Quantitative Structure - Activity Relationships Study of Carbonic Anhydrase Inhibitors Using Logistic Regression Model

Sahebjamee, Hassan*+; Yaghmaei, Parichehre

Department of Biology, Science and Research Branch, Islamic Azad University, Tehran, I.R. IRAN

Abdolmaleki, Parviz

Department of Biophysics, Faculty of Science, Tarbiat Modares University, P.O. Box: 14115/175 Tehran, I.R. IRAN

Foroumadi, Ali Reza

Department of Medicinal Chemistry, Faculty of Pharmacy, Tehran University of Medical Sciences, Tehran, I.R. IRAN

ABSTRACT: *Binary Logistic Regression (BLR) has been developed as non-linear models to establish quantitative structure- activity relationships (QSAR) between structural descriptors and biochemical activity of carbonic anhydrase inhibitors. Using a training set consisted of 21 compounds with known ki values, the model was trained and tested to solve two-class problems as active or inactive on the basis of the predicted value for IC50. Many quantitative descriptors were generated to express the physicochemical properties of 21 compounds with optimized structures. After filtration of these descriptors, 39 of descriptors for carbonic anhydrase (CA, EC 4.2.1.1) isozyme IX (CAIX) and 45 for isozymeXII (CAXII) remained and were selected for QSAR study. Logistic regression was then used to non-linearly select the most important descriptors and to develop a model for prediction of IC50. To evaluate the performance of the established models, Jjackknife and self consistency tests were performed during implementation of the two model-building methods. The applied indices including accuracy, sensitivity, and specificity were 85%, 82% and 100% for CAIX and also 71%, 68% and 80% for CAXII respectively. The primary advantage of such an approach is the reduction of redundant variables and the consequent improvement in the efficiency of modeling.*

KEY WORDS: *Quantitative structure-activity relationship, Binary logistic regression, Carbonic anhydrase inhibitors.*

INTRODUCTION

Carbonic anhydrases comprise an ever-present family of metallo-enzymes found in prokaryotes and eukaryotes, which catalyze the reversible hydration of carbon dioxide to the bicarbonate ion and a proton ($CO_2+H_2 \rightarrow HCO_3^-+H^+$). At least 14 dissimilar CA isozymes are currently recognized in humans, and many of them are targets for the design of inhibitors with potential use as anti-glaucoma, anti-obesity, or anti-cancer drugs [1-5]. It has newly been confirmed that such tumors related to CAs (mainly CA IX) may be of great significance as markers of tumor progression.

* *To whom correspondence should be addressed.*
+ *E-mail: sahebjamei.hassan386@gmail.com*
1021-9986/13/2/19 11/$/3.10

This is generally due to their induction by hypoxia, a clinically significant factor of cancer biology that considerably affects cure result and illness progression [6-8]. Hypoxia is associated with acidification of extracellular location that facilitates tumor invasion and CAIX is supposed to play a function in this process through its catalytic activity [9]. In fact, through the Hypoxia Inducible Factor (HIF) cascade, it directs to a strong overexpression of CA IX/XII in numerous tumors. CA IX is a transmembrane protein with a recommended function both in retaining acid-base balance and in intercellular communication. It consists of domain that is unique among the CAs, a greatly active CA catalytic domain, a particular transmembrane region and a short intracytoplasmic tail [9].

In fact, following an extensive evaluation, many sulfonamides were found to act as carbonic anhydrase inhibitors. A great number of sulfanilamide derivatives were synthesized, characterized, tested, and are broadly used in medical drug as pharmacological causes with a wide variety of biological actions [3,10-12]. During the last few years, *Supuran et al.* [13-15] have widely studied different aromatic sulfonamides as potent carbonic anhydrase inhibitors. The majority of the potent CAIs studied up to now belongs to the aromatic/heterocyclic sulfonamide or sulfamate classes, although compounds integrating other zinc-binding groups have also been studied. In fact, such derivatives directly bind by way of the deprotonated sulfonamide/sulfamate moiety to the catalytically vital Zn (II) ion of the enzyme active site, also participating in a multitude of polar and hydrophobic interactions with amino acid residues of the active site cavity. Clinically used sulfonamide/sulfamate CAIs demonstrate potencies in the low nanomolar range against the physiologically relevant isozymes, such as CA I, II, V, and IX [16]. Lately *Czewski et al.* [17] have reported the strong inhibition of tumor-associated isozymes IX and XII with some S-substituted 4-chloro-2-mercapto-6-methyl-benzenesulfonamides.

Several mathematical modeling approaches, however, could aid takeing out the most effective quantitative structural parameters which are descriptors, Descriptors determine the inhibition activity of the molecules. These methods have thus been employed in QSAR studies of a great number of biologically significant molecules [18]. QSAR is a powerful methodology useful in screening a great library of probable drugs for selectivity and potency. In this methodology the molecular structure is primary encoded for yielding molecular descriptors, which are numerical values corresponding to topological, geometry, and/or electronics. This methodology was first applied to substituted benzene sulfonamide by *Jaiswal et al.* [19]. They have the potential to reduce the time and attempt required to develop new molecules by reducing the requirement for expensive and time consuming trial-and-error experiments [20].

Recently, nonlinear methods such as logistic regression has been working. These new applications in chemistry by *Mohebi et al.* [21] and medicine by *Kumar et al.* [22], were created. In this study, we applied Binary Logistic Regression (BLR) as non-linear model to investigate the QSAR in benzenesulfonamides derivative inhibitors of CA. We used logistic regression for selecting more effective descriptors and to obtain an equation for prediction of inhibition activity. BLR is a form of regression which is used when the dependent variable is a dichotomy and the independent variables are continuous, categorical or both [23]. We intended to establish a logistic regression model to predict the activity the carbonic anhydrase inhibitors using data extracted from *Czewski et al.* [17]. The distinction is performed through establishing the discrimination rules. The rules will be estimated during the training procedure and can be used to allocate the new cases into the previously defined classes [24]. The performance of the established model was then evaluated using the common statistical index including accuracy, sensitivity and specificity. We have carried out BLR model to develop a rapid and reliable method to predict the activity carbonic anhydrase inhibitors.

EXPERIMENTAL SECTION

Dataset

A data set containing 21 derivatives of benzenesulfonamides was used in this study. The CA inhibitory activities of these compounds toward two different isoenzymes (hCA IX, and XII) were taken from the literature [17]. Although in small compounds, the S-substituted 4-chloro-2-mercapto-5-methylbenzenesulfonamides core (Fig. 1) has undergone small structural changes, it is conserved in many others. The structures of all compounds are shown in Fig. 2.

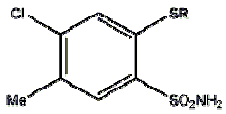
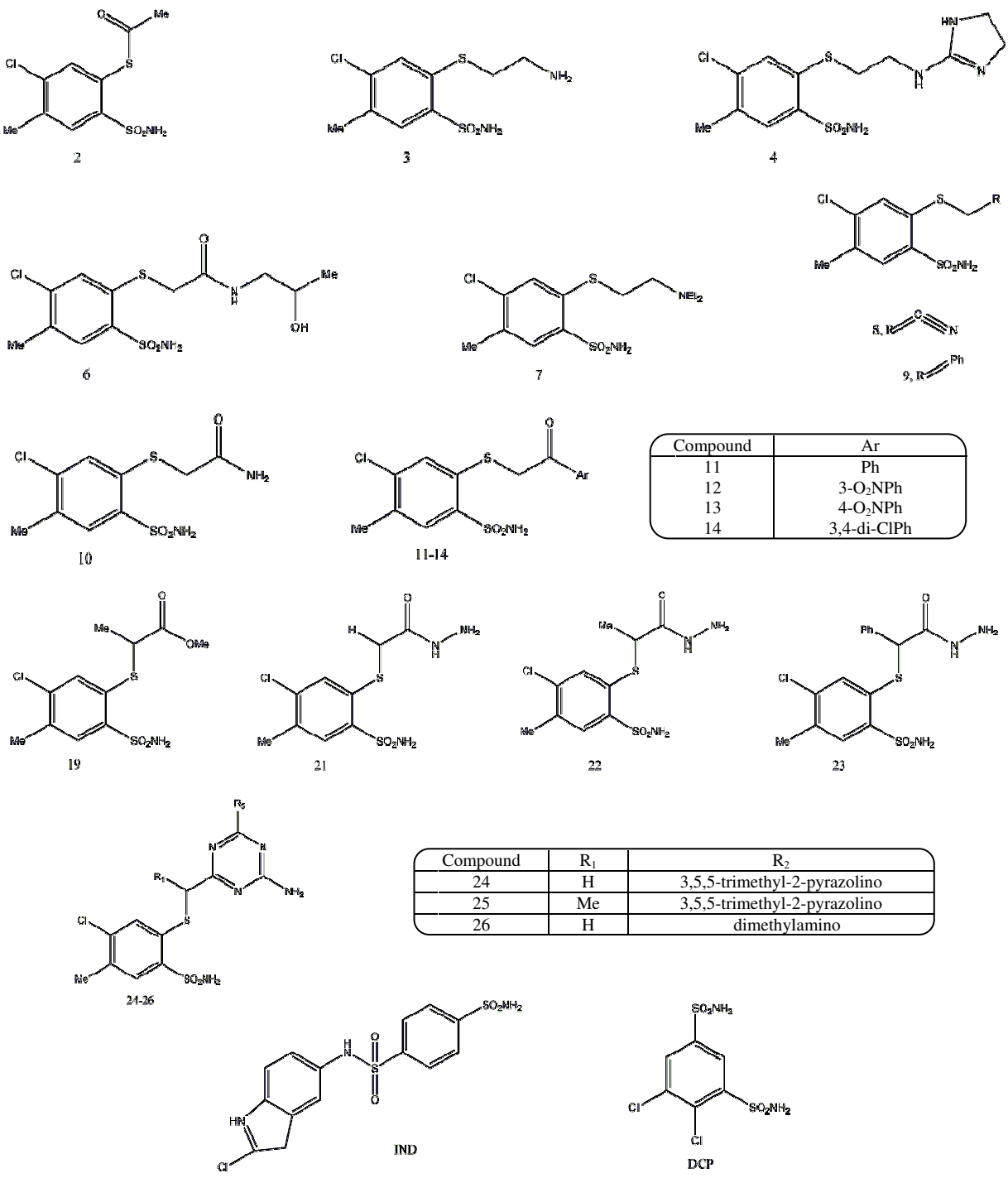


Fig. 1: S-substituted 4-chloro-2-mercapto-5-methylbenzenesulfonamides.



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

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

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

Table 1: DATA CAIX.

Compound	Constant values	GATS1e	ATS2m	S3k	MATS4v	C002	OUT values	Predicted activity class	Real activity class
10	1718.267	-1402.17	-265.16	0	0	0	1.39	1	1
11	989.281	-954.015	0	0	0	32.064	50.38	1	1
12	325.268	-360.361	0	15.086	0	0	59.97	1	1
13	325.268	-360.361	0	15.086	0	0	59.97	1	1
14	323.883	-358.809	0	15.019	0	0	21.89	1	1
19	1718.267	-1402.169	-265.16	0	0	0	1.39	1	1
2	211.357	-256.045	0	15.675	0	0	.99	1	1
21	1718.268	-1402.17	-265.16	0	0	0	1.39	1	1
22	1718.268	-1402.17	-265.16	0	0	0	1.39	1	1
23	1718.268	-1402.17	-265.16	0	0	0	1.39	1	1
24	-129.53	0	0	0	-563.345	80.337	-45.81	0	0
25	326.487	-361.618	0	15.117	0	0	-71.21	0	0
26	328.848	-364.241	0	15.226	0	0	-125.1	0	0
3	1732.168	-1409.712	-271.276	0	0	0	2.2	1	0
4	1718.267	-1402.169	-265.16	0	0	0	1.39	1	1
6	1718.267	-1402.169	-265.16	0	0	0	1.39	1	1
7	1732.168	-1409.712	-271.276	0	0	0	2.2	1	0
8	1718.267	-1402.169	-265.16	0	0	0	1.39	1	1
9	1718.267	-1402.17	-265.16	0	0	0	1.39	1	1
dcp	1328.333	0	-1867.14	0	-2310.68	0	-200.1	0	0
ind	339.436	-316.12	0	0	0	0	6.25	1	0

Table 2: DATA CAX II.

Compound	Constant values	GATS1e	MATS1e	GATS2e	RBF	CIC5	ATS4e	CIC3	GATS3v	JGT	OUT values	Predicted activity class	Real activity class
10	328.27-	0	609.695	328.752	-37.9	215.21	0	0	0	0	1.39	1	1
11	-2054.2	-1897.31	0	0	0	300.92	3599.6	0	0	0	83.59	1	1
12	447.05	-556.21	0	0	0	0	0	273.4	0	0	-3.16	0	0
13	388.83	-476.076	0	0	0	0	0	223.1	0	0	-12.62	0	0
14	-1696.9	0	1559.93	0	-1158.8	427.39	1486.9	0	0	0	-12.19	0	0
19	-328.27	0	609.675	328.752	-837.9	215.21	0	0	0	0	1.39	1	1
2	-331.34	0	614.452	331.277	-841.24	216.92	0	0	0	0	29.3	1	1
21	-328.27	0	609.674	328.752	-837.9	215.21	0	0	0	0	1.39	1	1
22	-328.27	0	609.674	328.752	-837.9	215.21	0	0	0	0	1.39	1	1
23	282.86	-239.724	0	0	-543.84	154.79	0	0	760.01	0	350	1	0
24	-270.45	0	534.423	276.569	-791.45	200.99	0	0	320.002	0	240	1	0
25	-332.72	0	614.679	332.572	-841.64	217.06	0	0	804.5	0	360	1	0
26	-4061.6	0	0	1677.49	0	678.15	0	0	1143	-453.4	472.14	1	0
3	-328.27	0	609.675	328.752	-837.9	215.21	0	0	0	0	1.39	1	1
4	-328.27	0	609.675	328.752	-837.9	215.21	0	0	0	0	1.39	1	1
6	-328.27	0	609.675	328.752	-837.9	215.21	0	0	0	0	1.39	1	1
7	-328.27	0	609.675	328.752	-837.9	215.21	0	0	0	0	1.39	1	1
8	-328.27	0	609.675	328.752	-837.9	215.21	0	0	0	0	1.39	1	1
9	282.86	-239.72	0	0	-543.84	154.79	0	0	0	0	2.2	1	0
Dcp	344.45	-679.753	-818.45	0	0	241.04	0	0	335.09	-496.8	-18.44	0	0
ind	-799	0	1236.34	474.764	0	429.38	0	0	0	0	-7.88	0	1

The biological data were obtained as IC_{50} (drug concentration, in nM). The greater is this value, the weaker is the inhibition activity of the compound. Since we had a relatively small data set, we trained our models only to solve two-class problems. In other words, models could categorize the compounds into active and inactive classes. For this purpose, compounds CAIX were labeled as active for log (ki) < 50 and inactive for log (ki) >50. According to this classification, seventeen compounds were active and the remaining seven were inactive. Also, compounds CAXII were labeled as active for log (ki) <18 and inactive for log (ki) >18.The structures, biological data and activation class of the compounds, are shown in Fig. 2 and Table 1, 2 respectively.

Descriptor Calculation and Selection

The chemical structure of molecules was set up with the Hyperchem program (Hypercube Inc.) and optimized

with molecular mechanics using an MMX force field, and then with semiempirical quantum chemical calculations using an AM1 Hamiltonian. Topological, constitutional, and functional group descriptors were calculated for each molecule by the Dragon software (Milano Chemometrics and QSAR Research Group, www.disat.unimib.it/chm/).

Totally 644 descriptors were generated 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 $\log(k_i)$ were regarded redundant and removed.

Finally, since highly correlated descriptors provide approximately identical information, performing a pair wise correlation, one of two descriptors was randomly removed if their correlation coefficient exceeded 0.75. After these three steps, the number of descriptors was reduced to 39 for CAIX and 45 for CAXII. Different types of numerical descriptors were generated to describe each compound. These descriptors are categorized in autocorrelation of a topological structure [25], geometrical, MoRSE [26-28] and GETAWAY [26-28] groups. The experimental index of log (ki) was reported for every compound as a measure of inhibition.

Since, we aimed to describe the structural requirements for inhibitory activity toward different isoenzymes of CA, the quantitative relationships between inhibitory activity, as dependent or predicted variable, and molecular descriptors, as dependent or predictor variables were obtained by binary logistic regression.

Model development

In order to evaluate the efficiency of Logistic Regression (LR) in obtaining QSAR models, cross-validation method was used. The original data set were partitioned into two subsets, one of which for model establishing, and the other for predicting examination. The most prevailing of this method is jackknife procedure. When the original data set is small, jackknife procedure is used. In jackknife, all samples are used to estimate testing case. The jackknife proceeds by removing one case (testing case) at a time from the original data set and the training is done using the remaining case. Then the testing case is examined by the resulting model of remaining case. This procedure is repeated until all cases within the data set are tested.

Logistic regression models

This model is used only when the dependent variable is dichotomous, that is, there are only two possible answers for the dependent variable. Let the dependent variable be Y. Since it is dichotomous, it takes on 0 or 1 for failure and success, respectively. The logistic regression model can be expressed as follows:

$$OUT = \log\left[\frac{P}{1-P}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \qquad (1)$$

Where p is the probability of Y = 1, β_0 is a constant, and β_1-β_n are coefficients associated with the independent variables $x_1 - x_n$. The ratio p / (1- p) takes on values between 0 and plus infinity. Consequently, the logarithm of this ratio (here called OUT) is a continuous variable that takes on values between minus infinity and plus infinity. Using this equation, the value of OUT is determined. Then a cutoff should be taken to recode the OUT values into two possible states of dependent variable. The value of such cutoff is a matter of case and should be optimized.

We used Backward-Wald Binary Logistic routine in SPSS program to develop models of this kind. This routine selects the most important descriptors first and develops a logistic regression model. As a Jackknife routine, the training procedure of model development was repeated 21 times. Every time, one compound was removed from the training set and was used as a testing case. In every training procedure of the logistic regression, an index called -2log likelihood was minimized through many iteration. For every value of -2log likelihood the model suggested coefficients for independent variables as well as a constant. The set of coefficients and constant related to the lowest value of -2log likelihood were retained and selected to build up an equation. After testing all equations by the relative testing case, the equations were unified using the averages of the retained coefficients and constants. Then all 21 cases were tested using the obtained unique equation. After testing with both jackknife equations and the unique equation, the cutoff used for recoding the OUT values into activity classes of inhibition was optimized. All models were evaluated using some statistical Indices.

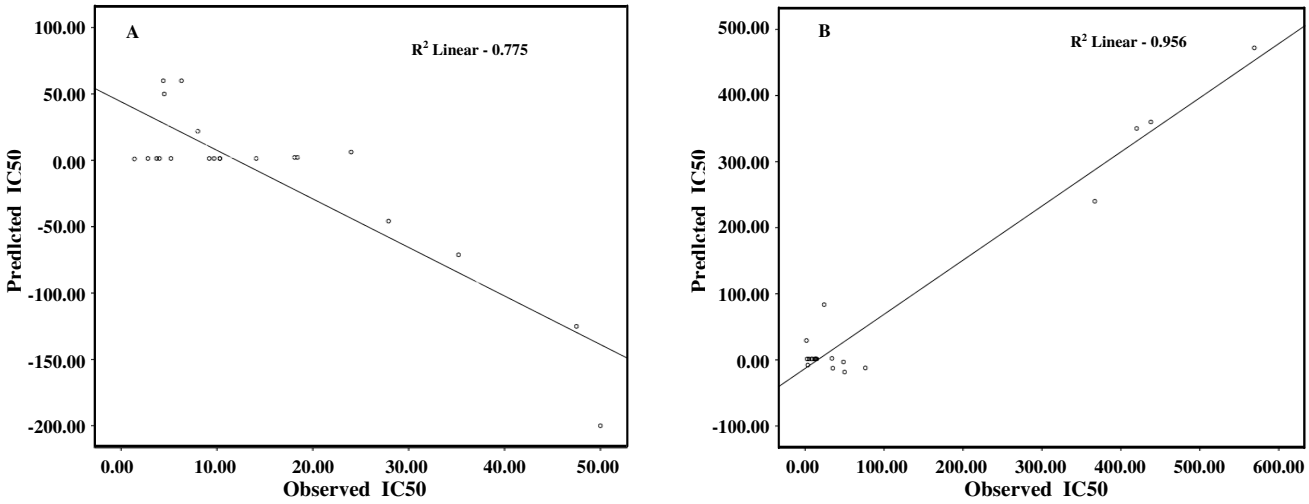


Fig.3: Plot of predicted activity against the experimental activity by unique equation: (A) CAIX and (B) CAXI1.

In the first stage of this method, logistic regression serves as a non-linear model on the dataset to select significant parameters through the 'Self-consistency Test'. This test is an examination to evaluate the performance of the established model on the same database which is used to train the model. In fact it is an index showing the ability of the model in extracting the relationship between the descriptors and the dependent variable. When the self-consistency test was performed, each descriptor in the dataset concerned is in turn identified using the rule parameters derived from the same dataset, the so-called training dataset. As a new approach, the jackknife technique was then applied to train and test the BLR model. In this method, all but one case in the database is used to train the BLR model. The single case that is left out is then used to test the model. This procedure is repeated until each case in the database is used once as the testing case. It therefore provided 21 different simulations. Although this method is extremely time-consuming, it is useful especially for the small database such as ours. By doing so, we hopefully expected that we got a unique conclusion on whole of the data set.

Performance measures

All models were evaluated using some statistical indices. Before calculating such indices, TP_i (number of cases that were truly recognized as active), TN_i (number of cases that were truly recognized as inactive), FP_i (number of cases that were wrongly recognized as active), and FN_i (number of cases that were wrongly recognized as inactive) should have been counted. Using the following formulas which have been previously reported in the published material, we calculated the prediction accuracy, sensitivity, specificity and probability of correct prediction for the output of the binary logistic regression models.

Prediction Accuracy

$$(Acc') = [(TP_i + TN_i)/t] \times 100 \quad (2)$$

Where

$$t = TP_i + TN_i + FP_i + FN_i \quad (3)$$

Sensitivity

$$(S^i_n) = [TP_i/(TP_i + FN_i)] \times 100 \quad (4)$$

Specificity

$$(S^i_p) = [TN_i/(TN_i + FP_i)] \times 100 \quad (5)$$

Statistical analysis was performed using SPSS 18 for Windows (SPSS Inc., Chicago, USA).

RESULTS AND DISCUSSSION

It would be found from Tables 1 and 2 that the number of significant parameters in this comparison is 5 for CAIX and 9 for CAXII. As a result of the logistic regression analysis for CAIX, the structure and activity of compounds were most effectively related by three descriptors of GATS1e, ATS2m and S3K respectively. The structure and activity of compounds for CAXII were most effectively related by four descriptors of GATS2e, MATS1e, RBF and CIC5 respectively.

Moreover, the high value of the cross-validated correlation coefficient indicates the predictivity of the proposed QSAR model. The plot of cross-validated calculated activity is represented in Fig. 3, which shows the scattering of data around a straight line with slope and intercept close to one and zero.

Table 3: Performance Comparison between two tests of selfconsistency and ackknife.

Enzyme	Performance measures	Logistic regression
CAIX	Accuracy (%)	85%
	Specificity (%)	100%
	Sensitivity (%)	82%
CAXII	Accuracy (%)	71%
	Specificity (%)	80%
	Sensitivity (%)	68%

Table 4: Definition of the finally selected set of descriptors.

Name	Type	Description	Ref
MATS1e	2D autocorrelations	Moran auto correlation –lag1/weighted by atomic sanderson electronegativitites	[26,28]
GATS2e	2D autocorrelations	Geary autocorrelation –lag2/weighted by atomic Sanderson electronegativitites	[26,28]
GATS1e	2D autocorrelations	Geary autocorrelation –lag1/weighted by atomic Sanderson electronegativitites	[26,28]
ATS2m	2D autocorrelations	Broto moreau autocorrelation of a topological structure –lag2/weighted by atomic masses	[26,28]
CIC5	Information indices	Complementary Information Content index (neighborhood symmetry of 5-order)	[26,27]
RBF	Constitutional indices	Rotatable bond fraction Constitutional indices	[26]
S3K	Topological indices	3-path Kier alpha-modified shape index	[28]

In 21 training procedures which were performed during the jackknife procedure these descriptors (GATS1e=19, ATS2m=12, S3K=12 for CAIX and GATS2e=14 MATS1e=15, RBF=15, CIC5=19 for CAXII) were selected as more determining factors. Names, types and definition of the descriptors which were and used in the final model are shown in Table 4. In such Jackknife procedure, 21 equations were suggested for predicting the inhibition class of all compounds. The results of these equations are shown in Tables 1 and 2. The optimal cutoff for the OUT values was found to be 0.693 for CAIX and to be 0.287 for CAXII. So, the compounds with predicted OUT < 0.693 were regarded as inactive and with OUT > 0.693 as active. Compounds the result of self-consistency test was evaluated by the performance measures. Through jackknife simulations we obtained 21 different equations, each suggesting a few descriptors as effective parameters in determining the activity class of cases. To obtain a unified equation, we used the averages of coefficients and constant values of equations. The following equations as the unified equations were obtained:

$$OUT_{CAIX} = 1059-1008.69\times GATS1e-389\times ATS2m+15.226\times S3K \quad (6)$$

$$OUT_{CAXII} = -503.892+426.764\times GATS2e+615.623\times MATS1e+818.765\times RBF+273.727\times CIC5 \quad (7)$$

The results shown in Table 3 are obtained according to the output of the model. Our results indicated that the BLR was able to classify correctly only 85% and 71% of cases for CAIX and CAXII, respectively. Also the obtained specificity and sensitivity for BLR model were 100%, 82% for CAIX and 80%, 68% for CAXII.

So the above-mentioned equations were found as a reliable predictive model to classify the inhibitors. Since the descriptors with greater coefficients are more determining in regression equations, we can conclude that according to this equation, the most important descriptor is S3K for CAIX and the more important descriptors are GATS2e, MATS1e, RBF and CIC5 for CAXII. In above equation for CAIX, the 2D-AUTO class descriptors, ATS2m (Broto-Moreau autocorrelation of a topological structure of lag 2 weighted by atomic masses) and GATS1e (Geary autocorrelation of lag 1 weighted by atomic Sanderson electronegativities), contribute negatively to the activity.

Thus molecules leading to lower value of such topological lags (paths) weighted by relevant atomic property would be favorable to the binding activity.

Regarding the obtained results, some descriptors of types 2D autocorrelation and topological indices seem to be of great significance respectively in directing the binding site inhibitors toward a greater inhibitory effect on the natural function of carbonic anhydrase.

The S3K is a kier alpha-modified shape descriptor. Although kier shape descriptors is defined in terms of the number of graph vertices and the number of paths with lengh m (m=1,2,3) in the H-depleted molecular graph, the alpha-modified shape descriptors take into account the dissimilar shape part of heteroatoms and hybridization states. Values are larger when the molecular branching is missing or, as in the training set under study, when it is positioned at the extremities of a graph [28]. The Rotatable Bond Fraction (RBF) represents a proportion between the number of rotatable bonds and the total number of bonds of any type in the molecule [26]. ATS2m describes the spatial distribution of the atomic mass [29].Topological indices are 2D descriptors which take into account the internal atomic arrangement of compounds and encode in numerical form information about molecular size, shape, branching, presence of heteroatoms, and multiple bonds [30]. GETAWAY descriptors are used to compare molecules or even conformers taking into account their molecular shape, size, symmetry and atom distributions. They are, moreover, independent of molecule alignment [28]. MATS1e the path connecting a pair of atoms has length 6 and involves the atomic Sanderson electronegativities as weighting scheme [31]. There are small differences between the 2D-autocorrelation descriptors of type ATSd, GATSd and MATSd; but in general, they explain how the considered property is spread along the topological structure [32].

Usually speaking, misclassifications in the jackknife procedures may have two explanations. First, the testing case may be a typical pattern of inhibition. Therefore, it plays such a key role in training procedure, that if removed, the model will not be efficiently trained. Second, the testing case may have unique structural properties in our database. In other words, the specific case has not been presented in the training set. Therefore, it cannot be recognized by the developed model during the testing procedure. The second supposition seems more reasonable when the size of data set is small such as ours (n = 21).

It was found that the kind of ligand – receptor interactions varies form one isoenzyme to another, which implies that by considering these interactions, it is possible to design selective ligands in the direction of a specified CA isozyme. But the non-specificity of sulphonamide derivatives to isozymes, leads to a range of side effects [22]. Others have done studies with different methods which showed that inhibitors of this isozymes carbonic anhydrase and structural descriptors vary [33-35].

As a routine in current pharmacology, QSAR methods provide important information for designing more effective drugs. Linear and non-linear classifying models were used in QSAR studies. Such classifiers can then be used by chemists as well as pharmacologists to design more effective drugs [36, 37]. QSAR studies on the inhibitory activity of a set of sulfonamide derivatives in the direction of two different isoenzymes of CA helped us to find the structural requirement of the sulfonamide ligands for binding to the isoenzymes. To obtain the effect of the structural parameters of the benzenesulfonamides on their inhibitory activity on CAs, QSAR analysis with different types of molecular descriptors was operated. The molecular descriptors which are important for a specific activity of the molecule, are determined and optimized.

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 two active and inactive classes efficiently. This study proved the capability of logistic regression to deal with this problem. This method correctly selected more efficient descriptors. Also it suggested a very easy equation to predict the inhibition activity of molecules. Three of the chief attractions of logistic discrimination are: (i) the model is easy and few distributional assumption are made. (ii) It is applicable with either continuous or discrete predictor variables, or both. (iii) It is very easy to use with fewer computational demands.

Our results are expected to be used in the future design of compounds with potentially higher inhibition activity against carbonic anhydrase especially with the aim of designing potent drugs for the treatment of cancer. For future studies we plan to increase the number of compounds used in data set and also to generate more descriptors. Consequently, we believe that the risk of statistical errors will be decreased. Also developing models that are able to classify compounds into three or more activity groups

will be possible. It is also possible to combine the presented parameter selection techniques with other non-algorithmic models in order to reach more accurate results.

CONCLUSIONS

Non-linear models allow the estimation of inhibition effect of benzenesulfonamides carbonic anhydrase inhibitors. The equations comprise minimum of easily and quickly computable descriptors.

The QSAR model obtained for CAIX suggested that by decreasing the molecular mass and the atomic sanderson electronegatives of a molecule and molecular branching have will improve inhibitory activity. On the other hand, for the benzenesulfonamide derivatives to CAXII, it was found that a higher value of Geary autocorrelation of lag 2 weighted by Sanderson electronegativity (GATS2e), Moran autocorrelation of lag 1 weighted by Sanderson electronegativity(MATS1e), Rotatable Bond Fraction (RBF), Complementary Information Content index (CIC5) are helpful in improving the inhibition activity of a compound pertaining to CAXII. The QSAR model obtained can be used to predict inhibition activity of inhibitors of carbonic anhydrase with similar structure. The latter finding proposes the possibility of designing selective ligands toward a specified CA isozyme by taking such interactions into consideration [33]. The results obtained offers very good regression models that hold good prediction ability.

The BLR model opens a new pathway in computing and characterization of inhibitory activity on CA IX/XII of the S-substituted 4-chloro-2-mercapto-5-methylbenzenesulfonamides.

Acknowledgments

We like to thank Mr. Ebrahim Barzeghary for his technical help and useful comments.

Received : Dec. 16, 2011 ; Accepted : Sep. 4, 2012

REFERENCES

[1] Fiore A.D., Simone G.D., Menchise V., Pedone C., Supuran C.T., Carbonic Anhydrase Inhibitors: X-ray Crystal Structure of a Benzenesulfonamide Strong CA II and CA IX Inhibitor Bearing a Pentafluorophenylaminothioureido Tail in Complex with Isozyme II, *Bioorganic & Medicinal Chemistry Letters,* **15**, p.1937 (2005).

[2] Guzel O., Salman A., Supuran C.T., Synthesis of 2,4,6-Trimethylpyridinium Derivatives of 2-Hydrazinocarbonyl)-3-Aryl-1H-Indole-5-Sulfonamides Acting as Potent Inhibitors of the Tumor-Associated Isoform IX and XII, *Bioorganic & Medicinal Chemistry Letters*, **19**, p.2931 (2009).

[3] Hemmateenejad B., Miri R., Jafarpour M., Tabarzad M., Exploring QSAR for the Inhibitory Activity of a Large Set of Aromatic /Heterocyclic Sulfonamides toward Four Different Isoenzymes of Carbonic Anhydrase, *QSAR Comb. Sci*., **10**, p. 1065 (2007).

[4] Turkmen H., Durgun M., Yilmaztekin S., Emul M., Innocenti A., Vullo D., Scozzafava A. Supuran, C.T., Carbonic Anhydrase Inhibitors. Novel Sulfanilamide/Acetazolamide Derivatives Obtained by the Tail Approach and Their Interaction with the Cytosolic Isozymes I and II, and the Tumor-Associated Isozyme IX, *Bioorganic & Medicinal Chemistry Letters*, **15**, P. 367 (2005).

[5] Carta F., Maresca A., Scozzafava A., Vullo D., Supuran C.T., Carbonic Anhydrase Inhibitors. Diazenylbenzenesulfonamides are Potent and Selective Inhibitors of the Tumor-Associated Isozymes IX and XII over the Cytosolic Isoforms I and II, *Bioorganic & Medicinal Chemistry*, **17**, p. 7093 (2009).

[6] Fukumura D., Chen Y., Gohongi T., Seed B., Jain R.K., Hypoxia and Acidosis Independently Up-Regulate Vascular Endothelial Growth Factor Transcription in Brain Tumors *in Vivo*, *Cancer Res*., **61**, p. 6020 (2001).

[7] Giatromanolaki A., Koukourakis M.I., Sivridis E., Pastorek J., Wykoff C.C., Gatter K. C., Harris A.L., Expression of Hypoxia-Inducible Carbonic Anhydrase-9 Relates to Angiogenic Pathways and Independently to Poor Outcome in Non-Small Cell Lung Cancer, *Cancer Res.*, **61**, p. 7992 (2001).

[8] Chia S.K., Wykoff C.C., Watson P. H., Han C., Leek R.D., Pastorek J., Gatter K.C., Ratcliffe P., Harris A.L., Prognostic Significance of a Novel Hypoxia-Regulated Marker, Carbonic Anhydrase ix, in Invasive Breast Carcinoma, *J. Clin. Oncol*., **19**, p. 3660 (2001).

[9] Vullo D., Franchi M., Gallori E. Supurana, C.T., Carbonic Anhydrase Inhibitors: Inhibition of the Tumor-Associated Isozyme IX with Aromatic and Heterocyclic Sulfonamides, *Bioorganic & Medicinal Chemistry Letters*, **13**, p. 1005 (2003).

[10] Winum J., Cecchi A., Montero J., Inoccenti A., Carbonic Anhydrase Inhibitors. Synthesis and Inhibition of Cytosolic/Tumor-Associated Carbonic Anhydrase Isozymes I, II, and IX with Boron-Containing Sulfonamides, Sulfamides, and Sulfamates: Toward Agents for Boron Neutron Capture Therapy of Hypoxic Tumors, *Bioorg. Med. Chem*, **15**, p. 3302 (2005).

[11] Mincione F., Starnotti M., Masini E., Bacciottini L., Scrivanti C., Casini A., Vullo D., Scozzafava A., Supuran C.T., Carbonic Anhydrase Inhibitors: Design of Thioureido Sulfonamides with Potent Isozyme II and XII Inhibitory Properties and Intraocular Pressure Lowering Activity in a Rabbit Model of Glaucoma, *Bioorg. Med. Chem. Lett*., **15**, p. 3821 (2005).

[12] Garaj V., Puccetti L., Fasolis G., Winum J.Y., Montero J.L., Carbonic Anhydrase Inhibitors: Novel Sulfonamides Incorporating 1,3,5-Triazine Moieties as Inhibitors of the Cytosolic and Tumour-Associated Carbonic Anhydrase Isozymes I, II and IX, *Bioorg. Med. Chem*., **15**, p. 3102 (2005).

[13] Thakur A., Thakur M., Khadikar P.V., Supuran C.T., Sudele P., QSAR Study on Benzenesulphonamide Carbonic Anhydrase Inhibitors: Topological Approach Using Balaban Index, *Bioorg. Med. Chem*., **12**, p. 789 (2004).

[14] Rusconi S., Innocenti A., Vullo D., Mastrolorenzo A., Scozzafava A., Supuran C.T., Carbonic Anhydrase Inhibitors. Interaction of Isozymes I, II, IV, V, and IX with Phosphates, Carbamoyl Phosphate, and the Phosphonate Antiviral Drug Foscarnet, *Bioorg. Med. Chem. Lett*., **14**, p. 5763 (2004).

[15] Scozzafava A., Menabuoni L., Mincione F., Supuran C.T., Carbonic Anhydrase Activators: High Affinity Isozymes I, II, and IV Activators, Incorporating a β-Alanyl-histidine Scaffold, *J. Med. Chem*., **45**, p. 1466 (2002).

[16] Abdel-Hamid M.K., Abdel- Hafez A.A., El-Koussi N.A., Mahfouz N.M., Innocenti A., Supuran C.T., Design, Synthesis, and Docking Studies of New 1, 3, 4-Thiadiazole-2- Thione Derivatives with Carbonic Anhydrase Inhibitory Activity, *Bioorganic & Medicinal Chemistry*, **15**, p. 6975 (2007).

[17] Czewski S., Innocenti A., Ski J.S., Kornick A., Brzozowski Z., Supuran C.T., Carbonic Anhydrase Inhibitors: Inhibition of Human Cytosolic Isozymes I and II and Tumor-Associated Isozymes IX and XII with S-Substituted 4-Chloro-2-Mercapto- 5-Methyl-Benzenesulfona- Mides, *Bioorganic & Medicinal Chemistry*, **16**, p. 3933 (2008).

[18] Bruno-Blanch L., Galvez J., Garcia-Domenech R., Topological Virtual Screening: A Way to Find New Anticonvulsant Drugs from Chemical Diversity, *Bioorg. Med. Chem. Lett*., **13**, p. 2749 (2003).

[19] Jaiswal M., Khadikar P.V., Supuran C.T., Topological Modeling of Lipophilicity, Diuretic Activity, and Carbonic Inhibition Activity of Benzene Sulfonamides: a Molecular Connectivity Approach, *Bioorganic & Medicinal Chemistry Letters*., **14**, p. 5661 (2004).

[20] Debnath A.K., Ghose A.K, Viswanadhan V.N., Eds., In "Combinatorial Library Design and Evaluation: Principles Software Tools and Applications in Drug Discovery", Marcel Dekker: New York, p. 73 (2001).

[21] Mohebi A., Sobbif A., Autamated Analysis of Pressure Build up Tests by Phase Redistribution, *Iraninan Journal of Chemistry and Chemical Engineering (IJCCE)*, **15**(1), p. 38 (1996).

[22] Kumar S., Singh V., Tiwari M., Quantitative Structure Activity Relationship Studies of Sulfamide Derivatives as Carbonic Anhydrase Inhibitor: As Antiglaucoma Agents, *Medicinal Chemistry*, **3**, p. 379 (2007).

[23] Hosmer D.W., Lemeshow S., "Applied Logistic Regression", Wiley, New York (1989).

[24] Abdolmaleki P., Mokhtari Dizagi M., Vahead M.R., Gity M., Logistic Discriminant Anlysis of Breast Cancer Using Ultrasound Measurements, *Iran. J. Radiat. Res*., **2**, p. 1 (2004).

[25] Caballero J., Garriga M., Fernandez M., 2D Autocorrelation Modeling of the Negative Inotropic Activity of Calcium Entry Blockers Using Bayesian-Regularized Genetic Neural Networks, *Bioorgan. Med. Chem*., **14**, p. 3330 (2006).

[26] Todeschini R., Consonni V. "Handbook of Molecular Descriptors", Wiley-VCH (2000).

[27] Hemmer M.C., Steinhauer V., Gasteiger J., Deriving the 3D Structure of Organic Molecules from Their Infrared Spectra, *Vib. Spectrosc.*, **19**, p. 151 (1999).

[28] Todeschini R., Consonni V., "Molecular Descriptors for Chemoinformatics", WILEY-VCH; John wiley (2009).

[29] Puta R., Xua Q.S., Massarta D.L., Vander Heydena Y., Multivariate Adaptive Regression Splines (MARS) in Chromatographic Quantitative Structure-Retention Relationship Studies, *Journal of Chromatography A*., **1055**, p. 11 (2004).

[30] Melagraki G., Afantitis A., Sarimveis H., Igglessi-Markopouloua O. Supuran C.T., QSAR Study on Para-Substituted Aromatic Sulfonamides as Carbonic Anhydrase II Inhibitors Using Topological Information Indices, *Bioorganic & Medicinal Chemistry*, **14**, p. 1108 (2006).

[31] Rohan G., Jyoti P., Neetu S., Yogesh P. A., Arun K.G., Insights Through Molecular Modeling Into the Structural Requirement of Phenyl(6-Phenylpyrazin-2-yl)Methanone Derivatives as Aldose Reductase Inhibitors, *International Journal of Drug Design and Discovery*., **2**, p. 575 (2011).

[32] Liane S., Maykel P., Marta T., 2D-Autocorrelation Descriptors for Predicting Cytotoxicity of Naphthoquinone Ester Derivatives Against Oral Human Epidermoid Carcinoma, *Bioorganic & Medicinal Chemistry*, **15**, p. 3565 (2007).

[33] Deeb O., Goodarzi M., Quantum Chemical QSAR Models to Distinguish Between Inhibitory Activities of Sulfonamides Against Human Carbonic Anhydrases I and II and Bovine IV Isozymes, *Chem Biol Drug Des.,* **79**, p. 514 (2011).

[34] Jantschi L., Daniela Bolboaca S., Modelling the Inhibitory Activity on Carbonic Anhydrase IV of Substituted Thiadiazole- and Thiadiazoline-Disulfonamides: Integration of Structure Information, *Rev Electron Biomed / Electron J Biomed*, **2**, p. 22 (2006).

[35] Clare B.W., Supuran C.T., A Perspective on Quantitative Structure-Activity Relationships and Carbonic Anhydrase Inhibitors, *Expert Opin Drug Metab Toxicol*., **2**, p. 113 (2006).

[36] Mattioni B.E., Jurs P.C., Development of Quantitative Structure-Activity Relationship and Classification Models for a Set of Carbonic Anhydrase Inhibitors, *J. Chem. Inf. Comput. Sci*., **42**, P. 94 (2002).

[37] Bakken G.A., Jurs P.C., QSARs for 6-Azasteroids as Inhibitors of Human Type 1 5r-Reductase: Prediction of Binding Affinity and Selectivity Relative to 3-BHSD, *J. Chem. Inf. Comput. Sci*., **41**, p. 255 (2001).