Bull. Chem. Soc. Ethiop. **2010**, 24(3), 317-325. ISSN 1011-3924
Printed in Ethiopia

QSAR STUDY OF BENZIMIDAZOLE DERIVATIVES INHIBITION ON ESCHERICHIA COLI METHIONINE AMINOPEPTIDASE

Zahra Garkani-Nejad* and Fereshteh Saneie

Department of Chemistry, Faculty of Science, Vali-e-Asr University, Rafsanjan, Iran

(Received November 19, 2009; revised March 28, 2010)

ABSTRACT. The paper describes a quantitative structure-activity relationship (QSAR) study of IC_{50} values of benzimidazole derivatives on escherichia coli methionine aminopeptidase. The activity of the 32 inhibitors has been estimated by means of multiple linear regression (MLR) and artificial neural network (ANN) techniques. The results obtained using the MLR method indicate that the activity of derivatives of benzimidazoles on Co^{II}-loaded escherichia coli methionine aminopeptidase depend on different parameters containing topological descriptors, Burden eigen values, 3D MoRSE descriptors and 2D autocorrelation descriptors. The best artificial neural network model is a fully-connected, feed forward back propagation network with a 5-4-1 architecture. Standard error for the training set using this network was 0.193 with correlation coefficient 0.996 and for the prediction set standard error was 1.41 with correlation coefficient 0.802. Comparison of the quality of the ANN with different MLR models showed that ANN has a better predictive power.

KEY WORDS: QSAR, Artificial neural network, Multiple linear regression, Molecular descriptors, Escherichia coli methionine aminopeptidase

INTRODUCTION

Methionine aminopeptidases (MetAPs) are target proteins for the development of both anticancer and antibacterial compounds [1]. MetAPs are ubiquitous metal dependent enzymes involved in the N-terminal processing of proteins [2] and have been characterized for *Escherichia coli*, *Salmonella typhimurium*, baker's yeast, humans, and other species [3]. Protein synthesis is normally initiated with the AUG (adenine, uracil, guanine) triplet coding for methionine (in the cytosol of eukaryotes) or N-formyl-methionine [4], (in prokaryotes, mitochondria [5] and chloroplasts). For a significant fraction of the intracellular proteins [6], the amino-terminal methionine is removed by MetAPs after the initiation of translation. This process is mostly dependent on the adjacent residue and essential. Thus MetAPs play a key role in the functional regulation of proteins. The physiological importance of MetAP activity is underscored by the non-viability of organisms where all MetAP genes have been deleted or all MetAP gene products are inhibited. This has been shown for *Escherichia coli*, *Salmonella typhimurium*, and *Saccharomyces cerevisiae*.

Quantitative structure activity relationship (QSAR) studies, as one of the most important areas in chemometrics, give information that is useful for molecular design and medicinal chemistry [7-9]. QSAR models are mathematical equations constructing a relationship between chemical structures and biological activities. These models have another ability, which is providing a deeper knowledge about the mechanism of biological activity. In the first step of a typical QSAR study one needs to find a set of molecular descriptors with the higher impact on the biological activity of interest [10-13]. A wide range of descriptors has been used in QSAR modeling. These descriptors have been classified into different categories, including constitutional, geometrical, topological, quantum chemical and so on. There are several variable selection methods including multiple linear regression (MLR), genetic algorithm (GA), partial least squares (PLS), and so on [11-13]. MLR yields models that are simpler and easier to

*Corresponding author. E-mail: garakani@mail.vru.ac.ir, Tel. +98-391-3202416

interpret than PLS, because these methods perform regression on latent variables that do not have physical meaning. Due to the colinearity problem in MLR analysis, one may remove the collinear descriptors before MLR model development. The quantitative structure-activity relationship studies (QSAR) [14, 15] represent one of the most effective computational approaches for inspection of inhibition mechanism.

For many years, it has been assumed that similar molecules tend to have similar activities, leading to activity landscapes. Very similar molecules may possess very different activities leading to what can be called activity cliffs. An activity cliff is defined by the ratio of the difference in activity of two compounds to their distance of separation in a given chemical space. The existence of such activity cliffs is not entirely surprising since molecular recognition plays a crucial role in determining activity.

In the present work, both artificial neural network (ANN) and multiple linear regression (MLR) techniques have been used for modeling of the observed IC_{50} values of derivatives of benzimidazoles on Co^{II}-loaded escherichia coli methionine aminopeptidase, (MetAps). In this study, 26 compounds are used as training set. For these molecules precise activity data are available. The adequacy of the developed QSAR models was examined using the prediction set which includes 6 compounds. The correlation coefficient (R), standard error (SE) and R^2cv were employed to judge the validity of the models.

EXPERIMENTAL

Biological data

The activity data were taken from reference [16]. The data set for active molecules consist of 32 inhibitors of derivatives of benzimidazoles on escherichia coli methionine aminopeptidase. This set was divided in to a training set and a prediction set. Usually, the molecules included in these sets are selected randomly. In the present work, the training set includes 26 compounds with precise activity data, and the prediction set includes 6 compounds whose imprecise activity data were reported. The chemical names for the training and prediction sets are given in Tables 1 and 2. The values of IC_{50} were used as the dependent variable.

Descriptor generation

The second step in developing the model was the numerical description of the molecular structures by defining descriptors. These descriptors were responsible for encoding important features of the structures. A large number of molecular descriptors were calculated using Hyperchem [17] and Dragon [18] softwares. Some chemical parameters including molecular volume, molecular surface area, hydrophobicity (Log P), hydration energy and molecular polarizability were calculated using Hyperchem software. Dragon software calculated different functional groups, topological, geometrical and constitutional descriptors for each molecule.

Regression analysis

Because of the large number of considered descriptors, a stepwise multiple linear regression procedure based on the forward-selection and backward-elimination methods was used for inclusion or rejection of descriptors in the screened models. In order to avoid of overestimations or difficulties in interpretation of the resulting models, pairs of variables with $r \geq 0.90$ were classified as inter correlating ones, and only one of these was included in the screened models. Many models were generated by using this method. However, an ideal model is one that has high correlation coefficient (R), low standard deviation, least numbers of independent variables,

high ability for prediction and high F statistic value [19]. F statistic shows the mean squares between treatments to the residuals. The best selected MLR model is presented in Table 3.

Artificial neural network approach

The ANN is a computer-based program in which a number of processing elements, also called neurons, are interconnected by links in a netlike structure forming 'layers'. A variable value is assigned to every neuron [20]. The neuron can be one of three different kinds: (i) input neurons, which receive their values by direct assignation and are associated with independent variables, with the exception of the bias neuron, to form the input layer; (ii) hidden neurons, which collect values from other neurons giving a result that is passed to a non-input neuron; and (iii) output neuron, which collect value from other units, and correspond to different dependent variables to form the output layer. The links between units have associated values, named weights that condition the values assigned to the neurons. There exist additional weights assigned to bias values that act as neuron value offsets. The weights are adjusted through a training process in order to minimize network error. Commonly neural networks are adjusted, or trained, so that a particular input leads to a specific target output. The characteristics of the ANN have been found to be suitable for data processing, in which the functional relationship between the input and the output is not previously defined. Structure-activity relationships are often nonlinear and very complex and neural networks are able to approximate any kind of analytical continuous function.

In the present work, MATLAB [21] software package was used for implementing fully connected, three-layer, feed-forward computational neural networks with back-propagation training and non-automatic regularization. The ANN approach was used for obtaining nonlinear models for the activities of the studied benzimidazoles, expecting to improve linear models performance. ANNs used here had variable architectures: (i) the number of neurons in the input layer was equal to the number of variables, (ii) the number of hidden neurons was optimised, and (iii) one neuron was placed in the output layers. Feed-forward network weights and bias values were updated according to gradient descent algorithm with momentum and an adaptive learning rate. Molecular descriptors giving the best linear models were used as networks inputs and the activities as target outputs [22].

Input-layer network with a sigmoidal transfer function was selected. Before the learning network was applied, the input vector and output values were normalized between 0.1 and 0.9. Sigmoidal transfer function has minimum and maximum values of 0 and 1, respectively. The normalizing of output values between 0.1 and 0.9 allows the network to slightly exceed the minimum and maximum values that were given in the original data file.

Testing for chance correlations

Part of validating the models is to check for the possibility of chance correlations. This can be done by performing the entire sequence of computations over but with the dependent variables scrambled. This scrambling destroys any relationship between the descriptors and the dependent variable. No model that exceeds chance performance should be found. The results obtained are compared to the results achieved with the actual computations to demonstrate that the actual results were achieved by finding relationships rather than by finding chance correlations.

RESULTS AND DISCUSSION

Multiple regression analysis

MLR was performed on the compounds described in Tables 1 and 2. We have included all 26 molecules of the training set for the model generation. We used Hyperchem software for drawing the compounds. After screening of the descriptors from Dragon software and submission of some parameters to the regression routine, a few suitable models were obtained.

Table 1. Chemical names along with the observed and calculated IC_{50} values of benzimidazole derivatives on Co^{II}-loaded escherichia coli methionine aminopeptidase (MetAps)–training set.

No.	Compound name	IC_{50}		
		Observed	MLR	ANN
1	Thiabendazole	0.472 ± 0.06	1.17	0.587
2	Carbedazim	8.956 ± 1.319	8.99	8.943
3	2-Pyridin-3yl-1H-benzoimidazol	0.540 ± 0.028	1.69	0.565
4	1H-Benzimidazol-2-yl-thiourea	0.574 ± 0.082	0.36	0.225
5	(1H-Benzimidazol-2-yl)- thiazol-2-yl-amin	1.343 ± 0.274	0.88	0.738
6	2-(1H-Benzoimidazol-2-yl)-pyridin-3-ol	0.777 ± 0.033	1.11	1.082
7	1-Methyl-2-thiazol-4-yl-1H-benzoimidazole	0.497 ± 0.007	0.20	0.485
8	1-Benzyl-2-thiazol-4-yl-1H-benzoimidazole	0.461 ± 0.013	0.14	0.644
9	1-Benzyl-2-pyridin-2-yl-1H-benzoimidazole	0.992 ± 0.13	1.06	0.941
10	5-Methyl-2-thiazol-4-yl-1(3)H-benzoimidazole	7.157 ± 0.065	5.39	7.101
11	5-Methyl-2-pyridin-2-yl-1(3)H-benzoimidazole	2.086 ± 0.286	1.91	1.958
12	5,6-Dimethyl-2-pyridin-2-yl-1H-benzoimidazole	2.433 ± 0.284	2.73	2.499
13	5-tert-Butyl-2-thiazol-4-yl-1H-benzimidazole	3.906 ± 0.582	3.74	3.826
14	5-tert-Butyl-2-pyridin-2-yl-1H-Benzoimidazole	2.598 ± 0.376	2.09	2.531
15	5-Nitro-2-thiazol-4-yl-1(3)H-benzoimidazole	1.218 ± 0.128	1.24	1.219
16	2-Pyridin-2-yl-3H-benzoimidazol-5-ylamine	0.967 ± 0.024	1.55	1.361
17	6-Fluoro-2-pyridin-2-yl-1H-benzoimidazole	1.511 ± 0.192	1.71	1.535
18	5-Chlor-2-thiazol-4-yl-1H-benzimidazol	5.153 ± 1.023	5.35	5.244
19	5-Chloro-2-pyridin-2-yl-1(3)H-benzoimidazole	1.695 ± 0.147	1.84	1.676
20	Phenyl-(2-thiazol-4-yl-1H-benzoimidazol-5-yl)- methanone	2.397 ± 0.470	2.39	2.385
21	2-Pyridin-2-yl-3H-benzoimidazole-5-carbonitrile	0.431 ± 0.018	0.72	0.559
22	2-Thiazol-4-yl-1H-imidazo[4,5-b] pyridine	0.078 ± 0.007	-1.25	0.071
23	2-Pyridin-2-yl-1H-imidazo[4,5-b] pyridine	0.105 ± 0.001	-0.38	-0.061
24	2-Thiazol-4-yl-1H-imidazo[4,5-c] pyridine	1.724 ± 0.158	2.67	1.657
25	2-Pyridin-2-yl-1H-imidazo[4,5-c] pyridine	0.550 ± 0.081	0.75	0.648
26	2-Pyridin-2-yl-1H-imidazo[4,5-b] pyrazine	0.240 ± 0.041	0.75	0.307

Table 2. Chemical names along with the observed and calculated IC_{50} values of benzimidazole derivatives on Co^{II}-loaded escherichia coli methionine aminopeptidase (MetAps)–prediction set.

No.	Compound name	IC_{50}		
		Observed	MLR	ANN
27	2-Pyrazin-2-yl-1H-benzoimidazol	4.591 ± 0.389	2.56	3.116
28	5-Nitro-2-pyridin-2-yl-1(3)H- benzoimidazole	0.162 ± 0.019	-0.42	-0.430
29	5-Amino-2-thiazol-4-yl-1H- benzimidazole	3.390 ± 0.242	3.98	4.872
30	5-Fluor-2-thiazol-4-yl-1H-benzimidazol	4.307 ± 0.635	2.93	3.135
31	Phenyl-(2-pyridin-2-yl-3H-benzoimidazol-5-yl)-methanone	0.403 ± 0.045	1.84	1.116
32	8-Pyridin-2-yl-7(9)H-purine	0.376 ± 0.023	1.81	1.599

SPSS [23] software have been used for model processing and among obtained models, the best one was selected and presented in Table 3. The statistical parameters for this model are shown in Table 5 (model No. 13). Most of descriptors in this model depend on electronegativity [24], polarizability [25] and topological distance. BeHe2, the highest positive eigen value n.2 of the non-diagonal elements in Burden matrix [26], that is weighted by Sanderson electronegativity, derived from BCUT (Burden-CAS-University of Texas eigen values) indices. Burden matrix [27] is a modified connectivity matrix. The non-diagonal elements M_{wk} are 1 if k = d_{ij} and 0 otherwise, where k is the lag defined as the topological distance d between the atom pair i-j and may have a value between 0-8. Thus for a given k, the non-diagonal element M_{ij} will be unity if the atoms i and j are apart by a topological distance k and zero otherwise. As can be seen from Table 3, this descriptor has a negative coefficient in the MLR model and indicates the effect of topology of the compounds on activity.

MATS8e and MATS8p are the Moran autocorrelation descriptors of lag (topological distance) 8 that are weighted by Sanderson electronegativity and polarizability, respectively [28]. Mor31m is the 3D Morse descriptor (molecule representation of structures based on electron diffraction) that is weighted by mass [29]. Regression coefficient for this descriptor is positive and indicates that with increasing of Mor31, the activity increases. The last descriptor is redial centric information index (ICR) derived of topological descriptor. As can be seen from Table 3, selected descriptors in the model depend on topological distances and electronic interactions of benzimidazole derivatives. Of course, the effect of benzimidazole derivatives on escherichia coli methionine aminopeptidase are very complex and obtained model could be used only for this family of compounds. Also, selected descriptors show that electronic parameters such as electronegativity, polarizability, and also, topological distances are important parameters that can affect on inhibitor activities of benzimidazole derivatives.

Table 3. The best MLR model for the prediction of IC_{50} values of benzimidazole derivatives on Co^{II}-loaded escherichia coli methionine aminopeptidase (MetAps).

Descriptor	Notation	Regression coefficient
The highest positive eigen value n.2 of the non-diagonal elements in Burden matrix weighted by Sanderson electronegativity	BEHe2	-30.518
The Moran autocorrelation descriptor of lag 8 (topological distance) weighted by Sanderson electronegativity	MATS8e	-4.568
3D Morse descriptor weighted by mass	Mor31m	4.906
The Moran autocorrelation descriptor of lag 8 (topological distance) weighted by polarizability	MATS8p	5.739
Radial centric information index derived of topological descriptor	ICR	4.198
Constant		105.630

Based on the correlation matrix (Table 4), it can be generalized that there is no significant correlation between the selected descriptors. The colinearity threshold in QSAR/QSPR studies is usually considered 0.9, i.e. descriptors with $R > 0.9$ are selected as collinear.

Table 4. Correlation matrix for the inter-correlation of structural descriptors and their correlation with the activity.

Descriptors	BEHe2	MATS8e	Mor31m	MATS8p	ICR
BEHe2	1				
MATS8e	0.069	1			
Mor31m	0.115	-0.16	1		
MATS8p	0.382	0.137	0.146	1	
ICR	0.722	0.196	0.414	0.487	1

Table 5. Cross-validation compounds and quality of the proposed models for MLR and ANN.

Model No.	Model	Prediction set deleted compounds	MLR			ANN		
			R	R^2cv	Std. Error	R	R^2cv	Std. Error
1	Training	1, 5, 6, 27, 7, 8	0.945	0.893	0.824	0.967	0.932	0.590
	Prediction		0.736	0.541	1.226	0.958	0.898	0.517
2	Training	9, 10, 11, 12, 13, 14	0.930	0.864	0.856	0.954	0.907	0.635
	Prediction		0.958	0.919	0.687	0.992	0.979	0.3105
3	Training	15, 28, 29, 16, 30, 17	0.935	0.873	0.908	0.973	0.944	0.539
	Prediction		0.867	0.752	0.880	0.867	0.690	0.880
4	Training	18, 19, 20, 31, 21, 22	0.939	0.881	0.859	0.958	0.915	0.649
	Prediction		0.835	0.698	1.18	0.980	0.950	0.428
5	Training	23, 24, 25, 26, 32, 2	0.909	0.827	0.846	0.943	0.885	0.617
	Prediction		0.970	0.940	0.944	0.986	0.966	0.641
6	Training	3, 4, 1, 5, 6, 27	0.953	0.909	0.764	0.967	0.932	0.5856
	Prediction		0.512	0.262	1.540	0.945	0.865	0.588
7	Training	6, 27, 13, 28, 29, 30	0.946	0.894	0.791	0.987	0.974	0.351
	Prediction		0.837	0.701	1.163	0.816	0.583	1.228
8	Training	6, 27, 18, 21, 24, 26	0.955	0.911	0.726	0.974	0.947	0.501
	Prediction		0.805	0.649	1.429	0.910	0.785	1.000
9	Training	13, 28, 29, 30, 18, 31	0.930	0.865	0.876	0.975	0.949	0.484
	Prediction		0.891	0.793	1.067	0.891	0.741	1.067
10	Training	6, 27, 13, 28, 24, 25	0.953	0.908	0.757	0.977	0.953	0.487
	Prediction		0.713	0.509	1.483	0.856	0.665	1.095
11	Training	29, 30, 18, 31, 24, 26	0.953	0.908	0.739	0.983	0.964	0.415
	Prediction		0.747	0.558	1.509	0.889	0.738	1.039
12	Training	6, 13, 29, 18, 31, 24	0.934	0.872	0.884	0.974	0.947	0.507
	Prediction		0.977	0.955	0.446	0.921	0.810	0.820
13	Training	27, 28, 29, 30, 31, 26	0.959	0.920	0.693	0.996	0.992	0.193
	Prediction		0.718	0.516	1.64	0.802	0.554	1.41

The consistency and reliability of a method can be explored using the cross-validation technique. The leave-multiple-out (LMO) cross-validation was carried out for MLR and ANN methods. Therefore, M represents a group of randomly selected data points (i.e. 6 molecules) which would be left out at the beginning and would be foretold by the model developed by the use of the remaining data points. Table 5 shows cross-validation results for training and prediction sets along with statistical parameters. The cross validation results confirm the reliability of the selected model.

Artificial neural network analysis

MATLAB version 7.1 was used for implementing fully connected, three-layer, and feed-forward computational neural networks with back-propagation training and non-automatic regularization with a 5-4-1 architecture. Input-layer-network with a sigmoid transfer function was selected. Before the learning network was applied, the input vector and output values were normalized between 0.1 and 0.9. The optimum number of nodes in the hidden layer in this network was 4. For The evaluation of the prediction power of the ANN, the trained ANN was used to predict the IC_{50} values of the molecules included in the prediction set. Molecular descriptors giving the best linear models (MLR) were used as networks input. Standard error for the training set using this network was 0.193 with R = 0.996 and $R^2_{\text{(cross validation)}}$ = 0.992. For the prediction set standard error was 1.41, R = 0.802 and $R^2_{\text{(cross validation)}}$ = 0.554 (model No. 13 in Table 5). The ANN calculated values of the biological activity for the training and prediction sets are shown in Tables 1 and 2. As can be seen from these two tables, the results for the ANN model are better than the MLR model.

Figure 1 shows the plot of experimental IC_{50} values for these compounds against the calculated values. The most of compounds in the training set are on the line. It indicates that the predicted values of IC_{50} are in agreement with the experimental values. But, for the prediction set, whose imprecise activity data were reported, errors are higher than the training set.

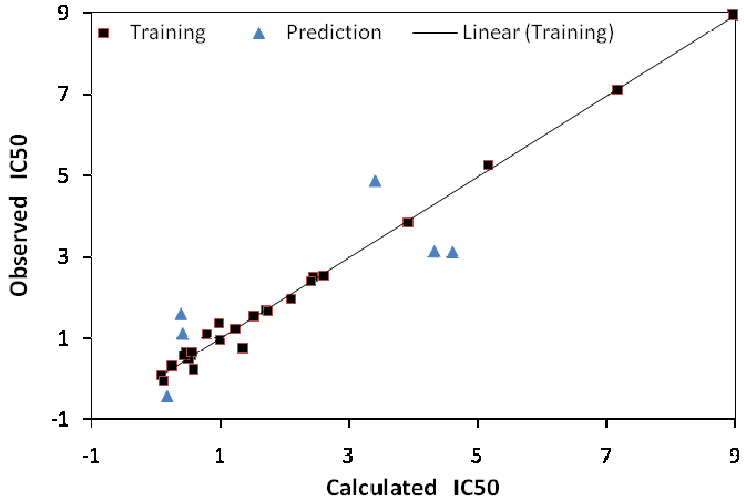

Figure 1. Plot of experimental IC_{50} values against the calculated values of IC_{50}.

Figure 2 shows the plot of residuals versus experimental IC_{50} values of derivatives of benzimidazoles on Co^{II}-loaded escherichia coli methionine aminopeptidase (MetAps). The propagation of residuals on both sides of zero indicates that no systematic error exists in the development of the neural network.

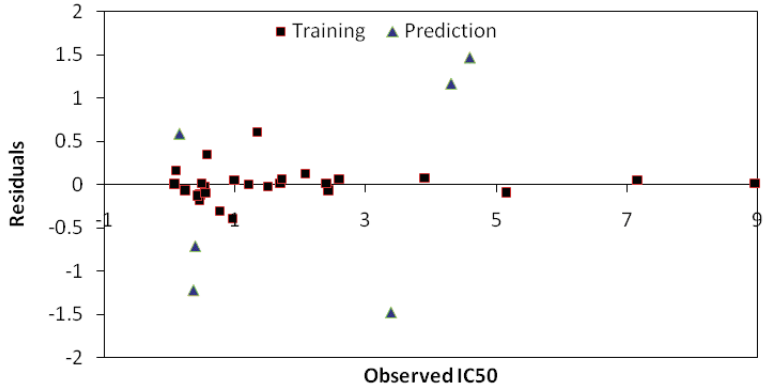

Figure 2. Plot of residuals versus experimental IC_{50} values.

From the above discussion, it is indicated that the molecular descriptors of drugs are key factors in influencing the value of IC_{50}. The selected five descriptors are the most important descriptors for the construction of QSAR model and the prediction of IC_{50} with satisfied results.

In order to ensure the robustness of the ANN model, the Y-randomization test was performed in this contribution. The dependent variable vector (IC_{50}) was randomly shuffled and a new QSAR model was developed using the original independent variable matrix. The new QSAR model is expected to have low R^2 and SE values. Several random shuffles of the Y-vector were performed and the results are given in Table 6. The R^2 and SE values indicate that the good results for the ANN model are not due to a chance correlation or structural dependency of the training set.

Table 6. Regression coefficient (R^2) and SE values for Y-randomization tests.

Iteration	SE	R^2
1	3.851	0.148
2	4.276	0.092
3	2.669	0.285
4	2.801	0.230
5	3.598	0.167

CONCLUSIONS

In the present study, first a set of descriptors was calculated to build a QSAR model able to describe the activity of 32 benzimidazole derivatives. Chemometric methods were successfully used for modeling and predicting the inhibitor activity of these compounds. An artificial neural network provides an accurate QSAR model. In addition, ANN model is able to detect relationships between depend (IC50) and independent (descriptors) variables.

The results of this work indicate that the ANN is a promising tool for a nonlinear approximation. ANN offers a method for solving complex technological and scientific problems. It is a good approach for predicting the expected activity of drugs and aiding in drug design.

REFERENCES

1. Dummitt, B.; Micka, William, S.; Chang, Y. *J. Biol. Chem.* **2005**, 280, 14356.
2. Meinnel, T.; Serero, A.; Giglione, C. *J. Biol. Chem.* **2006**, 387, 839.
3. Liao, Y.; Jeng, J.; Wang, C.; Wang, S.; Chang, S. *Protein Sci.* **2004**, 13, 1802.
4. Timmer, I.; Enoksson, M.; Wild Fang, E.; Zhu, W.; Igarashi, Y.; Denault, J.; Ma, Y.; Dummitt, B.; Chang, Y.; Eroshkin, A.; Tao, W.; Salvesen, G. *J. Biochem.* **2007**, 407, 41.
5. Costantini, P.; Jacotot, E.; Decaudin, D.; Kroemer, G. *J. Natl. Cancer Inst.* **2000**, 92, 1042.
6. Schuetz, R.; Kuepfer, L.; Sauer, U. *Mol. Syst. Biol.* **2007**, 3, 1.
7. Schmidi, H. *Chemom. Intell. Lab. Sys.* **1997**, 37, 125.
8. Hansch, C.; Kurup, A.; Garg, R.; Gao, H. *Chem. Rev.* **2001**, 101, 619.
9. Wold, S.; Trygg, J.; Berglund, A.; Antii, H. *Chemom. Intell. Lab. Syst.* **2001**, 58, 131.
10. Horvath, D.; Mao, B. *QSAR. Comb. Sci.* **2003**, 22, 498.
11. Putta, S.; Eksterowicz, J.; Lemmen, C.; Stanton, R. *J. Chem. Inf. Comput. Sci.* **2003**, 43, 1623.
12. Gupta, S.; Singh, M.; Madan, A.K. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 272.
13. Consonni, V.; Todeschini, R.; Pavan, M. *J. Chem. Inf. Comput. Sci.* **2002**, 42, 693.
14. Winkler, D.A. *Briefings in Bioinformatics* **2002**, 3, 73.
15. Guha, R.; Serra, J.R.; Jurs, P.C.; *J. Mol. Graph. Model.* **2004**, 23, 1.
16. Schiffmann, R.; Neugebauer, A.; Klein, Christian, D. *J. Med. Chem.* **2006**, 49, 511.
17. *Hyperchem software*, Version 7.0., Hyper Cube Inc.: Florida; **2002**.

18. Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M. Software Dragon ; 2003; available at http://disat.unimib.it/chm/Dragon.htm
19. Jalali-Heravi, M.; Parastar, F. *J. Chem. Inf. Comput. Sci.* **2000**, 40, 147.
20. Basheer, A.; Hajmeer, M.*; J. Microbiol. Methods* **2000**, 43, 3.
21. *MATLAB*, Mathworks Inc.; Version 7.1.; **2005**; available at www.mathworks.com.
22. Caballero, J.; Zampini, F.M.; collina, S.; Fernandez, M. *Chem. Biol. Drug. Des.* **2007**, 69, 48.
23. *SPSS*, Version 10; **2004**; available at http://www.spssscience.com.
24. Varekova, R.S.; Jirouskova, Z.; Vanek, J.; Suchomel, S.; Koca, J. *Int. J. Mol. Sci.* **2007**, 8, 572.
25. Carrasco, R.; Padron, A.; Galvez, J. *J. Pharm. Sci.* **2004**, 7, 19.
26. Burden, R. *J. Chem. Inf. Comput. Sci.* **1989**, 29, 225.
27. Pearlman, S.; Smith, M. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 28.
28. Eroglu, E.; Turkmen, H. *J. Mol. Graph. Model.* **2007**, 26, 701.
29. Saize-Urra, L.; Gonzale, M.P.; Teijeira, M. *Bioorg. Med. Chem.* **2007**, 14, 7347.