

Interactions between Activin-Like Kinase 5 (ALK5) Receptor and its Inhibitors and the Construction of a Docking Descriptor-Based QSAR Model

Malihe Ebrahimi, Taghi Khayamian* and Sajjad Gharaghani

Department of Chemistry, Isfahan University of Technology, Isfahan, Iran

Neste estudo, simulações de dinâmica molecular (MD) e *docking* foram realizadas a fim de investigar o sítio de ligação e as interações de 61 inibidores com o receptor de kinase-5 do tipo ativina (ALK5). Uma simulação MD foi realizada sobre o receptor para obter sua conformação em água. A análise de *docking* revelou que interações do tipo ligação hidrogênio e hidrofóbicas desempenham papéis importantes no complexo inibidor-ALK5. Estas interações foram confirmadas por cristalografia de raios X. Além disso, para estudar a estabilidade da conformação do receptor, uma segunda simulação MD sobre o complexo foi realizada em um ambiente aquoso. O raio de giro para o complexo mostrou que a conformação de ALK5 não muda na presença do inibidor. Foram calculados 134 descritores obtidos da estrutura molecular e *docking*; os mais viáveis foram usados em relações quantitativas entre estrutura química e atividade biológica (QSAR). A regressão de vetores-suporte baseada em mínimos quadrados (LS-SVR) apresentou um modelo confiável com $Q^2 = 0,837$ e $R = 0,917$. Finalmente, os tipos de interações e propriedades dos descritores foram usados para propor novos inibidores.

In this study, molecular docking and molecular dynamics (MD) simulations were conducted to investigate both the binding site and interactions of 61 inhibitors with activin-like kinase-5 (ALK5) receptor. A MD simulation was performed on the receptor to obtain receptor conformation in a water environment. Docking analysis revealed that hydrophobic and hydrogen bonding interactions play important roles in ALK5-inhibitor complex. These interactions were confirmed by X-ray crystallography. Furthermore, to study receptor conformation stability, a second MD simulation on complex was performed in an aqueous environment. Radius of gyration for complex showed that the ALK5 conformation did not change in the presence of the inhibitor. 134 descriptors emerging from docking and molecular structure were calculated and the most feasible ones were used in quantitative structure -activity relationships (QSAR). The LS-SVR (least squares support vector regression) gave reliable model with $Q^2 = 0.837$ and $R = 0.917$. Finally, the types of interactions and properties of the descriptors were used to propose new inhibitors.

Keywords: ALK5, docking, molecular dynamics simulation, least squares support vector regression, QSAR

Introduction

The main member of the transforming growth factor β (TGF- β) superfamily was discovered over 20 years ago in normal rat kidney fibroblasts.¹ TGF- β is composed of two large branches represented by the prototypic TGF- β s, bone morphogenetic proteins (BMPs), and growth differentiation factors (GDFs).² The TGF- β signaling plays key roles in the regulation of a variety of physiological processes including cell proliferation and differentiation as well as development of fibrosis in various organs such as kidney,

heart, lung, and liver, as well as in the immunity response.³⁻⁵ Deregulation of TGF- β signaling causes a variety of human diseases including cancer,⁶ pancreatic diseases,⁷ and hematological malignancies.⁸ TGF- β s signal is initiated through a heterotetrameric receptor complex that consists of two transmembrane serine/threonine kinase receptor, 'type I' receptor (TbRI) or activin-like kinase 5 (ALK5) and 'type II' (TbRII), which is a transmembrane serine/threonine kinase receptor.⁹ Upon TGF- β binding, TbRII phosphorylates ALK5 in the GS region. The activated ALK5 in turn phosphorylates Smad2/Smad3 proteins at the C-terminal SSXS-motif, thereby causing dissociation from the receptor and heteromeric complex formation with

*e-mail: taghi@cc.iut.ac.ir

Smad4. This complex is delivered to the nucleus where it regulates the transcription of specific genes, giving rise to cell differentiation, proliferation, apoptosis, migration, and extracellular matrix production.¹⁰⁻¹² Clearly, inhibition of phosphorylation of Smad2/Smad3 by ALK5 could reduce TGF- β -induced pathological fibrosis.¹³ Studies have shown that several small molecule inhibitors block ALK5 catalytic activity. Ren *et al.*¹⁴ developed 3D pharmacophore models on 106 ALK5 inhibitors reported in the literature.¹⁴ They used the best quantitative pharmacophore model (Hypo1) as a 3D search query on the known ALK5 inhibitors and, finally, selected several compounds for further *in vitro* and *in vivo* assay studies. Geldenhuys and Nakamura¹⁵ presented a 3D-QSAR (quantitative structure-activity relationship) model, comparative molecular field analysis (COMFA), for 23 compounds of ALK5 inhibitors, all of which share a similar central core consisting of imidazole with a pyridine ring.¹⁵ Docking results showed that hydrogen bonding groups play an important role in affinity for kinase. Jin *et al.*,¹⁶ Gellibert *et al.*¹⁷ and Yakymovych *et al.*¹⁸ have reported different classes of ALK5 inhibitors. Modeling methods such as QSAR models have been found to be valuable in further optimization and development of novel inhibitors in drug design.¹⁹

Docking descriptors are calculated based on the major interactions between inhibitors and receptor such as hydrophobic, hydrogen bonding, salt bridge, close contact and π interactions. Thus, the obtained QSAR model from these descriptors could be very useful for rational drug design. It is the objective of the present study to identify the interactions between the ALK5 receptor and these inhibitors. A structure-based QSAR model is also developed to represent relationship between descriptors originating from docking and the activities of the inhibitors. Finally, new potent inhibitors will be proposed based on the results obtained on inhibitor-receptor interactions and the properties of the descriptors.

Experimental

Preparation of data set

The data set (Table 1) containing the inhibitory activities in the logarithmic scale ($\text{pIC}_{50} = \log 1/\text{IC}_{50}$) was obtained from Jin *et al.*,¹⁶ Gellibert *et al.*¹⁷ and Yakymovych *et al.*¹⁸ The basic structures of these inhibitors are shown in Scheme 1. The data set was randomly divided into two groups, the training and the test sets, consisting of 49 and 12 inhibitors, respectively. The training set was used for model generation and the test set for model evaluation. The structure of the inhibitors was drawn in the Hyperchem package version 7.0

(Hypercube, Inc.).²⁰ Energy minimization of these inhibitors was accomplished using the molecular mechanics method with Polak-Ribiere algorithm up to a root mean square gradient of $0.1 \text{ kcal mol}^{-1}$. The optimized molecular structures were used for docking studies.

The initial 3D structure of ALK5 in the complex with 4-((5,6-dimethyl-2-(2-pyridyl)-3-pyridyl)oxy)-n-(3,4,5-trimethoxyphenyl)pyridin-2-amine at 1.85 \AA (PDB ID: 2WOT) was obtained from Protein Data Bank (PDB).²¹ The crystal structure of the receptor contained one chain with a length of 306 amino acids and a structure weight of 35459.98 Da.

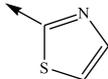
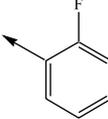
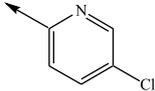
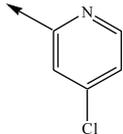
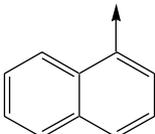
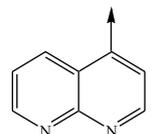
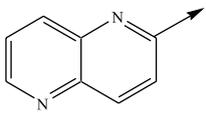
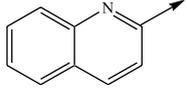
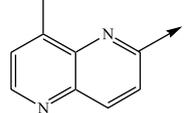
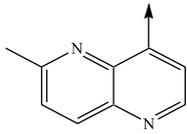
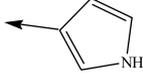
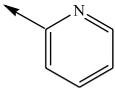
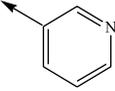
Molecular dynamics simulation

Molecular dynamics simulation was performed in two steps. In the first step, the structure of ALK5 was simulated in a water box. In the second, the result of docking of the most potent inhibitor, i.e. 3-(6-methylpyridin-2-yl)-4-([1,2,4] triazolo[1,5- α]pyridin-6-yl) pyrazoles-1-carbothioic acid (4-methoxy-phenyl)-amide (inhibitor **14a**), was loaded into the MD simulation to obtain receptor conformation. The topology parameters of inhibitor **14a** were built using the Dundee PRODRG2.5 server (beta).²² The MD simulation process was carried out by the GROMACS 4.5.1 package using the GROMOS-96 force field.²³ Water molecules were added using a simple point charge (SPC216) model.²⁴ Counterions of Na^+ were added by replacing solvent molecules in order to neutralize the system. The system was then placed in a cubic box with the dimensions $8.63 \times 8.63 \times 8.63 \text{ nm}^3$ containing 63413 atoms in total. Periodic boundary conditions were applied in the xyz space. Initially, energy minimization was performed before implementing the position restraint procedure. Then, NVT and NPT simulations were carried out. The NVT simulation was performed at a constant 310 K with a Nose-Hoover thermostat. Once the temperature was stabilized, the NPT simulation was performed using the Parrinello-Rahman pressure coupling under a pressure of 1 bar. The particle mesh Ewald (PME) method interaction was used.²⁵ For numerical integrations, the velocity verlet algorithm was used²⁶ and the initial atomic velocities were generated using a Maxwellian distribution at the given absolute temperature.²⁷ Finally, the full system was subjected to 6000 ps MD at 310 K under 1 bar pressure. The MD simulation and result analyses were performed on the open SUSE 11.3 Linux on an Intel Core 2 Quad Q6600 2.4 GHz with 4 GB of RAM.

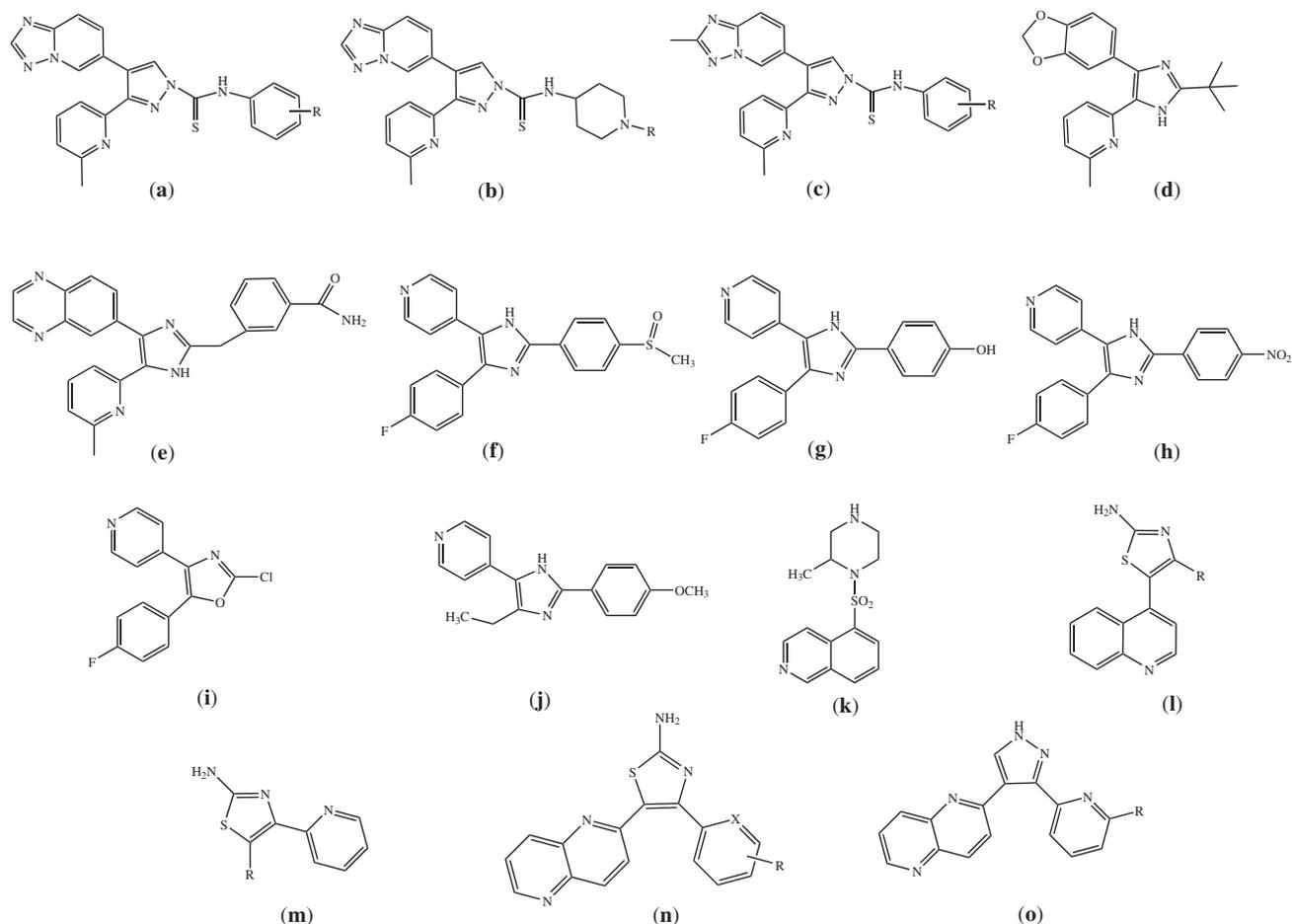
Molecular images and evaluation of molecular dynamics

All molecular images and animations were produced using molegro virtual docker (MVD)²⁸ and visual molecular

Table 1. Inhibitors and the predicted pIC_{50} by the LS-SVR model (see scheme 1)

Compound	R	X	pIC_{50}	Predicted pIC_{50}	Compound	R	X	pIC_{50}	Predicted pIC_{50}
1a ^t	H	-	8.55	8.87	47l		-	5.93	5.47
2a ^t	m-F	-	8.35	8.55	48l		-	5.89	6.46
3a	m-Cl	-	8.62	8.85	49l		-	5.75	6.52
4a ^t	m-Me	-	8.40	7.64	50l		-	5.99	5.98
5a	m-CF ₃	-	8.52	8.31	51m ^t		-	4.96	4.86
6a	m-OMe	-	8.65	8.38	52m		-	5.10	5.96
7a	m-OCF ₃	-	8.00	8.44	53m		-	7.39	7.26
8a	m-CN	-	8.35	8.12	54m		-	5.45	6.28
9a	m-CONH ₂	-	8.46	8.45	55m		-	7.52	7.11
10a	p-F	-	8.39	8.20	56m		-	6.42	6.79
11a	p-Cl	-	8.41	7.64	57n	6-CH ₃	N	7.60	6.82
12a ^t	p-Me	-	8.18	8.25	58n ^t	4-F	CH	7.38	6.79
13a	p-CF ₃	-	8.52	8.05	59n	3-Cl	CH	7.64	7.15
14a ^t	p-OMe	-	9.24	8.55	60o	H	-	7.52	6.98
15a	p-OCF ₃	-	8.48	8.73	61o	CH ₃	-	7.64	7.31
16a	p-CN	-	8.53	8.96	N.i [*]		-	-	10.00
17a	p-CONH ₂	-	7.81	7.94	N.i ^{**}	H	-	-	7.96
18a	o-F	-	8.27	8.46	N.i ^{**}	F	-	-	7.70
19a	o-Cl	-	8.55	8.22					
20a	o-Me	-	8.40	7.90					
21a	o-OMe	-	8.24	7.96					
22a	o-OCF ₃	-	8.40	8.48					
23a	3,4-F ₂	-	8.67	8.52					
24a ^t	3,5-F ₂	-	8.49	8.32					
25a	3,4-Cl ₂	-	8.42	8.36					
26a	3,5-Cl ₂	-	8.29	7.96					
27a	3,4-Me ₂	-	8.45	8.02					
28a	3,5-Me ₂	-	8.40	8.66					
29a	3,5-(CF ₃) ₂	-	8.38	8.90					
30a	3,4-(OMe) ₂	-	8.36	7.99					
31a	3,5-(OMe) ₂	-	8.60	8.27					
32b ^t	Me	-	7.40	7.85					
33b	Benzyl	-	7.02	7.72					
34c ^t	H	-	6.99	7.70					
35c	m-CN	-	7.07	7.64					
36c	m-CONH ₂	-	7.17	7.70					
37d	-	-	7.26	7.35					
38e	-	-	7.76	7.17					
39f	-	-	6.15	6.85					
40g	-	-	7.69	7.41					
41h	-	-	6.46	6.85					
42i ^t	-	-	5.30	4.59					
43j	-	-	4.52	5.35					
44k	-	-	3.70	3.76					
45l		-	6.14	5.72					
46l ^t		-	5.40	6.39					

^tThe compounds in test set; LS-SVR, least squares support vector regression; N.i^{*}: New inhibitor with basic structure of inhibitor a; N.i^{**}: New inhibitor with basic structure of inhibitor b.



Scheme 1. Basic structures of ALK5 inhibitors.

dynamics (VMD)²⁹ packages. Schematic two-dimensional representations of the docking results were performed using LIGPLOT.³⁰

Molecular docking

Docking was performed by AutoDock 4.0 program using the Lamarckian genetic algorithm (LGA) method.³¹ Autodock is a suitable program for actual docking simulation.³² The most potent inhibitor reported in Table 1, inhibitor **14a**, was docked into the structure extracted from MD trajectories of the receptor. All inhibitors were subsequently docked into the binding site obtained from the receptor and conformation of the inhibitors with the lowest binding free energy was used to calculate molecular descriptors.

Descriptor calculation

Descriptors were calculated based on inhibitor-receptor interactions using AutoDockTools and BINANA (BINding ANALyzer). AutoDockTools are capable of calculating

8 types of energy values that consist of: (i) estimated free energy of binding ($E_{FreeBind}$); (ii) estimated inhibition constant (k_i); (iii) final intermolecular energy ($E_{InterMol}$), which is the sum of the following three types of energy (iv) vdW + Hbond + desolv Energy (E_{vHD}); (v) electrostatic energy (E_{Elec}); (vi) final total energy (E_{FTot}); (vii) torsional free energy (E_{Tors}); (viii) unbound system's energy (E_{Unb}),³³ and the Gasteiger charge descriptor. Then, the docking conformer at its minimum $E_{FreeBind}$ was loaded into BINANA to calculate the descriptors. BINANA is a python-implemented algorithm to characterize the binding of inhibitor-receptor complex. The BINANA accepts receptor and inhibitor files generated by AutodockTools in the PDBQT format.³⁴ BINANA descriptors consist of (i) close contacts, BINANA determines all ligand and protein atoms that come within 2.5 and 4 Å of each other; (ii) electrostatic interactions; (iii) hydrophobic contacts, which is the sum of the six possible classifications: α -sidechain, α -backbone, β -sidechain, β -backbone, other-sidechain, and other-backbone; (iv) hydrogen bonds, which allow 12 possible categorizations: α -sidechain-ligand, α -backbone-ligand, β -sidechain-ligand, β -backbone-ligand, other-

sidechain-ligand, other-backbone-ligand, α -sidechain-receptor, α -backbone-receptor, β -sidechain-receptor, β -backbone-receptor, other-sidechain-receptor and other-backbone-receptor; (v) salt bridges, and (vi) π interactions.³⁵ Therefore, 117 descriptors may be calculated from the interactions of inhibitors with the receptor in BINANA. Also, 7 descriptors were obtained from Hyperchem that consist of surface area, volume, hydration energy, log P, refractivity, mass and polarizability.

Theory of least squares support vector regression

Support vector regression (SVR) was originally introduced by Cortes and Vapnik.³⁶ Least squares support vector regression (LS-SVR) is an alternative method of SVR described in Suykens *et al.*³⁷ LS-SVR always fits a linear relation, $y = wx + b$, between the independent variable (x) and the dependent one (y), where w is a weight vector and b is the bias term. Consider a given training set $\{x_i, y_i\}$, $i = 1, 2, \dots, N$, with input data $x_i \in \mathbb{R}$ and output data $y_i \in \mathbb{R}$. The error in the prediction for each sample i is defined as:

$$y_i = (\langle x_i, w \rangle + b) = e_i \quad i = 1, 2, \dots, N \quad (1)$$

As in SVR, it is necessary to minimize a cost function C containing a penalized regression error, as in equation 2:

$$\min c(w, e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^N e_i^2 \quad (2)$$

Subject to

$$y = W^T \varphi(x_i) + b = e_i \quad i = 1, 2, \dots, N \quad (3)$$

The parameter γ , which has to be optimized by the user, gives the relative weight of this part as compared to the first part.³⁸ The Lagrangian for this problem is defined as:

$$L(w, b, e; \alpha) = \frac{1}{2} W^T W + \frac{1}{2} \gamma \sum_{i=1}^N e_i^2 - \sum_{i=1}^N \alpha_i (\langle x_i, w \rangle + b + e_i - y_i) \quad (4)$$

where α refers to Lagrange multipliers, which can be either positive or negative due to the equality constraints in the LS-SVR algorithm. Optimum conditions can be obtained by setting all derivatives of $L(w, b, e; \alpha)$ equal to zero. Upon eliminating the variables w and e , one gets the following set of linear equations:

$$\sum_{j=1}^N \alpha_j \langle x_j, x_i \rangle + b + \gamma^{-1} \alpha_i = y_i \quad i = 1, 2, \dots, N \quad (5)$$

$$\sum_{j=1}^N \alpha_j = 0 \quad (6)$$

The training samples are only present in equation 5 as the inner product. This allows this standard inner product to be substituted with the most popular choice of the kernel types, i.e. the radial basis kernel function (RBF), defined as:

$$K(x, x_i) = \exp \left\{ - \left| x - x_i \right|^2 / \sigma^2 \right\} \quad (7)$$

where σ is the radial basis function of the kernel parameter. RBF function can also handle the nonlinear relationship between the input and the output.³⁹ The free LS-SVM toolbox (LS-SVM V-1.5, Suykens, Leuven, Belgium) was used in MATLAB Version 7.6 to derive all the LS-SVM models.⁴⁰

Applicability domain

In order to obtain a reliable prediction for the test compounds, the exploration of applicability domain (AD) for a QSAR model is necessary. Here a leverage approach was used to verify the prediction reliability. A simple measure of the applicability domain of the model is its leverage h_i , which is defined as:

$$h_i = x_i (X^T X)^{-1} x_i^T \quad (i = 1, \dots, n) \quad (8)$$

where x_i is the descriptor row-vector of the query compound i and X is the $n \times k$ matrix of the training set (k is the number of model descriptors, and n is the number of training set samples). The warning leverage (h^*) is, generally, fixed at $3(k+1)/n$, where k is the number of model descriptors and n is the number of training compounds.⁴¹⁻⁴³

Results and Discussion

Molecular dynamics simulation of the ALK5 receptor

To provide conformation of ALK5 in a water environment, a 6000 ps MD simulation was carried out on ALK5 in a water box. The stability of the system (protein, water, and ions) was tested by means of the root mean square deviations (RMSD's) of protein with respect to protein's initial structure. The RMSD values of ALK5 were plotted from 0 to 6000 ps which imply that the RMSD of the system reaches equilibrium and oscillates around 0.45 nm after 3800 ps of the simulation onset. The average RMSD value of protein backbone was calculated to be 0.45 ± 0.03 nm. The RMSD value indicates that ALK5

conformation reached equilibrium after 3800 ps in a water environment. The equilibrated conformation of the receptor was used for docking.

Molecular docking

In order to determine the binding site in ALK5 receptor, blind docking was performed on receptor with the most potent inhibitor (**14a**). In blind docking the grid map set to $92 \times 92 \times 92 \text{ \AA}$ along the x, y and z axes with 0.375 \AA grid spacing. The center of grid map set to 9.218, 1.277 and 12.584 \AA . The grid maps were obtained using AutoGrid. The obtained binding site from blind docking was used for docking of all inhibitors. In these dockings, a grid map of the size $60 \times 60 \times 60 \text{ \AA}$ with a grid-point spacing of 0.375 \AA was generated. The maps were centered on the inhibitor's

binding site in a center grid box $8.994, 5.928$ and 12.693 \AA for the x, y and z centers of the receptor, respectively. For conformational search, docking calculations were carried out using the Lamarckian genetic algorithm (LGA) method with default parameters. The residues of the binding site obtained were Ala230, Ala350, Asp351, Gly286, His283, Ile211, Leu260, Leu278, Leu340, Lys232, Ser287, Ser280, Tyr282, Val219 and Val231 in the cavity of the receptor. The LIGPLOT software was employed to show the hydrophobic and hydrogen bonding interactions. These interactions are shown in Figure 1. The hydrogen bonds (Figure 1) are generated between the His238 residue and the inhibitors. Furthermore, in the X-ray crystal structure of inhibitor 61 with ALK5,¹⁷ there is a hydrogen bond between N1 nitrogen of the inhibitor 61 and the backbone NH of His283. His 283 is a key residue in most reported ALK5

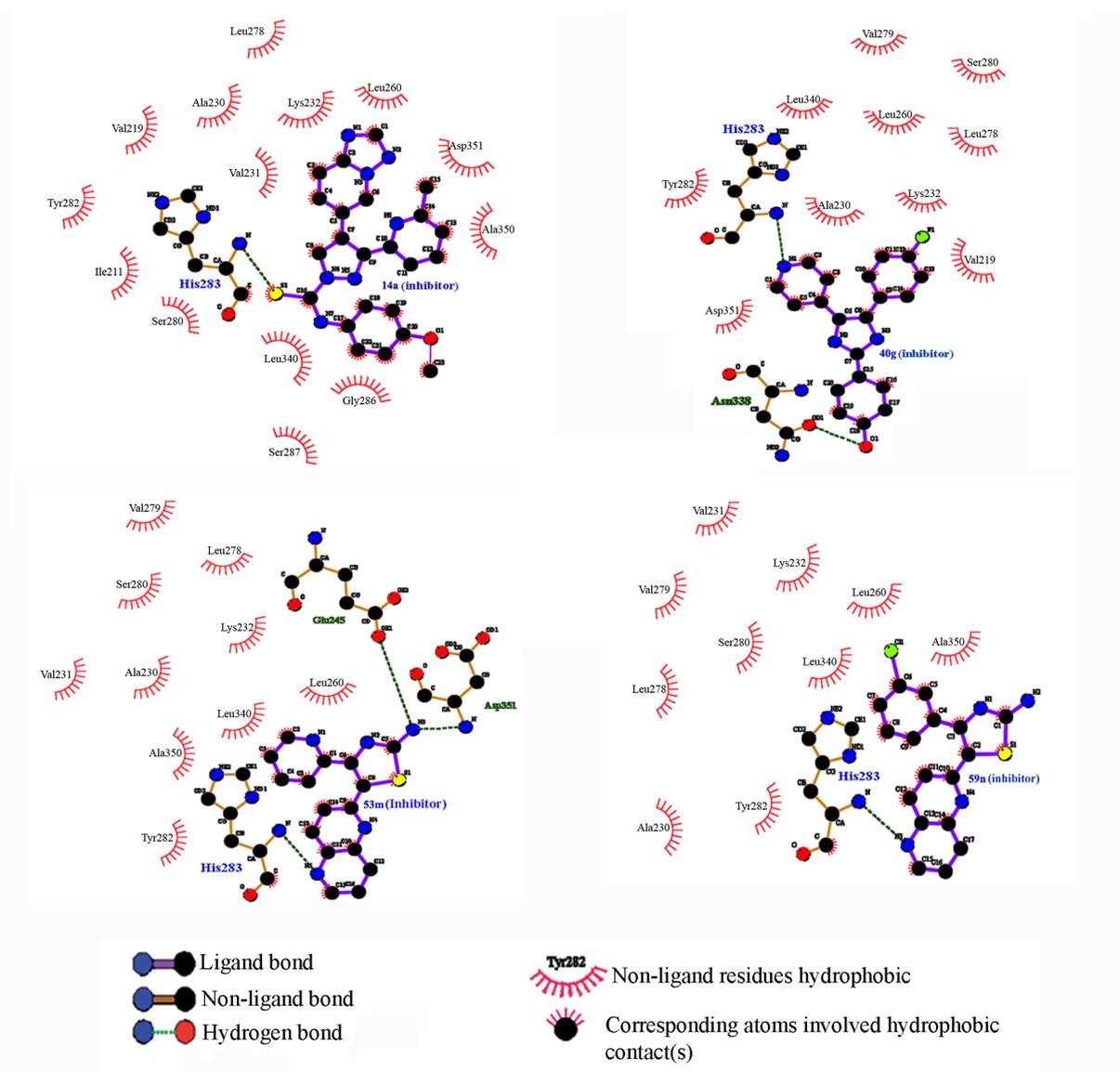


Figure 1. Two-dimensional schemes of interactions between inhibitors of **14a**, **40g**, **53m** and **59n** with ALK5 generated by LIGPLOT.

complexes by X-ray in protein data bank (PDB ID: 2WOT and 2WOU). Therefore, obtained binding site from docking is in agreement with X-ray crystallography. Results of molecular docking indicated that the interactions between the inhibitor and ALK5 are dominated by hydrophobic and hydrogen bond. The hydrophobic interactions are generated by the close contacts between the non-polar amino acid side chains of the receptor and the lipophilic groups of the inhibitor. It may be suggested that more hydrophobic interactions at the binding site could improve the inhibitory activity of the inhibitors.

Molecular dynamics simulation of the ALK5-inhibitor complex

In order to determine the effect of inhibitor on the receptor conformation we decided to perform a MD simulation. The simulation procedure was assessed by carrying out a 6000 ps MD simulation on the **14a**-ALK5 complex in a water box. Trajectory stability was investigated by analyzing RMSD. The average RMSD values of the protein backbone of the complex and ALK5, Figure 2, were calculated to be 0.47 ± 0.05 nm and 0.45 ± 0.03 nm, respectively. Furthermore, to investigate the conformational changes of the receptor, radius of gyration (R_g) values for the complex and ALK5 were also compared. The average values of R_g , Figure 3, for the complex and ALK5 were calculated to be 0.60 ± 0.05 nm and 0.59 ± 0.03 nm, respectively. These results indicate that the receptor conformation has not changed in the presence of inhibitor **14a**. The stability of the receptor conformation in the presence of the inhibitor confirmed the docking results, since the hydrophobic intermolecular interactions between the inhibitor and the receptor are weak interactions and these interactions do not affect the receptor conformation.

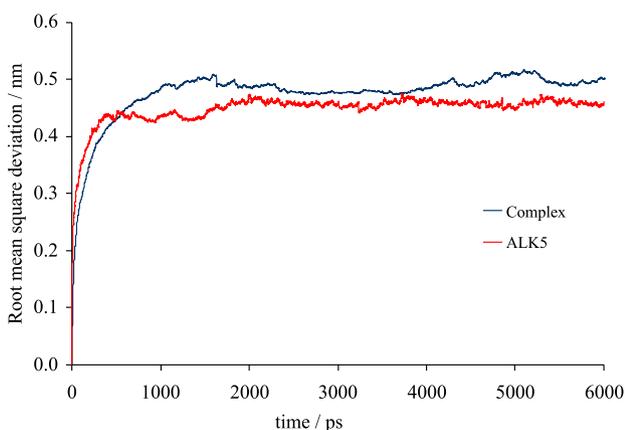


Figure 2. RMSD values of protein backbone for ALK5 and ALK5 complex during 6000 ps MD simulation.

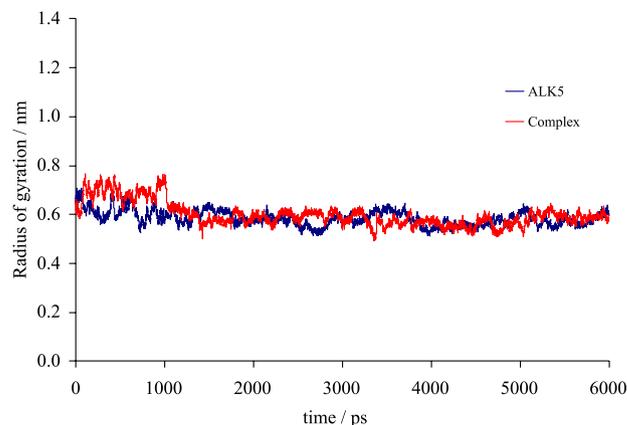


Figure 3. Time evolution of the radius of gyration (R_g) during 6000 ps of MD simulation of ALK5 and ALK5 complex.

Descriptor calculation and selection

As already mentioned, 132 descriptors were calculated for the inhibitors based on inhibitor-receptor interactions using AutoDockTools, BINANA and Hyperchem software. Once the descriptors had been calculated, those that remained constant for all the molecules were eliminated and pairs of variables with a Pearson correlation coefficient > 0.80 in bivariate correlations were classified as intercorrelated, one of which was eliminated. Subsequently, the stepwise multiple linear regression was carried out on the training set for selecting the most effective descriptors by SPSS software (version 13; SPSS Inc.). Since the number of descriptors should be five times less than that of the molecules in the training set,⁴⁴ eight descriptors were selected for constructing the QSAR model.

These descriptors were volume (V), estimated free energy of binding (ΔG), hydrophobic contacts carbon-carbon β -sidechain ($HC\beta$. C-C), summed electrostatic energy acceptor-carbon atoms (SEE. A-C), summed electrostatic energy acceptor-nitrogen atoms (SEE. A-N), summed electrostatic energy nitrogen-accepting nitrogen atoms (SEE. N-NA), summed electrostatic energy acceptor nitrogen-acceptor sulfur atoms (SEE. NA-SA) and number of close contacts between carbon and acceptor oxygen atoms in distance of 4.0 \AA of each other (NCC4. C-OA). The correlation matrix obtained for the selected descriptors are given in Table 2. It can be observed that the correlation coefficient between each pair of the descriptors is less than 0.68.

LS-SVR model

The selected descriptors were used as input for constructing the LS-SVR model. In order to generate

Table 2. The correlation matrix of the selected descriptors

	ΔG	V	NCC4. C-OA	SEE. A-C	SEE. A-N	SEE. N-NA	SEE. NA-SA	HC β . C-C
ΔG	1	-0.678	-0.126	-0.116	-0.232	-0.103	-0.483	-0.112
V		1	-0.356	0.245	0.121	-0.097	0.565	-0.05
NCC4. C-OA			1	0.005	-0.045	-0.163	0.168	0.140
SEE. A-C				1	0.229	-0.013	0.133	-0.223
SEE. A-N					1	-0.121	0.115	-0.405
SEE. N-NA						1	0.007	0.332
SEE. NA-SA							1	-0.155
HC β . C-C								1

a LS-SVR model, an appropriate set of regularization parameters, γ , and kernel parameters such as σ^2 for the radial basis function (RBF) kernel should be chosen. In this work, the radial basis kernel function was used because it is a non-linear function that can decrease the computational complexity of the training procedure.^{45,46} The next step in the construction of the LS-SVR model is optimization of its parameters, including γ and σ^2 , which were found to be 191.4 and 214.8, respectively, using the grid search method. The values of correlation coefficient (R) for training and test set in the LS-SVR model as a non-linear model were found to be 0.929 and 0.917, respectively.

LS-SVR model validation

The generated QSAR model was validated by root mean square error (RMSE) and cross-validation (Q^2) of the test set. RMSE was calculated as in equation 9,

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (y_{\text{ex},i} - y_{\text{pre},i})^2}{N}} \quad (9)$$

where N is the number of data patterns in the independent data set, $y_{\text{pre},i}$ designates the predicted value, and $y_{\text{ex},i}$ is the experimental value of one data point i . The values of RMSE for the training and test set in LS-SVR model were calculated to be 0.453 and 0.560, respectively. As recommended by Tropsha and co-workers,^{47,48} several statistical parameters were used for evaluating the significance and predictive power of QSAR models. A proper and predictive model should have a Q^2 value larger than 0.5 and a correlation coefficient of prediction (r^2) above 0.6. The Q^2 value is calculated from equation 10,

$$Q^2 = 1 - \frac{\sum_{i=1}^{\text{pre}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{\text{pre}} (y_i - \bar{y}_{\text{tr}})^2} \quad (10)$$

where y_i and \hat{y}_i are the measured and predicted values of the dependent variable (over the test set), respectively, and \bar{y}_{tr} is the average value of the dependent variable for the training set. The summation runs over all compounds in the test set.

The calculated statistical parameters of LS-SVR model are presented in Table 3. As can be seen in Table 3, the statistical parameters for the LS-SVR model show the excellent performance of the model. Table 1 presents the predicted values for inhibitory activity. The predicted LS-SVR values were plotted *versus* their experimental values, Figure 4, indicating a good agreement between the experimental and predicted values.

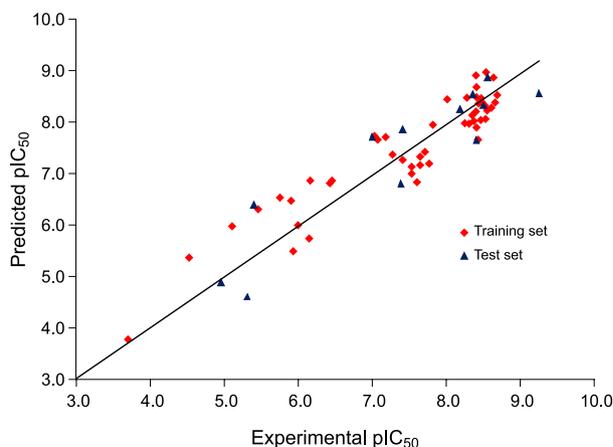


Figure 4. Plot of predicted inhibitory activity *versus* experimental values using LS-SVR model.

Moreover, in order to assess the predictive power of QSAR model, correlation coefficients between predicted and experimental values of inhibitory activity of compounds from an external prediction set (r_p^2), correlation coefficients for regressions through the origin (predicted *versus* experimental inhibitory activities, or experimental *versus* predicted activities, i.e., r_p^2 or $r_p'^2$, respectively) and the slope of the regression lines through the origin (k and k' , respectively) were also calculated. As suggested by Tropsha, a suitable QSAR model should satisfy all

of the following conditions (i) $Q^2 > 0.5$, (ii) $r_p^2 > 0.6$, (iii) r_{op}^2 or r_{op}^2 is close to r_p^2 , such that $[(r_p^2 - r_{op}^2)/r_{op}^2]$ or $[(r_p^2 - r_{op}^2)/r_{op}^2] < 0.1$, (iv) $0.85 \leq k \leq 1.15$ or $0.85 \leq k' \leq 1.15$. These values were calculated to be 0.837, 0.841, 0.009, 0.007, 0.993 and 1.000 for the LS-SVR model.

In addition, another statistic for external validation (r_m^2) was calculated as $r_m^2 = r_p^2 * [1 - (r_p^2 - r_0^2)^{1/2}]$ as suggested by Roy and Roy.⁴⁹ For a model with good external predictability, the value for r_m^2 should be > 0.5 . The value of r_m^2 was found to be 0.766 in the LS-SVR model. The results are given in Table 3. According to recommendation of Tropsha and Golbraikh,^{47,48} if difference between R^2 and Q^2 did not exceed 0.3, there is no over-fitting in the model. In the present work, the difference between R^2 and Q^2 was 0.004, indicating no over-fitting in the LS-SVR model.

Table 3. Statistical parameters of the external test set for LS-SVR model

Parameter	LS-SVR
Q^2	0.837
r_p^2	0.841
$[(r_p^2 - r_{op}^2)/r_{op}^2]$	0.009
$[(r_p^2 - r_{op}^2)/r_{op}^2]$	0.007
r_m^2	0.766
r_m^2	0.776
K	0.993
K'	1.000

Applicability domain

The predictive reliability of LS-SVR model was carried out using applicability domain. In this study, the value of warning leverage was calculated ($h^* = 0.55$). A leverage value greater than h^* for the training set ($h > h^*$) means that the inhibitor seriously influences on the model, while for the test set, it means that the prediction is the result of substantial extrapolation of the model and could not be reliable. To visualize the applicability domain of LS-SVR model, the standardized residuals *versus* leverage values were plotted and the results are shown in Figure 5. A compound with cross-validated standardized residual greater than 3σ (3 standard deviation units) is recognized as Y outlier. Figure 5 shows that all predictions are reliable for LS-SVR model and there is no outlier compound both for training and test sets, indicating the reliability of the predictions.

The results of statistical parameters (Table 3) and applicability domain represented that the LS-SVR model had a high predictive ability and was able to fulfill all the criteria to be accepted as a predictive model for the inhibitory activity.

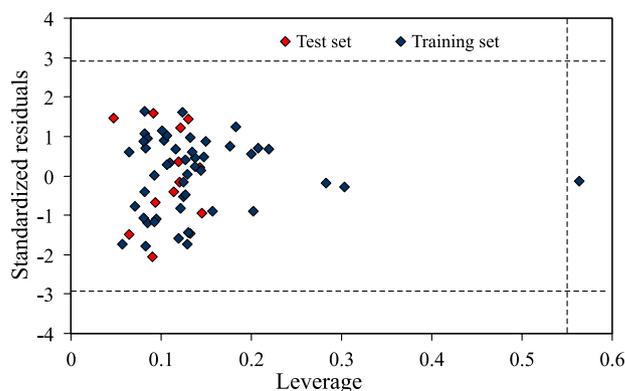


Figure 5. Plot of standardized residuals *versus* leverages. Horizontal dash lines represent $\pm 3\sigma$, vertical dash line represents warning leverage ($h^* = 0.55$).

Sensitivity analysis

In order to determine the relative importance of each descriptor in the LS-SVR model, a developed sensitivity analysis approach was used to rank the importance of the input descriptors of the LS-SVR model.⁵⁰ The performed sensitivity analysis is based on consecutive removal of descriptors by zeroing the specific connection weights for an input descriptor in the LS-SVR model. According to this method, the differences between the root-mean-square error (RMSE) of the complete model's prediction and the RMSE obtained when the i th variable is excluded from the trained model ($RMSE_i$) reveals the importance of the i th input descriptor:

$$Rmddiff = RMSE_i - RMSE \quad (11)$$

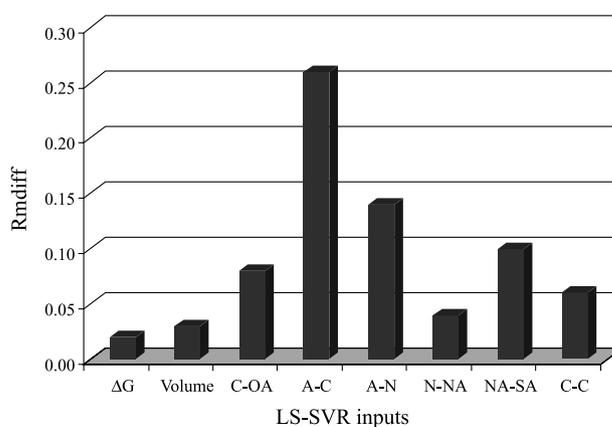


Figure 6. Result of sensitivity analysis for the LS-SVR model.

It is obvious that the most important descriptor is the one that leads to the highest value of Rmddiff. The values of Rmddiff for the LS-SVR model are plotted in Figure 6. As shown in this Figure, the order of importance is SEE. A-C > SEE. A-N > SEE. NA-SA > NCC4. C-OA > HC β .

C-C > SEE. N-NA >V > ΔG. Four descriptors consisting of SEE. A-C, SEE. A-N, SEE. NA-SA and SEE. N-NA are summed electrostatic energies. For each atom-type pair of atoms that come within 4.0 Å of each other, a summed electrostatic energy is calculated using the Gasteiger partial charges assigned by AutoDockTools:

$$V_{(a,b)} = \sum_{(a_i, b_i)} \frac{q_{a_i} q_{b_i}}{r_{a_i b_i}} \quad (12)$$

where $V_{(a,b)}$ is the summed electrostatic interaction energy of all the atoms of types a and b, respectively, q_{a_i} is the partial atomic charge of an atom of type a, q_{b_i} is the partial atomic charge of an atom of type b, and $r_{a_i b_i}$ is the distance between a and b atoms.³⁴ Another descriptor in the model is numbers of atom-type pair counts at a distance of 4.0 Å from C-OA. This descriptor shows the number of contacts between carbon-oxygen atoms of the receptor and the inhibitor located at a distance of 4.0 Å.³⁵ In this distance there are the van der Waals interactions between carbon atoms with oxygen. The Van der Waals attractions operate over only a very limited range of interaction distances between 3 and 6 Å.^{51,52} The next descriptor that represents a significant effect on the inhibitory activity of the inhibitors is hydrophobic contacts (β-sidechain) of carbon-carbon interaction between the receptor and the inhibitors obtained from BINANA. Volume is another important descriptor in interaction between receptor with inhibitor which can affect on the inhibitory activity of inhibitors. The final descriptor in this work is free energy of binding defined as follows:

$$\Delta G_{\text{binding}} = \Delta G_{\text{vdw}} + \Delta G_{\text{elec}} + \Delta G_{\text{hbond}} + \Delta G_{\text{desolv}} + \Delta G_{\text{tors}} + \Delta G_{\text{intermol}} \quad (13)$$

where, ΔG_{vdw} = van der Waals or Lennard-Jones potential, ΔG_{elec} = electrostatic factor with distance-dependent dielectric, ΔG_{hbond} = H-bonding potential with directionality, ΔG_{desolv} = charge-dependent variant of volume-based atomic solvation, ΔG_{tors} = torsional energy based on the number of rotatable bonds and $\Delta G_{\text{intermol}}$ = intermolecular energy of protein and inhibitor molecules.⁵³

Proposed potent inhibitors

New potent inhibitors were proposed based on the inhibitor-receptor interactions and descriptor properties. A new inhibitor was proposed based on the structure of inhibitor a. In this inhibitor, R was replaced by a pyrrole ring in para position, which exhibited the highest activity in this category of the inhibitors. In addition, two new inhibitors

were proposed based on inhibitor b with *substituents* H and F in the R position with the highest activity. The values of predicted pIC₅₀ for these inhibitors are given in Table 1.

Conclusions

In this work molecular docking was carried out to explore the binding site. Docking analysis showed that hydrogen bonding and hydrophobic interactions are important factors in the interactions between the inhibitors and receptor. After docking of the most potent inhibitor with receptor, the residues of the binding site obtained were Ala230, Ala350, Asp351, Gly286, His283, Ile211, Leu260, Leu278, Leu340, Lys232, Ser287, Ser280, Tyr282, Val219 and Val231 in the cavity of the receptor. X-ray crystallography of ALK5-inhibitor 61 complex and ALK5-inhibitor complexes with PDB ID: 2WOT and 2WOU in protein data bank confirmed the results of the docking studies. Molecular dynamics Simulation of the complex demonstrated that the receptor conformation was not changed in the presence of the inhibitor.

Finally, 3 new potent inhibitors were proposed based on the results of the inhibitor-receptor interactions and properties of the descriptors.

Acknowledgments

The authors acknowledge to the Research Council of Isfahan University of Technology and the Center of Excellence in Chemistry of Isfahan University of Technology for the financial support of this work. Our thanks also go to Dr. Ezzatollah Roustazadeh from ELC of IUT for his assistance in editing the final English version of this manuscript.

References

1. Roberts, A. B.; Anzano, M. A.; Lamb, L. C.; Smith, J. M.; Sporn, M. B.; *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78*, 5339.
2. Massague', J.; *Annu. Rev. Biochem.* **1998**, *67*, 753.
3. Plum, J.; De Smedt, M.; Leclercq, G.; Vandekerckhove, B.; *J. Immunol.* **1995**, *15*, 5789.
4. Wahl, S. M.; Swisher, J.; McCartney-Francis, J.; Chen, W.; *J. Leukocyte Biol.* **2004**, *76*, 15.
5. Wu, D. T.; Bitzer, M.; Ju, W.; Mundel, P.; Bottinger, E. P.; *J. Am. Soc. Nephrol.* **2005**, *16*, 3211.
6. Bierie, B.; Moses, H. L.; *Nat. Rev. Cancer.* **2006**, *6*, 506.
7. Rane, S. G.; Lee, J. H.; Lin, H. M.; *Cytokine Growth Factor Rev.* **2006**, *17*, 107.
8. Dong, M.; Blobe, G. C.; *Blood* **2006**, *107*, 4589.
9. Massague, J.; *Nat. Rev. Mol. Cell Biol.* **2000**, *1*, 169.

10. Derynck, R.; Zhang, Y.; Feng, X. H.; *Cell* **1998**, *95*, 737.
11. Yingling, J. M.; Blanchard, K. L.; Sawyer, J. S.; *Nat. Rev. Drug Discovery* **2004**, *3*, 1011.
12. Derynck, R.; Zhang, Y. E.; *Nature* **2003**, *425*, 577.
13. Li, X.; Wang, L.; Long, L.; Xiao, J.; Hu, Y.; Li, S.; *Bioorg. Med. Chem. Lett.* **2009**, *19*, 4868.
14. Ren, J. X.; Li, L. L.; Zou, J.; Yang, L.; Yang, J. L.; Yang, S. Y.; *Eur. J. Med. Chem.* **2009**, *11*, 4259.
15. Geldenhuys, J. W.; Nakamura, H.; *Bioorg. Med. Chem. Lett.* **2010**, *20*, 1918.
16. Jin, C. H.; Krishnaiah, M.; Sreenu, D.; Subrahmanyam, V. B.; Rao, K. S.; Mohan, A. V. N.; Park, C. Y.; Son, J. Y.; Sheen, Y. Y.; Kim, D. K.; *Bioorg. Med. Chem. Lett.* **2011**, *21*, 6049.
17. Gellibert, F.; Woolven, J.; Fouchet, M. H.; Mathews, N.; Goodland, H.; Lovegrove, V.; Laroze, A.; Nguyen, V. L.; Sautet, S.; Wang, R.; Janson, C.; Smith, W.; Krysa, G.; Boullay, V.; de Gouville, A. C.; Huet, S.; Hartley, D.; *J. Med. Chem.* **2004**, *47*, 4494.
18. Yakymovych, I.; Engström, U.; Grimsby, S.; Heldin, C. H.; Souchelnytskyi, S.; *Biochemistry* **2002**, *41*, 11000.
19. Wang, F. F.; Li, Y.; Ma, Z.; Wang, X.; Wang, Y. H.; *J. Mol. Model.* **2012**, *18*, 295.
20. HyperChem; *HyperChem Release 7.0 for Windows*, Hypercube, Inc., Florida, USA, 2002.
21. Goldberg, F. W.; Ward, R. A.; Finlay, R.; Powell, S. J.; Debreczeni, J. E.; Norman, R. A.; Roberts, N. J.; Dishington, A. P.; Gingell, H. J.; Wickson, K.; Roberts, A. L.; *Med. Chem.* **2009**, *52*, 7901.
22. Schuttelkopf, A. W.; Van Aalten, D. M.; *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 1355.
23. Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C.; *J. Comput. Chem.* **2005**, *26*, 1701.
24. Rivail, L.; Chipot, C.; Maignet, B.; Bestel, I.; Sicsic, S.; Tarek, M.; *J. Mol. Struct.* **2007**, *817*, 19.
25. Darden, T.; York, D.; Pedersen, L.; *Chem. Phys.* **1993**, *98*, 10089.
26. Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R.; *Chem. Phys.* **1982**, *76*, 637.
27. Fleagle, R. G.; Businger, J. A.; *An Introduction to Atmospheric Physics*, Academic Press: New York, 1963.
28. <http://www.molegro.com/index.php> accessed in July 2012.
29. Humphrey, W.; Dalke, A.; Schulten, K.; *J. Mol. Graphics Model.* **1996**, *14*, 33.
30. Wallace, A. C.; Laskowski, R. A.; Thornton, J. M.; *Protein Eng.* **1995**, *8*, 127.
31. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J.; *J. Comput. Chem.* **1998**, *19*, 1639.
32. Park, H.; Lee, J.; Lee, S.; *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 549.
33. Garg, A.; Tewari, R.; Raghava, G. P.; *BMC. Bioinformatics* **2010**, *11*, 125.
34. Sanner, M. F.; *J. Mol. Graphics Modell.* **1999**, *17*, 57.
35. Durrant, J. D.; McCammon, J. A.; *J. Mol. Graphics Modell.* **2011**, *29*, 888.
36. Cortes, C.; Vapnik, V.; *Mach. Learn.* **1995**, *20*, 273.
37. Suykens, J. A. K.; Gestel, T. V.; Brabanter, J. D.; Moor, B. D.; Vandewalle, J.; *Least-Squares Support Vector Machines*, World Scientific: Singapore, 2002.
38. Gestel, T. V.; Suykens, J. A. K.; Baesens, B.; Viaene, S.; Vanthienen, J.; Dedene, G.; De Moor, B.; Vandewalle, J.; *Mach Learn.* 2004, *54*, 5.
39. Aizerman, M. A.; Braverman, E. M.; Rozonoer, L. I.; *Automat. Remote Control.* **1964**, *25*, 821.
40. Venkatesan, S. K.; Shukla, A. K.; Dubey, V. K.; *J. Comput. Chem.* **2012**, *31*, 2463.
41. Eriksson, L.; Jaworska, J.; Worth, A. P.; Cronin, M. T. D.; McDowell, R. M.; Gramatica, P.; *Environ. Health. Perspect.* **2003**, *111*, 1361.
42. Tropsha, A.; Gramatica, P.; Gombar, V. K.; *QSAR Comb. Sci.* **2003**, *22*, 69.
43. Gramatica, P.; *QSAR Comb. Sci.* **2007**, *26*, 694.
44. Salt, D. W.; Ajmani, S.; Crichton, R.; Livingstone, D. J.; *J. Chem. Inf. Model.* **2007**, *47*, 143.
45. Gu, T.; Lu, W.; Bao, X.; Chen, N.; *Solid State Sci.* **2006**, *8*, 129.
46. Ensafi, A. A.; Hasanpour, F.; Khayamian, T.; *Talanta* **2009**, *79*, 534.
47. Tropsha, A.; Gramatica, P.; Gombar, V. K.; *QSAR Comb. Sci.* **2003**, *22*, 69.
48. Golbraikh, A.; Tropsha, A.; *J. Mol. Graphics Modell.* **2002**, *20*, 269.
49. Roy, P.; Roy, K.; *QSAR Comb. Sci.* **2008**, *27*, 302.
50. Fatemi, M. H.; Jalali-Heravi, M.; Konuze, E.; *Anal. Chim. Acta* **2003**, *486*, 101.
51. Garret, R. H.; Grisham, C. M.; *Biochemistry*, Marry Frinch: Boston, 2010.
52. Patil, R.; Das, S.; Stanley, A.; Yadav, L.; Sudhakar, A.; Varma, A. K.; *Plos One* **2010**, *25*, 12029.
53. Singh, S.; Gupta, S. K.; Nischal, A.; Khattri, S.; Nath, R.; Pant, K. K.; Seth, P. K.; *Hepat. Mon.* **2011**, *11*, 803.

Submitted: July 14, 2012

Published online: December 5, 2012