Chem Biol Drug Des 2012; **79:** 514–522

Research Article

doi: 10.1111/j.1747-0285.2011.01309.x

Quantum Chemical QSAR Models to Distinguish Between Inhibitory Activities of Sulfonamides Against Human Carbonic Anhydrases I and II and Bovine IV Isozymes

Omar Deeb[1,*], Mohammad Goodarzi[2,3] and Padmaker V. Khadikar[4]

[1]*Faculty of Pharmacy, Al-Quds University, P.O. Box 20002 Jerusalem, Palestine*
[2]*Faculty of Sciences, Department of Chemistry, Islamic Azad University, Arak Branch, P.O. Box 38135-567 Arak, Markazi, Iran*
[3]*Young Researchers Club, Islamic Azad University, Arak Branch, P.O. Box 38135-567 Arak, Markazi, Iran*
[4]*Research Division, Laxmi Fumigation and Pest Control Pvt. Ltd., 3, Khatipura, Indore 452 007, Madhya Pradesh, India*
**Corresponding author: Omar Deeb, deeb2000il@yahoo.com*

Received 14 March 2011, revised 2 December 2011 and accepted for publication 12 December 2011

Linear and nonlinear quantitative structure activity relationship models for predicting the inhibitory activities of sulfonamides toward different carbonic anhydrase isozymes were developed based on multilinear regression, principal component-artificial neural network and correlation ranking-principal component analysis, to identify a set of structurally based numerical descriptors. Multilinear regression was used to build linear quantitative structure activity relationship models using 53 compounds with their quantum chemical descriptors. For each type of isozyme, separate quantitative structure activity relationship models were obtained. It was found that the hydration energy plays a significant role in the binding of ligands to the CAI isozyme, whereas the presence of five-membered ring was detected as a major factor for the binding to the CAII isozyme. It was also found that the softness exhibited significant effect on the binding to CAIV isozyme. Principal component-artificial neural network and correlation ranking-principal component analysis analyses provide models with better prediction capability for the three types of the carbonic anhydrase isozyme inhibitory activity than those obtained by multilinear regression analysis. The best models, with improved prediction capability, were obtained for the hCAII isozyme activity. Models predictivity was evaluated by cross-validation, using an external test set and chance correlation test.

Key words: carbonic anhydrase isozymes and inhibitors, correlation ranking-principal component analysis, principal component-artificial neural network, quantitative structure activity relationship, quantum chemical descriptors

Thousands of different aromatic and heterocyclic sulfonamides carbonic anhydrase (CA, EC 4.2.1.1) inhibitors were synthesized in the exploration of diverse pharmacological agents (1,2), but the number of amino acid/oligopeptide derivatives among them is unexpectedly small. Accordingly, a series of 53 compounds were synthesized by Mincione *et al.* (3) and investigated for their inhibitory activity against physiologically relevant CA isozymes, such as CAI, II, and IV. The syntheses involved the reaction of 26 aromatic and heterocyclic sulfonamides containing amino, imino, hydrazine, or hydroxyl groups with *N*-tert-butoxycarbonyl- c-aminobutyric acid (Boc-GABA) in the presence of carbodiimide derivatives. The resulting water-soluble compounds were assayed as inhibitors of the cytosolic isozymes hCAI and II, and the membrane-bound form bCAIV, which were involved in important physiological processes, for example, respiration and transport of CO_2/bicarbonate between metabolizing tissues and lungs, pH and CO_2 homeostasis, electrolyte secretion in a variety of tissues/organs, biosynthetic reactions (such as gluconeogenesis, lipogenesis, and ureagenesis), bone resorption, calcification, tumorigenicity, and many other physiologic/pathologic processes (1,4–8).

In quantitative structure activity relationship (QSAR) models, the correlation between experimental values of the activity and descriptors reflecting the molecular structure of the compounds is obtained. To achieve a significant correlation, it is essential that proper descriptors are used. A wide variety of molecular descriptors are used in QSAR models (9). 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 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 (10,11). An ANN is formed from artificial neuron arranged in layers, connected 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 the literature, there have been a number of QSAR studies of sulfonamides using quantum chemical (12–17) and topological (18–23) descriptors and 3-D approach of CoMFA and CoMSIA (24,25). Recently, a QSAR study for the inhibitory activity of the transmembrane CA isozyme XIV with sulfonamides using PRECLAV software has been carried out by Khadikar *et al.* (26). The obtained QSAR equations pointed out the fact that the CA inhibitory activity decreased for unsubstituted (at the organic scaffold) aromatic/heteroaromatic sulfonamides, but was favored by the presence of alkyl groups substituting the scaffold, which led to a higher internal topological diversity, as well as by the presence of condensed aromatic rings in the structure of these enzyme inhibitors.

Recent QSAR studies on this class of compounds revealed that the CA inhibitory activities of such compounds could be modeled successfully using connectivity and indicator indices (18) as well as quantum chemical descriptors (17). Incited by such studies, and in continuation with the previous studies (27–29), a combination of these descriptors was used in this study for modeling inhibitory activities against all the three isozymes, that is, hCAI, hCAII, and bCAIV. Where, in this study, linear, multiple linear regression (MLR) and correlation ranking-principal component regression (CR-PCR), and nonlinear, principal component-artificial neural networks (PC-ANN), methods were applied with the aim of obtaining the most appropriate models for predicting isozyme selectivity of the compounds.

Materials and Methods

Software

Geometry optimization was performed by HyperChem[a] (Version 7.0; Hypercube, Inc) at the Austin model 1 (AM1), semiempirical method level. An AM1 optimization was chosen as it was developed and parameterized for common organic structures. Descriptors were calculated using HyperChem[a] and Dragon[b] software (Milano Chemometrics and QSPR Group). SPSS[c] Software was used for the simple MLR analysis. Principal component analysis (PCA), PC-ANN, and CR-PCR were performed in the MATLAB[d] environment.

Chemical data and descriptors

Table S1 in the Supporting Information shows the structural details of sulfonamides used in this study while Table S2 in the Supporting Information includes their inhibitory activities, log K_i(hCAI), log K_i(hCAII), and log K_i(bCAIV) (18). Chemical structures of these compounds were obtained from HyperChem software and optimized on AM1 semiempirical level. The Optimization was preceded by the Polak-Rebiere algorithm to reach 0.01 root mean square gradient. In this study, 36 quantum chemical descriptors including indicator descriptors were calculated using HyperChem and Dragon software, and these descriptors are the following: HOMO, LUMO, EN, HD, SOF, EPH, HE, HF, volume, mass, pol, ref, SA(appr.), logP, DMx, DMy, DMz, DMt, qpos, qneq, Qpos, Qneq, Qtot, Qmean, Q^2, RPCG, RNCG, SPP, TE1, TE2, PCWTe, LDip, 1vv, Ip1, Ip2 and Ip3. Table S3 in the Supporting Information shows a brief description of these descriptors used in this study.

Three sets of the calculated descriptors and the three types of CA isozyme activities such as I, II, and IV were gathered in a separate data matrix D_i with a dimension of ($m \times n$), where m and n being the number of molecules and the number of descriptors, respectively. In each group, the calculated descriptors were examined for the presence of constant or near-constant values for all molecules and those detected were removed. To decrease the redundancy that existed in the descriptor data matrix, the correlation among descriptors was examined and the detected collinear descriptors [i.e., coefficient of determination (R^2) ≥ 0.95] were removed from the data matrix. Then, the different sets of activities and the molecular descriptors were subjected to the MLR, PCA, PC-ANN, and CR-PCR analyses as described later.

MLR analysis

Multiple linear regression analysis was employed to model the inhibitory activities for each type of CA isozymes [log K_i(hCAI), log K_i(hCAII) and log K_i(bCAIV)] relationships with the set of quantum chemical descriptors. Multiple linear regression analysis was performed using the method of maximum-R^2 with stepwise selection and elimination of variables (30) on a training set composed of around 75% of the data set. After the model has been proposed using the training set, its predictivity was tested by making predictions against the test set (around 25% of the data). Data division was performed according to the PCA as it is discussed later.

Principal component analysis

The different sets of activities and the molecular descriptors were subjected to PCA. The PCs were calculated by singular value decomposition method in MATLAB environment (MathWork Inc). Before applying the multivariate analysis methods, and owing to the quality of data, a previous treatment of the data is essential.

Orthogonal transformation of the descriptors by PCA was performed to account for some collinearity between the descriptors. It is essential to carry out a previous treatment of the data such as scaling and centering before applying the regression analysis combined with feature extraction (i.e., PCR and PC-ANN). The outcome of projection methods depends on the normalization of the data. Descriptors with small absolute values slightly contribute to overall variances; such biases toward other descriptors with higher absolute values originate biased PCs. Therefore, equal weights are allocated to each descriptor, with appropriate scaling, so that the significant variables in the model are in focus. The descriptors are standardized to unit variance and zero mean (autoscaling), to give all variables the same significance. Application of the PCA on the calculated descriptors and activity data matrix resulted in 37 factors or principal components (PC1–PC37).

Consequently, each of the three data matrices that contain all the descriptors with each of the CA isozyme I, II, and IV inhibitory activities was subjected to PCA separately, and the first two principal components (PCs) outcome from each separate PCA were plotted against each other. Figure 1 represents the factor spaces of the descriptors and the CAI inhibitory activity where each point represents one compound. Figure 1 shows that compounds **39** and **40** are out-

liers implying that these two compounds behave differently from the rest of compounds investigated in this study. Hence, these compounds were excluded in the regression analyses applied in this study.

To examine the final model's performance, a homogenous set of 14 molecules (around 25% of the data set) were selected as prediction samples from the points in the resulted plot. These samples were selected based on descriptors spaces obtained from plotting the first and second PCs as it was described earlier, (see Figure 1). Among the points in the resulting plot, homogenous sets were selected with the 75% portion for the training (or calibration) set.

Artificial neural network

In contrast to MLR, the ANN is capable of recognizing highly nonlinear relationships. The flexibility of ANN enables it to discover more complex relationships in experimental data, when it is compared with the traditional statistical models. The PC-ANN analysis was proposed by Gemperline *et al.* (31), to improve the training speed and decrease the overall calibration error. In this method (31), as a preliminary treatment, the input data (i.e., molecular descriptors) were normalized so as to have zero mean and unity variance as it was mentioned earlier. It should be noticed that for each MLR model, a separate ANN model was developed so that the input's descriptors were the subsets selected by the stepwise MLR methods. In the case of each MLR model, a feedforward neural network with back-propagation of error algorithm was constructed to model the property structure relationships between the descriptors, on the one hand, and the activity data of sulfonamides, on the other hand. More details about the model development in ANN and the network architecture are explained in references (32–34). Overfitting problem or poor generalization capability happens when a neural network overlearns during the training period. A too well-trained model may not perform well on unseen data set because of its lack of generalization capability. 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 regression between the network output and the property was calculated for the two sets individually. The training function 'trainscg' in MATLAB was used to train the network. To find the models with lower errors, the ANN algorithm was run many times, each time run with different geometry and/or initial weights.

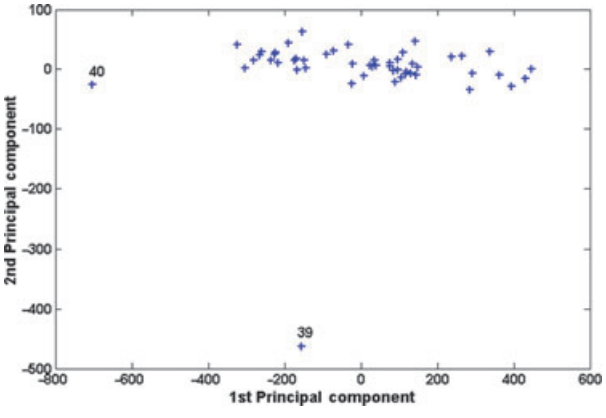

Figure 1: First and second principal components for the factor spaces of the descriptors and CAI inhibitory activity data.

Correlation ranking-principal component regression

In this approach, the best set of factors was selected by the correlation ranking (CR) procedures. In the CR-PCR, the correlation between each one of the extracted PCs with the inhibitory activities for each type of CA isozymes [log K_i(hCAI), log K_i(hCAII) and log K_i(bCAIV)], separately, was determined first. For each of the isozyme inhibitory activities, the resulting regression model was used to predict the activity of the test set compounds. The square of the correlation coefficient between the predicted and actual activities (R_P^2, that is, the amount of the variances in the activity, which can be explained by each PC) was calculated, and this quantity was used as a measure of the correlation ability of each PC. Then, the PCs were ranked in the order of decreasing correlation and entered to the regression model one after another. This procedure is well illustrated in the study by (35–37). That is, the stepwise entrance of the PCs to the PCR models was based on their decreasing correlation with the desired activity. Some statistical parameters such as the squared of the correlation coefficient (R^2), squared of the leave-one-out cross-validation correlation coefficient (R_{CV}^2), and the root mean square error (RMSE) were calculated to estimate the quality of the resulted models. Different models were obtained for each inhibitory activity type of CA isozymes [log K_i(hCAI), log K_i(hCAII), and log K_i(bCAIV)].

Results and Discussion

The sulfonamides (18) used in this study were investigated for their inhibitory activities of log K_i(hCAI), log K_i(hCAII), and log K_i(hCAIV), (see Table S1 and S2 in the Supporting Information), using MLR, PC-ANN, and CR-PCR analyses. The results of these analyses are discussed later.

MLR analysis

Quantitative structure activity relationship studies on the inhibitory activity of a set of sulfonamide derivatives toward three different isozymes of CA helped us to find the structural requirement of the sulfonamide ligands for binding to the isozymes. Table 1 shows the regression models suggested from MLR analysis. Generally, it was found that the type of ligand – receptor interactions – differs from one isozyme to another, which implies that by considering these interactions, it is possible to design selective ligands toward a specified CA isozyme (38–41).

The QSAR model obtained for CAI isozyme indicated that the hydration energy plays a significant role in the binding of ligands to the CAI isozyme. Molecules with higher hydration energies were found to bind to the receptor strongly. This proposes the presence of polar residues in the binding pocket of the isozyme. In addition, it was found that the first-order valence connectivity index ($^1\chi^v$), a

Table 1: Regression models suggested by multiple linear regression analysis

Activity	Regression model
hCAI	log K_i(hCAI) = 7.240 (±0.508) − 0.845 (±0.092) × $^1\chi^v$ + 0.114 (±0.018) × TE1 + 2.181 (±0.737) × HE − 0.066 (±0.037) × DMz
hCAII	log K_i (hCAII) = 3.097 (±0.194) − 0.280 (±0.056) × $^1\chi^v$ − 0.634 (±0.139) × Ip3 − 0.047 (±0.023) × DMy + 0.024 (±0.010) × TE1 + 0.033 (±0.017) × DMx − 0.224 (±0.141) × Ip2
bCAIV	log K_i (hCAIV) = 7.141 (±0.768) − 0.509 (±0.046) × $^1\chi^v$ + 0.139 (±0.029) × TE1 − 11.193 (±3.637) × SOF − 0.441 (±0.111) × Ip2 − 0.229 (±0.100) × Qtot + 0.068 (±0.033) × HF

molecular connectivity descriptor that is based exclusively on bonding and branching patterns, plays a role in the binding to the receptor. It was found that higher molecular connectivity tends to block the drug binding to the receptor.

On the other hand, for the binding of sulfonamides to CAII isozyme, it was found that the presence of five-membered ring plays the most significant role. The presence of such rings in the sulfonamide derivatives blocks their binding to the receptor. Similar to what was found for the binding of sulfonamides to CAI isozyme, the first-order valence connectivity index ($^1\chi^v$) plays a role in the binding to the receptor where higher molecular connectivity tends to block the drug binding to the receptor.

For the CAIV isozyme, the obtained QSAR model showed the extremely significant role of softness, so that the more polarizable is the ligand, the stronger it binds to the receptor. This QSAR model suggested the importance of acid–base interactions in the binding of sulfonamides to the isozyme. Again, the molecular connectivity was obtained as another controlling factor in ligand–receptor binding for this QSAR model. The effect of this descriptor was found to be in the same order of that found for CAII isozyme. Another significant descriptor in this model is the Ip2 (indicator of the halogen presence in the sulfonamide moieties) where it was found that the presence of halogen in the sulfonamide moieties blocks the binding to the CAIV isozyme.

Finally, Table 1 shows that the first-order valence connectivity index ($^1\chi^v$) and the topographic electronic descriptor (TE1) play a role in the binding to the receptor for all the CA isozyme types, I, II and IV, although the contribution of the former descriptor to the binding of the receptor for the CA isozymes is more important than that of the latter.

Table 2 shows that the lowest root mean square error (RMSE) of prediction and calibration ($RMSE_C$ = 0.327 and $RMSE_P$ = 0.379) are obtained for the regression model of the log K_i(hCAII). The calibration and cross-validation coefficients of determination (R^2_C and R^2_{CV}, respectively) obtained for this data set are both 0.718 while the prediction coefficient of determination (R^2_P) is 0.701. The linear relationships found by MLR analysis provide models with good cross-validation parameters. The coefficient of determination of prediction is close to the coefficient of determination of calibration, which is good evidence that the models are not overfitted. For the model to be overfitted, it is to be expected that the fitted relationship will appear to perform less well on a new unseen data set (prediction set) than on the data set used for fitting (calibration set). In particular, the value of the prediction coefficient of determination will shrink relative to the original calibration data, which is not the case here.

Table 2: Regression and cross-validation parameters of the models suggested by the MLR, PC-ANN, and CR-PCR analyses

	hCAI	hCAII	bCAIV
MLR analysis			
$RMSE_C$	0.576	0.327	0.379
$RMSE_P$	0.733	0.379	0.388
R^2_C	0.778	0.718	0.704
R^2_P	0.745	0.701	0.820
R^2_{CV}	0.778	0.718	0.704
PC-ANN analysis			
$RMSE_C$	0.732	0.255	0.343
$RMSE_P$	1.014	0.270	0.488
R^2_C	0.647	0.831	0.774
R^2_P	0.661	0.838	0.755
R^2_{CV}	0.323	0.779	0.572
CR-PCR analysis			
$RMSE_C$	0.541	0.282	0.305
$RMSE_P$	0.624	0.310	0.329
R^2_C	0.785	0.781	0.748
R^2_P	0.878	0.838	0.869
R^2_{CV}	0.726	0.719	0.664

PC-ANN, principal component-artificial neural networks; CR-PCR, correlation ranking-principal component regression; MLR, multiple linear regression.

Table S4 in the Supporting Information shows the results for randomization test performed to investigate the probability of chance correlation for the models obtained using the MLR analysis. The low value of the coefficients of determination obtained from the randomization test suggests that the QSAR models discussed earlier have not been obtained by chance.

An effective approach to improve the predictive power of simple linear equations was suggested in (42). This can be achieved by adjusting the data to higher-order fitting polynomials to generalize first-order multivariate formulas. Therefore, the obtained models were further investigated using the PC-ANN analysis as discussed later.

Principal component-artificial neural networks

The inputs of the ANN were the subset of the descriptors used in different MLR models (Table 1). A three-layered feedforward ANN model with back-propagation learning algorithm (43) was employed. First, the nonlinear relationship between the subset of descriptors selected by stepwise selection-based MLR (Table 1) and the inhibitory activities of log K_i(hCAI), log K_i(hCAII), and log K_i(hCAIV) values was preceded by ANN models with similar structure. The number of hidden layer's nodes was set 6 for all models, and the number of nodes in the input layer was the number of PCs extracted for each subset of descriptors.

Then, to optimize the performance of the suggested ANN models, we trained the ANN using different number of hidden nodes starting from 2 to 15 hidden nodes. The selection of the optimal number of hidden nodes was made according to the following two criteria:

- The assessment of the predictive ability of a multivariate calibration model is based on the determination of the minimum prediction error (44).
- Large numbers of hidden nodes often draw attention to the possibility of having overfitted model (45).

Therefore, the RMSE of prediction is the parameter considered to decide on the optimal model. When deciding on the optimal number of hidden nodes for each model, the models obtained using small numbers of hidden nodes were favored over those obtained using large number of hidden nodes,

Figure S1 in the Supporting Information shows plots of $RMSE_P$ against number of hidden nodes for the models obtained for the three isozyme activities. This figure shows the optimal number of hidden nodes obtained for the three isozyme inhibitor activities I, II, and IV are 9, 8, and 6 hidden nodes, respectively. Changing the number of hidden nodes affects the prediction accuracy of the model much more than it does for the calibration (or generalization) accuracy of the models (see Table S5 in the Supporting Information). It can be noticed from Figure S1 in the Supporting Information that the $RMSE_P$ for the ANN model obtained for the CAI isozyme inhibitor activity model is larger than that obtained for the CAII and CAIV inhibitor activities. The results of the optimal models obtained by the PC-ANN analysis are given in Table 2. This table shows that the lowest $RMSE_P$ and $RMSE_C$ (0.270 and 0.255, respectively) are obtained for modeling the log K_i (hCAII) data set. The R^2_C and R^2_{CV} obtained for this data set are 0.831 and 0.779, respectively while the R^2_P for this model is 0.838. Following the same argument used in discussing the MLR models, it is to be concluded that the PC-ANN model is not overfitted as the coefficient of determination of prediction is larger than the coefficient of determination of calibration.

Generally, the nonlinear relationships according to ANN analysis provide models with better regression coefficients and cross-validation parameters compared with MLR analysis. Table S6 in the Supporting Information shows the results for randomization test performed to investigate the probability of chance correlation for the optimal models obtained by the PC-ANN analysis. The low values of the coefficients of determination and high values of RMSE obtained from the chance correlation test prove that the ANN models have not been obtained by chance.

Correlation ranking-principal component regression

The R^2_P of each PC with the inhibitory activity was used for ranking the extracted PCs. The same approach was applied to each of the log K_i(hCAI), log K_i(hCAII), and log K_i(hCAIV) inhibitory activities, separately. Figure 2 and Table S7 in the Supporting Information show the ranking of the PCs obtained by PCR for each inhibitory activity according to their R^2_P values. Then, these PCs were entered to the PCR models successively, according to their R^2_P values. The number of PCs in the regression models suggested from PCR analysis for the PCs extracted is varied between 1 and 37. The evolution of R^2_P and $RMSE_P$ as the function of number of PC entered to the regression model is plotted in Figure S2A,B in the Supporting Information, respectively. This figure shows that the R^2_{CV} values are increasing while the $RMSE_P$ values are decreasing with increasing model number (where more PCs are added to the regression model) until it reaches some plateau. This figure also shows that by successive addition of PCs to the inputs of the PCR, the model performance increased for up to 9, 6, and 7 PCs for the inhibitory activities of log K_i(hCAI), log K_i(hCAII), and log K_i(hCAIV), respectively. The corresponding PCs used in each regression model for each of the CA I, II, and IV isozyme inhibitory activities are (PC18 + PC9 + PC24 + PC1 + PC7 + PC14 + PC2 + PC10 + PC13), (PC17 + PC13 + PC10 + PC9 + PC24 + PC1), and (PC10 + PC1 + PC17 + PC18 + PC9

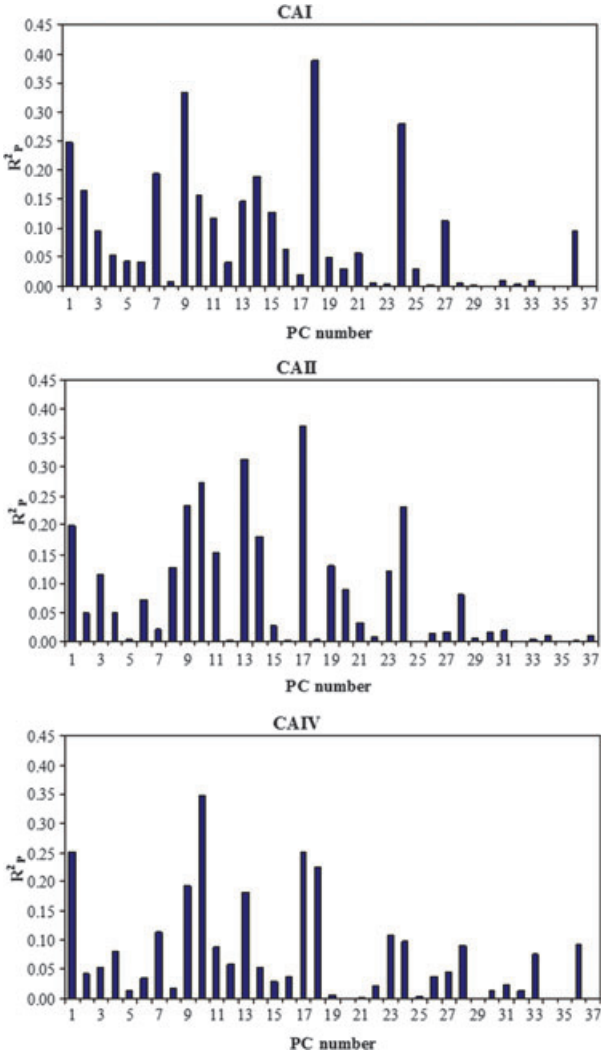

Figure 2: Principal components extracted from the CA inhibitory activities: log K_i(hCAI), log K_i(hCAII), and log K_i(bCAIV) datum using the principal component analysis approach ranked according to their R^2_P values.

+ PC13 + PC7 + PC23), respectively. The predictive abilities of the models were not enhanced significantly by adding more PCs to the regression models. Table 2 shows the regression and cross-validation parameters obtained when employing the CR-PCR analysis for each of the inhibitory activities of log K_i(hCAI), log K_i(hCAII), and log K_i(hCAIV). This table shows that the regression model obtained for the CAII isozyme inhibitory activity has the lowest RMSE values ($RMSE_C$ = 0.282 and $RMSE_P$ = 0.310). The R_C^2, R_{CV}^2, and R_P^2 obtained for this model are 0.781, 0.719, and 0.838, respectively. Again, the larger coefficient of determination of prediction compared with that of calibration indicates that the model is not overfitted.

Generally, the linear relationships according to CR-PCR analysis provide models with good cross-validation parameters. Table S8 in the Supporting Information shows the results for randomization test performed to investigate the probability of chance correlation for the optimal model using the CR-PCR analysis. The low values of the coefficients of determination and high RMSE values obtained from the chance correlation test prove that the QSAR models obtained by the CR-PCR analysis are better than those obtained by chance.

In summary, both the first-order valence connectivity index ($^1\chi^v$) and the topographic electronic descriptor (TE1) play a role in the binding to the receptor for all the CA isozyme types: I, II and IV. However, the contribution of the ($^1\chi^v$) is more important than that of the TE1. Furthermore, the models equations in Table 1 suggest that the inhibition activities of the CA isozyme types I, II, and IV increase with decreasing the first-order valence connectivity index ($^1\chi^v$) values and with increasing the topographic electronic descriptor (TE1) values. The nonlinear relationship between the subset of descriptors selected by stepwise selection-based MLR (Table 1), and the inhibitory activities of log K_i(hCAI), log K_i(hCAII) and log K_i(hCAIV) values were preceded by ANN models with similar structure. However, for the PC-ANN as well as for the CR-PCR methods, we used the number of PCs extracted for each subset of the descriptors used in different MLR models. Therefore, the comparison between the descriptors used in these methods will not be direct.

Comparing the linear, MLR and CR-PCR, and nonlinear, PC-ANN, methods applied in this study, one can see that the PC-ANN analysis provides models with better prediction ability than those obtained by MLR and CR-PCR analysis, considering the CAII isozyme inhibitory activity model. Nevertheless, the CR-PCR analysis provides models with the lowest prediction and calibration RMSE values for the three types of isozyme inhibitory activities in general.

The linear MLR analysis provides models with higher regression coefficients and cross-validation parameters, compared with CR-PCR and PC-ANN analyses, for the CAIV isozyme inhibitory activity while the CR-PCR analysis provides models with higher regression coefficients and cross-validation parameters, compared with MLR and PC-ANN analyses, for the CAI isozyme inhibitory activity. Furthermore, the nonlinear PC-ANN analysis provides models with higher regression coefficients and cross-validation parameters compared with MLR and CR-PCR analyses, for the CAII isozyme inhibitory activity. In summary, CR-PCR analysis provides models with better prediction capability for the three types of the CA isozyme inhibitory activity. The hCAII isozyme models obtained using the different statistical analysis applied in this study are superior over the models obtained for hCAI and bCAIV isozyme inhibitory activities.

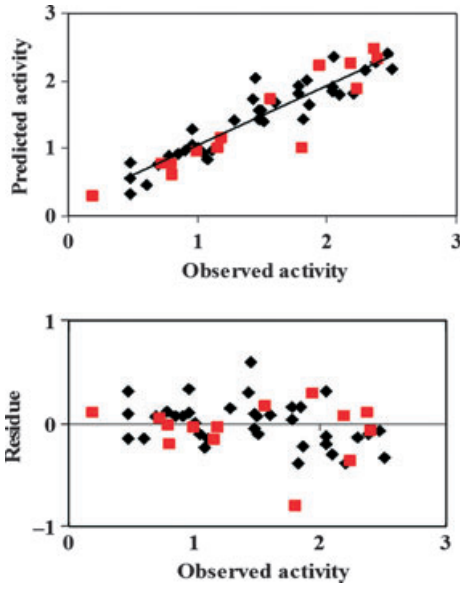

Figure 3: Plot of the predicted log K_i(hCAI), log K_i(hCAII), and log K_i(bCAIV) against observed ones and their residuals obtained from the correlation ranking-principal component regression analysis. Black diamonds and red squares indicate data from the calibration and prediction sets, respectively.

Table S2 in the Supporting Information shows the observed, predicted inhibitory activities for the three types of CA isozymes as well as their residues as obtained from the MLR, PC-ANN, and CR-PCR analyses. Figure 3 shows the predicted and observed inhibitory activities obtained by the CR-PCR analysis for the CAII isozyme inhibitory activity. The results obtained in this study agree with the results reported in (17), in signifying the importance of acid–base interactions in the binding of sulfonamides to the isozyme.

Conclusions

A QSAR analysis has been performed on three types of CA isozyme inhibitory activities for 53 sulfonamides using MLR, PC-ANN, and CR-PCR analyses. It was observed that the interaction between the ligand and receptor varies from one type of CA isozyme to another. The latter finding proposes the possibility of designing selective ligands toward a specified CA isozyme by taking such interactions into consideration (38–41). The results obtained offers very good regression models that hold good prediction ability. Correlation ranking-principal component regression analysis provides models with better prediction capability for the three types of the CA isozyme inhibitory activity while PC-ANN analysis provides models with better prediction capability for the hCAII isozyme activity. Generally, the models obtained for modeling the hCAII isozyme inhibitory activity are superior over those obtained for modeling the hCAI and bCAIV isozyme inhibitory activities.

The QSAR model obtained for CAI indicated that the molecules with higher hydration energies were found to bind to the receptor strongly. In addition, it was found that higher molecular connectivity tends to block the drug binding to the receptor. On the other hand, for the binding of sulfonamides to CAII isozyme, it was found that the presence of such rings in the sulfonamides blocks their binding to the receptor. For the CAIV isozyme, the obtained QSAR model illustrates the extremely significant role of softness, so that the more polarizable is the ligand, the stronger it binds to the receptor.

It was found that higher molecular connectivity tends to block the drug binding to the receptor for the three types of CA isozyme inhibitory activities. The results obtained show that linear and nonlinear regression analyses are useful tools to distinguish between the inhibitory activities of sulfonamides toward different CA isozyme types I, II, and IV.

References

1. Supuran C.T., Scozzafava A., Conway J., editors (2004) Carbonic Anhydrase – Its Inhibitors and Activators. Boca Raton, USA: CRC Press; p. 1–363.
2. Supuran C.T. (2010) Carbonic anhydrase inhibitors. Bioorg Med Chem Lett;20:3467–3474.
3. Mincione G., Menabuoni L., Briganti F., Mincione F., Scozzafava A., Supuran C.T. (1999) Carbonic anhydrase inhibitors. Part 79: synthesis of topically acting sulfonamides incorporating GABA moieties in their molecule, with long-lasting intraocular pressure-lowering properties. Eur J Pharm Sci;9:185–199.
4. Supuran C.T. (2008) Carbonic anhydrases: novel therapeutic applications for inhibitors and activators. Nat Rev Drug Discov;7:168–181.
5. Mincione F., Menabuoni L., Supuran C.T. (2004) Clinical applications of the carbonic anhydrase inhibitors in ophthalmology. In: Supuran C., Scozzafava A., Conway J., editors. Carbonic Anhydrase, Its Inhibitors and Activators. Boca Raton, USA: CRC Press; p. 243–254.
6. Clare B.W., Supuran C.T. (2004) QSAR studies of sulfonamide carbonic anhydrase inhibitors. In: Supuran C.T., Scozzafava A., Conway J., editors. Carbonic Anhydrase, Its Inhibitors and Activators. Boca Raton, USA: CRC Press; p. 149–182.
7. Supuran C.T., Scozzafava A., Casini A. (2004) Development of sulfonamide carbonic anhydrase inhibitors (CAIs). In: Supuran C.T., Scozzafava A., Conway J., editors. Carbonic Anhydrase, Its Inhibitors and Activators. Boca Raton, FL, USA: CRC Press; p. 67–148.
8. Supuran C.T., Scozzafava A., Casini A. (2003) Carbonic anhydrase inhibitors. Med Res Rev;23:146–189.
9. Todeschini R., Consonni V. (2000) Handbook of Molecular Descriptors in Methods and Principles in Medicinal Chemistry. Mannhold R., Kubinyi H., Timmerman H., Editors. Weinheim, Germany: Wiley-VCH. p. 1–667.
10. Despagne F., Massart D.L. (1998) Tutorial review: neural networks in multivariate calibration. Analyst;123:157R–178R.
11. Zupan J., Gasteiger J. (1999) Neural Networks in Chemistry and Drug Design. Weinheim, Germany: Wiley-VCH.
12. Clare B.W., Supuran C.T. (2000) Carbonic anhydrase inhibitors. Part 86. A QSAR study on some sulfonamide drugs which lower intra-ocular pressure, using the ACE non-linear statistical method. Eur J Med Chem;35:859–865.
13. Clare B.W., Supuran C.T. (1999) Carbonic anhydrase inhibitors. Part 61. Quantum chemical QSAR of a group of benzenedisulfonamides. Eur J Med Chem;34:463–474.
14. Supuran C.T., Clare B.W. (1999) Carbonic anhydrase inhibitors – Part 57: quantum chemical QSAR of a group of 1,3,4-thiadiazole- and 1,3,4-thiadiazoline disulfonamides with carbonic anhydrase inhibitory properties. Eur J Med Chem;34:41–50.
15. Eroglu E., Türkmen H. (2007) A DFT-based quantum theoretic QSAR study of aromatic and heterocyclic sulfonamides as carbonic anhydrase inhibitors against isozyme, CA-II. J Mol Graph Model;26:701–708.
16. Eroglu E., Türkmen H., Güler S., Palaz S., Oltulu O. (2007) A DFT-based QSARs study of acetazolamide/sulfanilamide derivatives with carbonic anhydrase (CA-II) isozyme inhibitory activity. Int J Mol Sci;8:145–155.
17. Hemmateenejad B., Miri R., Jafarpour M., Tabarzad M., Shamsipur M. (2007) Exploring QSAR for the inhibitory activity of a large set of aromatic/heterocyclic sulfonamides toward four different isozymes of carbonic anhydrase. QSAR Comb Sci;26:1065–1075.
18. Agrawal V.K., Singh J., Khadikar P.V., Supuran C.T. (2006) QSAR study on topically acting sulfonamides incorporating GABA moieties: a molecular connectivity approach. Bioorg Med Chem Lett;16:2044–2051.
19. Melagraki G., Afantitis A., Sarimveis H., Iglessi-Markopoulou O., Supuran C.T. (2006) QSAR study on para-substituted aromatic sulfonamides as carbonic anhydrase II inhibitors using topological information indices. Bioorg Med Chem;14:1108–1114.
20. Agrawal V.K., Banerji M., Gupta M., Singh J., Khadikar P.V., Supuran C.T. (2005) QSAR study on carbonic anhydrase inhibitors: water-soluble sulfonamides incorporating β-alanyl moieties, possessing long lasting-intra ocular pressure lowering properties-a molecular connectivity approach. Eur J Med Chem;40:1002–1012.
21. Khadikar P.V., Sharma V., Karmarkar S., Supuran C.T. (2005) QSAR studies on benzene sulfonamide carbonic anhydrase inhibitors: need of hydrophobic parameter for topological modeling of binding constants of sulfonamides to human CA-II. Bioorg Med Chem Lett;15:923–930.
22. Agrawal V.K., Sharma R., Khadikar P.V. (2002) QSAR studies on carbonic anhydrase inhibitors: a case of ureido and thioureido derivatives of aromatic/heterocyclic sulfonamides. Bioorg Med Chem;10:2993–2999.
23. Khadikar P.V., Clare B.W., Balaban A.T., Supuran C.T., Agrawal V.K., Singh J., Joshi A.K., Lakwani M. (2006) QSAR modeling of carbonic anhydrase-I, -II and -IV inhibitory activities: relative correlation potential of six topological indices. Rev Roum De Chim;51:703–717.
24. Huang H., Pan X., Tan N., Zeng N., Ji C. (2007) 3D-QSAR study of sulfonamide inhibitors of humancarbonic anhydrase II. Eur J Med Chem;42:365–372.
25. Sethi K.K., Verma S.M., Prasanthi N., Sahoo S.K., Parhi R.N., Suresh P. (2010) 3D-QSAR study of benzene sulfonamide analogs

as carbonic anhydrase II inhibitors. Bioorg Med Chem Lett;20:3089–3093.
26. Singh S., Khadikar P.V., Scozzafava A., Supuran C.T. (2009) QSAR studies for the inhibitory of the transmembrane carbonic anhydrase isozyme XIV with sulfonamides using PRECLAV software. J Enz Inhib Med Chem;24:337–349.
27. Deeb O., Hemmateenejad B. (2007) ANN-QSAR model of drug-binding to human serum albumin. Chem Biol Drug Des;70:19–29.
28. Deeb O., Hemmateenejad B., Jaber A., Garduno-Juarez R., Miri R. (2007) Effect of the electronic and physicochemical parameters on the carcinogenesis activity of some sulfa drugs using QSAR Analysis based on genetic-MLR and genetic-PLS. Chemosphere;67:2122–2130.
29. Deeb O., Goodarzi M. (2010) Predicting the solubility of pesticide compounds in water using QSPR methods. Mol Phys;108:181–192.
30. Chaterjee S., Hadi A.S., Price B.. (2000) Regression Analysis by Examples, 3rd edn New York: Wiley.
31. Gemperline P.J., Long J.R., Gregoriou G. (1991) Nonlinear multivariate calibration using principal components regression and artificial neural networks. Anal Chem;63:2313–2323.
32. Deeb O., Drabh M. (2010) Exploring QSARs of some analgesic compounds by PC-ANN. Chem Biol Drug Des;76:255–262.
33. Hemmateenejad B., Shamsipur B. (2004) Quantitative structure-electrochemistry relationship study of some Organic compounds using PC-ANN and PCR. Internet electron J Mol Des;3:316–334.
34. Hemmateenejad B., Safarpour M.A., Miri R., Nesari N. (2005) Toward an optimal procedure for PC-ANN model building: prediction of the carcinogenic activity of a large set of drugs. J Chem Inf Model;45:190–199.
35. Deeb O. (2010) Correlation ranking and stepwise regression procedures in principal components artificial neural networks modeling with application to predict toxic activity and human serum albumin binding affinity. Chemometr Intell Lab Syst;104:181–194.
36. Shamsipur M., Ghavami R., Sharghi H., Hemmateenejad B. (2008) Highly correlating distance/connectivity-based topological indices 5. Accurate prediction of liquid density of organic molecules using PCR and PC-ANN. J Mol Graph Model;27:506–511.
37. Feng J., Lurati L., Ouyang H., Robinson T., Wang Y., Yuan Y., Young S. (2003) Predictive toxicology: benchmarking molecular descriptors and statistical methods. J Chem Inf Comput Sci;43:1463–1470.
38. Supuran C.T., Nicolae A., Popescu A. (1996) Carbonic anhydrase inhibitors. Part 35. Synthesis of Schiff bases derived from sulfanilamide and aromatic aldehydes: the first inhibitors with equally high affinity towards cytosolic and membrane-bound isozymes. Eur J Med Chem;31:431–438.
39. Supuran C.T., Popescu A., Ilisiu M., Costandache A., Banciu M.D. (1996) Carbonic anhydrase inhibitors. Part 36. Inhibition of isozymes I and II with Schiff bases derived from chalkones and aromatic/heterocyclic sulfonamides. Eur J Med Chem;31:439–447.
40. Supuran C.T., Scozzafava A., Popescu A., Bobes-Tureac R. (1997) Carbonic anhydrase inhibitors. Part 43. Schiff bases derived from aromatic sulfonamides: towards more specific inhibitors for membrane-bound versus cytosolic isozymes. Eur J Med Chem;32:445–452.
41. Popescu A., Simion A., Scozzafava A., Briganti F., Supuran C.T. (1999) Carbonic anhydrase inhibitors. schiff bases of some aromatic sulfonamides and their metal complexes: towards more selective inhibitors of carbonic anhydrase isozyme IV. J Enz Inhib Med Chem;14:407–423.
42. Krenkel G., Castro E.A. (2003) Optimized calculation on the inhibitory of carbonic anhydrase isozymes I and II by some phenyl and pyridyl substituted sulfanilamide Schiff's bases. Mol Med Chem;1:13–20.
43. Rumelhart D.E., Hinton G.E., Williams R.J. (1986) Learning representations by back-propagating errors. Nature;323:33–536.
44. Martens H., Naes T. (1989) Multivariate Calibration. Chichester, UK: John Wiley.
45. Derks E.P.P.A., Buydens L.M.C. (1998) Training aspects. Chemometr Intell Lab Syst;41:171–184.

Notes

[a]HyperChem Release 7.0, HyperCube, Inc, available at: http://www.hyper.com.
[b]DRAGON 5.0 Evaluation version, available at: http://www.talete.mi.it/products/dragon_description.htm.
[c]SPSS version 13.0, SPSS, Inc.
[d]MATLAB [Version 7.0.1 (R14)], Mathworks, Inc, available at: http://www.mathworks.com.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Plot of the RMSE versus the number of hidden nodes used in the ANN analysis for modeling the CA isozyme types I, II, and IV.

Figure S2. The evolution of (A) R^2_{CV} and (B) $RMSE_P$ as the function of number of PC entered successively to the PCR as employed in the CR-PCR analysis.

Table S1. Structural details of sulfonamides used in this study.

Table S2. The observed (Obs.) and predicted (Pred.) inhibitory activities and their residues (Res.) for the following three types of CA isozyme: log K_i(hCAI), log K_i(hCAII), and log K_i(bCAIV), for the series of sulfonamides used in the present study as obtained from MLR, PC-ANN, and CR-PCR analyses.

Table S3. Brief description of the descriptors used in the present study.

Table S4. Coefficients of determination and cross-validation parameters for chance correlation investigation of the optimal models suggested by the MLR analysis.

Table S5. Coefficient of determination and cross-validation parameters for optimizing the number of hidden nodes for the ANN

models for the inhibitory activities: log K_i(hCAI), log K_i(hCAII), and log K_i(bCAIV).

Table S6. Coefficients of determination and cross-validation parameters for chance correlation investigation of the optimal models suggested by the PC-ANN analysis.

Table S7. PCs ranking according to decreasing R_P^2 values for the inhibitory activities: log K_i(hCAI), log K_i(hCAII), and log K_i(bCAIV).

Table S8. Coefficients of determination and cross-validation parameters for chance correlation investigation of the optimal models suggested by the CR-PCR analysis.