Quantitative Characterization of the Chemical Space Governed by Human Carbonic Anhydrases and selenium-containing derivatives of solfonamides

Behnam Rasti*

*Department of Microbiology, Faculty of Basic Sciences, Lahijan Branch,
Islamic Azad University (IAU), Lahijan, Guilan, Iran*

Due to the fact that different isoforms of carbonic anhydrase play distinct physiological roles, their diseases/disorders involvement are different as well. Involvement in major disorders such as glaucoma, epilepsy, Alzheimer's disease, obesity and cancers, have turned carbonic anhydrase into a popular case study in the field of rational drug design. Since carbonic anhydrases are highly similar with regard to their structures, selective inhibition of different isoforms has been a significant challenge. By applying a proteochemometrics approach, herein the chemical interaction space governed by acyl selenoureido benzensulfonamides and human carbonic anhydrases is explored. To assess the validity, robustness and predictivity power of the proteochemometrics model, a diverse set of validation methods was used. The final model is shown to provide valuable structural information that can be considered for new selective inhibitors design. Using the supplied information and to show the applicability of the constructed model, new compounds were designed. Monitoring of selectivity ratios of new designs shows very promising results with regard to their selectivity for a specific isoform of carbonic anhydrase.

Keywords: Carbonic anhydrase. Proteochemometric modeling. Grid Independent Descriptors. Z-scales. Isoform selectivity.

INTRODUCTION

Crucial and diverse physiologic processes such as lipogenesis, ureagenesis, gluconeogenesis, acid–base balance and oncogenesis are mediated by the members of the carbonic anhydrase (CA) superfamily. CAs play their critical roles through catalysing a simple reaction of interconversion between bicarbonate and carbon dioxide (Equation 1) (Supuran *et al.*, 2004).

$$CO_2 + H_2O \rightarrow HCO_3^- + H^+ \qquad \textit{Equation (1)}$$

There are seven classes of CA known to date, α-, β-, γ-, δ-, ξ-, η- and θ, of which the class of α is found in vertebrates (Supuran *et al.*, 2004). Alpha-CA consists of sixteen isoforms distributed in different tissues. Isoform variation and diverse localisations has made CA a vital target for the treatment of many disorders. As an example, isoform II with cytosolic localisation is a target for glaucoma, while isoforms IX and XII localised in transmembrane are involved in tumorigenesis and therefore are major cancer therapy targets (Ditte *et al.*, 2011; Purkerson, Schwartz, 2007; Henry, 1996). Various physiological roles in addition to the structural resemblance of different isoforms cause a great necessity for developing isoform-selective inhibitors, to avoid unwanted side effects. Inhibitors of CAs (CAIs) can be put in two major classes of sulfonamide and non-sulfonamide derivatives (Supuran *et al.*, 2003). The latter itself is divided into different sub-groups such as polyamines (Carta *et al.*, 2010), phenols (Carta *et al.*, 2013), thiols (Wischeler *et al.*, 2010) and coumarin derivatives (Bonneau *et al.*, 2013), all showing inhibitory mechanisms different than that

***Correspondence:** Behnam Rasti. Department of Microbiology. Faculty of Basic Sciences. Lahijan Branch, Islamic Azad University (IAU). Lahijan, Guilan, 44169-39515, Iran. E-mail: rasti@liau.ac.ir. ORCID: 0000-0002-5061-3142

of sulfonamide derivatives. Sulfonamide derivatives with an unsubstituted functional group ($-SO_2NH_2-$) have the highest inhibitory potency among sulfonamide derivatives, suggesting that the sulfonamide group is highly essential for the inhibitory mechanism of sulfonamide derivatives (Supuran *et al.*, 2004). To sum up, given the importance of CA and non-selectivity of many current inhibitors, the design of compounds with selective inhibitory properties is of great essence. To investigate the selectivity, the proteochemometrics (PCM) approach is shown to be highly effective. PCM can be considered as an extended form of the traditional quantitative structure activity relationship (QSAR) models in which, interaction space of different ligands across multiple receptors is simultaneously modelled (Prusis *et al.*, 2001). Major protein families including G protein-coupled receptors (Lapinsh *et al.*, 2005), proteases (Lapinsh *et al.*, 2008), kinases (Subramanian *et al.*, 2013), aromatase (Simeon *et al.*, 2016), thymidylate synthase (Rasti, Shahangian, 2018), carbonic anhydrase (Rasti, Karimi-Jafari, Ghasemi, 2016; Rasti *et al.*, 2017a;2018), phosphodiesterase (Rasti *et al.*, 2017b), Dihydrofolate Reductase (Hariri *et al.*, 2019), and antimicrobial inhibitors (Rasti *et al.*, 2019) have been so far investigated using PCM, and valuable information has been successfully gathered. The inhibitory mechanism of acyl selenoureido benzensulfonamides against human carbonic anhydrases (hCAs) has been previously studied (Angeli *et al.*, 2017). However, the previously proposed

inhibitors barely show selectivity towards a specific isoform of hCA. Therefore, herein, the chemical interaction space governed by the acyl selenoureido benzensulfonamides and hCAs is explored. Applying PCM, compound structural descriptors capable of discriminating between hCA isoforms were found. Utilizing these discriminative descriptors, novel acyl selenoureido benzensulfonamides with better selectivity towards different hCAs were then designed.

METHODS

Interaction data

A data set of 64 complexes (hCAs-acyl selenoureido benzensulfonamides) along with their bioactivities (Ki) (Angeli *et al.*, 2017) was applied for the purpose of the modeling. A schematic workflow representing the steps of PCM modeling is summarized in figure 1. Besides the popularity of sulfonamide derivatives as CAIs and their clinical involvement in major disorders such as glaucoma and obesity, the following facts were also considered for dataset selection: *i*) the selected sulfonamides are a novel group of derivatives belong to the acyl selenoureido benzensulfonamides; *ii*) the majority of these acyl selenoureido benzensulfonamides are hardly selective for a specific isoform of hCA. The chemical scaffold of these compounds are represented in figure 2. Biological activities were subjected to negative logarithmic transformation (pKi) to give more spread data points.
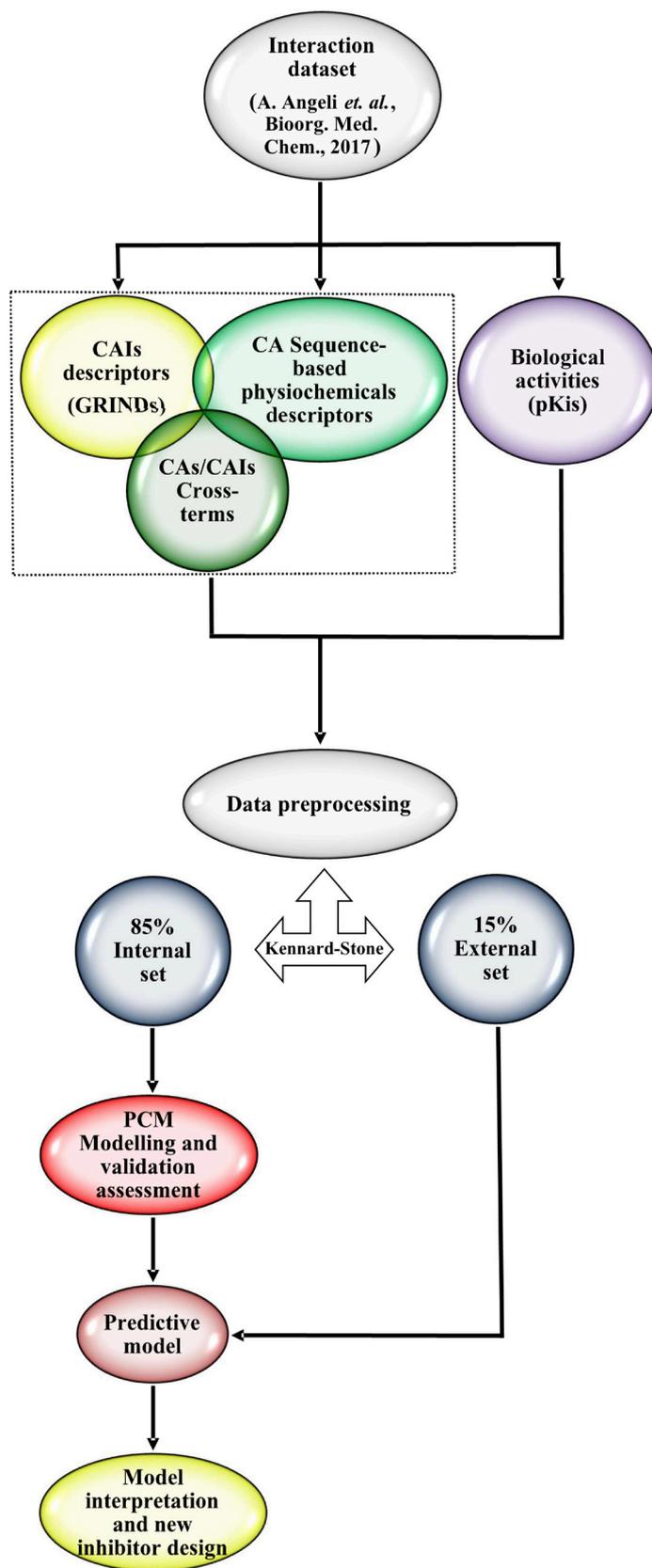
FIGURE 1 - Workflow of PCM modelling.
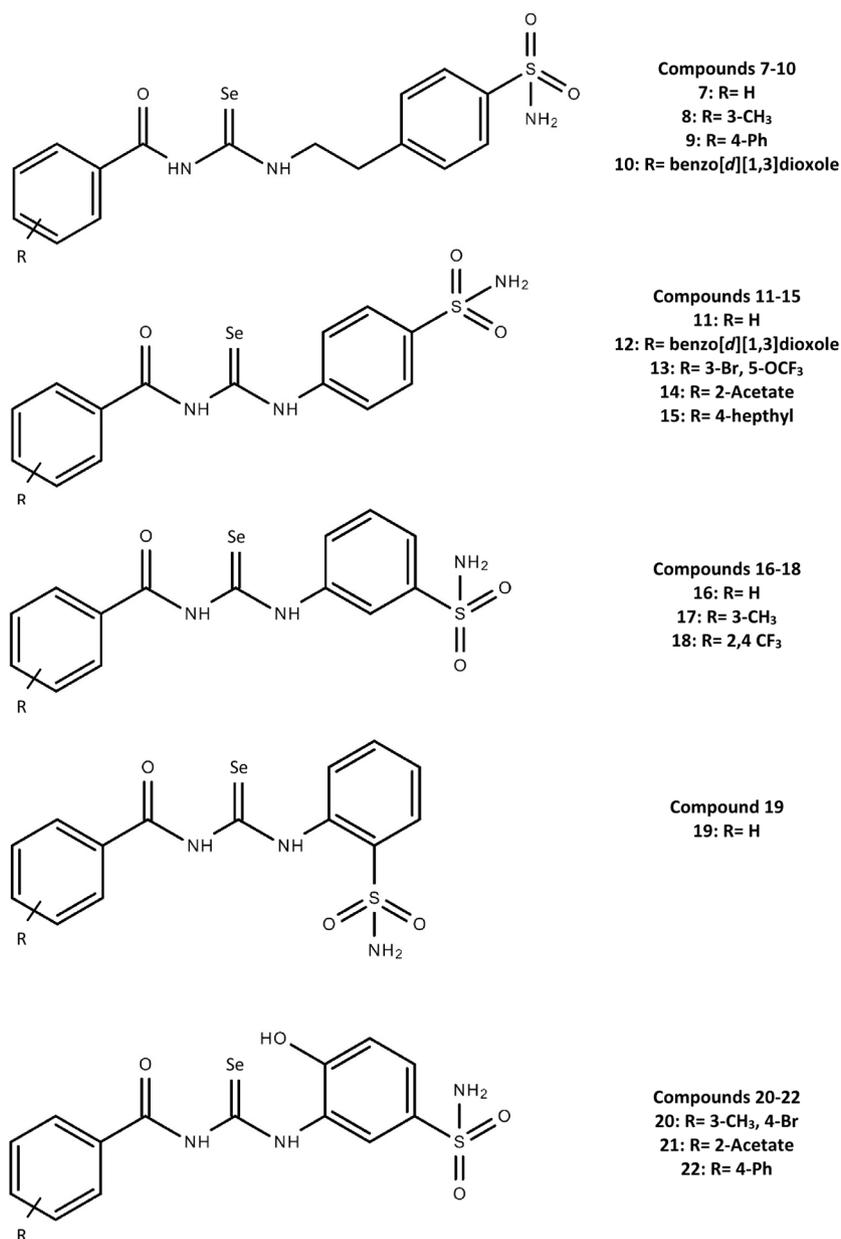
FIGURE 2 - Chemical scaffolds of acyl selenoureido benzensulfonamides as CA inhibitors.

Descriptors of organic compounds

Chemical structures were drawn in SYBYL7.3 (Version, S., 6.9, Tripos Associates, St. Louis, Mo, 2001) and the optimisation was performed applying Tripos force field with a distance-dependent dielectric and the Powell conjugate gradient algorithm with convergence criterion of 0.001 kcal/mol Å. Partial atomic charges and Grid Independent Descriptors (GRINDs) were calculated using Gasteiger-Huckel method and Pentacle 1.05 software, respectively (Pastor *et al.*, 2000). Regarding the nature of the Grid Independent Descriptors, the final information provided by the GRINDs approach is directly correlated to the structures of inhibitors. In the present study, the following probes were applied to calculate the MIFs: DRY (hydrophobic probe), N1 (representing H-bond donor), O (representing H-bond acceptor) and TIP (representing molecular shape). Grid points distances, the number of

filtered nodes for each MIF and the smoothing window were adjusted to 0.5 Å, 100 and 0.4 Å, respectively. Utilizing these parameters, 73 descriptors were calculated for each MIF-MIF multiplication. Considering the application of four types of MIF, a total number of 730 descriptors were generated with regard to each structure. All descriptors showing the same value for all compounds were removed.

Descriptors of carbonic anhydrases

Carbonic anhydrase is a metalloenzyme with a zinc ion positioning in the center of protein cavity. To identify the key positions interacting with chemical compounds, a cutoff of 10 Å from the location of the Zn^{2+} was applied. Residues falling within the 10 Å cutoff were considered as ligand interacting residues. Of the 55 amino acids, 26 positions were found to be mutated as represented in figure 3. Regarding the proteins, the sequence based Z-scale descriptors are the most widely used descriptors in PCM modeling (Hellberg *et al.*, 1987; Sandberg *et al.*, 1998). Z-scales are based on a PCA analysis originally performed by Hellberg *et al.* on physicochemical data gathered from experiments like NMR and thin-layer chromatography (TLC). Using 26 properties derived for 87 natural and unnatural amino acids, the original three-component Z-scales were then extended to five-membered Z-scales by Sandberg *et al.*. Z1, Z2 and Z3 are the three first principal components, representing the largest variations of physicochemical properties. The fourth and fifth components, Z4 and Z5, are more difficult to interpret due to properties such as electronegativity, heat of formation, electrophilicity and hardness. Z1, Z2, Z3, Z4 and Z5 are representing lipophilicity, size/polarizability, electronic and electrostatic properties, respectively.


```
                       1111111111111111111111111122222222222222
                 12236666669999990001112223344444479999900000000004444
                 576890245679123456567789012151234560467890123456789 4567
CA I     WYWQSPVHSFHNFQFHFHSEHELHVAHLALAVIGVKYGSLTHPPLYESVTWNNRP
CA II    WYWQSPNHAFNEIQFHFHSEHELHLVHFVLAVLGIKYGSLTTPPLLECVTWNWRP
CA IX    WYWQSPNHSVQTLQLHLHSEHEIHVVHVLLAVLAAEYGSLTTPPCAQGVIWNFRA
CA VII   WYWQSPNHSVQDKQFHFHSEHELHLVHFALAVVGVKYGSLTTPPLSESVTWNFRP
         ******  *     * * ***** *  *  ***     ***** **   * ** *
```

FIGURE 3 - Structure and Sequence alignment of hCA. Upper panel represents the structure of hCA II (PDB ID: 2VVB). Major cavity non-conserved residues are indicated in illustration. Lower panel shows the sequence alignment of hCA isoforms. Conserved positions are marked by asterisks. All the numberings are based on the sequence numbers of hCA II.

Feature selection

GRINDs were subjected to genetic algorithm (GA) (Rogers, Hopfinger, 1994) to select the best fitted structural features. A population size of 256, a generation size of 100, a mutation rate of 0.01 and double crossing over with the rate of 0.8 were utilised as GA parameters. The Equation 2, representing predictive squared correlation coefficient (Q^2), was used as fitness function. 120 most fitted GRINDs were selected by the algorithm.

$$Q^2 = 1 - \frac{PRESS}{\sum_{i=1}^{n} (y_i - y_m)^2}$$

$$PRESS = \sum_{i=1}^{n} \left(y_{predicted,\ i} - y_i\right)^2$$

Equation (2)

where the is the data values not used to construct the CV model, the is the mean of the experimental bioactivities, and PRESS is the predictive residual sum of the squares.

Ligand-receptor cross-term descriptors

In PCM models, interactions between ligands and receptors are represented by cross-terms. Herein, cross-terms were calculated by multiplying descriptors of proteins by those of ligands. Mean-centering and scaling to unit variance was performed for all descriptors prior to modelling.

Data splitting

Considering 16 compounds and 4 carbonic anhydrase isoforms, the final chemical space being investigated consists of 64 receptor-ligand complexes. Prior to the modelling, the dataset was split (Kennard *et al.*, 1969) into two internal validation (nearly 85% of dataset) and external test sets (nearly 15% of dataset).

Multivariate modelling

Partial least squares (PLS) is an extension of PCA that simultaneously projects descriptors and biological activities to PLS components and finds linear relationships between them. PLS is capable of making reliable predictive models even in the presence of collinearity and redundancy of variables. Regression coefficients derived from the PLS regression equation reveals the direction and magnitude of the influence of X on Y. Regarding the protein-ligand system, the PLS regression equation can be expressed as follows:

$$y = \bar{y} + \sum_{p=1}^{P} coeff_p D_p + \sum_{l=P+1}^{P+L} coeff_l D_l + \sum_{p=1}^{P} \sum_{l=P+1}^{P+L} coeff_{p,l} D_p D_l$$

Equation (3)

where, and are regression coefficients, and and stand for receptor and ligand descriptors, respectively. The PLS Toolbox 3.5 was applied to conduct PLS regression.

Validation of model performance

The internal validation set was subjected to training and venetian blinds cross-validation. The external set was then used to validate the predictability of the constructed models. An applicability domain (AD) analysis was also performed on internal and external sets 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 (Gramatica., 2007). The William's plot was built by plotting the standardised residuals of compounds versus their leverage values. In addition, Y-scrambling approach was applied to assess the robustness of the PCM models. The response vector (Y) was randomly scrambled for 100 times, while the X-data were kept untouched. Each time a new model was generated with the scrambled Y and the correspondent R^2 and Q^2 were calculated. The R^2 and Q^2 values of the original model together with the R^2 and Q^2 values of 100 scrambled 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. It has been already shown that the $R^2_{intercept}$ of < 0.3 and the $Q^2_{intercept}$ of < 0.05 are acceptable for a robust model (Eriksson *et al.*, 2003).

Contribution of ligand properties for receptor selectivity

Unlike traditional QSAR, PCM applies compounds and proteins chemical spaces and their intercepts (cross-terms) to show the ability of compounds and proteins for making non-covalent interactions. Cross-terms, which are considered as a representation of interactions between compounds and proteins, can be used to assess the significance of a compound descriptor in selective inhibition of different receptors (Equation 4).

$$\Delta y_{rec.l} = \frac{dy}{dD_l} = coeff_l + \sum_{p=1}^{P} coeff_{p.l} D_p$$

Equation (4)

where *rec* is a receptor with descriptors D_1, D_2 …. D_p, and shows the change in the selectivity of the lth descriptor of compounds for this particular receptor. Moreover, and are regression coefficients, and and stand for receptor and ligand descriptors, respectively.

RESULTS AND DISCUSSION

PCM modeling of carbonic anhydrase activity

Prior to construct the PCM models, GRINDs were subjected to genetic algorithm to select the best fitted structural descriptors. Combinations of GRINDs and Z-scales were used to build PCM models. Compared to GRINDs/Z-scales-5 based model (model 2: $R^2 \sim 0.98$, $Q^2_{CV} \sim 0.63$, $Q^2_{EXT} \sim 0.6$), GRINDs/Z-scales-3 based model showed better metrics (model 1: $R^2 \sim 0.97$, $Q^2_{CV} \sim 0.70$, $Q^2_{EXT} \sim 0.79$). As mentioned earlier, Z1, Z2 and Z3 are the first three PCs, representing the largest variations of physicochemical properties. Therefore, herein, it seems that introducing more variance (Z4 and Z5) is not in favour of the modelling. Therefore, GRINDs/Z-scales-3 based model was considered for further analyses. Robustness and likelihood of reliable prediction were confirmed for the final model by Y-scrambling approach and AD analysis, respectively. Both $R^2_{intercept}$ and $Q^2_{intercept}$ show acceptable values of ~ 0.1 and ~ 0, confirming the robustness of the PCM model. As is illustrated in the Williams plot (Figure 4), almost all compounds are located within the boundaries of applicability domain, suggesting a well-defined AD for the proposed PCM model. There are only two compounds (one belongs to the internal set and the other belongs to the external set) falling outside the ± 3 standardised residual ranges.
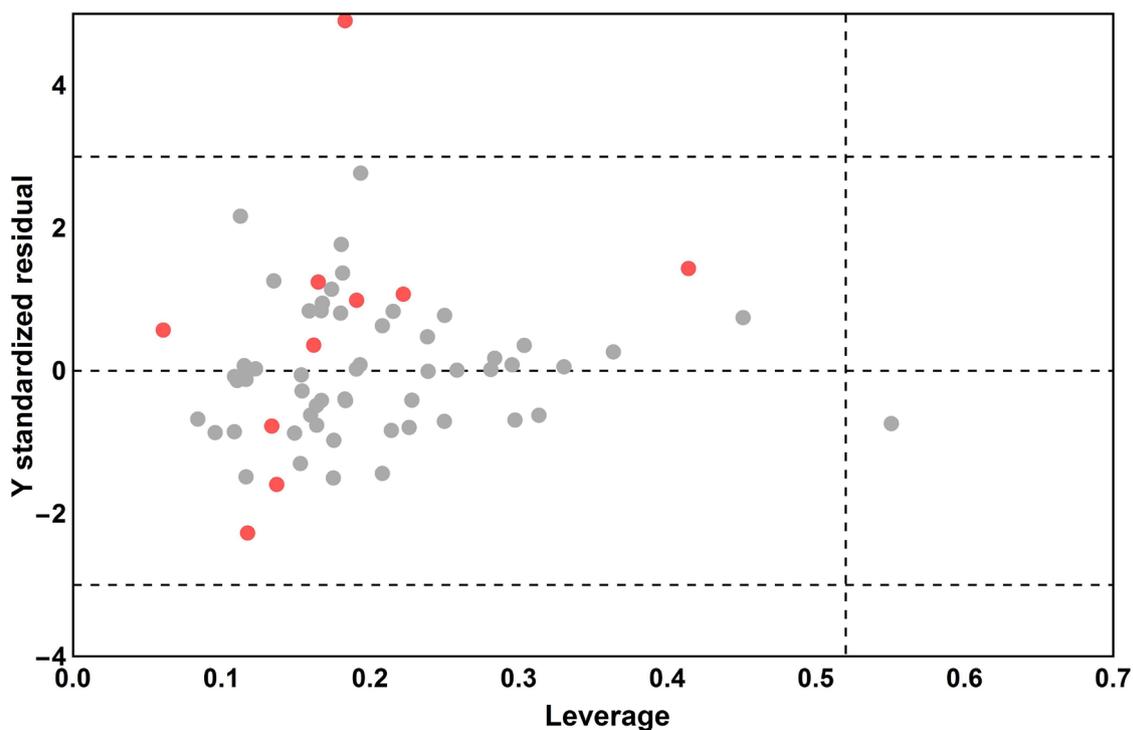
FIGURE 4 - Evaluation of applicability domain using William plots for PCM model. Internal and external set compounds are indicated by gray and red circles, respectively.

Assessing the significance of compound descriptors in receptor selectivity

Applying PCM, discriminative information with regard to receptors chemical space was entered into the models. Consequently, specific contributors showing high selectively for a particular isoform of carbonic anhydrase were found. Equation (4) was used to assess the significance of compound descriptors in receptor selectivity. Figure 5 illustrates $\Delta y_{rec,l}$ values for structural descriptors of compounds, selected by genetic algorithm. The vertical dashed separators represent a distance range from 0 Å to 29.2 Å. Inspecting figure 5 with regard to DRY-DRY descriptors reveals their positive values at short and medium range distances in case of isoform IX, in particular. This finding suggests that the presence of hydrophobic groups, with certain distances, in the structure of compounds strongly favours isoform IX inhibition. A similar situation applies to N1-N1 descriptors with regard to isoform VII. Their positive values at short range distances suggests that the presence of HBA (H-bond acceptor) moieties, with short range distances, in the structure of compounds is strongly in favour of isoform VII inhibition. GRINDs with highly discriminative selectivity (coefficient) towards a specific isoform of carbonic anhydrase are circled and discussed in details below.
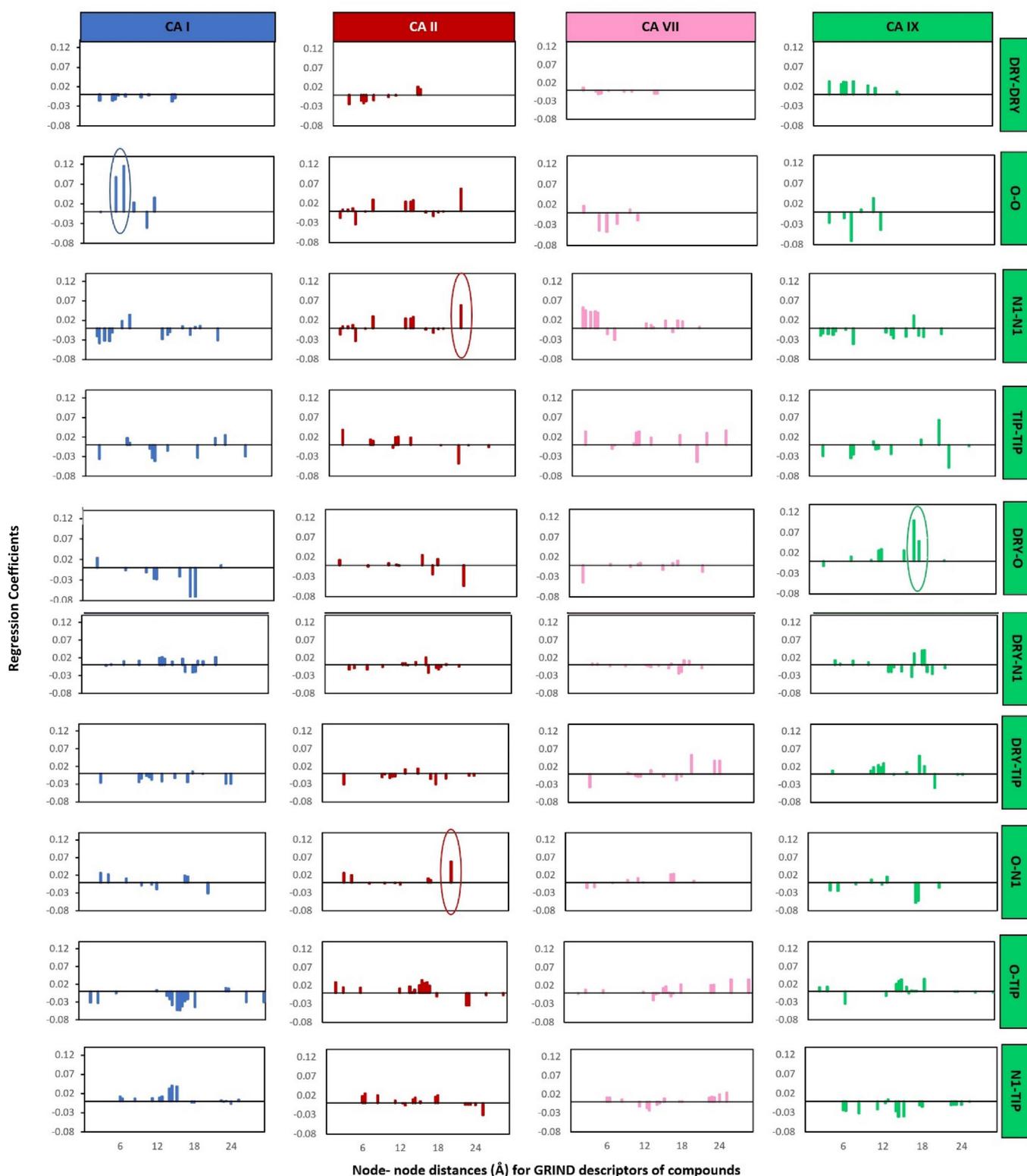
FIGURE 5 - Overall representation of GRINDs contribution in receptor selectivity. Y-axes indicate the regression coefficients. Intervals separated by dashed vertical separators represents the node-node distance range of 0 Å to 29.2 Å for each type of GRIND. GRINDs which are highly discriminative with regard to a certain receptor are circled.

O-O descriptors: Inspecting O-O descriptors indicates two regions at distances of 5.2 Å and 6.4 Å (Figure 5, blue oval) with significantly positive coefficients for isoform I, in particular. In case of other isoforms, the correspondent descriptors seem to be at expense of compound-protein interactions, since they have highly negative PLS coefficients. Subsequently, this finding suggests that emplacement of HBD (H-bond donor) moieties, within the correspondent distances, can substantially turn the selectivity of compounds in favour of receptor I.

N1-N1 and O-N1 descriptors: A close look at N1-N1 and O-N1 descriptors reveals that they have highly positive PLS coefficients for isoform II with regard to distances of 21.2 Å and 20.4 Å, respectively (Figure 5, red ovals). In case of other isoforms, however, the correspondent descriptors are either at expense of or barely involved in compound-protein interactions. Therefore, it can be concluded that emplacement of dual moieties of either HBA/HBA or HBD/HBA, within the correspondent distances, can be in favour of selective inhibition of isoform II.

DRY-O descriptors: Observing DRY-O descriptors reveals that there are two regions at distances of 16.8 Å and 17.6 Å with highly PLS coefficients for isoform IX, in particular (Figure 5, green oval). Since the correspondent descriptors in other isoforms are not involved in their inhibition, compounds with dual hydrophobic/HBD moieties, within the correspondent distances, are expected to act as selective inhibitors for isoform IX.

New inhibitor design

To show the applicability of the GRINDs, discovered to highly discriminate the investigated isoforms, some of current inhibitors were modified and their selectivity specification toward CA isoforms were monitored. Structural modifications were exerted based on the discovered discriminative contributors to change the selectivity ratio of the compounds. Original structures and their modified versions are represented in figure 6. Compounds 7, 11 and 19 were chosen to test the applicability of the GRINDs discovered to be discriminative with respect to isoforms II, I and

IX, respectively (the discovered GRINDs are already discussed above). The main reason why these compounds were chosen, other than non-selectivity for their correspondent receptors, underlies the fact that their original structures lack the discovered discriminative GRINDs. Therefore, any possible inhibitory alternation raised by the modified structures was surely caused by the newly added GRINDs. As is illustrated in figure 6, compound-7 (Comp_7) lacks both O-N1 and N1-N1 descriptors at distances of 20 Å and 21.2 Å, respectively. To generate the missing descriptors, a $-CF_3$ moiety was added to the phenyl ring (Mod_7). The modifications were expected to increase selectivity of compound-7 for CA II. Compound-11 (Comp_11) which lacks O-O descriptors at distances of 5.2 Å and 6.4 Å, was selected to test the effect of these GRINDs on selectivity of CA I. To generate the missing descriptors, $-NH_2$ moieties were substituted in positions 3 and 6 of the phenyl ring (Mod_11). The modifications were expected to change the selectivity in favor of CA I. Finally, compound-19 (Comp_19) which lacks DRY-O descriptor at distances of 16.8 Å and 17.6 Å, was modified through adding a $-CH_3$ group to the phenyl ring (Mod_19). The modified compound was then expected to show better selectivity against CA IX. Three modified compounds were then put in the test set and their activities were predicted using the constructed PCM model. The activity ratios of designed compounds for CA isoforms are presented in Table I. As it is indicated, the results are very promising with regard to all three modified compounds. In the majority of cases, introducing the missing discriminative GRINDs has altered the selectivity ratios as expected. Significant examples where the ratio was reversed for two isoforms are as follows: *i)* Comp_7 shows selectivity ratio of ~ 2.6 for CA II/CA IX, while it's modified version (Mod_7) shows selectivity ratio of ~ 0.14 for CA II/CA IX. *ii)* Comp_19 with selectivity ratios of ~ 4.24 and ~ 2.83 for CA IX/CA II and CA IX/CA VII, respectively. After modification (Mod_19) the selectivity ratios have been inverted in favour of CA IX, by values of ~ 0.95 and ~ 0.59 respectively. Other significant favourable changes belong to selectivity ratios of CA I/CA IX, CA II/ CA I and CA IX/CA I with ~ 8.95, ~ 5.6 and ~ 4.62 folds, respectively. In case of recent descriptors; however, the selectivity ratios

are not inverted. In three cases (CA I/CA II, CA I/CA VII, CA II/CA VII) the selectivity ratios are not changed enough to be commented on. Taken together, it seems that applying the discriminative structural contributors captured by the constructed PCM model, can increase the selectivity of a compound for a specific receptor.
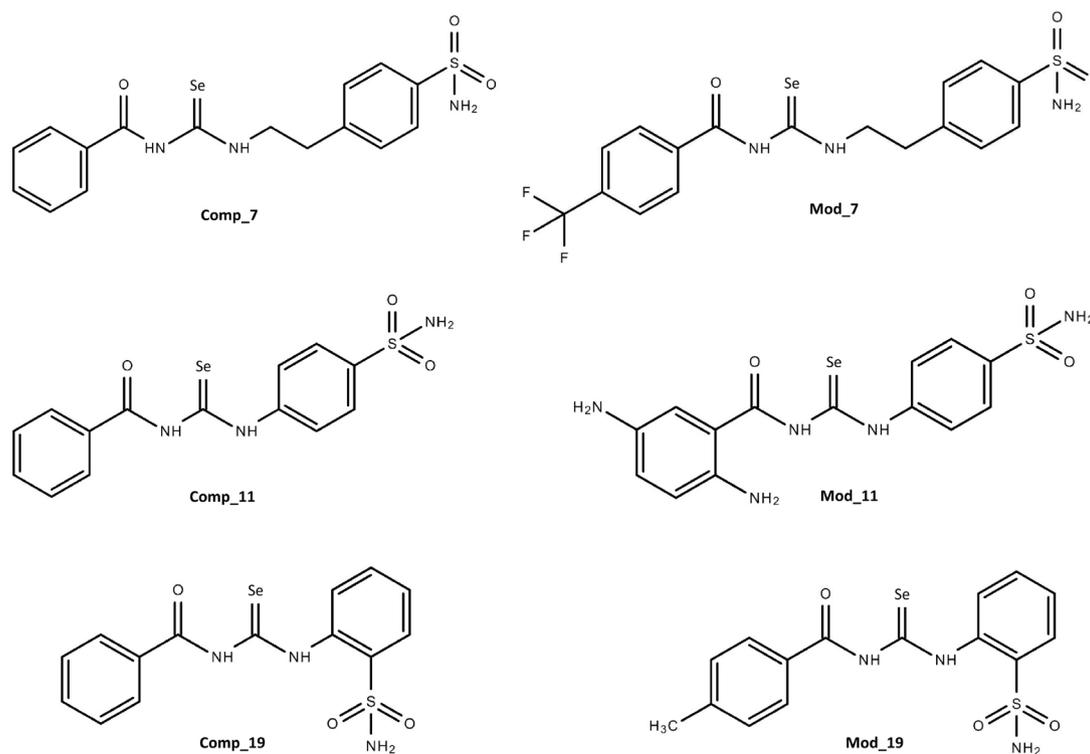
FIGURE 6 - Original and modified compounds used to explore the impact of discriminative GRINDs on CAs selectivity**.** Original and modified structures are indicated by Comp and Mod tags, respectively.

TABLE I - Experimental/experimental and predicted/predicted ratio values of Kis for original (Comp) and modified (Mod) compounds, respectively

CA II/CA I	CA II/CA VII	CA II/CA IX	
~ 0.28	~ 5.84	~ 2.60	Comp_7
~ 0.05	~ 5.20	~ 0.14	Mod_7

CA I/CA II	CA I/CA VII	CA I/CA IX	
~ 7.77	~ 7.94	~ 16.05	Comp_11
~ 7.60	~ 7.60	~ 1.79	Mod_11

CA IX/CA I	CA IX/CA II	CA IX/CA VII	
~ 4.62	~ 4.24	~ 2.83	Comp_19
~ 1.00	~ 0.95	~ 0.59	Mod_19

CONCLUSIONS

Due to diverse biological engagements of CA isoforms, availability of isoform-selective inhibitors is of great importance. Complementary physicochemical/structural contributors of ligands and receptors mediate their interactions. Using a proteochemometrics approach, information extracted from ligands-receptors chemical interaction spaces can be applied to investigate selective inhibition of a protein isoform by a set of compounds. Herein, a set of sequence based descriptors and their combinations together with 3D-based GRINDs were used to assess the impact of sequence based descriptors on modelling. Robustness and predictivity power of models were validated using a diverse set of approaches; and based on the obtained results, GRINDs/Z-scales-3 based model showed the best indices. Furthermore, major structural contributors with regard to hydrophobic and H-bond interactions have been captured by the PCM model and shown to be discriminative against different CA isoforms. Such contributors can be applied to design new potent inhibitors with better selectivity toward a specific isoform of carbonic anhydrase.

REFERENCES

Angeli A, Carta F, Bartolucci G, Supuran CT. Synthesis of novel acyl selenoureido benzensulfonamides as carbonic anhydrase I, II, VII and IX inhibitors. Bioorg Med Chem. 2017;25(13):3567-3573.

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.

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.

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.

Ditte P, Dequiedt F, Svastova E, Hulikova A, Ohradanova-Repic A, Zatovicova M, et al. Phosphorylation of carbonic anhydrase IX controls its ability to mediate extracellular acidification in hypoxic tumors. Cancer Res. 2011;71(24):7558-7567.

Eriksson L, Jaworska J, Worth AP, Cronin MT, 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.

Gramatica P. Principles of QSAR models validation: internal and external. Mol Inf. 2007;26(5):694-701.

Hariri S, Ghasemi JB, Shirini F, Rasti B. Probing the Origin of Dihydrofolate Reductase Inhibition via Proteochemometric Modeling. J Chemom. 2019;33(2):e3090.

Hellberg S, Sjöström M, Skagerberg B, Wold S. Peptide quantitative structure-activity relationships, a multivariate approach. J Med Chem. 1987;30(7):1126–1135.

Henry RP. Multiple roles of carbonic anhydrase in cellular transport and metabolism. Annu Rev Physiol. 1996;58:523-538.

Kennard RW, Stone LA. Computer aided design of experiments. Technometrics. 1969;11(1):137-148.

Lapinsh M, Eklund M, Spjuth O, Prusis P, Wikberg JE. Proteochemometric modeling of HIV protease susceptibility. BMC Bioinf. 2008;9:181.

Lapinsh M, Prusis P, Uhlén S, Wikberg JE. Improved approach for proteochemometrics modeling: application to organic compound—amine G protein-coupled receptor interactions. Bioinformatics. 2005;21(23):4289-4296.

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.

Prusis P, Muceniece R, Andersson P, Post C, Lundstedt T, Wikberg JE. PLS modeling of chimeric MS04/MSH-peptide and MC 1/MC 3-receptor interactions reveals a novel method for the analysis of ligand–receptor interactions. Biochim Biophys Acta. 2001;1544(1-2):350-357.

Purkerson J, Schwartz G. The role of carbonic anhydrases in renal physiology. Kidney Int. 2007;71(2):103-115.

Rasti B, Entezari Heravi Y. Probing the chemical interaction space governed by 4-aminosubstituted benzenesulfonamides and carbonic anhydrase isoforms. Res Pharm Sci. 2018;13(3):192–204.

Rasti B, Karimi-Jafari MH, 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.

Rasti B, Mazraedoost S, Panahi H, Falahati M, Attar F. New insights into the selective inhibition of the β-carbonic anhydrases of pathogenic bacteria Burkholderia pseudomallei and Francisella tularensis: a proteochemometrics study. Mol Divers. 2019;23(2):263-273.

Rasti B, Namazi M, Karimi-Jafari MH, 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. 2017a;36(4):1600102.

Rasti B, Schaduangrat N, Shahangian SS, Nantasenamat C. Exploring the origin of phosphodiesterase inhibition via proteochemometric modeling. RSC Adv. 2017b;7:28056-28068.

Rasti B, Shahangian SS. Proteochemometric Modeling of the Origin of Thymidylate Synthase Inhibition. Chem Biol Drug Des. 2018;91(5):1007-1016.

Rogers D, Hopfinger AJ. Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. J Chem Inf Comput Sci. 1994;34(4):854-866.

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.

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.

Subramanian V, Prusis P, Pietilä LO, Xhaard H, Wohlfahrt G. Visually interpretable models of kinase selectivity related features derived from field-based proteochemometrics. J Chem Inf Model. 2013;53(11):3021-3030.

Supuran CT, Scozzafava A, Casini A. Carbonic anhydrase inhibitors. Med Res Rev. 2003;23(2):146-189.

Supuran CT, Scozzafava A, Conway J. Carbonic anhydrase: its inhibitors and activators. Vol. 1. 2004: CRC Press.

Wischeler JS, Innocenti A, Vullo D, Agrawal A, Cohen SM, Heine A, et al. Bidentate zinc chelators for α-carbonic anhydrases that produce a trigonal bipyramidal coordination geometry. Chem Med Chem. 2010;5(9):1609-1615.