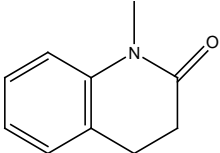
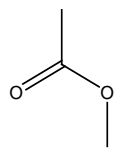
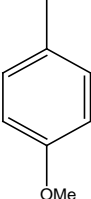
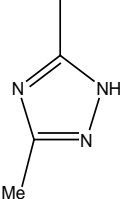
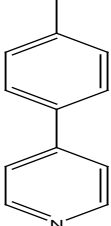
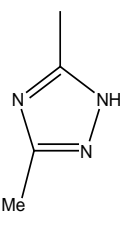
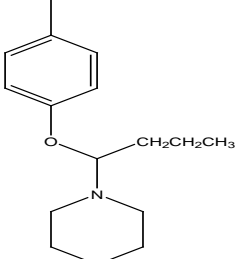
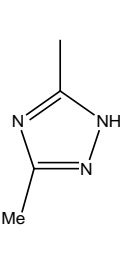
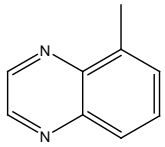
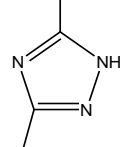
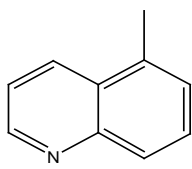
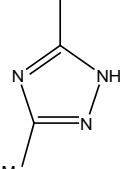
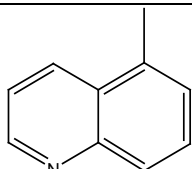
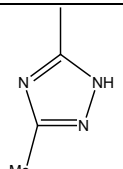
For any molecule to possess therapeutic relevance or property of the drug-likeness, it must obey "rule of five" given by Lipinski. It is an empirical approach conventionally employed for predicting drug like properties in a molecule and clearly states that molecules with a molecular weight greater than 500, log P greater than 5, more than 5 hydrogen bond donors, and more than 10 hydrogen bond acceptors exhibit poor pharmacokinetics properties in terms of absorption or permeation through biological membrane^[39]. This rule explains only the

molecular properties associated to the pharmacokinetics of molecules, i.e., the absorption, distribution, metabolism, and excretion (ADME) of bioactive compounds in a higher organism. It does not deal with any of the pharmacodynamics aspect of the molecules, which involves action of the drug on the cells or on microorganisms and other parasites within or on the body. The parameters

included in Lipinski's rule of five were calculated for the selected series of molecules and no compound was found to have violated the above mentioned set of rules given by Lipinski (Table 7). This explicitly indicates that all molecules exhibit adequate pharmacokinetic profile.

Table 4: structures of the training set compounds, used to build the model, along with their observed and predicted values.

Name of the compound	R ₁	R ₂	R ₃	-logIC ₅₀ (nm)	PREDICTED VALUES		
					MLR	PLS	ANN
4		Me		-2.8451	-2.79415	-2.84003	-2.83904
7		CF ₃		-4.58883	-4.68435	-5.00203	-4.57564
8		CN		-2.49136	-2.77477	-2.68022	-2.57305
9		CCH		-3.48001	-3.08216	-2.96259	-3.42912
15		Me		-2.41996	-2.36231	-2.38199	-2.31726
16		Me		-2.70243	-2.62228	-2.72982	-2.68266
18		Me		-2.35411	-2.43635	-2.50476	-2.39388
20		Me		-2.20952	-2.40596	-2.40306	-2.37617

Name of the compound	R ₁	R ₂	R ₃	-logIC ₅₀ (nm)	PREDICTED VALUES		
					MLR	PLS	ANN
21		Me		-2.04139	-2.19049	-2.18177	-2.13982
23		Me		-2.09691	-1.91241	-1.85852	-1.67053
27		Me		-2.66745	-2.82221	-2.90116	-2.88521
28		Me		-2.51055	-2.44376	-2.45017	-2.3823
29		Me		-2.29447	-2.38729	-2.49537	-2.37786
30		Me		-2.77379	-2.292	-2.32821	-2.34988
31		Me		-1.90309	-2.19487	-2.19489	-2.14963
32		Me		-2.15534	-1.73336	-1.7743	-1.98488

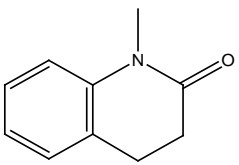
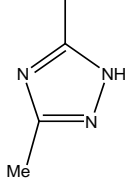
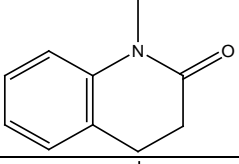
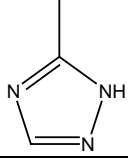
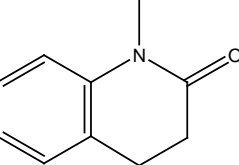
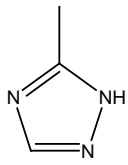
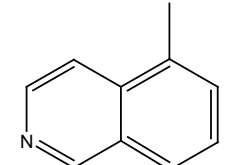
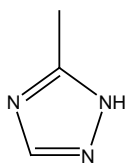
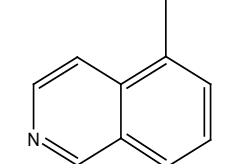
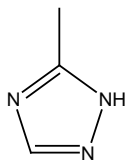
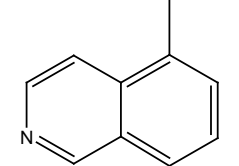
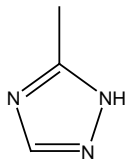
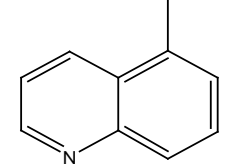
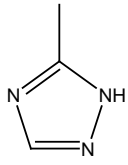
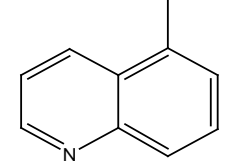
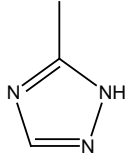
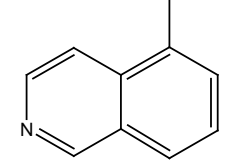
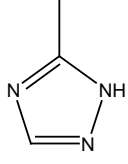
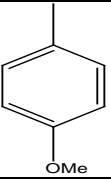
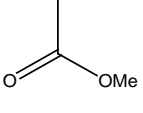
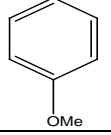
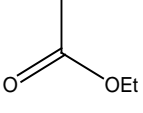
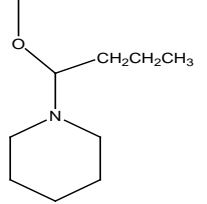
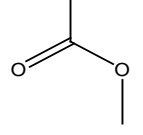
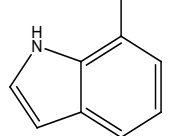
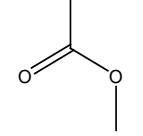
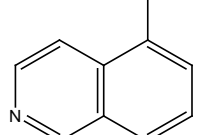
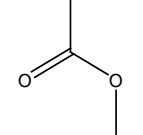
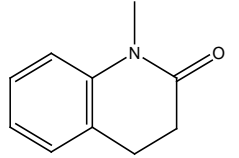
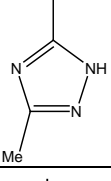
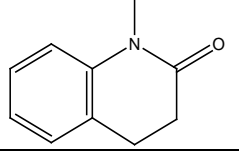
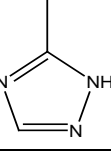
Name of the compound	R ₁	R ₂	R ₃	-logIC ₅₀ (nm)	PREDICTED VALUES		
					MLR	PLS	ANN
34		Me		-1.39794	-1.2301	-1.22606	-1.40955
44		Cl		-0.69897	-0.84762	-0.75832	-0.84101
45		CN		-0.77815	-0.77887	-0.63378	-0.60757
46		Br		-1.51851	-1.77444	-1.7184	-1.55662
47		Cl		-1.04139	-1.19444	-1.14901	-1.2058
48		CN		-0.8451	-0.92824	-0.82612	-0.84391
49		Br		-1.53148	-1.49011	-1.50711	-1.64672
50		Cl		-1.25527	-1.25601	-1.22153	-1.30553
51		CN		-1.07918	-1.04175	-0.95106	-0.93661

Table 5: Representing test set compounds, used to validate the model, along with their observed and predicted activities.

Name of the compound	R ₁	R ₂	R ₃	-Log IC ₅₀ (nm)	Predicted values		
					MLR	PLS	ANN
5		Et		-3.95952	-4.3446	-4.79824	-3.98806
6		cPr		-3.98318	-4.43786	-4.56025	-4.45665
17		Me		-1.99564	-2.40075	-2.47921	-2.36455
19		Me		-3.0086	-2.15079	-2.13414	-2.08324
22		Me		-1.79239	-2.1256	-2.10688	-2.0274
33		Me		-1.39794	-2.02173	-1.96388	-1.73991
43		Br		-1.44716	-1.21543	-1.14936	-1.1001

DISCUSSION

Verloop descriptors define the optimal volume and shape desirable to have a molecule aligned with the active site of the receptor and show binding affinity towards it. Verloop B2 (at R₂) and Verloop B4 (at R₃) are negatively contributing towards the model suggesting the increase in the width, hence, shape or volume of the molecule will lead to reduced biological activity of the compounds. Based on such information we can infer that those groups, contributing towards the substitution pattern of the molecule, which can increase the shape as well as volume distribution at the R₂ and R₃ positions is expected to make a significant contribution towards drop in activity profile of a molecule.

Molecular surface area defines the sum total of not only the total area that a molecule occupies but also the polar area of that molecule. And for a molecule to cross blood brain barrier (BBB), which is commonly through passive transport, apart from having an optimal logP (partition coefficient) value, it must have less molecular weight and also must be less polar. Also, in the present model, through the negative correlation of Molecular Surface area descriptor, this has been indicated that by decreasing the total surface area including the polar surface area, which is majorly occupied by nitrogen and oxygen atoms, the activity profile can be improved.

Furthermore, the shape of the molecule is responsible for its orientation and hence its interaction with the binding domain of the receptor whereas, the mass plays a crucial

role in ADME (absorption, distribution, metabolism, excretion) properties and in some cases, is also responsible for toxicity profile of the molecule. Therefore, a drug to possess both safety and efficacy, must have an optimal size and shape.

Kier Chiv4 (path/cluster) index (whole molecule) belongs to the molecular connectivity indices class of descriptors. Usually, steric hindrance caused by the bulky groups in the bonding process of a molecule with a receptor is undesirable, but in certain cases an optimal level of bulk and branching aids in better orientation of the compound thus enhancing the bonding interactions of a molecule with the receptor. A molecule is expected to exhibit strong interactions, with the binding domain of the receptor, only if a favourable substitution pattern along with optimal branching at proper position is present. As Kier Chiv4 (path/cluster) index is having the positive correlation with the bioactivity, increasing the bulk through replacements with bulky groups or through branching at this particular position will have positive impact on the biological activity profile of the compounds.

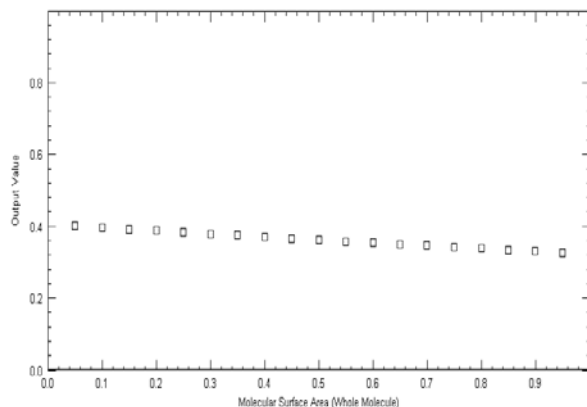


Figure 5: Dependency plot of Molecular Surface Area (Whole molecule) descriptor with Experimental activity.

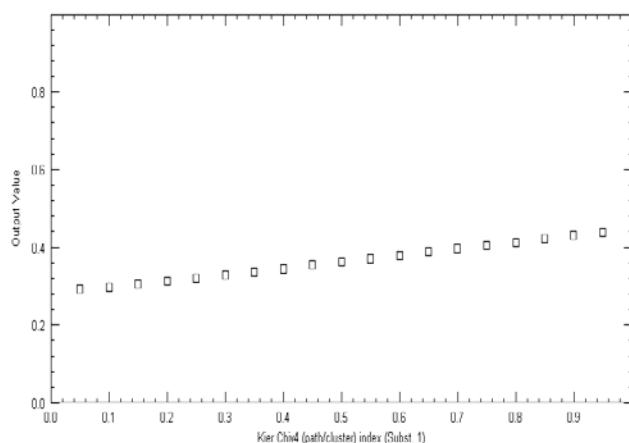


Figure 6: Dependency plot of Kier Chiv4 (path/cluster) index descriptor (Subst. 1) descriptor with Experimental activity.

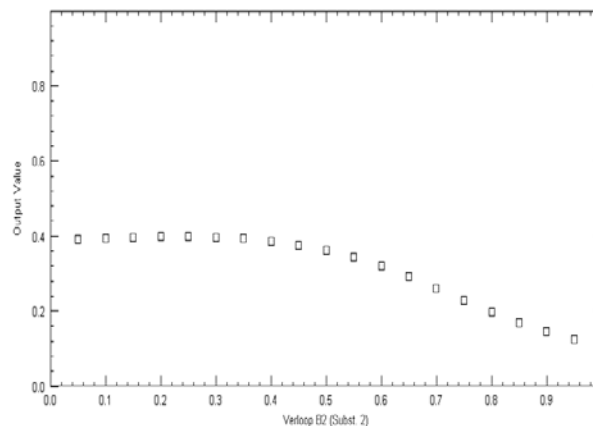


Figure 7: Dependency plot of Verloop B2 (Subst. 2) descriptor with Experimental activity.

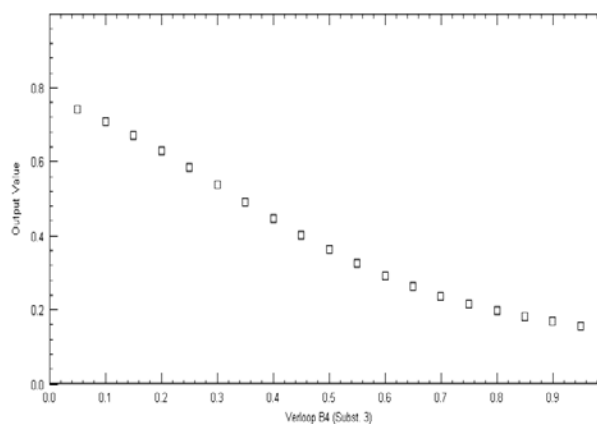


Figure 8: Dependency plot of Verloop B4 (Subst. 3) descriptor with Experimental activity.

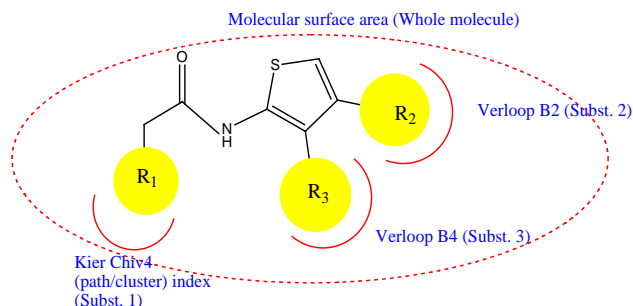


Figure 9: Relevance of descriptors on substitution pattern around a common nucleus.

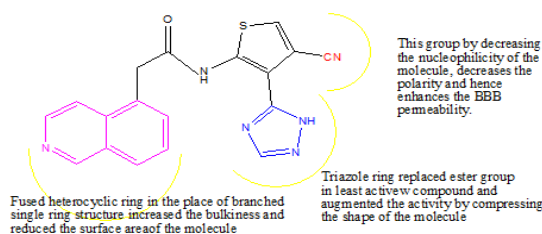


Figure 10: significance of the descriptors evaluated through comparison between the substitution around the most active compound (IC_{50} 7nm) with the least active compound (IC_{50} 38800nm).

Table 6: values of the descriptors for individual compounds.

Name of the compound	Molecular surface area (whole molecule)	Verloop B2 (subs 2)	Verloop B4 (subs 3)	KierChiv 4 (subs 3)
4	294.682	2.09779	2.02262	0.383284
5	320.969	2.00424	3.08551	0.383284
6	0	2.5611	3.22968	0.383284
7	339.552	2.42569	2.82775	0.383284
8	327.953	2.37467	1.65	0.383284
9	329.275	2.55983	1.65494	0.383284
15	344.27	2.0588	2.02749	0.621575
16	401.41	2.00014	2.02969	0.545889
17	430.474	2.02243	2.02537	0.698765
18	424.358	2.05526	2.02702	0.698765
19	300.876	2.0185	2.02128	0.621775
20	297.775	2.07737	2.02109	0.545329
21	304.974	2.01965	2.02615	0.613964
22	311.839	2.02691	2.0212	0.654437
23	327.768	2.05954	2.03903	0.807468
27	320.734	2.0564	2.03035	0.383284
28	341.324	2.11945	2.02349	0.621575
29	455.694	1.96899	2.02017	0.698765
30	317.397	1.95878	2.02986	0.545329
31	314.376	2.00834	2.02134	0.613964
32	311.326	1.72337	2.02345	0.613964
33	339.85	2.12043	2.02186	0.807468
34	322.256	1.65852	2.02295	0.807468
43	319.796	1.77131	1.9	0.807468
44	312.548	1.65059	1.8	0.807468
45	321.841	1.73201	1.65	0.807468
46	302.876	1.9402	1.9	0.654437
47	306.878	1.66433	1.8	0.654437
48	300.212	1.65371	1.65	0.654437
49	322.509	1.66649	1.9	0.613964
50	302.523	1.65473	1.8	0.613964
51	300.259	1.66791	1.65	0.613964

Table 7: Depicting values of various parameters constituting Lipinski's rule of five.

Name of the Compound	ADME weight(Whole molecule)	ADME H-bond Acceptors	ADME H-bond donors	ADME log P	ADME violations
4	319.4	4	1	2.1557	0
5	333.43	4	1	2.552	0
6	359.47	4	1	2.721	0
7	387.4	4	1	2.9138	0
8	344.41	5	1	1.896	0
9	343.42	4	1	2.2126	0
15	366.46	4	1	3.2458	0
16	413.53	5	1	2.4939	0
17	430.61	5	1	2.9088	0
18	416.58	5	1	2.5125	0
19	328.41	3	2	2.1816	0
20	341.41	5	1	1.585	0
21	340.42	4	1	2.4978	0
22	340.42	4	1	2.5636	0
23	358.44	4	1	1.6098	0
27	342.45	4	2	2.4641	0
28	389.51	4	2	3.5542	0
29	453.66	5	2	3.2172	0
30	364.46	5	2	1.8934	0
31	363.47	4	2	2.8062	0
32	349.44	4	2	2.6577	0
33	381.49	4	2	1.9182	0
34	367.46	4	2	1.7697	0
43	432.32	4	2	2.0943	0
44	387.87	4	2	1.8205	0
45	378.44	5	2	1.1675	0
46	414.3	4	2	3.0481	0
47	369.85	4	2	2.7743	0
48	360.42	5	2	2.1213	0
49	414.3	4	2	2.9823	0
50	369.85	4	2	2.7085	0
51	360.42	5	2	2.0555	0

In depth analysis of the derived descriptors and their correlation with the structural architecture of the molecules helped in surfacing of the interesting facts. Upon comparison of the least active molecule with that of the most active compound of the selected series we observed that in compound 7 (least active), R₁ is substituted with the methoxy benzene, R₂ with methyl group and R₃ with an ester group all of which increased the shape and volume of the molecule and eventually resulted in enhanced surface area. Whereas in the most active compound, compound 48, methoxy benzene is replaced with a fused heterocyclic substituent and ester group with a triazole ring both of which reduced the width and volume of the molecule, thus, resulting in profound increase in the JNK3 inhibitory activity. Interestingly, both ester and triazole ring have comparable molecular mass and, thus, no additional bulk was introduced to the molecule. Therefore, a simple augmentation in bulk or mass cannot be accounted for an improved activity profile. The possible reason for improved activity, perhaps, could be the flat structure of the triazole ring, 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. The shape of the triazole ring, hence, can be attributed for the better orientation of the molecule that capacitate it to show better interactions with the binding domain. In addition to this, compound 48 has lower molecular mass (360.42) than compound 7 (387.4) that may also be accounted for the discernible augmentation in the activity. Also, the compounds bearing chloro and cyano groups were found to be the most potent and those bearing bromo group exhibited comparatively less potency. The dramatic increase in the biological activity can clearly be attributed to the fact that chloro and cyano groups being highly electronegative, decreased the overall nucleophilicity of the molecule to a greater extent than the bromo group, which is comparatively less electronegative. The decreased nucleophilic character of the molecule eventually capacitated the molecule to accept the electron pairs and make covalent bonds with the receptor binding domain. A point must be made here, that positive correlation of Kier Chiv4 (path/cluster) index at R₁ indicates an increase in bulk at this position would lead to increase in activity but Molecular surface area, which is in negative correlation with the whole molecule predicts the decrease in activity with overall increase in volume or size of the molecule (Figure 10). Therefore, an optimal increase in the bulk or branching only at certain positions will bring about an increase in bioactivity of the molecules. The above results explicitly indicates that all the descriptors that entered the final 2D QSAR model were significant and their correlation with the biological activity concurred well to the substitution present in the structures of the selected series of the JNK3 inhibitors. The in-depth study of these physicochemical parameters has provided substantial insights to design better chemical scaffolds in term of selectivity and efficacy. This model has provided sufficient information to design new molecules and have ignited a hope that through incorporating the appropriate features, deduced to be important through these descriptors,

designing of JNK3 inhibitors with a better selectivity profile can be easily achieved.

CONCLUSION

Through the proposed study, an effort has been made to analyse the reported JNK3 inhibitors and to get insight into their structural architecture responsible for their specific JNK3 inhibitory activity, and thereby to suggest the beneficial or detrimental impact of the substitution pattern on the biological activity. The scrupulous evaluation of the chemical structures and the physicochemical descriptors derived from them capacitated us to understand the dependence of the biological activity on the structural architecture of the molecules. The major factor that govern the efficacy of a CNS targeting drug is its ability to cross Blood Brain Barrier (BBB) which in turn depends on optimal logP (partition coefficient), Molecular mass and polar surface area. The developed model explicitly indicated that an introduction of optimal bulk or mass distribution at certain positions, and not at any position, will lead to an increased activity profile of the selected set of compounds. Additionally, the shape of the molecules, that determines the efficacy with which a molecule aligns itself with the binding domain of the active site, was found to be a dominating factor in determining the potency of these JNK3 inhibitors. Also, decreasing the overall nucleophilic character, through replacement of electron donor groups with electronegative groups resulted in overall increase in the activity. Therefore, through this work, we concluded that when molecules are substituted with those groups, at certain positions, that increase the mass of the molecule but compresses its shape, hence, its surface area and those that decreases the nucleophilic character of these molecules can, possibly, contribute positively towards increasing the bioactivity and selectivity and, hence, exhibiting better JNK3 inhibitory activity. The proposed study, capacitated us to gain insights into the structural dependency of the biological activity and thereby to suggest possible replacements that, perhaps, would eventually lead to optimized JNK3 inhibitors with better activity as well as selectivity profile.

ACKNOWLEDGEMENT

My sincere thanks to Vice chancellor Banasthali University for providing all the required softwares and my lab mates for their cooperation.

REFERENCES

1. Kallunki, T., Su, B., Tsigelny, I., Sluss, H. K., Derijard, B., Moore, G., Davis, R., Karin, M., *Genes Dev.* 1994, 8, 2996 – 3007.
2. Gupta, S., Barrett, T., Whitmarsh, A. J., Cavanagh, J., Sluss, H. K., Derijard, B., Davis, R. J., *EMBO J.* 1996, 15, 2760 – 2770.
3. Pulverer, B. J., Kyriakis, J. M., Avruch, J., Nikolakaki, E., Woodgett, J. R., *Nature.* 1991, 353, 670 – 674.
4. Mohit, A. A., Martin, J. H., Miller, C. A., *Neuron.* 1995, 14, 67 – 78.
5. Martin, J. H., Mohit, A. A., Miller, C. A., *Brain Res Mol Brain Res.* 1996, 35, 47 – 57.
6. Derijard, B., Hibi, M., Wu, I. H., Barrett, T., Su, B., Deng, T., Karin, M., Davis, R. J., *Cell.* 1994, 76, 1025 – 1037.
7. Smeal, T., Binetruy, B., Mercola, D., Grover-Bardwick, A., Heidecker, G., Rapp, U. R., Karin, M., *Mol. Cell. Biol.* 1992, 12, 3507 – 3513.

8. Smeal, T., Binetruy, B., Mercola, D. A., Birrer, M., Karin, M. *Nature*. 1991, 354, 494 – 496.
9. Kyriakis, J. M., Banerjee, P., Nikolakaki, E., Dai, T., Rubie, E. A., Ahmad, M. F., Avruch, J., Woodgett, J. R., *Nature*. 1994, 369, 156 – 160.
10. Behrens, A., Sibilia, M., Wagner, E. F., *Nat. Genet.* 1999, 21, 326 – 329.
11. Kuan, C. Y., Yang, D. D., Samanta Roy, D. R., Davis, R. J., Rakic, P., Flavell, R. A., *Neuron*. 1999, 22, 667 – 676.
12. Chang, L., Jones, Y., Ellisman, M. H., Goldstein, L. S., Karin, M., *Dev. Cell*. 2003, 4, 521 – 533.
13. Sabapathy, K., Jochum, W., Hochedlinger, K., Chang, L., Karin, M., Wagner, E. F., *Mech. Dev.* 1999, 89, 115 – 124.
14. Yang, D. D., Kuan, C. Y., Whitmarsh, A. J., Rincon, M., Zheng, T. S., Davis, R. J., Rakic, P., Flavell, R. A., *Nature*. 1997, 389, 865 – 870.
15. Hunot, S., Vila, M., Teismann, P., Davis, R. J., Hirsch, E. C., Przedborski, S., Rakic, P., Flavell, R. A., *Proc. Natl. Acad. Sci. U.S.A.* 2004, 101, 665 – 670.
16. Borsello, T., Clarke, P. G. T., Hirt, L., Vercelli, A., Repici, M., Schorderet, D. F., Bogousslavsky, J., Bonny, C., *Nat. Med.* 2003, 9, 1180 – 1186.
17. Zhu, X., Raina, A. K., Rottkamp, C. A., Aliev, G., Perry, G., Boux, H., Smith, M. A., *J. Neurochem.* 2001, 76, 435 – 441.
18. Pei, J. J., Braak, E., Braak, H., G-Iqbal, I., Iqbal, K., Winblad, B., Cowburn, R. F., *J. Alzheimer's Dis.* 2001, 3, 41 – 48.
19. Morishima, Y., Gotoh, Y., Zieg, J., Barrett, T., Takano, H., Flavell, R., Davis, R. J., Shirasaki, Y., Greenberg, M. E., *J. Neurosci.* 2001, 21, 7551 – 7560.
20. Reynolds, C. H., Betts, J. C., Blackstock, W. P., Nebreda, A. R., Anderton, B. H., *J. Neurochem.* 2000, 74, 1587 – 1595.
21. Reynolds, C. H., Utton, M. A., Gibb, G. M., Yates, A., Anderton, B. H., *J. Neurochem.* 1997, 68, 1736 – 1744.
22. Kuo, L. H., Hu, M. K., Hsu, W. M., Tung, Y. T., Wang, B. J., Tsai, W. W., Yen, C. T., Liao, Y. F., *Mol. Biol. Cell.* 2008, 19, 4201 – 4212.
23. Tan, L., Schedl, P., Song, H. J., Garza, J., Konsolaki, M., *PLoS One*. 2008, 3, e3966. doi:10.1371/journal.pone.0003966.
24. Bowers, S., Truong, A. P., Neitz, R. J., Neitzel, M., Probst, G. D., Hom, R. K., Peterson, B., Galemno, R. A. Jr., Konradi, A. W., Sham, H. L., Tóth, G., Pan, H., Yao, N., Artis, D. R., Brigham, E. F., Quinn, K. P., Sauer, J. M., Powell, K., Ruslim, L., Ren, Z., Bard, F., Yednock, T. A., Griswold-Prenner, I., *Bioorg Med Chem Lett*. 2011, 21, 1838 – 1843.
25. Kubinyi, H., *Drug Discov. Today*. 1997, 2, 457 – 467.
26. Dessalew, N., Bharatam, P. V., *Eur J Med Chem*. 2007, 42, 1014 – 1027.
27. Dessalew, N., Patel, D. S., Bharatam, P.V., *J. Mol. Graph. Mod.* 2007, 25, 885 – 895.
28. Dalby, A., Nourse, J. G., Hounshell, W. D., Gushurst, A. K. I., Grier, D. L., Leland, B. A., Laufer, J., *J Chem Inf Comput Sci.* 1992, 32, 244 – 255.
29. Sadowski, J., Gasteiger, J., *Chem. Rev.* 1993, 93, 2567 – 2581.
30. Kovatcheva, A., Buchbauer, G., Golbraikh, A., Wolschann, P., *J. Chem. Inf. Comput. Sci.* 2003, 43, 259 – 266.
31. Paliwal, S. K., Singh, S., Kumari, S., Siddiqui, A. A., *Indian J Chem.* 2010, 49B, 554 – 560.
32. Paliwal, S. K., Pal, M., Siddiqui, A. A., *Med Chem Res.* 2010, 19, 475 – 489.
33. Luco, J. M., Ferretti, F. H., *J Chem Inf Comput Sci.* 1997, 37, 392 – 401.
34. Tarko, L., Ivanciuc, O., *Match Commun Math Comput Chem.* 2001, 44, 201 – 214.
35. Dessalew, N., *Acta Pharm.* 2009, 59, 31 – 43.
36. Dessalew, N., *QSAR Comb Sci.* 2008, 27, 901 – 912.
37. Hawkins, D. M., Basak, S. C., Mills, D., *J Chem Inf Comput Sci.* 2003, 43, 579 – 586.
38. Roy, P. P., Paul, S., Mitra, I., Roy, K., *Molecules.* 2009, 14, 1660 – 1701.
39. Lipinski, C. A., Lombardo, F., Dominy, D. W., Feeney, P. J., *Adv Drug Deliv Rev.* 1997, 23, 3 – 25.