Research in Pharmaceutical Sciences, June 2018; 13(3): 192-204
Received: November 2017
Accepted: January 2018

School of Pharmacy & Pharmaceutical Sciences
Isfahan University of Medical Sciences

Original Article

Probing the chemical interaction space governed by 4-amino-substituted benzenesulfonamides and carbonic anhydrase isoforms

Behnam Rasti[1*] and Yeganeh Entezari Heravi[2]

[1]*Department of Microbiology, Faculty of Basic Sciences, Lahijan Branch, Islamic Azad University (IAU), Lahijan, Guilan, I.R. Iran*
[2]*Department of Chemistry, Faculty of Science, University of Tehran, Tehran, I.R. Iran.*

Abstract

Isoform diversity, critical physiological roles and involvement in major diseases/disorders such as glaucoma, epilepsy, Alzheimer's disease, obesity, and cancers have made carbonic anhydrase (CA), one of the most interesting case studies in the field of computer aided drug design. Since applying non-selective inhibitors can result in major side effects, there have been considerable efforts so far to achieve selective inhibitors for different isoforms of CA. Using proteochemometrics approach, the chemical interaction space governed by a group of 4-amino-substituted benzenesulfonamides and human CAs has been explored in the present study. Several validation methods have been utilized to assess the validity, robustness and predictivity power of the proposed proteochemometric model. Our model has offered major structural information that can be applied to design new selective inhibitors for distinct isoforms of CA. To prove the applicability of the proposed model, new compounds have been designed based on the offered discriminative structural features.

Keywords: Carbonic anhydrases; Proteochemometrics; Selectivity; GRINDs; Z-scales

INTRODUCTION

Carbonic anhydrases (CAs) are of particular importance due to their diverse physiological roles and pathogenicity. The interconversion between bicarbonate (HCO_3^-) and carbon dioxide (CO_2) is the main responsibility of CAs as illustrated in reaction below (1-3).

$$CO_2 + H_2O \rightarrow HCO_3^- + H^+ \quad (1)$$

The reaction is involved in many physiological processes including respiration, acid-base balance, oncogenesis, proliferation, and biosynthetic reactions (1-3). To date, CAs are classified into seven categories, α, β, γ, δ, ξ, η, and θ (4) which are distributed in various organism, tissues, and cells. Alpha-class, found in vertebrates, possesses sixteen isoforms of which thirteen isoforms have biological activity due to the presence of zinc ion within their active sites (5). Distribution of these isoforms in various tissues has made CA a vital target for the treatment of many diseases. For instance, the cytosolic isoform II is a target for glaucoma, whereas isoforms IX and XII have significant impact on tumorigenesis and therefore are major targets for cancer therapy (1-3). Due to the structural resemblance of CA isoforms and their involvement in various diseases, availability of selective inhibitors to avoid unwanted side effects is of great necessity. Different classes of inhibitors including sulfonamide and non-sulfonamide compounds have been so far identified (6). The latter consists of compounds like phenols (7), thiols (8), coumarin derivatives (9), and polyamines (10), showing completely different inhibitory mechanism compared to sulfonamide derivatives. Among sulfonamide inhibitors, those with unsubstituted functional groups (SO_2NH_2) have the highest inhibitory potency due to the formation of hydrogen bonds with some cavity key residues such as glutamic acid (Glu106) and threonine (Thr199) (6,11). Therefore, it seems that the sulfonamide group is highly essential for the inhibitory mechanism of sulfonamide-derivatives inhibitors.

*Corresponding author: B. Rasti
Tel: +98-1342222605, Fax: +98-1342222605
Email: rasti@liau.ac.ir

Access this article online

Website: http://rps.mui.ac.ir

DOI: 10.4103/1735-5362. 228940

As a matter of fact, special properties of sulfonamide group have made sulfonamide derivatives the most potent inhibitors of CAs. Some unique features are as follows: (a) the monoprotonated nitrogen (NH-) containing negatively charged nitrogen has a great tendency to coordinate toward zinc ion. At the same time, NH- group donates hydrogen to the Oγ of Thr199, resulting in formation of a bridge with the carboxylate moiety of Glu106, *(b)* a hydrogen bond is formed between one of the oxygens of -SO_2NH_2 group and NH- of the backbone of Thr199 (12). To sum up, given the importance of this enzyme in vital physiological processes as well as the structural diversity of CA and non-selectivity of many current inhibitors, the design of new compounds with improved inhibitory properties as well as different mechanism of action is the main subject of matter. To investigate the selectivity, proteochemometrics (PCM) approach can be applied since it considers interaction space of different ligands across multiple receptors (13). PCM investigations have so far shed light on valuable information regarding major protein families such as G protein-coupled receptors (14), proteases (15), thymidylate synthase (16), cytochrome P450 (17), CA (18, 19) and phosphodiesterase (20). In the present study, we have developed a PCM model in which we applied different combinations of z-scale and molecular interaction field (MIF) based descriptors to investigate the chemical interaction space between six isoforms of CA and a series of sulfonamide-derivatives inhibitors. We found some major structural contributors that can help to design inhibitors with enhanced selectivity for the investigated isoforms. We also designed some compounds, based on the presented findings, to confirm the reliability and the value of the findings.

MATERIALS AND METHODS

Interaction data

Inhibition activity (Ki) of a set of benzenesulfonamide derivatives against isoforms I, II, VI, VII, XII, and XIII of human CA has been recently investigated by Rutkauskas *et al.* (21). We selected our data set based on the following facts: (a) sulfonamides and their derivatives are the most popular and well known CA inhibitors which are clinically used as antiglaucoma agents, antiobesity drugs and diuretics, (b) most of the inhibitors designed by Rutkauskas *et al.* is not efficiently selective for a specific CA isoform, showing the potential of causing side effects.

Descriptors of organic compounds

Structures of 28 organic compounds were drawn and optimized in SYBYL7.3 (SYBYL Molecular Modeling Software version 7.3, Tripos Associates St. Louis, MO). Tripos force field with a distance-dependent dielectric and the Powell conjugate gradient algorithm with convergence criterion of 0.001 kcal/mol Å were used for the optimization. Gasteiger-Huckel method was used to calculate partial atomic charges of all compounds. The final structures were subjected to Pentacle 1.05 software (22) and grid independent descriptors (GRINDs) were calculated for them using both ALMOND and AMANDA algorithms (23,24). Both algorithms work through three following steps: (a) calculating MIFs for different types of interactions, (b) node filtration process in which regions with greatest favorable interaction energy are selected. The main difference between ALMOND and AMANDA relates to the node filtration step. ALMOND uses a Fedorov-like optimization algorithm (25) to reduce the number of nodes, whereas AMANDA uses a pre filtering step in which all nodes failing an energy cutoff are primarily removed (24), and (c) finally, the chosen nodes are encoded into the descriptors. Pairs of interaction energies are multiplied and the greatest product is kept for each internode distance. The approach provides information that directly correlates with the structures of the molecules. We used the following probes to calculate MIFs: DRY (hydrophobic probe), N1 (H-bond donor (HBD)), O (H-bond acceptor/HBA), and TIP (representing molecular shape). Distance between grid points, the number of extracted nodes (for each MIF) and the smoothing window were set to 0.5 Å, 100 and 0.8 grid units (0.4 Å), respectively. Application of the mentioned parameters resulted in generation of 65

descriptors for each auto/cross MIF-MIF multiplication (auto/cross-correlograms). A total number of 650 descriptors were therefore obtained for each compound, since four types of MIF were applied. Descriptors showing same value for all the compounds were removed.

Proteins descriptors

To determine the cavity residues structure of human CA II, (1CA2) was obtained from RCSB Protein Data Bank (RCSB PDB). From the center of the zinc ion, a cutoff of 10 Å was applied to identify ligand interacting residues. All CA sequences were then aligned by ClustalW2 web server (26) and cavity amino acids were identified in correspondence positions for other isoforms (Fig. 1A). 26 non-conserved ligand interacting residues were encoded by both three (z1, z2, and z3) and five z-scale descriptors (z1, z2, z3, z4, and z5), resulting in total number of 78 (26 × 3) and 130 (26 × 5) descriptors, respectively (Fig. 1B). Z-scales are obtained by principal component analysis of 26 measured physicochemical properties of 87 natural/ artificial amino acids. Z1, z2 and z3 are the three first principal components, representing the largest variations of physicochemical properties. Z1, z2, z3, z4 and z5 are representing lipophilicity, size/ polarizability, electronic and electrostatic properties, respectively (27).

Feature selection

Feature selection was performed using genetic algorithm (GA) (28-31) to select the best fitted GRIND descriptors for the final modeling. GA-partial least square (GA-PLS) consists of the following steps: (a) generation of the initial population of chromosomes, (b) the squared predictive correlation coefficient (Q^2) is used as an index to evaluate the fitness of each chromosome in the population, and (c) reproduction of the population which involves processes such as crossing-over, mutation. Steps 2 and 3 continue up to the designated number of generations (29). PLS Toolbox 3.5 (Eigen vector Research, Inc, Manson, WA, USA) was used and GA with default parameters was applied on both ligand descriptors. The final numbers of 88, 92 GRINDs were selected by GA for AMANDA and ALMOND descriptors, respectively.

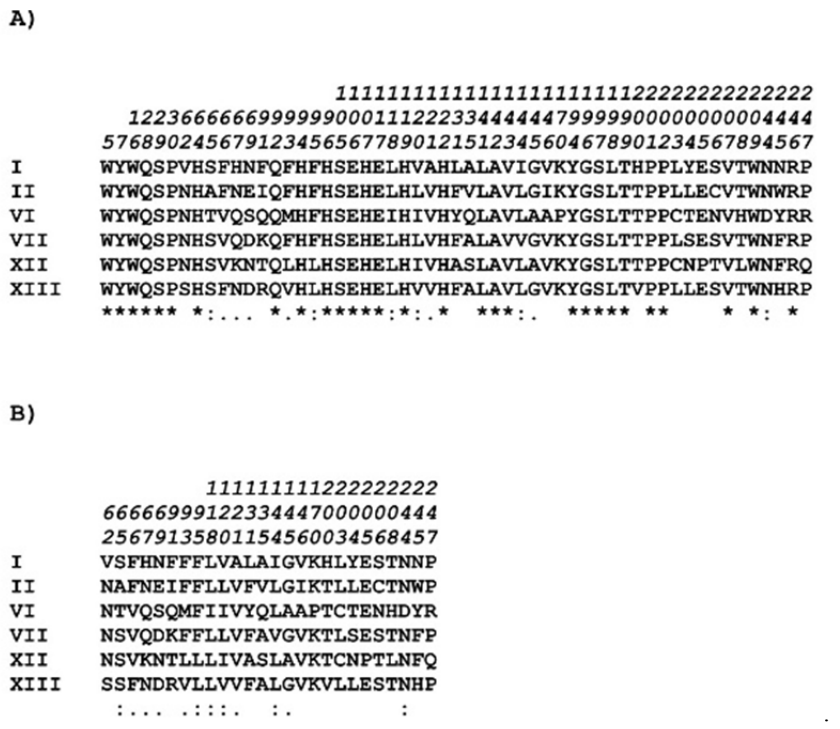


Fig. 1. Multiple sequence alignment of carbonic anhydarse isoforms. (A) multiple sequence alignment of cavity amino acids. Conserved positions are marked by asterisks. (B) multiple sequence alignment of 26 non-conserved residues. Isoform II is used as numbering reference.

Ligand-receptor cross-term descriptors

Protein-ligand cross-terms were calculated by production of protein descriptors with ligand descriptors. Therefore, the total number of cross-terms equals to 6864, 11440, 7176, and 11960 in case of AMANDA/3-z-scales, AMANDA/5-z-scales, ALMOND/3-z-scales, and ALMOND/5-zscales combinations, respectively. Mean-centering and scaling to unit variance was performed for all descriptors prior to modeling.

Multivariate modeling

Since PLS is capable of making reliable predictive models even in the presence of collinearity and redundancy of variables, we applied PLS approach to correlate descriptors and biological activities. PLS simultaneously projects descriptors and biological activities to PLS components and finds linear relationships between them. Regression coefficients derived from the PLS regression equation reveals the direction and magnitude of the influence of X on Y. Regarding a ligand-receptor system, the PLS regression equation is expressed as follows:

$$y = \bar{y} + \sum_{r=1}^{R} coeff_r x_r + \sum_{l=R+1}^{R+L} coeff_l x_l + \sum_{r=1}^{R} \sum_{l=R+1}^{R+L} coeff_{r.l} x_r x_l \quad (2)$$

where, $coeff_r, coeff_l$ and $coeff_{r.l}$ are regression coefficients, and x_r $(1 \le r \le R)$ and x_l $(R + 1 \le l \le R + L)$ stand for receptor and ligand descriptors, respectively. PLS Toolbox 3.5 was applied in the present study to conduct PLS regression.

Assessing validation of the models

To assess the power of the model for predicting the biological activity of a new compound, we randomly excluded a few complexes from the modeling process, so that they could not affect the PLS model. For the excluded set, the predicting power of the model has been reported as Q^2_{EXT}. Furthermore, we used the Kennard-Stone algorithm (32) to divide the remaining dataset into internal validation (nearly 75% of dataset) and external testing sets (nearly 25% of dataset). Cross-validation was performed on the internal validation set using the venetian blinds approach and the results are presented as R^2_{pred}. Moreover, internal and external sets were subjected to applicability domain (AD) analysis to estimate the likelihood of reliable prediction for compounds. The leverage value together with the William's plot is usually used to assess the AD of a QSAR model (33). The William's plot was built by plotting the standardized residuals of compounds versus their leverage values. Finally, the robustness of model was confirmed by applying the approach of Y-scrambling. The response vector (Y) was randomly scrambled for 100 times, while the X-data were left intact. Each time a new model with scrambled data was generated and its R^2 and Q^2 were calculated. The R^2 and Q^2 values of scrambled and unscrambled models were plotted versus correlation coefficients between original and scrambled Ys. Afterwards, a regression line was conducted and the intercepts for R^2 and Q^2 ($R^2_{intercept}$ and $Q^2_{intercept}$) were calculated. Previous works have shown that the $R^2_{intercept}$ of < 0.3 and the $Q^2_{intercept}$ of < 0.05 are acceptable for a robust model (34).

Contribution of ligand properties for receptor selectivity

The significance of a compound descriptor in selective inhibition of different receptors can be assessed by the following equation:

$$\Delta y_{rec.l} = \frac{dy}{dx_l} = coeff_l + \sum_{r=1}^{R} coeff_{r.l} x_r \quad (3)$$

where, *rec* is a receptor with descriptors x_1, x_2 …. x_R, and $\Delta y_{rec.l}$ shows the change in the selectivity of the l^{st} descriptor of compounds for this particular receptor. Moreover, $coeff_l$ and $coeff_{l.r}$ are regression coefficients, and x_r $(1 \le r \le R)$ and x_l $(R + 1 \le l \le R + L)$ stand for receptor and ligand descriptors, respectively.

RESULTS

PCM modeling

Combination of four types of MIF (DRY, N1, O and TIP) was used to calculate descriptors of compounds using both ALMOND and AMANDA algorithms. Prior to modeling, GRid-Independent Descriptors calculated by ALMOND, AMANDA were subjected to feature selection, performed by GA, to find the best fitted structural

descriptors. Different combinations of compound and receptor descriptor types were used to build PCM models. Validation parameters regarding different models are presented in Table 1. Descriptors of the compounds, protein cavity residues and their cross-terms were correlated to the pKis using PLS approach.

Assessing validity of the model

As is represented in Table 1, all models showed acceptable values for R^2 (0.8-0.93) and Q^2 (0.68-0.8). In case of external cross-validation (R^2_{pred}) (Fig. 2A) and predicting the biological activity of a new compound (Q^2_{EXT}) (Fig. 2B) however, the ALMOND/5-zscales based model showed much better values compared to other PCM models (Table 1). Therefore, we considered this PCM model for further analyses. Finally, the robustness of the model was confirmed by Y-scrambling approach (Table 1 and Fig. 3). As is shown in Fig. 3, both $R^2_{intercept}$ and $Q^2_{intercept}$ show acceptable values (nearly 0.1 and 0.01 respectively), confirming the robustness of the PCM model. Fig. 4 illustrates the Williams plot applied to analyze the AD of the PCM model. As shown in the Williams plot, almost all compounds were located within the boundaries of applicability domain, suggesting a well-defined AD for the proposed PCM model. There are only five compounds falling outside the ± 3 standardized residual ranges, all belong to the internal set.

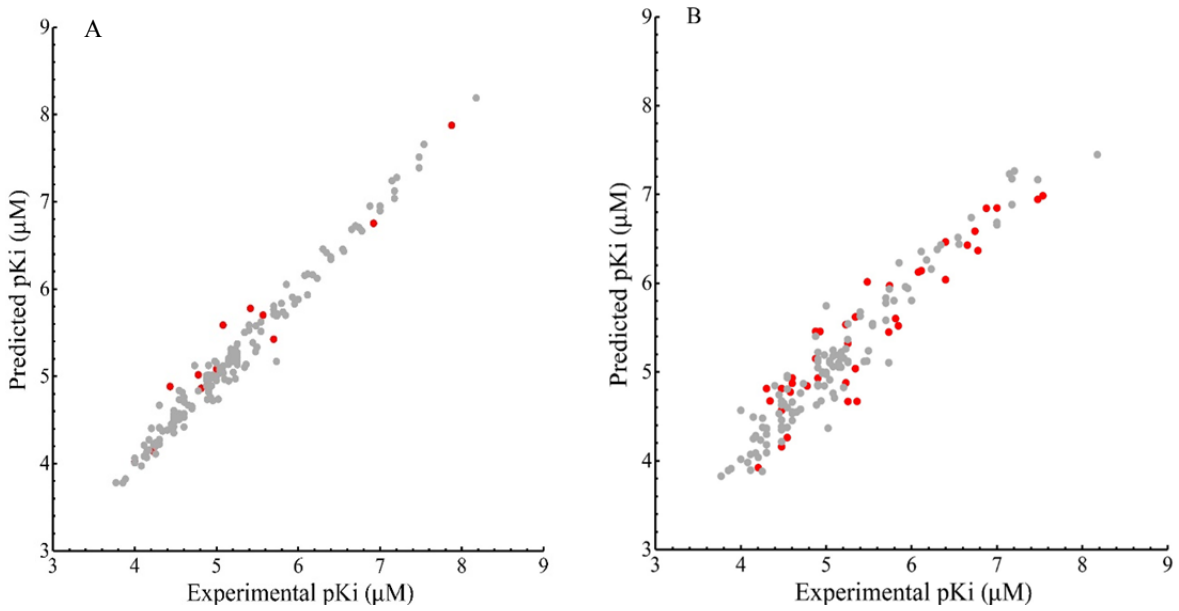


Fig. 2. Plot of experimental versus predicted pKi values for the PCM model. (A) corresponds to12 complexes which have been randomly left-out (red circles) to assess the power of the model for predicting the biological activity of a new compound (Q2EXT). (B) corresponds to 25% external test set (red circles) selected by Kennard-Stone algorithm to perform the external cross-validation (R2pred) test.

Table 1. Results of PLS modeling using combination of different descriptors

	Ligand descriptors		Protein descriptors							
Model	**ALMOND**	**AMANDA**	**3 z-scales**	**5 z-scales**	**R^2**	**Q^2**	**R^2_{pred}**	**Q^2_{EXT}**	**$R^2_{intercept}$**	**$Q^2_{intercept}$**
1	√	-	√	-	~0.88	~0.78	~0.80	~0.89	ND	ND
2	√	-	-	√	~0.93	~0.80	~0.87	~0.95	~0.1	~0.01
3	-	√	√	-	~0.80	~0.68	~0.71	~0.76	ND	ND
4	-	√	-	√	~0.83	~0.72	~0.74	~0.78	ND	ND

(PLS), partial least square

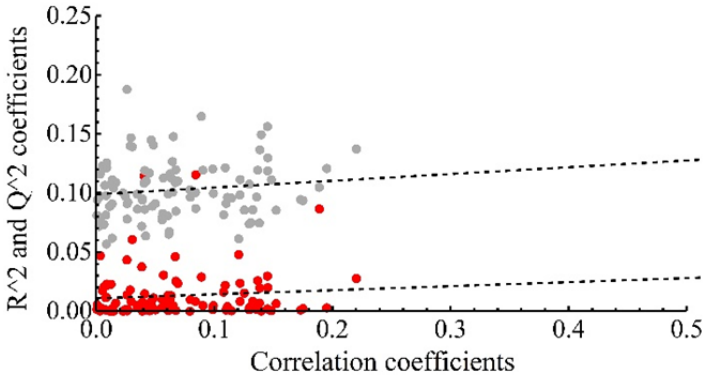


Fig. 3. Y-scrambling plot of pKi built based on the PCM model. The y-axis represents R2 (grey circles) and Q2 (red circles) coefficients for 100 models built based on randomly scrambled response data. The x-axis designates the correlation coefficient between original and permuted response data.

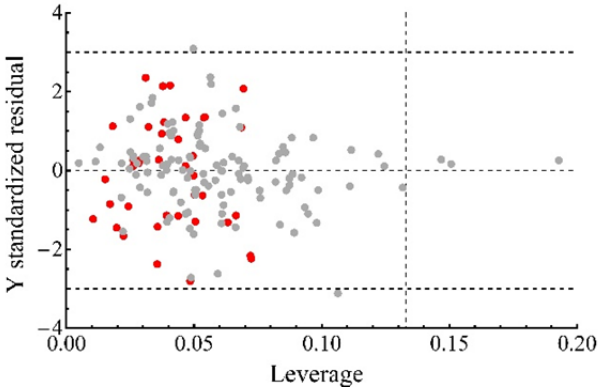


Fig. 4. Evaluation of applicability domain using William plots for PCM model. Compounds in the internal and external sets are shown by gray circles and red circles, respectively.

Analysis of ligand properties and their contribution to the receptor selectivity

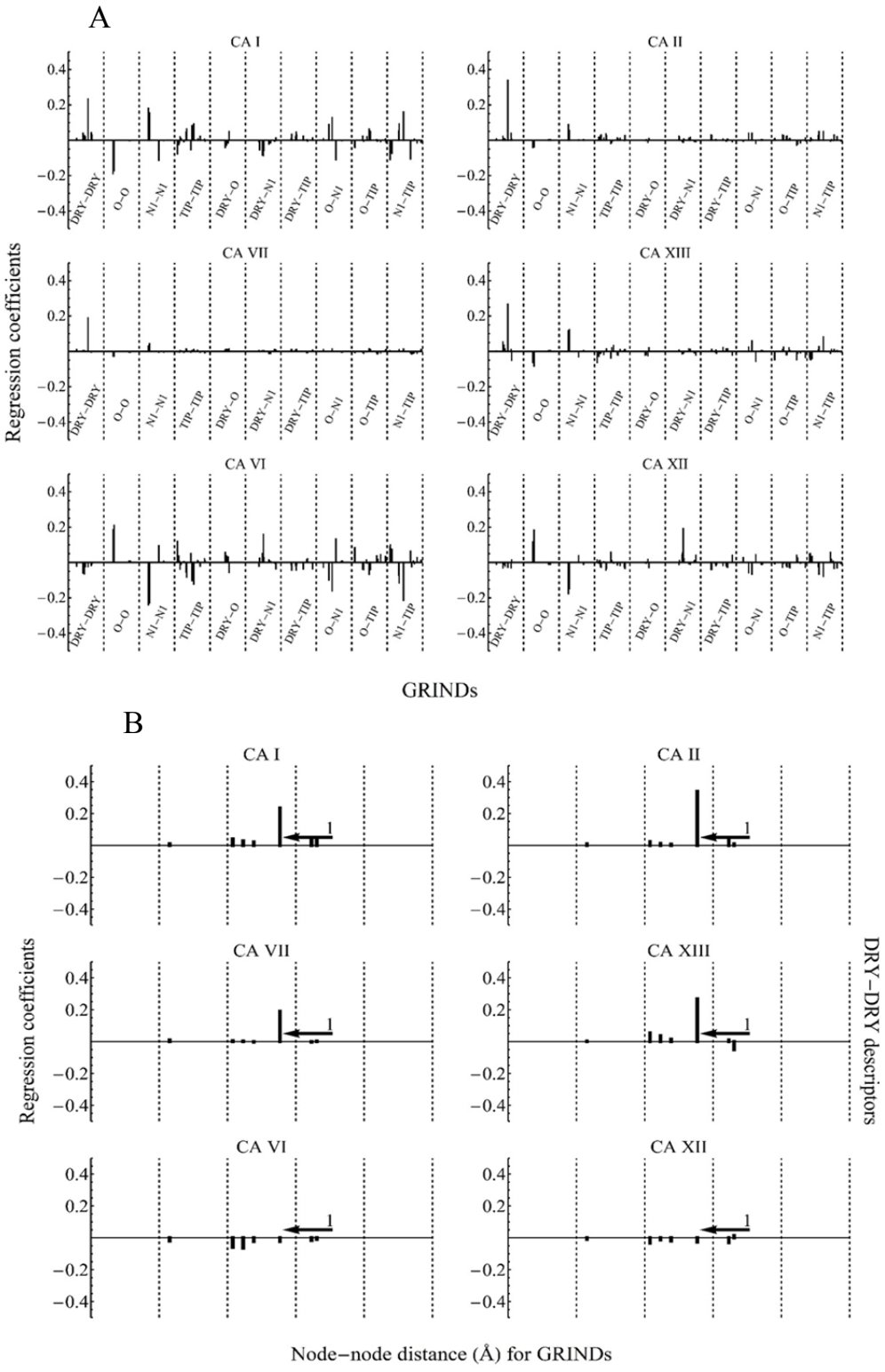
To assess the importance of ligand properties for receptor selectivity, equation (2) was used to extract the contribution of each GRIND from coefficients of cross-terms involving that GRIND. Fig. 5A illustrates $\Delta y_{rec.l}$ values for structural descriptors of ligands (selected by feature selection). The vertical separators represent a distance range from 0 Å to 26 Å.

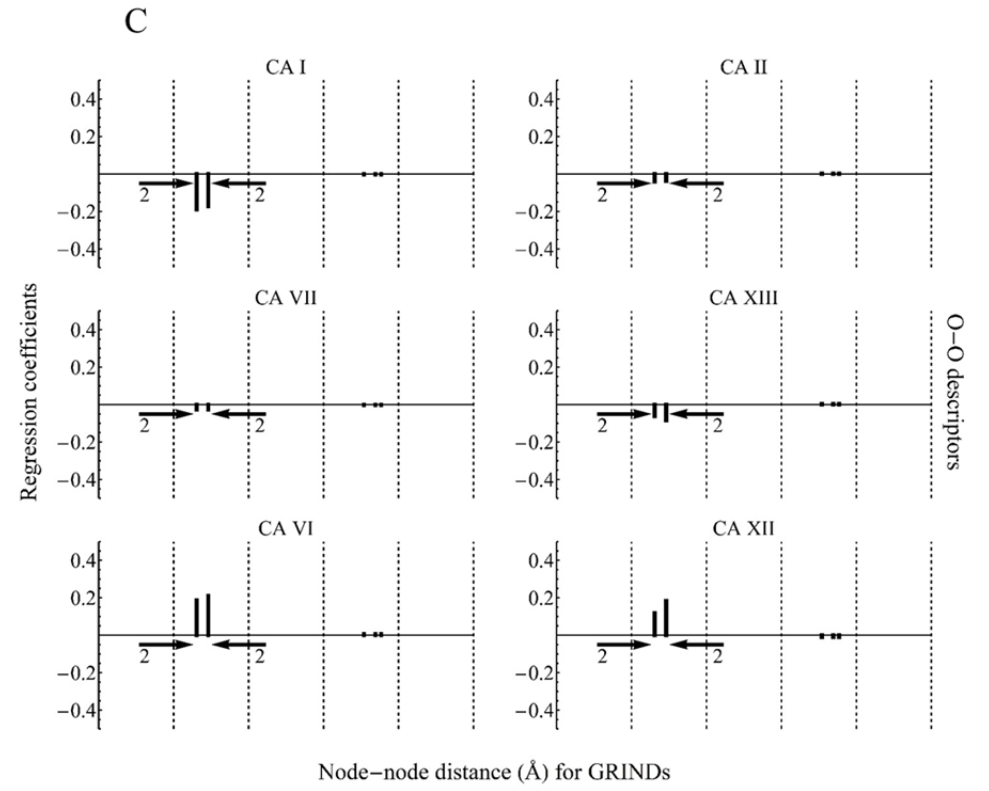
The most discriminative contributors are DRY-DRY descriptor at distance of 14.4 Å, O-O descriptor at distances of 6.8 Å and 7.6 Å, N1-N1 descriptor at distances of 6.8 Å and 7.6 Å and DRY-N1 descriptor at distance of 13.2 Å. These structural descriptors, showing highly discriminative behavior towards different isoforms of CA, are indicated in the parts B, C, D, and E of Fig. 5 and will be discussed.

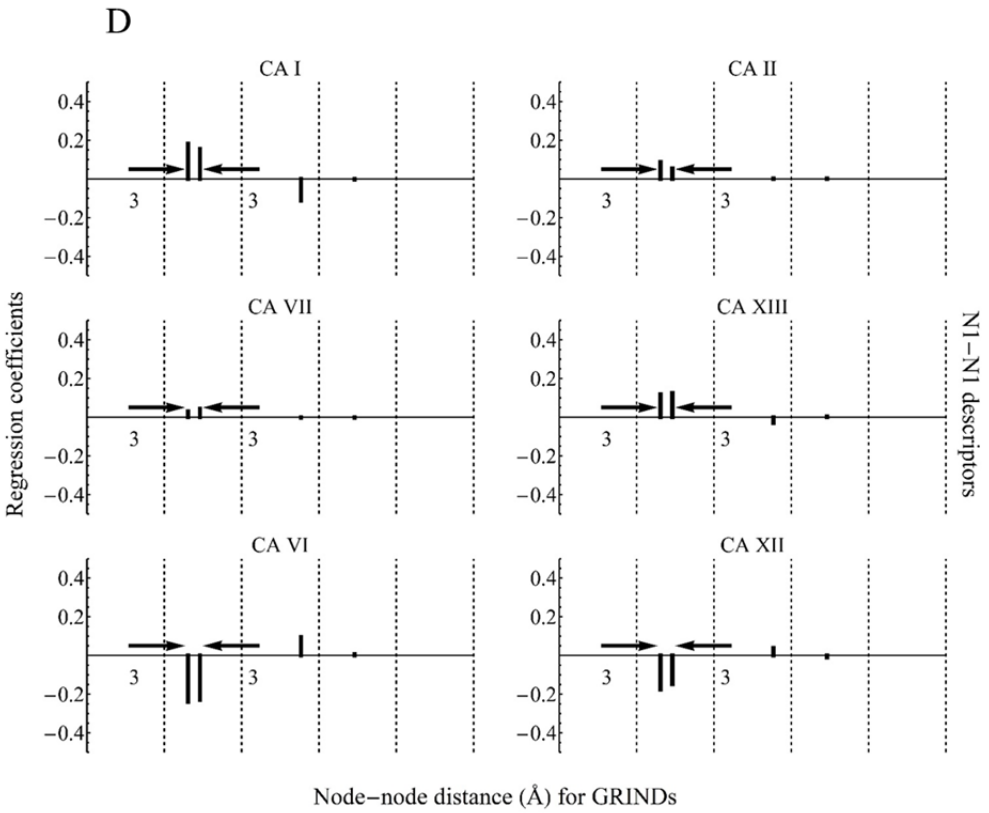
Applicability of the model

To show the applicability of the model, we modified some of current inhibitors (Fig. 6) and monitored their selectivity specification toward different isoforms of CA. Compounds with the most similar biological activity among different receptors were selected and structural modifications were carried out based on the proposed discriminative features. Fig. 6 also illustrates the modifications have been made in order to change the selectivity ratio of a compound for different receptors. As is shown in Fig. 6, compound 1 lacks the DRY-DRY descriptor at distance of 14.4 Å, while the modified version of the compound 1 has been given the mentioned descriptor. The modification is expected to be in favor of increasing the selectivity towards cytosolic isoforms. Compound 16 was selected to test the effect of descriptors N1-N1 and O-O at distances of 6.8 Å and 7.6 Å. By turning the

N-H moiety into the C=O, the O-O descriptors were removed on the one hand and N1-N1 descriptors were given to the compound 16 on the other hand. Both changes were made in favor of selective inhibition of cytosolic isoforms. Finally, compound 23 and compound 24 which lacked the DRY-N1 descriptor at distance of 13.2 Å were given the mentioned descriptor by elongating their hydrocarbon chains. The modification was made to change the selectivity ratio in favor of isoforms VI and XII. New compounds were then put in the test set and their activities were predicted based on the constructed PCM model.




C
CA I
CA II
CA VII
CA XIII
CA VI
CA XII
Regression coefficients
O–O descriptors
Node–node distance (Å) for GRINDs


D
CA I
CA II
CA VII
CA XIII
CA VI
CA XII
Regression coefficients
N1–N1 descriptors
Node–node distance (Å) for GRINDs

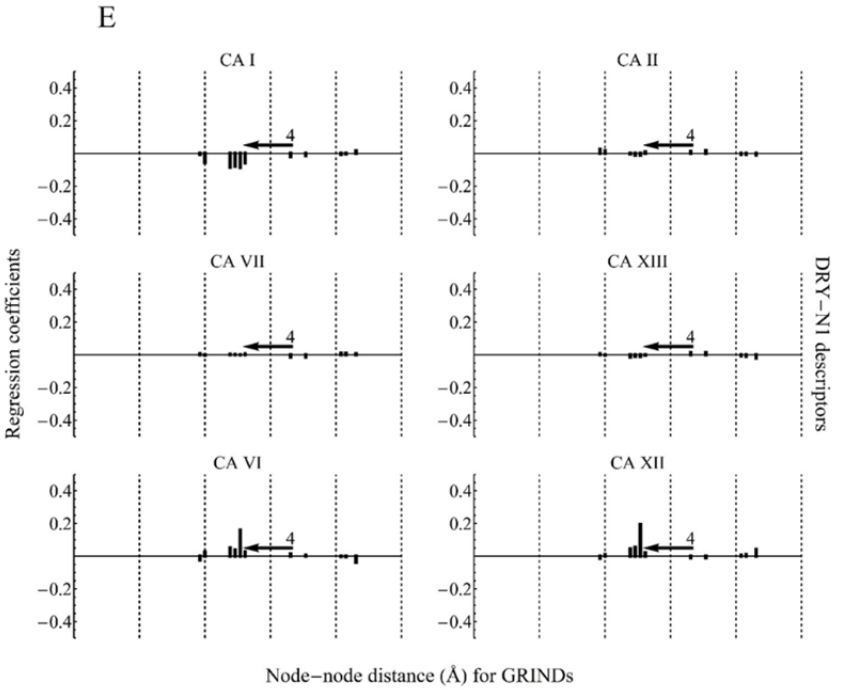


Fig. 5. Contribution of compound descriptors towards receptor-ligand interactions. (A) overall representation of GRINDs contribution in receptor selectivity. Y-axes indicate the regression coefficients related to the cross-terms. The interval between the vertical separators represents the node-node distance range of 0 Å to 26 Å for each particular GRIND. (B), (C), (D), and (E) represent DRY-DRY, O-O, N1-N1, and DRY-N1 descriptors, respectively. Within certain distances (marked by numbered arrows), the mentioned types of GRINDs are highly discriminative with regard to specific receptors. (GRIND), grid independent descriptors.

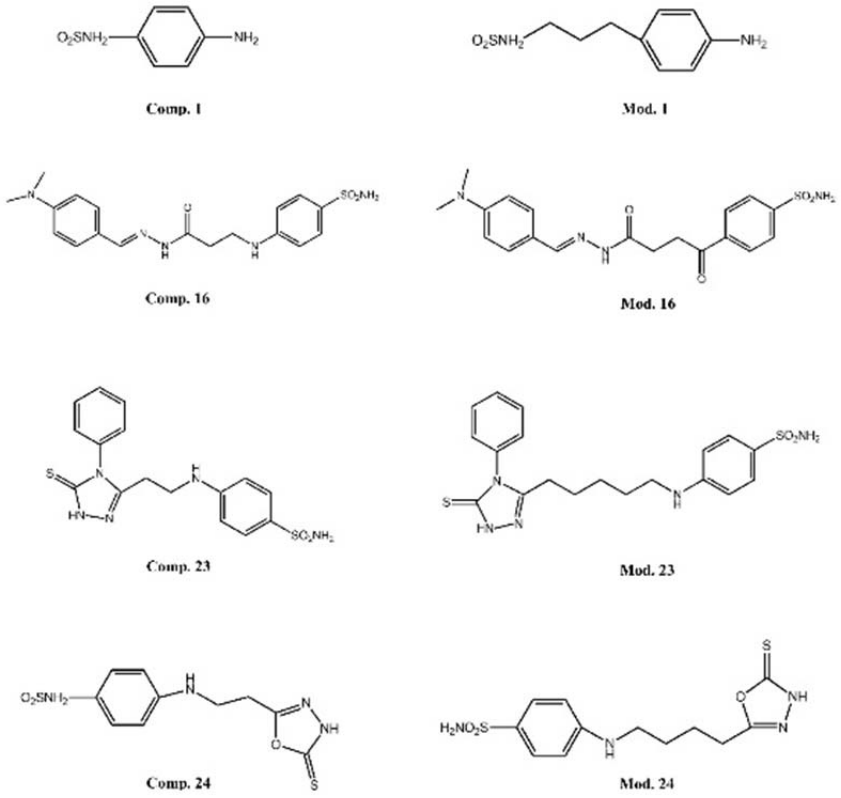


Fig. 6. Compounds applied to explore the effect of inhibitors differential descriptors on receptors selectivity. Comp. 1, Comp. 16, Comp. 23, and Comp. 24 are original compounds used in PCM modeling. Mod. 1, Mod. 16, Mod. 23, and Mod. 24 are modified versions of original compounds.

DISCUSSION

After the discovery of CA in 1930s, the enzyme is still of particular importance not only because of the isoform diversity but due to the major physiological roles played by the members of CA family. Furthermore, the extremely high turnover number (1×10^{-6} sec^{-1}) owned by CA has made it the most efficient enzyme known so far. Thus, any overexpression or functional defect can cause major disorders such as glaucoma and cancer. Sulfonamide derivatives are known as one of the most potent class of CA inhibitors. The sulfonamide moiety directly binds to the Zn^{+2} and form a tetrahedron geometry. Since different isoforms of CA are highly similar with regard to their structures, inhibition of one defect isoform can lead to inhibition of those showing natural physiological function. Therefore, accessing highly selective compounds to inhibit one specific isoform is of great importance.

Overall interpretations

Overall observation of Fig. 5A shows that coefficients related to receptors I, VI, XII, and XIII are significantly bigger than coefficients of receptors II and VII. Therefore, it seems that the selected features of compounds have higher impact on inhibition of isoforms I, VI, XII, and XIII than that of isoforms II and VII.

A closer look at the distribution pattern of coefficients reveals that investigated receptors can be put in two different groups, based on the contribution of the selected features to their inhibition. First group contains cytosolic receptors (I, II, VII, and XIII), while the second group consists of isoform XII (a transmembrane isoform) and isoform VI (a secreted isoform).

Finally, it can be concluded that these discriminative contributors can be used to design compounds capable of inhibiting either cytosolic CAs or isoforms VI and XII.

Cytosolic isoforms vs. isoforms VI and XII

Using receptors chemical space information, we entered discriminative information into the modeling process. Subsequently, interesting contributors were found that can be applied to design compounds with distinct selectively towards different isoforms of CA. A close look at Fig. 5B, Arrows 1 reveals a region (distance 14.4 Å) where DRY-DRY descriptor shows highly positive PLS coefficients with regards to cytosolic isoforms. In case of isoforms VI and XII however, the correspondent contributor is hardly involved in ligand-protein interactions. Compounds with hydrophobic moieties distanced nearly 14.4 Å from each other, are therefore expected to show higher selectivity towards cytosolic isoforms. Inspecting O-O descriptors (Fig. 5C, Arrows 2) indicates two regions (distances 6.8 Å and 7.6 Å) with significantly positive coefficients for isoforms VI and XII, in particular. The correspondent descriptors are at expense of ligand-protein interactions in cytosolic isoforms. The finding suggests that emplacement of HBD groups with correspondent distances in structures of compounds can substantially turn the selectivity in favor of receptors VI and XII. Interestingly, there is an exactly opposite condition with regards to N1-N1 descriptor (Fig. 5D, Arrows 3). In contrast to O-O descriptors at distances of 6.8 Å and 7.6 Å, N1-N1 descriptors at equivalent distances have positive coefficients for cytosolic isoforms, in particular. Therefore, it is expected that inhibitors having dual HBA moieties with distance of either 6.8 or 7.6 angstrom in their structures show enhanced selectivity towards cytosolic isoforms. Finally, another major contributor found by the model relates to DRY-N1 descriptors (Fig. 5E, Arrows 4). There are significant positive coefficients for DRY-N1 descriptor at distance of 13.2 Å with regards to isoforms VI and XII, whereas the correspondent descriptors do not contribute in interactions between compounds and cytosolic receptors. Based on the latter observation, we expect compounds with hydrophobic/HBA moieties, located in distance of about 13.2 Å from each other, to be more selective for isoforms VI and XII.

Applicability of the model

The results of new inhibitors design (Fig. 6) have been promising in case of all discriminative descriptors (Table 2).

Table 2. Experimental and predicted values of Ki for original (Comp) and modified version of the compounds (Mod).

Isoforms	Compound dissociation constants K_d (μM) for carbonic anhydrase isoforms							
	Comp. 1	**Comp. 16**	**Comp. 23**	**Comp. 24**	**Mod. 1**	**Mod. 16**	**Mod. 23**	**Mod. 24**
hCAI	100	37.10	10	12.60	0.27	2.60	49.80	143.50
hCAII	12.80	8.31	1.99	4.57	0.32	1.19	7.11	17.30
hCAVII	50.11	25.11	6.61	4.57	0.77	5.94	45.21	78.10
hCAXIII	100	56.23	10	26.30	1.36	9.63	35.69	152.70
hCAVI	56.23	50.11	11.20	33.11	9.62	25.09	20.45	48.11
hCAXII	67.60	33.11	6.61	5.88	31.80	21.53	9.30	6.66

Comparison of the Kis for compound 1 and compound 16 against different isoforms with that of modified compound 1 and compound 16 reveals two major points as follows: (a) the average affinity of the modified compounds have been substantially increased in case of all receptors, showing the importance of the structural descriptors captured by the PCM model, (b) the selectivity ratio of the inhibitors has been significantly changed in favor of cytosolic isoforms, as was expected. Significant examples with regard to compound 1 are selectivity ratios for isoform I/isoform VI, isoform I/isoform XII, isoform XIII/isoform VI and isoform XIII/isoform XII with experimental values of 1.778, 1.479, 1.778, and 1.479, respectively. Creating the DRY-DRY descriptor in modified version of compound 1 has resulted in turning these ratios in favor of cytosolic isoforms as follows: 0.028, 0.008, 0.141, and 0.042 (Table 2). A same trend can be observed in case of compound 16 with the most significant examples of selectivity ratios for isoform XIII/isoform VI and isoform XIII/isoform XII. The experimental values are 1.122 and 1.698, while the predicted selectivity ratios related to modified compound are 0.383 and 0.447, showing the higher selectivity of modified compound for cytosolic isoforms (Table 2). The two other compounds (23 and 24) were modified in a way that their selectivity ratios would be changed in favor of isoforms VI and XII. Comparing the experimental Kis of compound 23 and compound 24 with predicted Kis of their modified versions reveals that the trends were altered for the selectivity ratios as we expected (Table 2). Inspecting Table 2 shows that in case of both compounds the modifications we made caused an increase in the selectivity of the compounds for isoforms VI and XII. Significant examples are selectivity ratios for isoform VII/isoform XII with regard to both compound 23 and compound 24. The experimental values are 1 and 0.777, while the predicted selectivity ratios related to modified compound are 4.861 and 11.726, showing that selectivity of the modified compounds have been substantially increased for isoforms VI and XII (Table 2).

CONCLUSION

Since distinct isoforms of CA are engaged with different disorders, isoform-selective inhibitors seem necessary to avoid unwanted side effects. Chemical interaction space available for a receptor-inhibitor complex is governed by complementary structural/ physico-chemical features of the receptor and the compound. Using PCMs therefore, one can study the selective behavior of a group of common inhibitors toward different isoforms of a protein. In the present study, structural GRIND descriptors together with z-scales ones have been applied to build the PCM model. The most important structural features affecting the selective inhibition of isoforms have been screened applying GA-based variable selection. Robustness and predictivity power of the model were validated using several validation methods. Based on the obtained results, it seems that the structural features selected by variable selection approach are more involved in selective inhibition of isoforms I, VI, XII, and XIII than II and VII. Moreover, major structural features with regard to hydrophobic-hydrophobic and hydrophobic-HBD/HBA interactions have been shown to be discriminative among

different isoforms of CA. Such contributors can be considered to design new inhibitors with higher selectivity towards isoforms of CA.

ACKNOWLEDGEMENTS

This project was financially supported by Lahijan Branch of Islamic Azad University (IAU), Lahijan, I.R. Iran.

REFERENCES

1. Supuran CT. Carbonic anhydrases: novel therapeutic applications for inhibitors and activators. Nat Rev Drug Discov. 2008;7(2):168-181.
2. Henry RP. Multiple roles of carbonic anhydrase in cellular transport and metabolism. Annu Rev Physiol. 1996;58:523-538.
3. Henry RP, Swenson ER. The distribution and physiological significance of carbonic anhydrase in vertebrate gas exchange organs. Respir Physiol. 2000;121(1):1-12.
4. Kose LP, Gülçin İ, Özdemir H, Atasever A, Alwasel SH, Supuran CT. The effects of some avermectins on bovine carbonic anhydrase enzyme. J Enzyme Inhib Med Chem. 2016;31(5):773-778.
5. Nishimori I. Acatalytic CAs: Carbonic Anhydrase-Related Proteins. In: Supuran CT, Scozzafava A, Conway J, editors. Carbonic Anhydrase: Its Inhibitors and Activators. New York: CRC Press; 2004. pp. 12-14.
6. Supuran CT, Scozzafava A, Casini A. Carbonic anhydrase inhibitors. Med Res Rev. 2003;23(2): 146-189.
7. Carta F, Vullo D, Maresca A, Scozzafava A, Supuran CT. Mono-/dihydroxybenzoic acid esters and phenol pyridinium derivatives as inhibitors of the mammalian carbonic anhydrase isoforms I, II, VII, IX, XII and XIV. Bioorg Med Chem. 2013;21(6):1564-1569.
8. Schulze Wischeler J, Innocenti A, Vullo D, Agrawal A, Cohen S M, Heine A, *et al*. Bidentate Zinc chelators for alpha-carbonic anhydrases that produce a trigonal bipyramidal coordination geometry. ChemMedChem. 2010;5(9):1609-1615.
9. Bonneau A, Maresca A, Winum JY, Supuran CT. Metronidazole-coumarin conjugates and 3-cyano-7-hydroxy-coumarin act as isoform-selective carbonic anhydrase inhibitors. J Enzyme Inhib Med Chem. 2013;28(2):397-401.
10. Carta F, Temperini C, Innocenti A, Scozzafava A, Kaila K, Supuran CT. Polyamines inhibit carbonic anhydrases by anchoring to the zinc-coordinated water molecule. J Med Chem. 2010;53(15):5511-5522.
11. McKenna R, Supuran CT. Carbonic Anhydrase Inhibitor Drug Design. In: Frost SC, McKenna R, editors. Carbonic Anhydrase: Mechanism, Regulation, Links to Disease, and Industrial Applications. Netherlands: Springer Science+Business Media Dordrecht; 2013. pp. 291-323.
12. Lindahl M, Vidgren J, Eriksson E, Habash J, Harrop S, Helliwell J, *et al*. Crystallographic Studies of Carbonic Anhydrase Inhibition. In: Botre F, Gros G, Storey BT, editors. Carbonic Anhydrase: From Biochemistry and Genetics to Physiology and Clinical Medicine. Weinheim: VCH; 1991. pp. 111-118.
13. Prusis P, Muceniece R, Andersson P, Post C, Lundstedt T, Wikberg JE. PLS modeling of chimeric MS04/MSH-peptide and MC1/MC 3-receptor interactions reveal a novel method for the analysis of ligand-receptor interactions. Biochim Biophys Acta. 2001;1544(1-2):350-357.
14. Lapinsh M, Prusis P, Lundstedt T, Wikberg JE. Proteochemometrics modeling of the interaction of amine G-protein coupled receptors with a diverse set of ligands. Mol Pharmacol. 2002;61(6):1465-1475.
15. Prusis P, Lapins M, Yahorava S, Petrovska R, Niyomrattanakit P, Katzenmeier G, *et al*. Proteochemometrics analysis of substrate interactions with dengue virus NS3 proteases. Bioorg Med Chem. 2008;16(20):9369-9377.
16. Rasti B, Shahangian SS. Proteochemometric modeling of the origin of thymidylate synthase inhibition. Chem Biol Drug Des. 2017. DOI: 10.1111/cbdd.13163.
17. Simeon S, Spjuth O, Lapins M, Nabu S, Anuwongcharoen N, Prachayasittikul V, *et al*. Origin of aromatase inhibitory activity via proteochemometric modeling. PeerJ. 2016;4:e1979.
18. Rasti B, Karimi-Jafari M H, Ghasemi JB. Quantitative characterization of the interaction space of the mammalian carbonic anhydrase isoforms I, II, VII, IX, XII, and XIV and their inhibitors, using the proteochemometric approach. Chem Biol Drug Des. 2016;88(3):341-353.
19. Rasti B, Namazi M, Karimi-Jafari M, Ghasemi JB. Proteochemometric modeling of the interaction space of carbonic anhydrase and its inhibitors: an assessment of structure-based and sequence-based descriptors. Mol. Info. 2017;36(4). DOI: 10.1002/minf.201600102.
20. Rasti B, Schaduangrat N, Shahangian SS, Nantasenamat C. Exploring the origin of phosphodiesterase inhibition via proteochemometric modeling. RSC Adv. 2017;7(45):28056-28068.
21. Rutkauskas K, Zubrienė A, Tumosienė I, Kantminienė K, Kažemėkaitė M, Smirnov A, *et al*. 4-amino-substituted benzenesulfonamides as inhibitors of human carbonic anhydrases. Molecules. 2014;19(11):17356-17380.
22. Durán A, Zamora I, Pastor M. Suitability of GRIND-based principal properties for the description of molecular similarity and ligand-based virtual screening. J Chem Inf Model. 2009;49(9):2129-2138.
23. Pastor M, Cruciani G, McLay I, Pickett S, Clementi S. GRid-INdependent descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors. J Med Chem. 2000;43(17):3233-3243.

24. Durán Á, Martínez GC, Pastor. M. Development and validation of AMANDA, a new algorithm for selecting highly relevant regions in molecular interaction fields. J Chem Inf Model. 2008;48(9):1813-1823.
25. Fedorov VV. Theory of Optimal Experiments. New York: Academic Press; 1972. pp. 292.
26. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, *et al*. Clustal W and Clustal X version 2.0. bioinformatics. 2007;23(21):2947-2948.
27. Sandberg M, Eriksson L, Jonsson J, Sjöström M, Wold S. New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. J Med Chem. 1998;41(14):2481-2491.
28. Beasley D, Bull DR, Martin RR. An overview of genetic algorithms: Part 1, fundamentals. University Computing.1993;15(2):58-69.
29. Van de Waterbend H. Chemometric Methods in Molecular Design. In: Van de Waterbend, Timmerman H, Mannhold R, Krogsgaard-Larsen P, editors. Methods and Principles in Medicinal Chemistry. New York: VCH: Publishers; 1995. pp. 195-219.
30. Parish M, Hemmateenejad B, Miri R, Edraki N, Elyasi M, Khoshneviszadeh M, *et al*. Qsar study of N-acyl-3,5-bis(arylidene)-4-piperidones as cytotoxic agents. Res Pharm Sci. 2012;7(5):s557.
31. Karbakhsh R, Sabet R. Application of different chemometric tools in QSAR study of azolo-adamantanes against influenza A virus. Res Pharm Sci. 2011;6(1):23-33.
32. Kennard RW, Stone LA. Computer aided design of experiments. Technometrics. 1969;11(1):137-148.
33. Gramatica P. Principles of QSAR models validation: internal and external. Mol Inform. 2007;26(5): 694-701.
34. Eriksson L, Jaworska J, Worth AP, Cronin MTD, McDowell RM, Gramatica P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. Environ Health Perspect. 2003; 111(10):1361-1375.