Journal of the Serbian Chemical Society
JSCS-info@shd.org.rs • www.shd.org.rs/JSCS
*J. Serb. Chem. Soc. 80* (2) 187–196 (2015)
*JSCS–4709*
UDC 519.233.5:539.16:66.097.8:547.75
*Original scientific paper*

# Application of genetic algorithm – multiple linear regressions to predict the activity of RSK inhibitors

ZHILA MOHAJERI AVVAL[1], ESLAM POURBASHEER[1]*, MOHAMMAD REZA GANJALI[2] and PARVIZ NOROUZI[2,3]

[1]*Department of Chemistry, Payame Noor University, Tehran, Iran,* [2]*Center of Excellence in Electrochemistry, Faculty of Chemistry, University of Tehran, Tehran, Iran and* [3]*Biosensor Research Center, Endocrinology and Metabolism Molecular-Cellular Sciences Institute, Tehran University of Medical Sciences, Tehran, Iran*

(Received 25 May, accepted 19 June 2014)

*Abstract*: This paper considers the development of a linear quantitative structure–activity relationship (QSAR) model for predicting the ribosomal S6 kinase (RSK) inhibition activity of some new compounds. A dataset consisting of 59 pyrazino[1,2-$\alpha$]indole, diazepino[1,2-$\alpha$]indole, and imidazole derivatives with known inhibitory activities was used. The multiple linear regressions (MLR) technique combined with stepwise (SW) and the genetic algorithm (GA) methods as variable selection tools was employed. For more checking of the stability, robustness and predictability of the proposed models, internal and external validation techniques were used. Comparison of the obtained results, indicate that the GA-MLR model is superior to the SW–MLR model and that it is applicable for designing novel RSK inhibitors.

*Keywords:* QSAR; genetic algorithms; multiple linear regression; RSK inhibitors.

## INTRODUCTION

The RSK (90 kDa ribosomal S6 kinase) family comprises a group of highly related serine/threonine kinases that regulate diverse cellular processes, including cell growth, proliferation, survival and motility. This family consists of four human isoforms (RSK1-4), and single family member orthologues are also present in *Drosophila* and *Caenorhabditis elegans*.[1] RSK1-4 are a family of widely expressed Ser/Thr kinases characterized by two non-identical, functional kinase domains[2] and a carboxy-terminal docking site for extracellular signal-regulated kinases (ERKs).[3] Several sites both within and outside of the RSK kinase domain, including Ser380, Thr359, Ser363, and Thr573, are important for kinase activation.[4] RSK1-3 are activated *via* coordinated phosphorylation by mitogen-

* Corresponding author. E-mail: pourbasheer@ut.ac.ir
doi: 10.2298/JSC140523064A

activated protein kinases (MAPKs), autophosphorylation, and phosphoinositide-3-OH kinase (PI3K) in response to many growth factors, polypeptide hormones, and neurotransmitters.[4] RSK4 appears to demonstrate different pharmacology.

Metastasis, the spreading of cancer cells from a primary tumor to secondary sites throughout the body, is the primary cause of death for patients with cancer. New therapies that prevent invasion and metastasis in combination with current treatments could therefore significantly reduce cancer recurrence and morbidity. Metastasis is driven by altered signaling pathways that induce changes in cell–cell adhesion, the cytoskeleton, integrin function, protease expression, epithelial--to-mesenchymal transition and cell survival. The ribosomal S6 kinase (RSK) family of kinases is a group of extracellular signal-regulated kinase/mitogen-activated protein kinase (ERK/MAPK) effectors that can regulate these steps of metastasis by phosphorylating both nuclear and cytoplasmic targets.

However, the present understanding of the function of RSK in metastasis remains incomplete and is complicated by the fact that the four RSK isoforms perform non-redundant, sometimes opposing functions. Although some isoforms promote cell motility and invasion by altering transcription and integrin activity, others impair cell motility and invasion through effects on the actin cytoskeleton. The mechanism of RSK action depends on both the isoform and the cancer type. However, despite the variance in RSK-mediated outcomes, chemical inhibition of this group of kinases has proven effective in blocking invasion and metastasis of several solid tumors in preclinical models. RSKs are therefore a promising drug target for antimetastatic cancer treatments that could supplement and improve current therapeutic approaches.[5]

The experimental evaluation of the inhibitory activity of chemical structures is difficult, expensive and time-consuming, thus the development of a computational method for its prediction would be useful and of interest.[6–12] Among computational methods, the quantitative structure–activity relationship (QSAR) model has found diverse applications for predicting the properties/activities of a compound, including prediction of biological activity,[13,14] physical properties,[15,16] toxicity[17,18] and antiviral activity.[19,20] This model is a mathematical equation that can express the chemical properties or activities of compounds as a function of their various structural parameters (descriptors). The first step in building a QSAR model is the selection of a set of molecular descriptors that represent variation in the structural property of the molecules by a number. Thus, variable selection methods, as an inseparable part of the model development in QSAR studies, are used to select the best subset of descriptors. There are several variable selection methods, such as genetic algorithm (GA) and stepwise (SW). After selection of the most important descriptors, the model can be generated based on these selected descriptors using multiple linear regressions (MLR). The success of any QSAR model depends on the accuracy of the input data, selection

of appropriate descriptors and most importantly validation of the developed model.[21]The aim of this work was to develop a new QSAR model to predict the (RSK) inhibition activity of pyrazino[1,2-$\alpha$]indole, diazepino[1,2-$\alpha$]indole and imidazole derivatives.

## MATERIAL AND METHODS

### *Data set*

The data set consisting of 59 molecules of pyrazino[1,2-$\alpha$]indole, diazepino[1,2-$\alpha$]indole, and imidazole derivatives along with their experimental inhibitory activities were collected from the literature.[22,23] The chemical structures and their experimental values are presented in Table S-I of the Supplementary material to this paper. The half-maximal inhibitory concentration data ($IC_{50}$ / μM) was converted to the logarithmic scale p$IC_{50}$ (–log ($IC_{50}$ / M)) and then utilized for the subsequent QSAR analyses as the dependent variable. The whole dataset was randomly segregated into a training and a test set consisting of 48 and 11 compounds, respectively. The training set was used to construct a regression model, and the test set was used to evaluate the predictive ability of the obtained model.

### *Descriptors calculation*

The two-dimensional (2D) structures of the molecules were constructed using Hyperchem 7 and pre-optimization was performed using molecular mechanics force field (MM+) and the final optimization of the geometries was realized using a semi-empirical (AM1) procedure with a root mean square gradient of 0.01 kcal $mol^{-1}$. Dragon 2.1 software was employed to calculate 1497 molecular descriptors for all the studied chemical structures. In order to reduce redundant and non-useful information, constant or near-constant values and descriptors found to be highly correlated pair wise were removed in a pre-reduction step. Thereby, 398 molecular descriptors remained for the variable selection step.

### *Genetic algorithm (GA)*

The main challenges for QSAR practitioners are to find an appropriate set of descriptors and a suitable function that can accurately illustrate the experimental data. Nowadays, the genetic algorithm (GA) method, developed by Holland *et al*.,[24] is considered superior to other variable selection methods. It is a powerful optimization method that was inspired by evolutionary principles, including survival of the fittest, reproduction, crossover, and mutation. In this study, GA-MLR was used to build the QSAR model. The fitness function utilized herein was the leave-one-out (LOO) cross-validated correlation coefficient ($Q^2$). The GA program was written in Matlab 6.5.[25]

### K-*Means cluster analysis*

One of the main non-hierarchical clustering techniques is the *K*-means clustering, which is used in the division of a dataset into the training and test sets.[26] Ideally, this division is performed so as the points representing the training and test set are distributed within the whole descriptor space occupied by the entire data set.[27] In *K*-means cluster analysis (*K*-MCA), the clusters are started randomly and their means are calculated in the descriptor space. Molecules are reassigned to clusters the means of which are closer to the position of the molecules. This is followed by the selection of the test set molecules from each cluster since both test set and training set can represent all clusters and characteristics of the whole data set. In this study, the original data set was partitioned into four clusters based on *K*-means clustering. Then, about 20% of compounds of each cluster were chosen as members of the test set. The *K*-means clustering results are shown in Table S-II of the Supplementary material.

## RESULTS AND DISCUSSION

The whole data set was randomly partitioned into the training set of 48 compounds, and the test set of 11 compounds based on rule: the range of the experimental inhibitory activity values of both the training set and test set should be covered from the lowest to the highest. The training and test sets are indicated in Table S-1 of the Supplementary material. After the splitting of the datasets, the stepwise method was performed to select the main descriptors correlated to the activity based on the training set compounds. The six descriptors obtained by the SW–MLR linear model are as follows:

$$\begin{aligned} pIC_{50} = &-6.820(\pm1.315) + 5.757(\pm0.540)BELp5 + 32.252(\pm6.716)HATS6p + \\ &+ 1.362(\pm0.304)GATS8e - 5.443(\pm1.246)E3m + 0.154(\pm0.056)nHDon - \\ &- 0.305\ (\pm0.119)RDF130p \end{aligned} \quad (1)$$

whereby $N_{train}$ = 48, $R^2_{train}$ = 0.824, $R^2_{test}$ = 0.159, $R^2_{adj}$ =0.798, $F_{train}$ = 31.971, $F_{test}$ = 0.369, $RMSE_{train}$ = 0.395, $RMSE_{test}$ = 1.079, $Q^2_{LOO}$ = 0.767, $Q^2_{LGO}$ = = 0.745, $Q^2_{BOOT}$ = 0.751, where $N$ is the number of compounds, $R^2$ is the squared correlation coefficient, $R^2_{adj}$ is adjusted $R^2$, *RMSE* is the root mean square error, $F$ is the Fisher $F$ statistic and $Q^2_{LOO}$, $Q^2_{LGO}$ and $Q^2_{BOOT}$ are the squared cross-validation coefficients for leave one out, leave group out and bootstrapping, respectively.

The obtained statistical parameters indicate that the SW-MLR procedure produced good results for the training set, but it did not produce good results for the test set. Therefore, a genetic algorithm was used to select the best set of descriptors, and various models with various numbers of descriptors were obtained. To select the optimum number of descriptors with GA, the influence of the number of the descriptors was investigated from one to ten descriptors. Finally, a GA–MLR model with six selected descriptors was obtained. This model is described by the following equation:

$$\begin{aligned} pIC_{50} = &+2.713(\pm1.678) + 7.017(\pm1.080)MATS6e - 3.085(\pm0.666)MATS8e + \\ &+ 0.209(\pm0.071)RDF140u - 0.142(\pm0.057)RDF120m - \\ &- 1.016(\pm0.117)Mor04m + 31.823(\pm10.919)G3e \end{aligned} \quad (2)$$

whereby $N_{train}$ = 48, $R^2_{train}$ = 0.824, $R^2_{test}$ = 0.864, $R^2_{adj}$ = 0.798, $F_{train}$ = 31.90, $F_{test}$ = 3.143, $RMSE_{train}$ = 0.395, $RMSE_{test}$ = 0.483, $Q^2_{LOO}$ = 0.761, $Q^2_{LGO}$ = = 0.749, $Q^2_{BOOT}$ = 0.740.

The GA-MLR model was then used to predict the test set data and the prediction results are given in Table S-I of the Supplementary material. The predicted values of $pIC_{50}$ for the compounds in the training and test sets using the GA-MLR model are plotted against their experimental values in Fig. 1. As can be seen from Table S-I and Fig. 1, the prediction values are in good agreement with the experimental values.

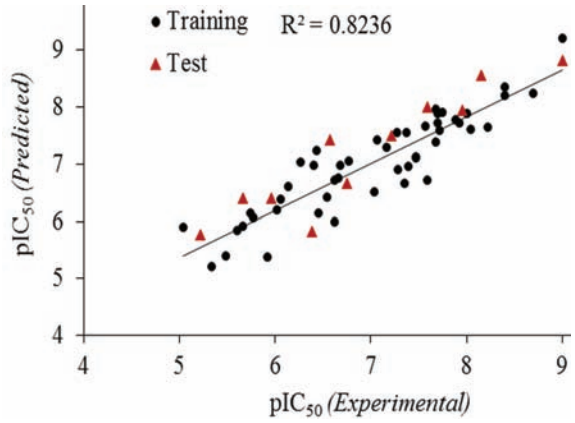


Fig. 1. The $pIC_{50}$ values predicted by GA–MLR modeling *vs*. the experimental $pIC_{50}$ values.

The six selected descriptors using the GA–MLR method were MATS6e, MATS8e, RDF140u, RDF120m, Mor04m and G3e. The multi-colinearity between the selected descriptors was checked by calculating their variation inflation factors (*VIF*), which can be calculated as follows:

$$VIF = \frac{1}{1-r^2} \tag{3}$$

where $r^2$ is the multiple correlation coefficient of the effect of one descriptor regressed on the remaining molecular descriptors. If *VIF* equals 1, then no inter--correlation exists for each variable; if *VIF* falls into the range of 1–5, the related model is acceptable; and if *VIF* is larger than 10, the related model is unstable and a recheck is necessary.[28] As can be seen from Table I, most variables had *VIF* values of less than 5, indicating that the obtained GA–MLR model has obvious statistical significance.

The correlation matrix for the six selected descriptors is given in Table II, from which, it could be seen that the correlation coefficient value of each of the two descriptors was less than 0.56, which meant that the selected descriptors were independent in the analysis.

In general, the assessment of stability and predictive ability of a model is an important step in the expression of the quality of a model. The ability of the GA--MLR model was verified by the leave-one-out and leave-group-out cross-validated correlation coefficients ($Q^2_{LOO}$ and $Q^2_{LGO}$). The internal predictive ability of the model was also verified using the bootstrap $Q^2_{BOOT}$ procedure, as is strongly recommended for QSAR modeling. The robustness of the proposed model and its predictive ability was guaranteed by the high value of $Q^2_{BOOT}$ based on the bootstrapping being repeated 5000 times. The cross-validation parameters for the MLR model are given in Eq. (2). The cross-validation results indicate that the obtained regression model has good internal and external predictive power.

TABLE I. Details of the constructed GA-MLR model; *SE* – standard error, *MF* – mean effect, *VIF* – variation inflation factor

| Descriptor | Coefficient | *SE* | *MF*[a] | *VIF*[b] | Chemical meaning |
|---|---|---|---|---|---|
| Constant | 2.713 | 1.678 | 0 | 0 | – |
| MATS6e | 7.017 | 1.080 | -0.18954 | 1.09685 | Moran autocorrelation - lag 6 / weighted by atomic Sanderson electronegativities |
| MATS8e | –3.085 | 0.666 | -0.21865 | 1.45634 | Moran autocorrelation - lag 8 / weighted by atomic Sanderson electronegativities |
| RDF140u | 0.209 | 0.071 | 0.02197 | 1.42781 | Radial Distribution Function - 14.0 / unweighted |
| RDF120m | –0.142 | 0.057 | -0.02503 | 1.47806 | Radial Distribution Function - 12.0 / weighted by atomic masses |
| Mor04m | –1.016 | 0.117 | 0.24938 | 1.50138 | 3D-MoRSE - signal 04 / weighted by atomic masses |
| G3e | 31.823 | 10.919 | 1.16187 | 1.65773 | 3rd component symmetry directional WHIM index / weighted by atomic Sanderson electronegativities |

[a]Mean effect; [b]variation inflation factors

TABLE II. Correlation coefficient matrix of the selected descriptors

| | MATS6e | MATS8e | RDF140u | RDF120m | Mor04m | G3e |
|---|---|---|---|---|---|---|
| MATS6e | 1 | 0 | 0 | 0 | 0 | 0 |
| MATS8e | –0.12 | 1 | 0 | 0 | 0 | 0 |
| RDF140u | 0.05 | –0.04 | 1 | 0 | 0 | 0 |
| RDF120m | 0.12 | 0.01 | 0.55 | 1 | 0 | 0 |
| Mor04m | 0.05 | 0.45 | 0.10 | –0.12 | 1 | 0 |
| G3e | –0.20 | 0.54 | –0.07 | –0.15 | 0.47 | 1 |

A Williams plot is used to visualize the applicability domain (AD) of QSAR models. It is a plot of the standardized residuals *vs*. the leverage values ($h$).[29] From the Williams plot (Fig. 2), it is obvious that there are three compounds (Nos. 38 and 55 in the training set and No. 50 in the test set) that have a leverage value higher than the warning $h^*$ value of 0.44 and thus, they could be considered as structural outliers. Fortunately, in these cases, the data predicted by the model are good for the three compounds and thus, they are "good leverage" chemicals. From Fig. 2, it is obvious that there are no outlier compounds with standard residuals $>3\delta$ for both the training and the test sets.

The *Y*-randomization test is performed to assess the robustness of a QSAR model by building several random models *via* shuffling the dependent variable vector ($pIC_{50}$), while keeping the independent variables as it is. The resultant random models are expected to have low $R^2$ and $Q^2_{LOO}$ values.[30] The results of *Y*-randomization tests are shown in Table III.

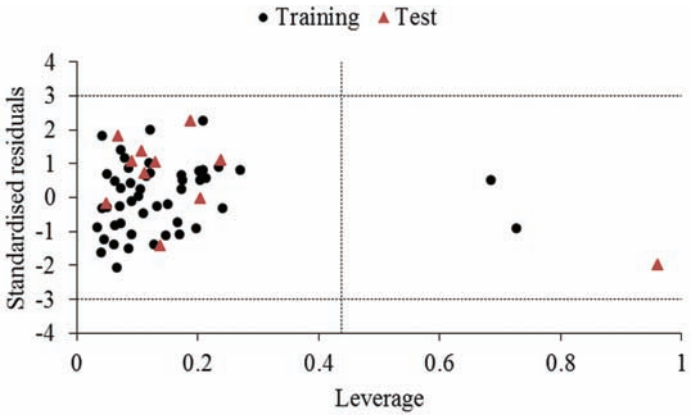


Fig. 2. The William plot of the GA–MLR model.

TABLE III. $R^2_{train}$ and $Q^2_{LOO}$ values after several *Y*-randomization tests

| No. | $Q^2$ | $R^2$ |
|---|---|---|
| 1 | 0.060 | 0.082 |
| 2 | 0.214 | 0.078 |
| 3 | 0.093 | 0.037 |
| 4 | 0.027 | 0.109 |
| 5 | 0.007 | 0.151 |
| 6 | 0.106 | 0.067 |
| 7 | 0.015 | 0.085 |
| 8 | 0.102 | 0.030 |
| 9 | 0.008 | 0.160 |
| 10 | 0.031 | 0.090 |

*Interpretation of descriptors*

Interpretation of descriptors contained in the best model (GA-MLR) provides useful chemical insights into the mechanism of the inhibitory activity. Thus, an acceptable interpretation of the QSAR results is provided below. The molecular descriptors selected by the genetic algorithm are given in Table I.

The first and second descriptors are MATS6e (Moran autocorrelation - lag 6 / weighted by atomic Sanderson electronegativities) and MATS8e (Moran autocorrelation - lag 8 / weighted by atomic Sanderson electronegativities) that belong to the 2D autocorrelations descriptors. In this descriptor, the Geary coefficient is a distance-type function, this function is any physicochemical property calculated for each atom of the molecule, such as atomic mass, polarizability, *etc*. Thus, the molecule atoms represent a set of discrete points in space and the atomic property is the function evaluated at these points.[31] In these descriptors (MATS6e and MATS8e), the weighting scheme is the atomic Sanderson electronegativities, which show that the electronegativities of the molecule atoms play the main role in these descriptors. MATS6e has a positive sign in equation, which indicates

that the $pIC_{50}$ value is directly related to this descriptor. By increasing the atomic Sanderson electronegativities for each molecule atoms, the value of this descriptor increased, causing an increase in its $pIC_{50}$ value. MATS8e displays a negative sign, which shows that the $pIC_{50}$ value is inversely related to this descriptor. Hence, it was concluded that increasing the value of this descriptor causes a reduction in the $pIC_{50}$ value.

The third and fourth descriptors appearing in the model are RDF140u (Radial Distribution Function - 14.0 / unweighted) and RDF120m (Radial Distribution Function - 12.0 / weighted by atomic masses). These descriptors belong to the radial distribution function (RDF) descriptors. RDF descriptors meet all the requirements for 3D structure descriptors; they are independent of the number of atoms (*i.e.*, the size of a molecule), and are unique regarding the three-dimensional arrangement of the atoms and are invariant against the translation and rotation of the molecules. Additionally, the RDF descriptors can be restricted to specific atom types or distance ranges to represent specific information in a certain three-dimensional structure space (*e.g.*, to describe the steric hindrance or the structure/activity properties of a molecule). Formally, the radial distribution function of an ensemble of *n* atoms can be interpreted as the probability distribution of finding an atom in a spherical volume of radius *R*.[32] RDF140u has a positive sign, and the positive sign indicates that the $pIC_{50}$ value was directly related to this descriptor. The negative sign of RDF120m suggests that the $pIC_{50}$ value was inversely related to this descriptor. Therefore, increasing the atomic masses of molecules leads to decrease in the $pIC_{50}$ value.

The next descriptor is Mor04m (3D-MoRSE - signal 04 / weighted by atomic masses) which is one of the 3D-MoRSE descriptors. 3D-MoRSE descriptors (3D Molecule Representation of Structures based on Electron diffraction) are derived from Infrared spectra simulation using a generalized scattering function.[33] Mor04m had a negative sign in the equation, which indicates that the $pIC_{50}$ value was inversely related to this descriptor.

The last descriptor is G3e (3rd component symmetry directional WHIM index / weighted by atomic Sanderson electronegativities) that belongs to the WHIM descriptors. Weighted holistic invariant molecular (WHIM) descriptors are geometrical descriptors based on the statistical indices calculated on the projections of atoms along the principal axes. The algorithm consists of performing a principal components analysis on the centered Cartesian coordinates of a molecule by using a weighted covariance matrix obtained from different weighting schemes for the atoms. Directional WHIM symmetry descriptors are related to the number of central symmetric atoms (along the $m^{th}$ component), the number of unsymmetrical atoms and the total number of molecule atoms.[34] As it is clear from Table I, G3e has a positive sign, illustrating a greater coefficient

value than that of the other descriptors. This descriptor had a significant effect on the $pIC_{50}$ value of the studied molecules.

From the above discussion, it could be concluded that the atomic Sanderson electronegativities, radial distribution function and atomic masses play an important role in the RSK inhibitory activities of the studied compounds.

## CONCLUSIONS

In this work, a linear QSAR model was presented for prediction of RSK inhibitors of pyrazino[1,2-$\alpha$]indole, diazepino[1,2-$\alpha$]indole and imidazole derivatives. The best subset of calculated descriptors was selected by use of stepwise and genetic algorithm methods. Validation of the model was performed using separation of the data into two independent sets, *Y*-randomization, cross-validation by LOO, LGO and bootstrap. The results indicated that the constructed GA-MLR model is a valid model with high statistical quality and low prediction errors. The proposed model could identify and provide an insight into some suggestions for the further design of new RSK inhibitors.

## SUPPLEMENTARY MATERIAL

Chemical structures, the observed $pIC_{50}$ values and *K*-means clustering of the compounds are available electronically from http://www.shd.org.rs/JSCS/, or from the corresponding author on request.

## И З В О Д

## ПРИМЕНА ГЕНЕТСКОГ АЛГОРИТМА – ВИШЕСТРУКЕ ЛИНЕАРНЕ РЕГРЕСИЈЕ ЗА ПРЕДВИЂАЊЕ АКТИВНОСТИ ИНХИБИТОРА RSK

ZHILA MOHAJERI AVVAL[1], ESLAM POURBASHIR[1],MOHAMMAD REZA GANJALI[2] и PARVIZ NOROUZI[2,3]

*[1]Department of Chemistry, Payame Noor University, Tehran, Iran, [2]Center of Excellence in Electrochemistry, Faculty of Chemistry, University of Tehran, Tehran, Iran and [3]Biosensor Research Center, Endocrinology and Metabolism Molecular-Cellular Sciences Institute, Tehran University of Medical Sciences, Tehran, Iran*

Развијен је линеарни QSAR модел за предвиђање RSK инхибиторне активности неких нових једињења. База података се састојала од 59 деривата пиразино[1,2-$\alpha$]индола, диазепино[1,2-$\alpha$]индола и имидазола са познатим инхибиторним активностима. Примењена је техника вишеструких линеарних регресија (MLR) комбинована са SW и GA алгоритмима. Оригинални сет података је подељен на тренажни (80 % података) и тест сет (20 % података). Добијени резултати указују на то да је GA-MLR модел бољи од SW-MLR модела, и да је употребљив за дизајнирање нових инхибитора RSK.

(Примљено 25. маја, прихваћено 19. јуна 2014)

## REFERENCES

1. R. Yves, Z. Xiaocui, P. R. Philippe, *Biochem. J.* **441** (2012) 553
2. T. L. Fisher, J. Blenis, *Mol. Cell. Biol.* **16** (1996) 1212
3. J. A. Smith, C. E. Poteet-Smith, K. Malarkey, T. W. Sturgill, *J. Biol. Chem.* **274** (1999) 289
4. K. N. Dalby, N. Morrice, F. B. Caudwell, J. Avruch, P. Cohen, *J. Biol. Chem.* **273** (1998) 1496
5. F. J. Sulzmaier, J. W. Ramos, *Cancer. Res.* **73** (2013) 6099

6. S. Vahdani, Z. Bayat, *Der. Chem. Sin.* **2** (2011) 235
7. M. Jalali-Heravi, M. Asadollahi-Baboli, *QSAR Comb. Sci.* **27** (2008) 750
8. A. Habibi-Yangjeh, E. Pourbasheer, M. Danandeh-Jenagharad, *Bull. Korean Chem. Soc.* **29** (2008) 833
9. A. Habibi-Yangjeh, E. Pourbasheer, M. Danandeh-Jenagharad, *Monatsh. Chem.* **140** (2009) 15
10. E. Pourbasheer, R. Aalizadeh, M. R. Ganjali, P. Norouzi, *Struct. Chem.* **25** (2014) 355
11. E. Pourbasheer, S. Riahi, M. R. Ganjali, P. Norouzi, *Mol. Divers.* **15** (2011) 645
12. S. Riahi, E. Pourbasheer, M. R. Ganjali, P. Norouzi, *Chem. Biol. Drug. Des.* **73** (2009) 558
13. O. Deeb, B. Hemmateenejad, *Chem. Biol. Drug. Des.* **70** (2007) 19
14. G. Ramírez-Galicia, R. Garduño-Juárez, B. Hemmateenejad, O. Deeb, M. Deciga-Campos, J. C. Moctezuma-Eugenio, *Chem. Biol. Drug. Des.* **70** (2007) 53
15. R. P. Verma, A. Kurup, C. Hansch, *Bioorg. Med. Chem.* **13** (2005) 237
16. M. H. Knaggs, C. McGuigan, S. A. Harris, P. Heshmati, D. Cahard, I. H. Gilbert, J. Balzarini, *Bioorg. Med. Chem. Lett.* **10** (2000) 2075
17. P. V. Khadikar, A. Phadnis, A. Shrivastava, *Bioorg. Med. Chem.* **10** (2002) 1181
18. V. Agrawal, P. Khadikar, *Bioorg. Med. Chem.* **9** (2001) 3035
19. W. G. Zhao, J. G. Wang, Z. M. Li, Z. Yang, *Bioorg. Med. Chem. Lett.* **16** (2006) 6107
20. V. Pandey, Z. Tusi, M. Tandon, M. Joshi, S. Bajpai, *Indian J. Chem., B* **42** (2003) 2583
21. E. Ibezim, P. Duchowicz, N. Ibezim, L. Mullen, I. Onyishi, S. Brown, E. Castro, *Afr. J. Basic Appl. Sci.* **1** (2009) 76
22. S. J. Boyer, J. Burke, X. Guo, T. M. Kirrane, R. J. Snow, Y. Zhang, C. Sarko, L. Soleymanzadeh, A. Swinamer, J. Westbrook, *Bioorg. Med. Chem.* **22** (2012) 733
23. T. M. Kirrane, S. J. Boyer, J. Burke, X. Guo, R. J. Snow, L. Soleymanzadeh, A. Swinamer, Y. Zhang, J. B. Madwed, M. Kashem, *Bioorg. Med. Chem. Lett.* **22** (2012) 738
24. J. H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, 1975
25. W. Zhu, G. Chen, L. Hu, X. Luo, C. Gui, C. Luo, C. M. Puah, K. Chen, H. Jiang, *Bioorg. Med. Chem.* **13** (2005) 313
26. E. Pourbasheer, A. Beheshti, H. Khajehsharifi, M. R. Ganjali, P. Norouzi, *Med. Chem. Res.* **22** (2013) 4047
27. Z. Cheng, Y. Zhang, W. Fu, *Med. Chem. Res.* **20** (2011) 1235
28. E. Pourbasheer, R. Aalizadeh, M. R. Ganjali, P. Norouzi, *Med. Chem. Res.* **23** (2014) 57
29. J. E. Riviere, J. D. Brooks, *Toxicol. Sci.* **119** (2011) 224
30. E. Pourbasheer, R. Aalizadeh, M. R. Ganjali, P. Norouzi, J. Shadmanesh, C. Methenitis, *Med. Chem. Res.* **23** (2014) 2264
31. E. Pourbasheer, R. Aalizadeh, M. R. Ganjali, P. Norouzi, A. Banaei, *Med. Chem. Res.* **23** (2014) 3082
32. T. Puzyn, J. Leszczynski, M. T. Cronin, *Recent advances in QSAR studies: methods and applications*, Springer Science & Business Media, Dordrecht, 2010
33. M. Nekoei, M. Mohammadhosseini, A. A. Gharahbagh, *Anal. Bioanal. Electrochem.* **1** (2009) 159
34. E. Pourbasheer, S. Riahi, M. R. Ganjali, P. Norouzi, *J. Enzyme Inhib. Med. Chem.* **25** (2010) 844.