ISSN : 0974 - 7532 Volume 7 Issue 8

Trade Science Inc.

Research & Reviews in

BioSciences

Regular Paper

RRBS, 7(8), 2013 [295-311]

Quantitative structure - Activity relationships study of carbonic anhydrase inhibitors using multinomial logistic regression model and artificial neural networks

Hassan Sahebjamee*[1], Parviz Abdolmaleki[2], Alireza Foroumadi[3], Parichehre Yaghmaei[1]

[1]Department of Biology, Science and Research Branch, Islamic Azad University, Tehran, (IRAN)

[2]Department of Biophysics, Faculty of Science, Tarbiat Modares University, P.O. Box: 14115/175, Tehran, (IRAN)

[3]Department of Medicinal Chemistry, Faculty of Pharmacy, Tehran University of Medical Sciences, Tehran, (IRAN)

E-mail: sahebjamei.hassan386@gmail.com

ABSTRACT

Multinomial logistic regression (MLR) and artificial neural networks (ANNs) were employed to seek the quantitative structure – activity relationships (QSARs) that correlate structural descriptors and inhibition activity of carbonic anhydrase IX inhibitors. Many quantitative descriptors (n=644) were generated to express the physicochemical properties of 132 compounds with optimized structures with known ki values. MLR were used to nonlinearly select different subsets of descriptors and develop nonlinear models for prediction of log (ki). The most significant parameters were then selected. A neural network model was then constructed and fed by the parameters selected by MLR. The networks have been trained and tested using the best subset selected by MLR. The best prediction model was found to be a 5-3-3 artificial neural network which was fed by the most frequently selected descriptors among these subsets. Cross-validation and a separate prediction set were used to evaluate the stability and prediction ability of the established models. Our results demonstrated that descriptors correlated to autocorrelations, topological properties were major determinants of inhibition activity of these compounds. Both methods were able to significantly describe and predict the CAIX inhibitory activity.

INTRODUCTION

Carbonic anhydrases (CAs, EC 4.2.1.1) catalyze the interconversion between carbon dioxide and the bicarbonate ion, and are thus involved in vital physiological processes[1]. At least four CA isozymes (CA IV, CA IX, CA XII, and CA XIV) are associated to cell membranes, with the enzyme active site generally oriented extracellularly. Some of these isozymes were shown to play crucial physiological roles such as CA IX may serve as markers for tumors and hypoxia[2-4]. The carbonic anhydrase inhibitors are widely studied in the last decade, due to their potential functions for the avoidance and treatment of a large number of diseases[5-7]. Supuran and his group have presented a great amount of work in this field[8,9]. So far, two main classes of CA inhibitors are recognized as the metal complexing anions and the unsubstituted sulfonamides. They bind to the Zn (II) ion of the enzyme by substituting the non-protein zinc ligand or by adding to the metal coordination sphere to generate trigonal-bipyramidal species[1]. ulfonamides, as the most important CA inhibitors, bind in the tetrahedral geometry of Zn (II) ion in deprotonated state; whereas metal-complexing anions are weak CAIs, with affinities generally in the millimolar range. The aromatic side chains of sulfonamide inhibitors interact with the hydrophobic amino acid residues in the binding site e.g. Phe131, Leu141, Val143, and Ala45 and stabilize

Regular Paper

the interaction[10].

The application of computational methods for designing biologically active compounds has recently introduced a new approach to modern drug discovery research. Computational methods can move forward the procedure of discovering new drugs by designing new compounds and predict their potency or activity. Different kinds of statistical models such as regression analyses and artificial intelligence based models such as neural networks can be used as techniques for feature selection to develop QSAR models. Indeed predictive ability of such models can be applied to estimate activity of the designed molecules before their synthesis and assay in laboratory. In this way, the research cost can be minimized. QSAR models are mathematical equations that provide comprehensive knowledge on the mechanism of biological activity of compounds by establishing a relationship between chemical structures and their biological activities. Molecular descriptors play an important role in developing QSAR models and finding a set of molecular features affect the biological activity of interest is the essential part of modeling procedure in QSAR analyses[11]. Molecular descriptors are determined through the production of the features which are numerical values corresponding to topological, geometric, constitutional or and quantum chemical features[10]. The derived relationships between molecular descriptors and activity are used to estimate the property of other molecules and/or finding the parameters affecting the biological activity.

Carbonic anhydrase inhibitors were studied by many authors through quantitative structure-activity relationships[12-16]. In this study, we used Multinomial Logistic Regression (MLR) and artificial neural networks (ANNs) as nonlinear models to search the QSAR benzenesulfonamides derivatives as inhibitors CAIX. Initially, we used MLR for selecting more effective descriptors and to obtain an equation for prediction of inhibition activity. Then, according to these results, we developed neural networks with different input patterns. MLR compares multiple groups through a combination of binary logistic regressions. The group comparisons are equivalent to the comparisons for a dummy-coded dependent variable, with the group with the highest numeric score used as the reference group. The mathematical adaptability of ANNs acclaims them as a powerful tool for pattern classification and building predictive models. A particular preference of ANNs is their ability to incorporate nonlinear dependencies between the dependent and independent variables without using a distinct mathematical function. There are, of course, a number of standard nonlinear techniques but one advantage of ANNs is that the form of the non-linear relationship does not need to be specified in advance. A disadvantage of the ANN approach is that it is difficult, perhaps impossible, to extract the relationship created in the modeling[17]. An ANN is formed from artificial neuron arranged in layers, linked with coefficients (or weights), which makes the neural structure. Neural networks do not need explicit formulation of the mathematical or physical relationships of the handled problem, which gives ANNs an advantage over traditional fitting methods for some chemical applications.

In this article, we used a large number of sulfonamide CAIX inhibitors to establish QSAR models with the predictive ability for the activity CAIX inhibitors.

MATERIALS AND METHODS

Data set

Structures of all compounds were drawn in HyperChem (Hypercube Inc.) software. Geometrical optimization was then performmed using the semi-empirical method of Austin Model 1 (AM1)[18]. Constitutional descriptors and topological indices were calculated utilizing Dragon software created by the Milano QSAR and Chemometrics Research Group (www.disat.unimib.it/chm/). In addition, Dragon calculates a large number of descriptors from the optimized three dimensional structures of the molecules. The total 644 descriptors extracted from compounds that were too many to be fitted in our models. So we had to reduce the number of descriptors through an objective feature selection which was performed in three steps. First, descriptors that had the same value for at least 80% of compounds within the dataset were removed. Next step, descriptors with correlation coefficient less than 0.3 with the dependent variable log (ki) were removed from the database. Finally, since highly correlated descriptors provide approximately identical information, performing a pair wise correlation and if their correlation coefficient exceeded 0.80, one of two descriptors was randomly removed. After these three steps, the number of descriptors was reduced to 45.

Although in many compounds, the sulfonamides core (Figure1) has undergone small structural changes, it is conserved in some others. The structures of all compounds are shown in Figure 2. Also the experimental index of log (ki) was reported for every compound as a measure of inhibition. The greater is this value, the weaker is the inhibition activity of the compound. In other words, models could categorize the compounds into active and medium and weak classes. For this purpose, compounds were labeled as active for range log (ki) between 1.4-15 and medium log (ki) 16-70 and weak log (ki) 72-500. According to this classification, 43 compounds were active and medium and 46 were weak. The list of the chemical name and value of their inhibitive ability in decadic logarithm of *K*I (in nM) of 132 sulfonamide compounds-that were used for model development-taken from the literature[19-25] are given in TABLE 1.

A set of 22 compounds were randomly removed from the dataset to be used as the prediction set (PSET). The remaining 110 compounds were used as the training set (TSET). The jackknife test, also called leave-one-out cross-validation (LOOCV), was applied to train and testing the linear discriminated and other models on the database. Through the jackknife procedure, one case (here called testing case) is left out from the database and the training procedure is done using the remaining cases; then the testing case is examined by the obtained model. This procedure is repeated until all cases are tested. As many simulations as the number of samples are made in each database and all cases are used in both the training and testing processes. In order to improve the results of the model and also to achieve a single equation, average coefficients of parameters were calculated and used to form a new equation. After testing with both jackknife equations and the single equation, a cut-off value was taken to recode the obtained values into two possible states of the dependent variable.

Software

A Pentium IV personal computer (CPU at 2.4 GHz) with windows XP operating system was used. Geometry optimization was performed by Hyperchem (version 7.0 Hypercube, Inc.) at the Austin model 1 (AM1). Dragon software was used for calculation of constitutional, topological, geometrical, and functional group descriptors. SPSS Software (SPSS Inc., Version 18) was used for the simple MLR analysis. ANN was performed in the MATLAB environment.

The most significant parameters were then selected

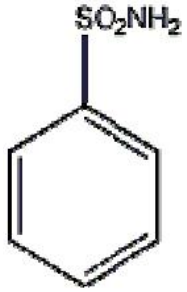


Figure 1: benzenesulfonamides

TABLE 1: Biological data OUT values and real activity class of compounds and the prediction of the model.

Compound	IC50(nm)[a]	real activity class[b]	OUT values class 1and3	Predicted activity class1and3	OUT values class 2and3	Predicted activity class2and3
10	10.30	1.00	1.16	1		
11	4.50	1.00	5.44	1		
11aa	6.70	1.00	4.74	1		
11bb	5.60	1.00	7.29	1		
11cc	4.80	1.00	5.84	1		
11dd	6.70	1.00	4.43	1		
11r	3.00	1.00	7.79	1		
12	6.30	1.00	7.83	1		
12aa	5.40	1.00	5.57	1		
12bb	4.80	1.00	7.89	1		
12cc	5.00	1.00	6.65	1		
12dd	6.40	1.00	5.36	1		
12ee	5.00	1.00	7.48	1		
12ff	5.20	1.00	6.61	1		

Regular Paper

Compound	IC50(nm)[a]	real activity class[b]	OUT values class 1and3	Predicted activity class1and3	OUT values class 2and3	Predicted activity class2and3
12gg	4.60	1.00	11.95	1		
13	4.40	1.00	6.26	1		
13aa	6.10	1.00	6.85	1		
14	8.00	1.00	-0.01	0		
19	2.80	1.00	1.16	1		
1a	6.40	1.00	2.71	1		
1b	6.00	1.00	2.44	1		
1c	4.90	1.00	2.39	1		
1d	6.60	1.00	2.69	1		
1e	5.40	1.00	6.38	1		
1f	3.50	1.00	4.63	1		
11s	12.00	1.00	4.2	1		
2	1.40	1.00	-3.03	0		
21	3.70	1.00	1.16	1		
22	5.20	1.00	1.16	1		
23	14.10	1.00	1.16	1		
2a	6.10	1.00	1.68	1		
2b	5.90	1.00	1.42	1		
2d	6.40	1.00	1.68	1		
2s	14.00	1.00	-0.38	0		
4	9.70	1.00	1.16	1		
6	10.30	1.00	1.16	1		
8	4.00	1.00	1.16	1		
8a	14.00	1.00	-0.37	0		
8d	6.00	1.00	-0.31	0		
8e	8.00	1.00	-4.34	0		
8q	8.00	1.00	-0.27	0		
9	9.20	1.00	1.16	1		
pentaf~1	15.00	1.00	-1.21	0		
11j	52.00	2.00			.16	.00
11k	37.00	2.00			-1.13	.00
11n	26.00	2.00			.44	.00
11p	21.00	2.00			1.19	1.00
11q	18.00	2.00			1.19	1.00
11y	24.00	2.00			-1.44	.00
12y	39.00	2.00			-3.49	.00
15y	38.00	2.00			1.02	1.00
16y	34.00	2.00			.82	1.00
17y	20.00	2.00			-4.36	.00
18y	31.00	2.00			1.80	1.00
19y	24.00	2.00			2.75	1.00
1m	40.00	2.00			1.19	1.00
1y	33.00	2.00			-7.51	.00
20y	16.00	2.00			-2.95	.00

Regular Paper

Compound	IC50(nm)[a]	real activity class[b]	OUT values class 1and3	Predicted activity class1and3	OUT values class 2and3	Predicted activity class2and3
24	27.90	2.00			-1.86	.00
25	35.20	2.00			-2.45	.00
25y	22.00	2.00			-3.88	.00
26	47.50	2.00			-1.42	.00
26y	26.00	2.00			-4.82	.00
2e	57.00	2.00			1.39	1.00
2f	63.00	2.00			.76	1.00
3	18.40	2.00			-4.84	.00
6b	48.00	2.00			-4.25	.00
6c	43.00	2.00			-5.67	.00
6y	33.00	2.00			-3.88	.00
7	18.10	2.00			-4.84	.00
7a	38.00	2.00			1.87	1.00
7b	42.00	2.00			1.47	1.00
7c	54.00	2.00			.43	.00
7d	26.00	2.00			1.89	1.00
7e	29.00	2.00			.40	.00
7h	64.00	2.00			-1.73	.00
7q	35.00	2.00			1.81	1.00
8b	31.00	2.00			1.92	1.00
8c	49.00	2.00			.98	1.00
8h	37.00	2.00			-.85	.00
8i	70.00	2.00			-1.58	.00
bza	47.00	2.00			.30	.00
dcp	50.00	2.00			-1.52	.00
ind	24.00	2.00			-.67	.00
sulthi~1	43.00	2.00			-2.04	.00
10y	285.00	3.00	-255.00	.00	-255.00	.00
11a	135.00	3.00	.21	.00	.21	.00
11b	112.00	3.00	-85.00	.00	-85.00	.00
11c	106.00	3.00	-80.00	.00	-80.00	.00
11d	83.00	3.00	-52.00	.00	-52.00	.00
11e	139.00	3.00	-96.00	.00	-96.00	.00
11f	79.00	3.00	-39.00	.00	-39.00	.00
11g	136.00	3.00	-101.00	.00	-101.00	.00
11h	73.00	3.00	-48.00	.00	-48.00	.00
11i	113.00	3.00	1.75	1.00	1.75	.00
24y	121.00	3.00	-88.00	.00	-88.00	.00
2y	238.00	3.00	-197.00	.00	-197.00	.00
3s	146.00	3.00	-117.00	.00	-117.00	.00
3y	294.00	3.00	-247.00	.00	-247.00	.00
4y	305.00	3.00	-276.00	.00	-276.00	.00
5y	103.00	3.00	-74.00	.00	-74.00	.00
6a	165.00	3.00	-141.00	.00	-141.00	.00
6d	178.00	3.00	-146.00	.00	-146.00	.00

Regular Paper

Compound	IC50(nm)[a]	real activity class[b]	OUT values class 1and3	Predicted activity class1and3	OUT values class 2and3	Predicted activity class2and3
6e	160.00	3.00	-132.00	.00	-132.00	.00
6f	280.00	3.00	-249.00	.00	-249.00	.00
6g	450.00	3.00	-311.00	.00	-311.00	.00
6h	500.00	3.00	-341.00	.00	-341.00	.00
6i	500.00	3.00	-372.00	.00	-372.00	.00
6j	500.00	3.00	-366.00	.00	-366.00	.00
6k	500.00	3.00	-375.00	.00	-375.00	.00
6m	72.00	3.00	-42.00	.00	-42.00	.00
7f	230.00	3.00	-187.00	.00	-187.00	.00
7g	100.00	3.00	-70.00	.00	-70.00	.00
7i	79.00	3.00	-41.00	.00	-41.00	.00
7j	85.00	3.00	-48.00	.00	-48.00	.00
7k	80.00	3.00	-46.00	.00	-46.00	.00
7m	135.00	3.00	-97.00	.00	-97.00	.00
7n	500.00	3.00	-390.00	.00	-390.00	.00
7o	120.00	3.00	-89.00	.00	-89.00	.00
7p	106.00	3.00	-86.00	.00	-86.00	.00
7y	245.00	3.00	-208.00	.00	-208.00	.00
8f	205.00	3.00	-175.00	.00	-175.00	.00
8g	89.00	3.00	-57.00	.00	-57.00	.00
8j	84.00	3.00	-55.00	.00	-55.00	.00
8k	78.00	3.00	-45.00	.00	-45.00	.00
8m	120.00	3.00	-95.00	.00	-95.00	.00
8n	500.00	3.00	-380.00	.00	-380.00	.00
8o	95.00	3.00	-66.00	.00	-66.00	.00
8p	81.00	3.00	-43.00	.00	-43.00	.00
8y	264.00	3.00	-225.00	.00	-225.00	.00
9y	269.00	3.00	-214.00	.00	-214.00	.00

[a]IC50 is an experimental index reported in nanomolars (nM), represent the inhibition activity of the molecule toward carbonic anhydrase.

[b]Compounds were regarded active (coded as 1) for IC50<15 and medium (coded as 2) for IC50<70 and weak (coded as 3) for IC50<500.

using multinomial logistic regression model. The rational underlying this study was to use MLR to build the most effective set of parameters which then were fed into a well-established neural network. MLR analysis of molecular descriptors was carried out using the stepwise strategy in SPSS.

Model development and evaluation

In the first stage, MLR serves as a non-linear model on the dataset to select significant parameters through the 'Self-consistency Test'. This test is an examination for the self-consistency of a prediction method. Then the ANNs, which act non-linearly in the last stage, were fed by the outputs of MLR to predict the activity CAIX inhibitors. The log (ki) of the compounds was used as the dependent variable in model development. Also the independent variables in each model were selected some quantitative descriptors. In neural network based QSAR models, we performed a leave 10 out procedure (cross-validation) to avoid any possible bias in selecting testing set individuals. The structures of all neural networks were optimized for minimum root mean square error (RMSE) as a performance benchmark.

Multinomial logistic regression model

The used multinomial logistic regression model is a

generalization of the logistic regression model. It is commonly used for data in which the dependent variable is polytomous, and independent variables are numerical or categorical predictors. As the binary dependent variable can always be interpreted as the occurrence or non-occurrence of characteristic, the logistic regression model is an expression of the form

$$\log(\Pr/1-\Pr) = \beta_0 + \sum_{i=1}^{n} \beta_i x_i, \quad (1)$$

where β_0 is the intercept and the β_i's denote the unknown logistic regression coefficients of x_i parameters; also Pr denotes the probability that characteristic will occur. The quantity on the left side of Equation (1) is called a logit. The model can be generalized in the case where the dependent variables unlike a binary logistic regression model, have more than two categories. For such a simple model, a multinomial logistic regression model with logit link can be represented as

$$\log\left(\frac{\Pr(c)}{\Pr(0)}\right) = \beta_0(c) + \sum_{i=1}^{4} \beta_i(c) x_i, \quad c = 1-4 \quad (2)$$

In this model, the same independent variable appears in each of the c categories, and the separate intercept, $\beta_0(c)$, and slopes (or logit coefficients), $\beta_i(c)$, are usually estimated for selected parameters in each contrast. A way to interpret the effect of independent variables, x_i, on the probability of being in category c, is to use predicted probabilities, Pr (c), for different values of x_i:

$$\Pr(c) = \frac{\exp\left(\beta_0(c) + \sum_{i=1}^{n} \beta_i(c) x_i\right)}{1 + \sum_{k=1}^{4} \exp\left(\beta_0(k) + \sum_{i=1}^{n} \beta_i(k) x_i\right)} \quad (3)$$

Then, the probability of being in the reference category, '0' (Type IV), can be calculated by subtraction:

$$\Pr(0) = 1 - \sum_{k=1}^{4} \Pr(k) \quad (4)$$

The class with the highest probability is the final prediction[26].

Artificial neural network model

Artificial neural networks (ANNs) are powerful non-algorithmic models used vastly for classifying different types of data. Being trained from a number of samples, a neural network will be capable of drawing non-linear boundaries to put the new unobserved samples into relevant classes. In this way, the selected variables from multinomial logistic regression model were used as input nodes for the ANN. this is supposed to reduce the number of input nodes, simplify the network structure and shorten the model building time.

We used feed-forward with backpropagation algorithm to train our networks. Using this algorithm, descriptors of training cases are fed into the network. The final outputs (O_{jk}) estimated by the network are compared with the real class of the cases (T_{jk}), producing a sum square error:

$$SSE = \sum_{K=1}^{M} \sum_{J=1}^{N} (T_{jk} - O_{jk})2 \quad (5)$$

Where M is the number of training cases and N is the number of output neurons. SSE is propagated back into the network to adjust the weights. the training cases are tested with new weights and the process is repeated. Through such process, the SSE is minimized[27]. We used the SSE as an index of network efficiency in optimizing the number of hidden neurons in networks. To do so, the number of hidden neurons was changed in every network in order to develop networks generating the minimal SSE. Finally, after such optimizing procedure, the number of hidden layer units reached[15].

We used three layer networks. Each unit in the input layer was fed by one independent variable which has been selected by multinomial logistic regression model. The final neural network architecture was consisted of 22 units in input layer, 15 units in hidden layer and 3 units in output layer. The activation function of hidden layer units was logsig. Training has been performed for 15000 epochs. The value of the learning rate parameter has been set to 08. The software used to build the neural networks was in-house written in the MATLAB programming language.

Artificial neural networks are powerful non-linear models used vastly for classifying different types of data. A neural network is composed of few layers of neurons. Neurons in adjusting layers are connected with relative quantitative weights. These weights are randomly chosen, and then are changed through the training procedure, so that the sum-square error (SSE) is minimized. We used three layer networks. Each neuron in the input layer was fed by one independent variable. A bias neuron was added to the input as well as hidden layers to avoid network collapse.

We used networks with a 5-x-3 structure. To obtain the best classification results, the number of neurons in the hidden layer and other parameters of the ANN structure, including learning rates, training function and training epochs were optimized through a trial-

Regular Paper

and-error procedure. Each neuron in the network was connected to all neurons in neighboring layer (s) through adjustable weights. Network training is the process of adjusting such weights somehow that the error is minimized. The number of input layer neurons is equal to the number of descriptors. We had three output layer neuron, while the number of hidden layer neurons was a matter of optimization. It is generally said that the ratio of training pairs to the whole network weights should be between 1 and 3[28]. Since the numbers of input and output neurons are constant, the approximate number of hidden neurons can be calculated using this rule. It is said that if the ratio is less than 1, then the network simply memorizes the train set or in other words, gets over-trained; while if it exceeds 3, then the network fails to find a relationship between dependent and independent variables. The descriptor values were first ranged between zero and one in order to ensure that some descriptors are not weighted more heavily than others due to their nature. The first layer only fed network with the descriptors, while in hidden and output layers, a sigmoid function acted on summation of incoming weights.

The data set was divided into two subsets: training (75%) and test sets (25%). the test set is used to test the trend of the prediction accuracy of the model trained at some point of the training process. then, the training set was used to optimize the network performance. the training function 'trainscg' in MATLAB was used to train the network. the rationale for just using these descriptors is their small number, which provides us with the best classifier. models through avoiding variable redundancy and overfitting problem of the network. Overfitting problem or poor generalization capability happens when a neural network overlearns during the training period. The optimal network was trained and tested using jackknife method.

The flexibility of ANN enables it to discover more complex relationships in experimental data, when it is compared with the traditional statistical models[29].

Model evaluation

Models were evaluated using some statistical indices. For calculating such indices, N_{TP} (true positive predictions), N_{TN} (true negative predictions), N_{FP} (false positive predictions), and N_{FN} (false negative predictions) were counted. The first index is called fraction correct (FC)[30], which shows the fraction of compounds correctly classified:

$$FC = ((N_{TN} + N_{TP})/N_{total}) \times 100 \quad (6)$$

False alarm rate (FAR)[30] represents the fraction of inactive compounds that were wrongly classified. A high FAR value increases the risk of detecting fake inhibitors by the model. So this is an important index to be noted when using an inhibition prediction model:

$$FAR = (N_{FP}/(N_{TN} + N_{TP})) \times 100 \quad (7)$$

Probability of detection (POD)[30] is another index representing the fraction of active compounds being truly classified. A high value of this index, guarantees not missing any active compound by the model; that is, it causes as many active compounds as possible to be classified correctly in the active class. POD is defined as:

$$POD = (N_{TP}/(N_{FP} + N_{TP})) \times 100 \quad (8)$$

This study clarified the efficiency of using the statistical model of multinomial logistic regression as a preprocessor in determining effective parameters. Moreover, the optimal structure of neural network can be simplified by a preprocessor in the first stage, thereby reducing the needed time for neural network training procedure in the second stage and the probability of over fitting occurrence decreased and a high precision and reliability obtained in this way[31].

RESULTS AND DISCUSSION

QSAR using MLR

Descriptors are divided into groups such as constitutional, topological, geometrical. Constitutional descriptors are related to the number of atoms and bonds in each molecule. Topological descriptors include valence and non valence molecular connectivity indices calculated from the hydrogen-suppressed formula of the molecule, encoding information about the size, composition, and the degree of branching of a molecule. Geometrical descriptors are calculated from 3-D atomic coordinates of the molecule and comprise moments of inertia, molecular volumes, molecular surface areas, and gravitation indices[32].

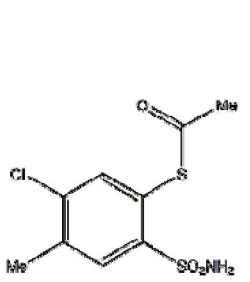
In order to obtain a unified equation, we ignored the misclassified cases. Then, we used the averages of coefficients and constant values of equations suggested for predicting the remaining cases in jackknife proce-

dure. The following equations were obtained:

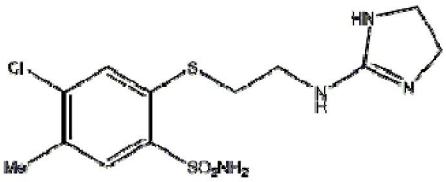
Y13 = -61+123× MSD+3.3× TI2+137.5× JGI4-20.5 MATS3e **(9)**

Y23= -32.5+71.6× MSD+2.1× TI2+66×JGI4+3× nDB +7.7×C028 -18× MATS4V **(10)**

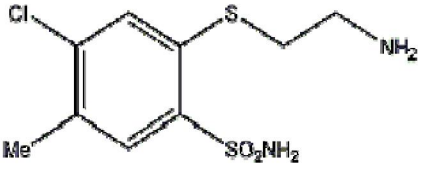
Since the descriptors with greater coefficients are more determining in regression equations, we can conclude that according to this equation, the most important descriptor are MSD and JGI4 and the least determining one is TI2 and nDB and C028. According to this equation, the most important descriptors among all the 2D-autocorrelation ones are weighted by the atomic mass, electronegativity, and van der Waals volumes. Topological descriptor helps to differentiate the molecules according mostly to their size, degree of branching, flexibility and overall shape. MSD, Mean square distance index (balaban) is contributing positively to the activity, which suggests that substituents have smaller branching will improve inhibitory activity[33-34]. *JGI4* is belongs to Galvez topological charge indices, which evaluate the charge transfer between pairs of atoms and hence the global charge transfer in the molecule. C-028 is the second descriptor, appearing in the model. It is one of the atom -centered fragment descriptors that describe each atom by its own atom type and the bond types and atom types of its first neighbors. The C-028 descriptor displays R-CR-X. This atom centered fragment descriptor is defined for each ring atom that has three neighbors. In this case, R-CR-X can be defined as a central carbon atom (C) on an aromatic ring that has one carbon neighbor (R) and one heteroatom neighbor (X) on the same aromatic ring and the third neighbor outside this ring is a carbon (R). The C-028 mean effect has a positive sign. Hence, it was concluded that by increasing the number of heteroatom (with R-CR-X format) in molecules the value of this descriptor increased. TI2 is Topological Second Mohar index. The number of double bonds (nDB) is equal to the number of non-



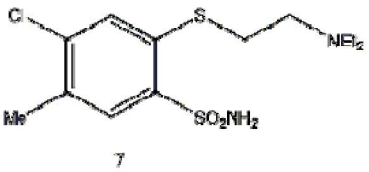
2



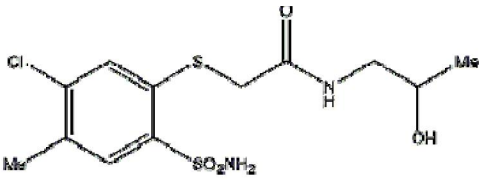
4



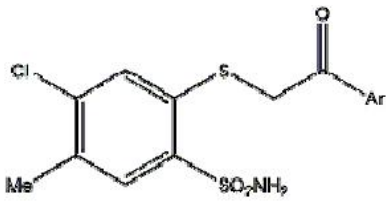
3



7

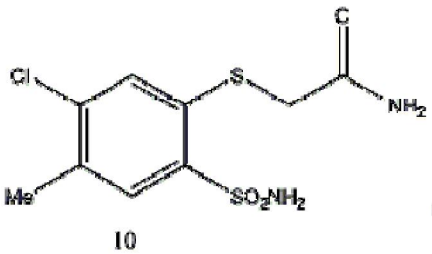


6

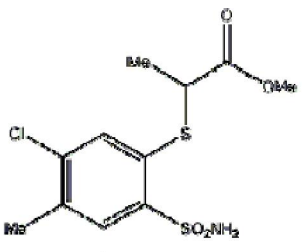


11-14

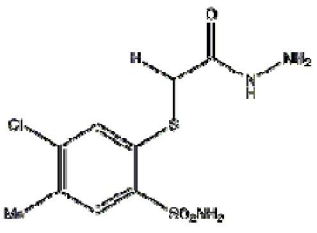
Compound	Ar
11	Ph
12	3-O_2NPh
13	4-O_2NPh
14	3,4-di-ClPh



10

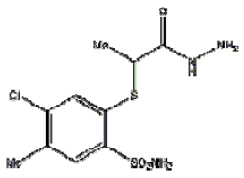


19

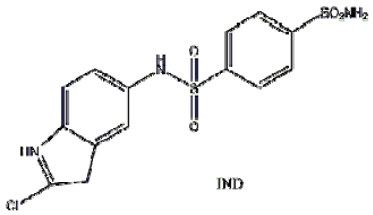


21

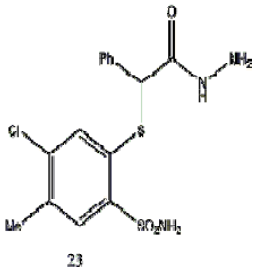
Regular Paper



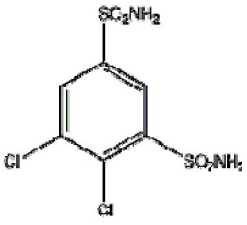
22



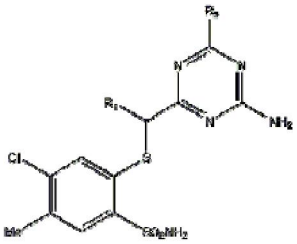
IND



23

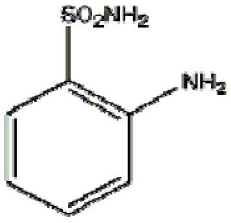


DCP



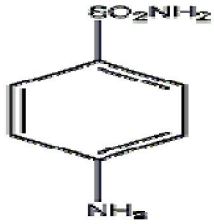
24-26

Compound	R_1	R_2
24	H	3,5,5-trimethyl-2-yrazolino
25	Me	3,5,5-trimethyl-pyrazolino
26	H	dimethylamino

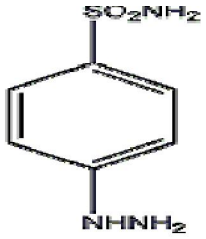


1y

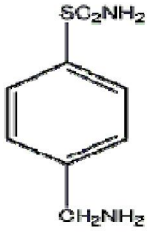
2y



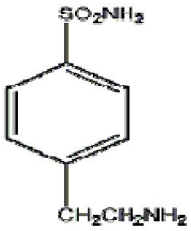
3y



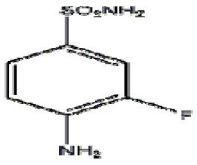
4y



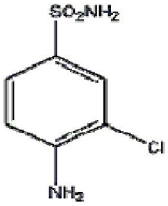
5y



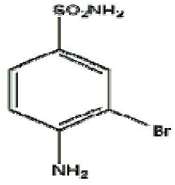
6y



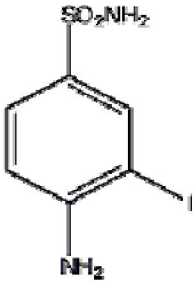
7y



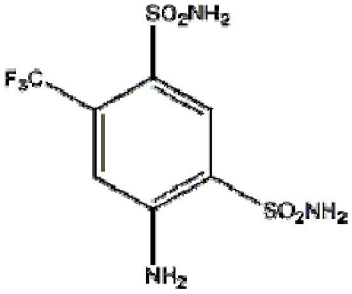
8y



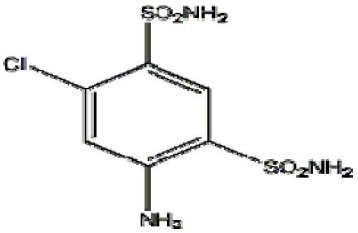
9y



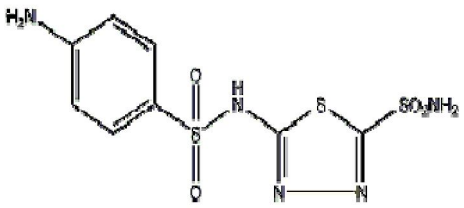
10y



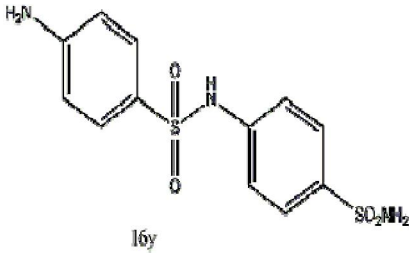
11y



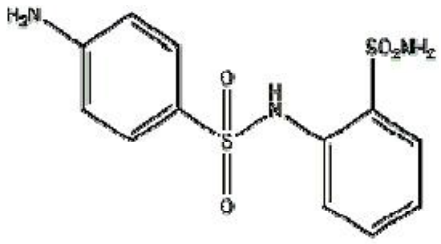
12y



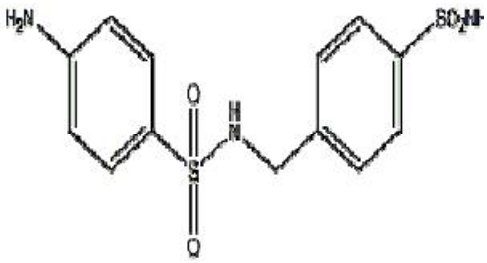
15y

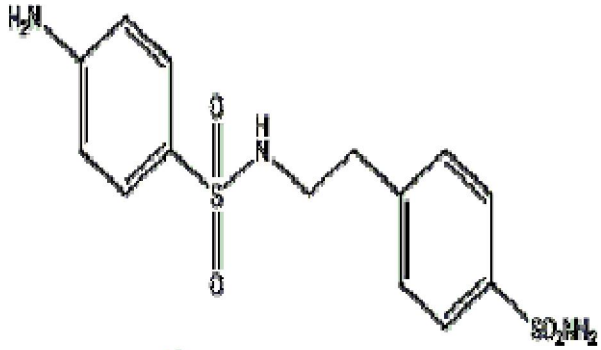


16y

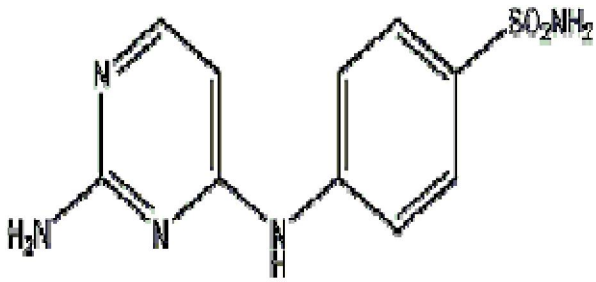


17y

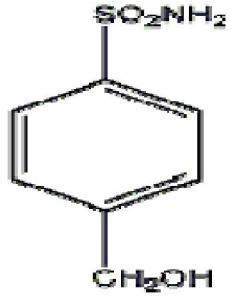




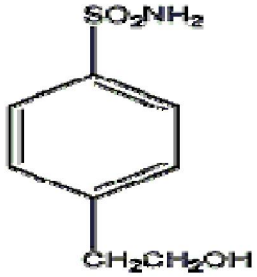
19y



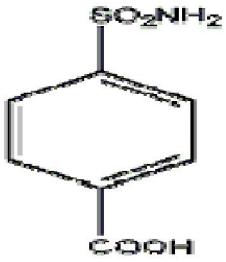
20y



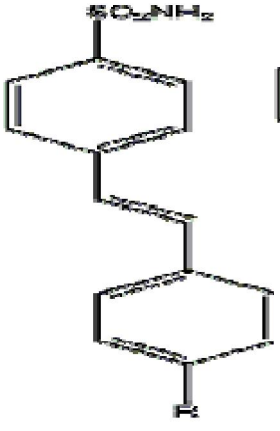
24y



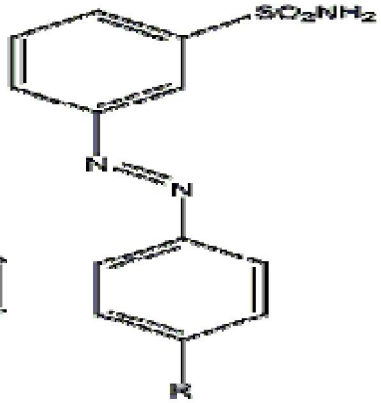
25y



26y

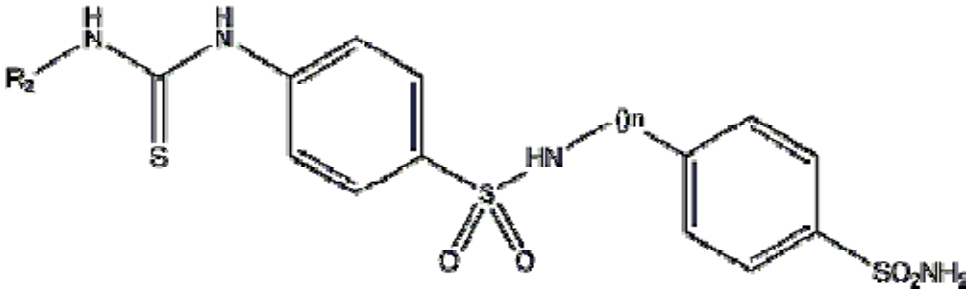


1



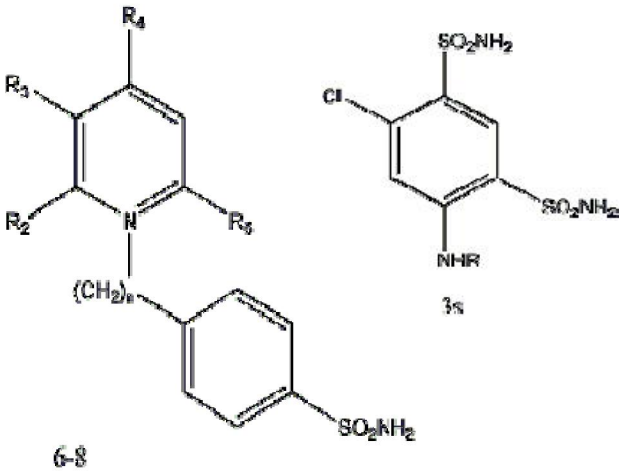
2

No.	R	CAIX
1a	OH	6.4
1b	NH2	6
1c	NHMe	4.9
1d	NMe_2	6.6
1e	$NHCH_2SO_3Na$	5.4
1f	$N(Me)CH_2SO_3Na$	3.5
2a	OH	6.1
2b	NH_2	5.9
2d	NMe_2	6.4
2e	$NHCH_2SO_3Na$	57
2f	$N(Me)CH_2SO_3Na$	63



Regular Paper

Inhibitor	R_2	CAIX
11aa	$Me_2NCH_2CH_2$	6.7
11bb	$[O(CH_2CH_2)_2N]$ CH_2CH_2	5.6
11cc	Me-$[N(CH_2CH_2)_2N]$	4.8
11dd	$[O(CH_2CH_2)_2N]$	6.7
12aa	$Me_2NCH_2CH_2$	5.4
12bb	$[O(CH_2CH_2)_2N]$ CH_2CH_2	4.8
12cc	Me-$[N(CH_2CH_2)_2N]$	5
12dd	$[O(CH_2CH_2)_2N]$	6.4
12ee	$PhCH_2CH_2$	5
12ff	2-Pyridyl- CH_2	5.2
12gg	4-$H_2NO_2SC_6H_4CH_2CH_2$	4.6
13aa	$Me_2NCH_2CH_2$	6.1



Compound	n	R	CAIX
11a	0	CH_3CO	135
11b	0	CF_3CO	112
11c	0	EtCO	106
11d	0	n-PrCO	83
11e	0	i-PrCO	139
11f	0	n-BuCO	79
11g	0	t-BuCO	136
11h	0	PhCO	73
11i	0	$MeSO_2$	113
11j	0	$PhSO_2$	52
11k	0	4-$AcNHC_6H_4SO_2$	37
11m	1	$PhSO_2$	40

Compound	R_2	R_3	R_4	R_6	CAIX
6a	Me	H	Me	Me	165
6b	Me	H	Ph	Me	48
6c	Et	H	Ph	Et	43
6d	n-Pr	H	Ph	n-Pr	178
6e	i-Pr	H	Ph	i-Pr	160
6f	Me	H	Ph	Ph	280
6g	Et	H	Ph	Ph	450
6h	n-Pr	H	Ph	Ph	>500
6i	i-Pr	H	Ph	Ph	>500
6j	n-Bu	H	Ph	Ph	>500
6k	Ph	H	Ph	Ph	>500
6m	Me	Me	Me	Me	72
7a	Me	H	Me	Me	38
7b	i-Pr	H	Me	Ph	42
7c	i-Pr	H	Me	i-Pr	54
7d	Me	H	Ph	Me	26
7e	Et	H	Ph	Et	29
7f	n-Pr	H	Ph	n-Pr	230
7g	i-Pr	H	Ph	i-Pr	100
7h	Me	H	Ph	Ph	64
7i	Et	H	Ph	Ph	79
7j	n-Pr	H	Ph	Ph	85
7k	i-Pr	H	Ph	Ph	80
7m	n-Bu	H	Ph	Ph	135
7n	t-Bu	H	Ph	Ph	>500
7o	Ph	H	Ph	Ph	120
7p	Ph	H	H	Ph	106
7q	Me	Me	Me	Me	35
8a	Me	H	Me	Me	14
8b	i-Pr	H	Me	Me	31
8c	i-Pr	H	Me	i-Pr	49
8d	Me	H	Ph	Me	6
8e	Et	H	Ph	Et	8
8f	n-Pr	H	Ph	n-Pr	205
8g	i-Pr	H	Ph	i-Pr	89
8h	Me	H	Ph	Ph	37
8i	Et	H	Ph	Ph	70
8j	n-Pr	H	Ph	Ph	84
8k	i-Pr	H	Ph	Ph	78
8m	n-Bu	H	Ph	Ph	120
8n	t-Bu	H	Ph	Ph	>500
8o	Ph	H	Ph	Ph	95
8p	Ph	H	H	Ph	81
8q	Me	Me	Me	Me	8

Compound	n	R	CAIX
11n	1	PhNH-C(=S)	26
11p	2	PhNH-C(=S)	21
11q	2	PhNH-C(=O)	18
11r	2	4-$H_2NO_2SC_6H_4NH$- C(=S)	3
11s	2	4-$H_2NO_2SC_6H_4CO$	12
2s	-	PhNH-C(=O)	14
3s	-	PhNH-C(=O)	146

Regular Paper

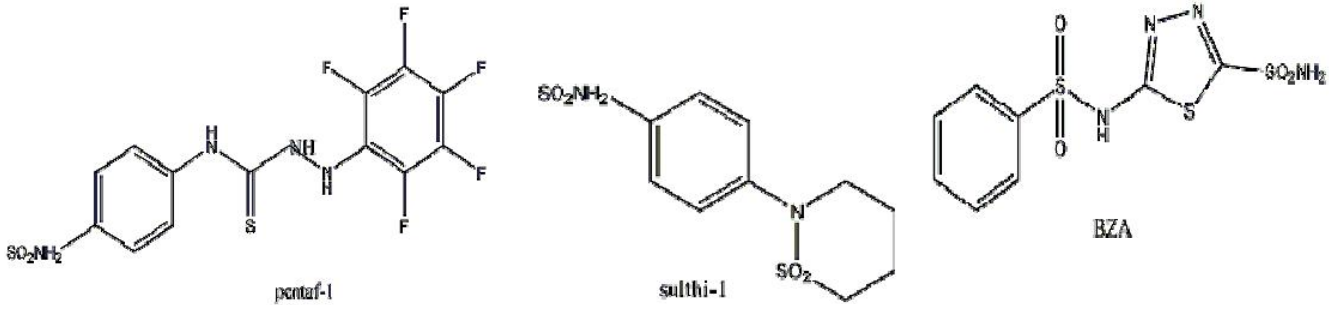


Figure 2 : Chemical structures of compounds used in our dataset.

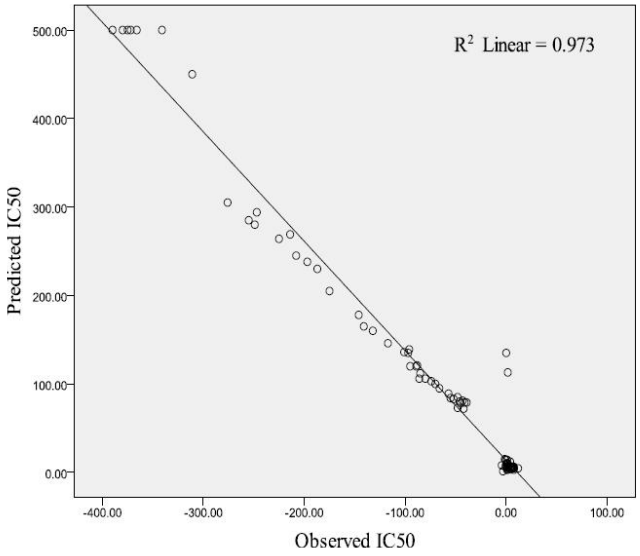


Figure 3 : Plot of predicted versus observed IC50 for class 1, 3

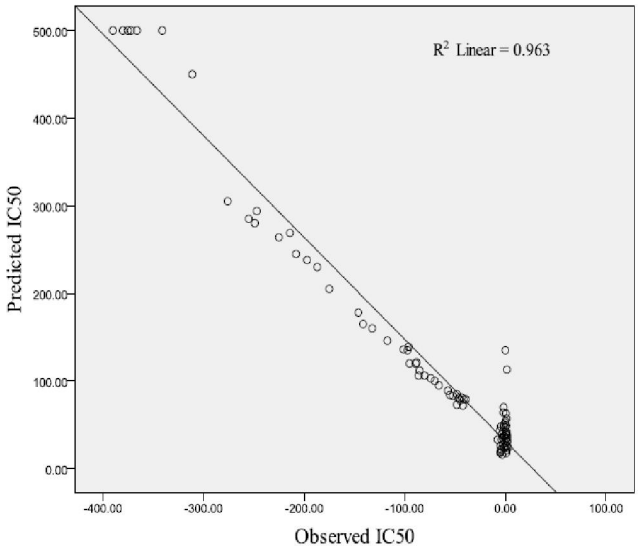


Figure 4 : Plot of predicted versus observed IC50 for class 2, 3

aromatic double bonds. MATS3e describes the autocorrelation of the atomic electronegativities by Moran autocorrelation of lag 3 weighted by atomic Sanderson electronegativities (*MATS3e*) and Moran autocorrelation of lag 4 weighted by van der Walls volumes (*MATS4v*). Similarly *MAST4v* the path connecting a pair of atoms has length 4 and involves the atomic van der Waals volumes as weighting scheme. Figure 3, 4 shows the plot of observed versus predicted Ki for both the training data and the test set.

QSAR using ANNs

We started with a network that was supplied by two descriptors that were selected by, multinomial logistic routine. This was a 2-x-3 network (two input neurons, x hidden and three output neuron). Then we added other descriptors and continued with two 3-x-3 as well as a 4-x-3 networks. Finally, we finished our model building with a 5-x-3 network using other two descriptors (MATS3e and MATS4V) that were supposed to be less effective. As can be seen from this TABLE 2, evaluating results of the networks showed that the networks supplied with five descriptor, relatively offered good predictive ability except MATS4V and MATS3e descriptors. When MATS4V and MATS3e were added to the input layer of this network the statistical indices were deteriorated, that supported the assumption expressing the deteriorate of MATS4V and MATS3e descriptors in their influence of the inhibition activity of compound. The last network was constructed in order to ensure the logistic judgment about the descriptors. As a MLR and ANN, the structure and activity of compounds were most effectively related by MSD and TI2 and JGI4 for class1, 3 and MSD and TI2 and JGI4 and nDB and C028 for class2, 3, respectively.

As was mentioned in Materials and methods, we totally invented 15 networks. The number of neurons in hidden layer was optimized in each network regarding the least SSE rate. The indices FC, FAR, and POD resulted by networks are separately illustrated in TABLE 3.

The optimal cutoff for the OUT values was found to be 0 for class 1, 3 and -.11 for class 2, 3. So, the compounds with predicted OUT < 0 were regarded as weak

Regular Paper

TABLE 2 : Evaluating the results obtained by several neural networks

Networks with different inputs[a]	FC (%)	FAR (%)	POD (%)
MSD-TI2-JGI4-nDB- C028-MATS4V-MATS3e	Class1,3=92 Class2,3=77	Class1,3=6 Class2,3=7	Class1,3=89 Class2,3=84
MSD-TI2-JGI4-nDB-MATS4V-MATS3e	Class1,3=91 Class2,3=75	Class1,3=6 Class2,3=7	Class1,3=88 Class2,3=83
MSD-TI2-JGI4- C028-MATS4V-MATS3e	Class1,3=89 Class2,3=78	Class1,3=3.7 Class2,3=4	Class1,3=92 Class2,3=9
MSD-JGI4-nDB- C028-MATS4V-MATS3e	Class1,3=88 Class2,3=71	Class1,3=8 Class2,3=3	Class1,3=85 Class2,3=92
MSD-JGI4-MATS4V-MATS3e	Class1,3=66 Class2,3=59	Class1,3=25 Class2,3=30	Class1,3=66 Class2,3=58
MSD-TI2-JGI4-MATS4V-MATS3e	Class1,3=84 Class2,3=82	Class1,3=12 Class2,3=12	Class1,3=80 Class2,3=80
TI2-JGI4-nDB- C028-MATS4V-MATS3e	Class1,3=92 Class2,3=75	Class1,3=2.4 Class2,3=2	Class1,3=95 Class2,3=92
MSD-TI2-nDB- C028-MATS4V-MATS3e	Class1,3=92 Class2,3=73	Class1,3=4 Class2,3=6	Class1,3=90 Class2,3=85
MSD-TI2-JGI4-nDB- C028-MATS3e	Class1,3=94 Class2,3=80	Class1,3=2 Class2,3=2	Class1,3=95 Class2,3=93
MSD-TI2-JGI4-nDB- C028-MATS4V	Class1,3=92 Class2,3=82	Class1,3=3 Class2,3=4	Class1,3=92 Class2,3=90
MSD-TI2-JGI4-nDB- C028	Class1,3=93 Class2,3=80	Class1,3=4 Class2,3=5	Class1,3=91 Class2,3=88
MSD-TI2-JGI4-Ndb	Class1,3=88 Class2,3=70	Class1,3=7 Class2,3=9	Class1,3=86 Class2,3=79
MSD-TI2-JGI4- C028	Class1,3=80 Class2,3=76	Class1,3=2 Class2,3=2	Class1,3=93 Class2,3=92
TI2-nDB- C028	Class1,3=74 Class2,3=59	Class1,3=12 Class2,3=15	Class1,3=77 Class2,3=65
MSD-TI2-JGI4	Class1,3=89 Class2,3=69	Class1,3=9 Class2,3=11	Class1,3=83 Class2,3=76

[a]**The networks differ in input neurons. Regarding the obtained results, we could conclude that MATS4V, MATS3e do not seem to good effects on the network decision. So these two descriptors were found to be the less reliable ones in predicting the inhibition activity of compounds.**

TABLE 3 : Prediction results obtained of two models.

Test	Performance measures	class1,3	class2,3
Multinomial logistic regression			
	FC (%)	86	65
	FAR (%)	1	11
	POD (%)	100	87
Neural networks			
	FC (%)	79	62
	FAR (%)	17	21
	POD (%)	78	74

and with OUT > 0 as active and OUT > -.11 as medium. The model with FC=86% FAR = 1%, POD = 100% for class1, 3 and FC=65% FAR = 11%, POD = 87% for class2, 3 were a very good predictive tool.

The sulfonamide derivatives used in this study belong to a wide variety of molecular family containing similar number of sulfonamide groups, similar number of aromatic rings, and similar number of heterocyclic rings. Names, types and definition of the descriptors suggested and used in the final model are shown in TABLE 4. Synthesis and inhibition assay of a larger number of inhibitors, under the same experimental conditions as the stud-

TABLE 4 : Definition of the finally selected set of descriptors

Name	Type	Description	Refrence
MSD	Topological indices	mean square distance index (Balaban)	[33,37]
TI2	2D matrix-based descriptors	second Mohar index	[33,37]
JGI4	2D autocorrelations	mean topological charge index of order 4	[33,37]
nDB	Constitutional indices	Number of double bonds	[33,37]
C028	R--CR--X	Atom-centred fragments	[33,37]
MATS3e	2D autocorrelations	Moran autocorrelation of lag 3 weighted by Sanderson electronegativity	[33,37]
MATS4v	2D autocorrelations	Moran autocorrelation of lag 4 weighted by van der Waals volume	[33,37]

ied compounds, will also be helpful in this context.

QSAR studies provide deeper insight into the mechanism of action of compounds that ultimately becomes of great importance in modification of the structure of compounds. In addition, QSAR studies also provide quantitative models, which permit prediction of activity of compounds prior to the synthesis[35]. Although the resulted models were not capable to estimate the exact value of ki for compounds, they had the ability to classify the compounds into three active and medium and weak classes efficiently. The study proved the capability of MLR and neural network to deal with this problem. The first method was easy and fast, and it correctly selected more efficient descriptors. On the other hand, the artificial neural network outperformed the Multinomial Logistic Regression method through establishing the non-linear association between evaluated descriptors and IC50.

In order to be comparable with results of similar QSAR attempts, we reported RMSE values in normal range of log (ki) as well. From an experimental point of view, CA is a complicated enzyme and is not so easy to assay. It has more than one isozyme. In this way, possible errors in reporting the ki values, due to assay difficulties, could adversely affect our QSAR results. In spite of these facts, the RMSE values resulted by our models are still good enough to make these models trustable in future predictions. This model has good statistical characteristics as evident from its R2=0.973 and R2=0.963 values.

Equations 4 and 5 reveal that a higher value of Balaban mean square distance index (MSD) and 4nd order mean topological charge index (JGI4) are advantageous to enhance the activity. On the other hand, a higher value number of double bonds (nDB), second Mohar index (TI2) and Counts for certain structural fragment, R-CH-X (descriptor C-028) are detrimental to the activity. Thus the descriptors identified for rationalizing the activity give paths to modulate the structure to a desirable biological end point. The topological (TOPO) class descriptors are based on a graph representation of the molecule and are numerical quantifiers of molecular topology obtained by the application of algebraic operators to matrices representing molecular graphs and whose values are independent of vertex numbering or labeling. They can be sensitive to one or more structural features of the molecule such as size, shape, symmetry, branching and cyclicity and can also encode chemical information concerning atom type and bond multiplicity. The 2D autocorrelations are molecular descriptors which describe how a considered property is distributed along a topological molecular structure. The 2D-AUTO descriptors have their origin in autocorrelation of topological structure of Broto-Moreau (ATS), of Moran (MATS) and of Geary (GATS). The computation of these descriptors involve the summations of different autocorrelation functions corresponding to the different fragment lengths and lead to different autocorrelation vectors corresponding to the lengths of the structural fragments. Also a weighting component in terms of a physicochemical property has been embedded in this descriptor. Atom centered fragments (ACF descriptors) are simple molecular descriptors defined as the number of specific atom types in a molecule and their calculation is based on the knowledge of the molecular composition and atom connectivity. Even descriptors that at the first look seem not to be related to the 3D molecular structure, like the number of double bonds or the number of CHR3 groups, in fact, do identify molecular sub-fragments that can be consider as 'structure making' factors. For example, the number of double bonds between two carbon atoms is related with the cis-trans isomerism or may show the existence of an aromatic ring. The number of double bonds may also be related with the hydrophobicity and reactivity of the considered compounds. Another significant structural element, which contains a double bond, is the carbonyl C=O group[36]. The structure-activity correlations obtained with the descriptors suggest that less branched and saturated structural templates would be better for the activity.

The structural insights obtained from the present study are expected to be useful in the future design of new compounds with potentially higher inhibition activity against carbonic anhydrase CAIX.

CONCLUSIONS

To achieve a significant correlation, it is essential that proper descriptors are used. A wide variety of molecular descriptors are used in QSAR models[37]. However, as the number of descriptors increases, the model becomes complicated, and its interpretation is difficult when many variables are used. Thus, the application of

Regular Paper

such techniques generally involves variable selection for building well-fitted models. Many different methods have been used to select the significant descriptors for calibration purposes. On the other hand, artificial neural networks (ANNs) are popular in QSAR models as a result of their success where complex nonlinear relationships exist among data[38-39]. The ANN model was primarily developed for predictive ability and classification. The descriptors identified in MLR analysis have highlighted the role of mean square distance index (MSD), topological charges (JGI4), certain structural fragments (C-028), the second Mohar index (TI2) have positive influence on the inhibitory effect. Based on the MLR equation, which indicates the dependence and the extent of influence of the descriptors to the inhibitory activity, various structural modifications can be proposed for designing of novel structures with desired characteristics.

The results of two QSAR models tell us that nonlinear selection methods and activity prediction models do better than their linear counterparts. This fact – that is also confirmed by other QSAR attempts[27,28,40] -is due to complicated relations between structure and activity of compounds. From the above discussion, it can be seen that both approaches are statistically meaningful. The results obtained show that nonlinear regression analyze is useful tools to distinguish between the inhibitory activities of sulfonamides toward CAIX isozyme.

REFERENCES

[1] B.Hemmateenejad, R.Miri, M.Jafarpour, M.Tabarzad; QSAR Comb.Sci., **26(10)**, 1065-1075 **(2007)**.

[2] Claudiu T.Supuran, A.Scozzafava; Exp Opin Ther Patents, **12**, 217-242 **(2002)**.

[3] Claudiu T. Supuran, A.Scozzafava; Exp Opin Ther Patents, **10**, 575-600 **(2000)**.

[4] Claudiu T.Supuran, A.Scozzafava; CRC Press LLC: Boca Raton, FL, ISBN 0-41, 30673-6 **(2000)**.

[5] M.Jaiswal, P.V.Khadikar, A.Scozzafava, Claudiu T. Supuran; Bioorganic & Medicinal Chemistry Letters, 3283-3290 **(2004)**.

[6] I.Nishimori, D.Vullo, A.Innocenti, A.Scozzafava, A.Mastrolorenzo, Claudiu T.Supuran; Bioorganic & Medicinal Chemistry Letters, **15**, 3828-3833 **(2005)**.

[7] A.Fiore, G.Simone, V.Menchise, C.Pedone, A.Casini, A.Scozzafava, Claudiu T.Supuran; Bioorganic & Medicinal Chemistry Letters, **15**, 1937-1942 **(2005)**.

[8] A.Innocenti, D.Vullo, A.Scozzafava, C.T.Supuran; Bioorganic & Medicinal Chemistry Letters, **18**, 1583-1587 **(2008)**.

[9] F.Sączewski, A.Innocenti, J.Slawinski, A.Kornicka, Z.Brzozowski, E.Pomarnacka, A.Scozzafava, C.Temperini, C.T.Supuran; Bioorganic & Medicinal Chemistry, 6 **(2008)**.

[10] T.Anupama; Journal of Engineering.Science and Management Education, **4**, 27-32 **(2011)**.

[11] B.Hemmateenejad, K.Javidnia, M.Nematollahia, M.Elyasia; J.Iran.Chem.Soc., 6(2), 420-435 **(2009)**.

[12] A.T.Balaban, S.C.Basak, A.Beteringhe, D.Mills, C.T.Supuran; Mol Divers, **8**, 401-412 **(2004)**.

[13] P.V.Khadikar, V.Sharma, S.Karmarkar, C.T.Supuran; Bioorg.Med.Chem.Lett., **15**, 923-930 **(2005)**.

[14] P.V.Khadikar, J.Singh, S.Singh, R.Mishra, C.T.Supuran, B.W.Clare, M.Lakhwani; Medicinal Chemistry, **4**, 30-66 **(2008)**.

[15] G.Melagraki, A.Afantitis, H.Sarimveis, O.Igglessi-Markopoulou, C.T.Supuran; Bioorg.Med.Chem., **14**, 1108-1114 **(2006)**.

[16] J.Singh, B.Shaik, S.Singh, V.K.Agrawal, P.V.Khadikar, O.Deeb, C.T.Supuran; Chemical Biology & Drug Design, **71**, 244-259 **(2008)**.

[17] J.David, T.Livingstonea and David; Manallack Neural Networks in 3D QSAR, QSAR Comb.Sci., 22 **(2003)**.

[18] M.J.S.Dewar, E.G.Zoebisch, E.F.Healy, J.J.P.J.Stewart; Am Chem.Soc., **107**, 3902-3909 **(1985)**.

[19] S.Czewski, A.Innocenti, J.S.ski, A.Kornick, Z.Brzozowski, C.T.Supuran; Bioorganic & Medicinal Chemistry, **16**, 3933-3940 **(2008)**.

[20] D.Vullo, M.Franchi, E.Gallori, J.Pastorek, A.Scozzafava, S.Pastorekovac, Claudiu T. Supuran; Bioorganic & Medicinal Chemistry Letters, **13**, 1005-1009 **(2003)**.

[21] H.Turkmen, M.Durgun, S.Yilmaztekin, Claudiu T. Supuran; Bioorganic & Medicinal Chemistry Letters, **15**, 367-372 **(2005)**.

[22] C.Fabrizio, M.Alfonso, S.Andrea, V.Daniela, Claudiu T. Supuran; Bioorganic & Medicinal Chemistry, **17**, 7093-7099 **(2009)**.

[23] L.Puccetti, G.Fasolis, A.Cecchi, Jean-Yves Andrea Scozzafava and T.Claudiu; Bioorganic & Medicinal Chemistry Letters, **15**, 2359-2364 **(2009)**.

[24] S.Pastorekova, A.Casini, A.Scozzafava, D.Vullo, Jaromir Pastoreka and Claudiu T. Supuran; Bioorganic & Medicinal Chemistry Letters, **14**, 869-873 **(2004)**.

[25] O.Zensoy, I.Nishimori, D.Vullo, L.Puccetti, Andrea Scozzafava and Claudiu T. Supuran; Bioorganic & Medicinal Chemistry, **13**, 6089-6093 **(2005)**.
[26] D.W.Hosmer, S.Lemeshow; Applied logistic regression, John Wiley & Sons Inc, New York, **(2000)**.
[27] S.H.Sadat Hayatshahi, P.Abdolmaleki, S.Safarian, K.Khajeh; Biochemical and Biophysical Research Communications, **338**, 137-1142 **(2005)**.
[28] S.H.Sadat Hayatshahi, P.Abdolmaleki, M.Ghiasib, S.Safarian; FEBS Letters, **581**, 506-514 **(2007)**.
[29] O.Deeb, M.Goodarzi, Padmaker V. Khadikar; Chem.Biol.Drug.Des., **(2012)**.
[30] M.Hajmeer, I.Basheer; Food Microbiol, **20**, 43-55 **(2003)**.
[31] M.Poursheikhali Asgary, S.Jahandideh, P.Abdolmaleki, A.Kazemnejad; **23(23)**, 3125-3130 **(2007)**.
[32] E.Eroglu; Int.J.Mol.Sci., **9**, 181-197 **(2008)**.
[33] R.Todeschini, V.Consonni; Molecular descriptors for chemoinformatics, WILEY-VCH; John wiley, **(2009)**.
[34] S.P.Gupta, S.Kumaran; J.Enzyme Inhib Med. Chem., **20**, 251-9 **(2005)**.
[35] L.Gupta, A.Patel, C.Karthikeyan, P.Trivedi; Journal of Current Pharmaceutical Research, **01**, 19-25 **(2010)**.
[36] A.Fedorowicz, L.Zheng, H.Singh, E.Demchuk; Int.J.Mol.Sci., **5**, 56-66 **(2004)**.
[37] R.Tadeschini, V.Consonni; Handbook of Molecular Descriptors, Wiley-VCH: Weinheim,Germany, **(2000)**.
[38] F.Despagne, D.L.Massart; Tutorial review: neural networks in multivariate calibration. Analyst, **123**, 157R-178R **(1998)**.
[39] J.Zupan, J.Gasteiger; Neural Networks in Chemistry and Drug Design.Weinheim, Germany: Wiley-VCH, **(1999)**.
[40] A.Gregory, C.Bakken and J.Peter; Chem.Inf. Comput.Sci., **41**, 1255-1265 **(2001)**.