



QSAR modeling of pyrazoline derivative as carbonic anhydrase inhibitors

Imad Hammoudan¹ · Mohammed Chafi¹

Received: 8 February 2023 / Accepted: 12 June 2023

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

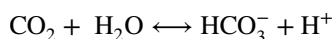
Abstract

The efficacy of 34 pyrazoline derivatives as carbonic anhydrase inhibitors was studied *in silico*. The quantum descriptors were calculated by the DFT/B3LYP method using the 6-31G(d) basis; the dataset was randomly divided into training and testing. By altering the compounds in the sets, four models were created, and they were then used to determine the predicted pIC₅₀ values for the six chemicals in the test set. According to the OECD guidelines for QSAR model validation and the Golbraikh and Tropsha's criteria for model approval, each created model was independently validated both internally and externally, along with YRandomization. Model 3 is chosen because it has higher R₂, R_{2test}, and Q_{2cv} values (R₂=0.79, R_{2test}²=0.95, Q_{2cv}²=0.64). Only one descriptor has a proportional influence on pIC₅₀ activity, but the other four descriptors have an inverse influence on pIC₅₀ because of the negative contribution coefficient. Given the descriptors of the model, we could propose new molecules with remarkable inhibitory activity.

Keywords QSAR · MLR · Carbonic anhydrase · DFT

Introduction

Carbonic anhydrase isozyme XII, a newly identified member of the carbonic anhydrase gene family, has been linked to von Hippel-Lindau gene-mediated carcinogenesis (Kivelä et al. 2000). These enzymes participate in buffering the pH of intra- and extracellular spaces by catalyzing the reversible hydration of carbon dioxide and water (Wang et al. 2013). Carbonic anhydrase (CA XII) is the most strongly expressed gene in response to hypoxia in human cancer cells (carcinomas: colorectal, breast, lung, etc.) (Hilvo et al. 2008). CA enzymes promote biosynthetic reactions in the body, as well as the production or expenditure of carbon dioxide and bicarbonate according to the following reaction (Yorulmaz and Eroğlu 2020):



Acetazolamide, methazolamide, ethoxzolamide, dichlorphenamide, dorzolamide, and brinzolamide are drugs for treating glaucoma by inhibiting the cytosolic isoform CAII. CA XII is considered a target for the treatment of certain cancerous tumors (Jaakkola et al. 2001). Research indicates that loss of CA XII function should be considered in individuals without CFTR mutations who have CF-like features in the sweat glands and lungs (Lee et al. 2016). Recent research has focused on inhibitors of CA XII associated with tumors. Some attempts have been made to synthesize selective inhibitors for use as drugs. CA XII has been validated as a marker for many hypoxic tumors and their inhibition reduces the growth of primary tumors and metastases (Matysiak et al. 2017). Zinc-binding compounds (depending on their binding mode) are effective for drug design; they can be like CAI inhibitors (Mishra et al. 2020). Many of the recent QSAR studies have shown a good connection between the ligands and CA XII through the relationship between the activity of the compounds and the descriptors (Eroğlu 2019). More than 1250 molecular descriptors were calculated using the reliable programs. Multiple linear regression equations were developed and validated using the validation technique (Alafeefy et al. 2015). QSAR models were built to explore the correlations between the molecular

Responsible Editor: Lotfi Aleya

✉ Mohammed Chafi
chafimoham@gmail.com

¹ LIPE, Higher School of Technology, University Hassan II of Casablanca, B.P 8012 Oasis, Casablanca, Morocco

descriptors calculated on 16 compounds and their experimental inhibitory activities on CA XII (Kumar and Roy 2020). Results found showed the characteristics of mercapto quinazolinone benzene sulfonamide derivatives against hCA XII (Gopinath and Kathiravan 2022). In our research, we will develop QSAR models to find a good correlation between the activity of pyrazoline molecular series and their descriptors derivative to inhibit CA XII.

Computational methods

Using the Chemdraw3D (Mendelsohn 2004) and Chemsketch software (Li et al. 2004), we were able to calculate the thermodynamic and topological descriptors, and using the Gaussian 09 (Gaussian_09_ReferenceManual.Pdf 2022) software, the quantum descriptors were calculated. The functional and the basis used in the quantum computation are, respectively, DFT (Rivero et al. 2015)/B3LYP (Zhang and Xu 2021) and 6-3G(d). Principal component analysis (PCA) (Josse et al. 2009) and multiple linear regression (MLR) (Zhou et al. 2017) were performed using ChemOffice 2015 and XLSTAT (Vidal et al. 2020). The relationship between pIC50 for 34 compounds and the descriptors was studied by the MLR statistical technique. We made the applicability domain for the chosen model using Matlab 2015 software. This enzyme is located outside the cell and has a very high catalytic activity and is also a multidomain protein that can be inhibited by pyrazoline sulfamate derivatives.

• Dataset and generation of molecular descriptors

Data on the activities against carbonic anhydrase of 34 pyrazoline derivatives were collected from the literature (Moi et al. 2019). IC50 is the inhibitory activity factor of the bioassay, which shows the required concentration of an inhibitor to achieving 50% inhibition of carbonic anhydrase replication. Table 1 shows the compounds studied with their activity. The experimental IC50 activity values were converted to the negative logarithm of the IC50 activity (pIC50). 11 quantum chemistry descriptors were calculated using the findings of DFT(B3LYP/6-31G(d)) computations (Table S1). Another 34 descriptors were calculated using the ChemOffice 3D and Chemscketch programs (Table S2 and S3, respectively).

• Principal component analysis (PCA)

Descriptors with a low correlation coefficient ($r \leq 0.15$) value concerning the dependent variable IC50 were excluded; hence, the interest of the principal component analysis (PCA). This allows us to select the input

data for the multiple linear regression studies. Nevertheless, the descriptors with a correlation coefficient higher than 0.95 are taken into account to reduce the uncertainty present in our data matrix. Thus, thanks to PCA, we were able to select 20 descriptors from 40 to use in the development of MLR models.

• Data splits and model development

We divided the dataset randomly (80% for the training set and 20% for the test set).

The multiple linear regression (MLR) method was used for training set regression. The transparency of linear regression analysis is a significant benefit; as a result, the algorithm is easily accessible and predictions can be made.

Model validation

The QSAR model's predictive power and fit were assessed through internal and external validation measures. Coefficient of determination R^2 , Fischer's value (F_{test}), mean square error of the model (MSE), variance inflation factor (VIF), leave-one-out cross-validation coefficient of determination Q_{cv}^2 , external test coefficient of determination R_{test}^2 , and Y-randomization parameters (R_{Rand}^2 and $R_{cv(\text{Rand})}^2$) are quality validation parameters. For a model to be valid, the proposed new molecules must be within the OECD range of applicability.

Results and discussion

Model development

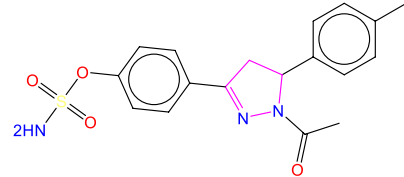
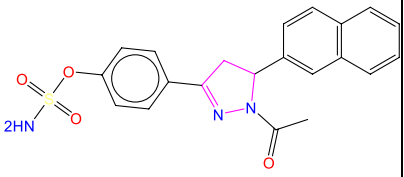
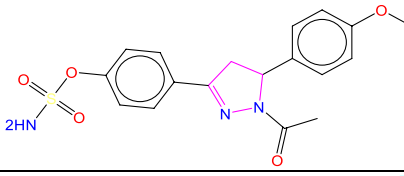
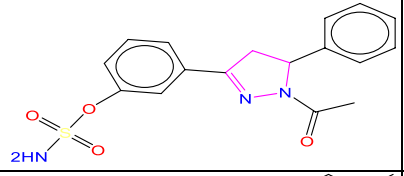
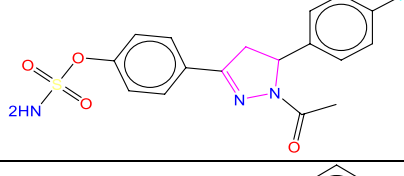
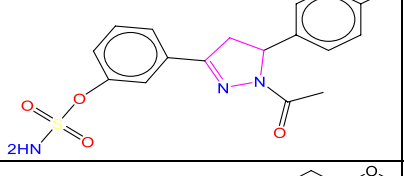
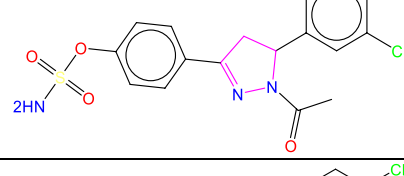
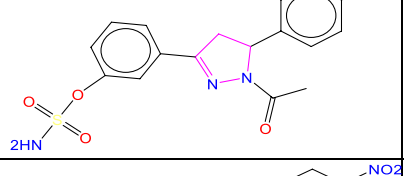
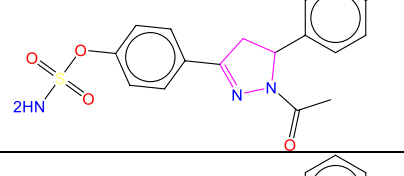
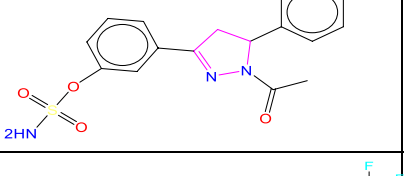
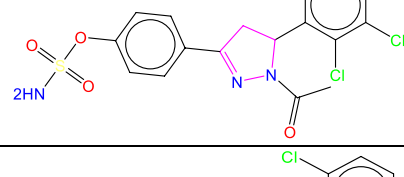
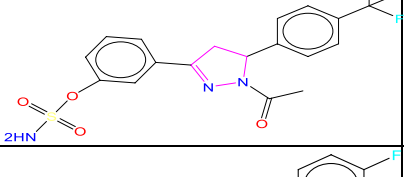
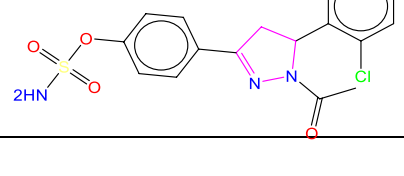
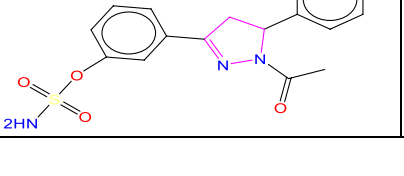
The dataset for the molecules was divided into two (28 for the training set and 6 molecules for the test set).

With acceptable levels of statistical parameters used to assess for internal and external validation of QSAR models, the equations models shown in Table 1 with the typical interpretation of the statistical symbols are statistically sound and predictive. High R^2 , R_{adj}^2 , Q_{cv}^2 , and R_{test}^2 values and low MSE values indicate that all of these models are statistically sound.

The cross-validation parameter Q_{cv}^2 has high values, indicating that the models are statistically robust. The R^2 test indicates a strong capacity of the models to predict future outcomes beyond the observed data (Fig. 1).

The values of R^2 and R_{adj}^2 are nearly close according to the results displayed in the Table 2, which means that the build models found do not contain many descriptors. That is

Table 1 pIC50 of compounds

Structure and pIC50	N ⁰	Structure and pIC50	N ⁰
	1 7.65		8 7.24
	2 7.37		9 7.258
	3 7.92		10 7.252
	4 8.00		11 7.31
	5 7.56		12 7.68
	6 7.58		13 8.05
	7 7.35		14 7.86

confirmed by low values of MSE. The external predictive capacity is high because the values of R^2_{test} were superior of 0.5. To test the robustness of the studied models, we calculated the value of Q^2_{cv} , and as we see from Table 2, the values of Q^2_{cv} were highly superior to 0.5.

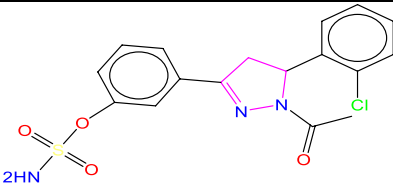
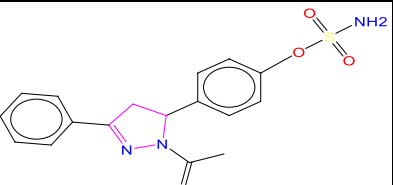
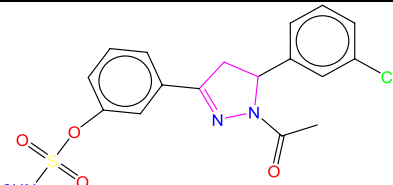
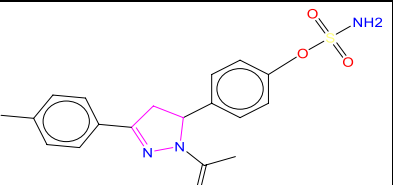
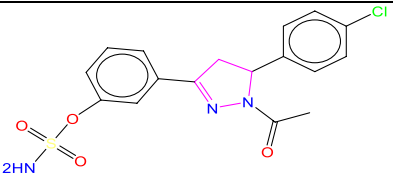
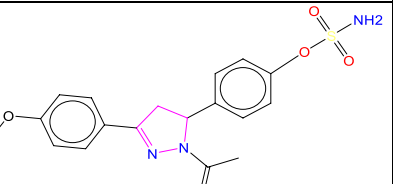
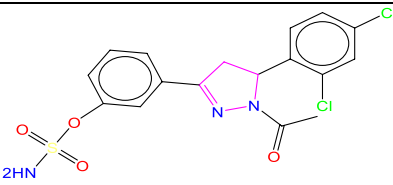
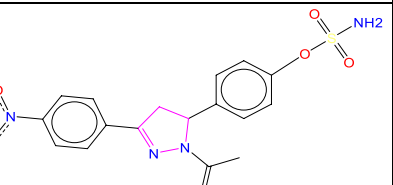
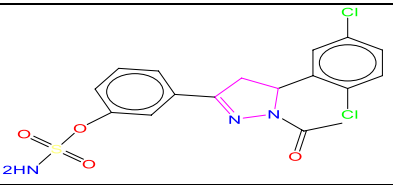
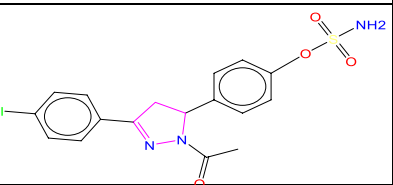
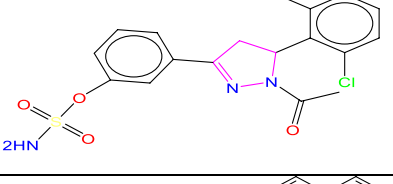
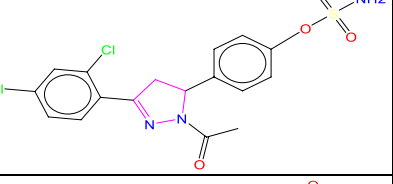
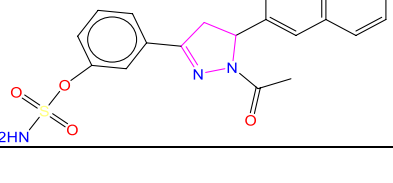
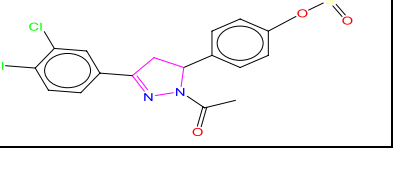
Applicability domain (AD)

The evaluation of the applicability domains of the four models shows that model 3 gives good values in the Williams diagram.

To avoid the found model not predicting the activity of another molecule, which does not belong to the dataset, we must call on the applicability domain. This allows us to determine the molecule that is out of the model QSAR.

Molecule 32 shown in Fig. 2 has a value of h_i higher than that of h^* means that this molecule is chemically different compared to the whole studied. All other molecules have h_i less than h^* even the molecules in the test set.

Table 1 (continued)

Structure and pIC50	N ⁰	Structure and pIC50	N ⁰
	15 7.12		22 7.97
	16 7.35		23 8.11
	17 8.03		24 8.02
	18 7.47		25 9.05
	19 7.46		26 7.53
	20 7.21		27 7.37
	21 6.33		28 7.88

Y-randomization test for model X

To further evaluate the constructed model, the calculations were repeated one hundred times with randomized activities for the same training set. For this, we have used the QSAR-Tools server available online at http://teqip.jdvu.ac.in/QSAR_Tools.

$Q_{cv}^2(Rand)$, R_{Rand}^2 , and cR_p^2 represent the average values; these values are respectively -1.608 , 0.185 , and 0.702 .

We find that all these values are less than the original model. We can say that this model was not found by chance.

The VIF values for the five descriptors (E_{HOMO-1} , Gap, Na, LogP, Ra) of the chosen model are respectively 2.057, 1.956, 2.485, 1.410, and 1.865. All of these values are less than 5 which means that this model is robust. The limit of Golbraikh and Tropsha is very important to check the reliability of the result found.

Table 1 (continued)

Structure and pIC50	N ^o	Structure and pIC50	N ^o
	29		32
	7.64		7.65
	30		33
	7.30		7.51
	31		34
	7.68		7.06

Fig. 1 Observed vs predicted activities by model 3 (train samples in blue color and test samples in orange color)

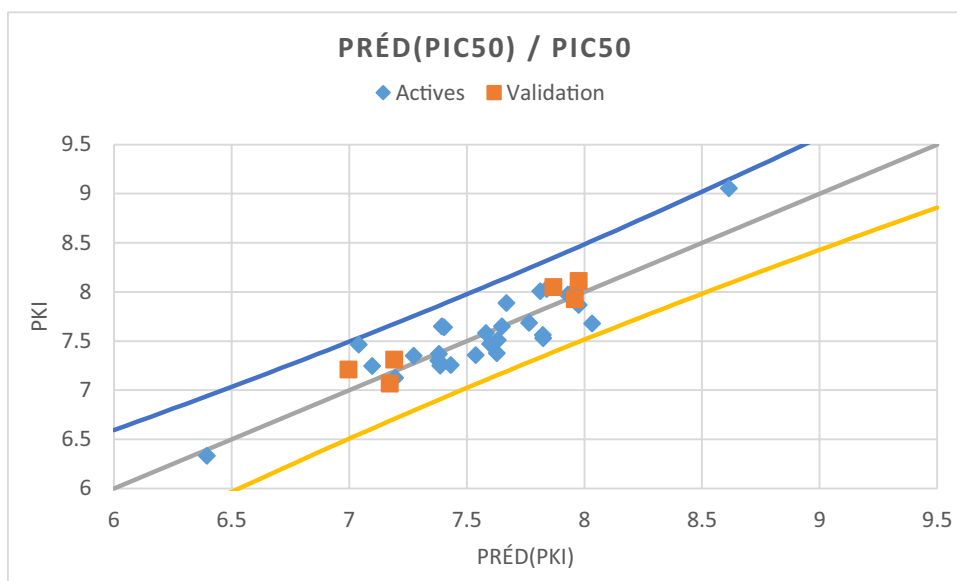


Table 2 Training and test set statistical parameters and model equations

Model equations	R^2	R^2_{adj}	MSE	R^2_{test}	Q^2_{cv}
1 $pKi = 14,720 - 0,7110MR + 1,867E-04PM2 + 6,088CMR$	0.72	0.68	0.04	0.67	0.64
2 $pKi = 15,260 - 0,680MR + 1,652E-04PM3 + 5,740CMR$	0.65	0.61	0.05	0.70	0.52
3 $pKi = 6,786 - 0,613E_{HOMO-1} - 0,101Gap - 0,132Na - 0,354LogP + 0,557Ra$	0.79	0.74	0.05	0.95	0.64
4 $pKi = 9,996 - 0,247Na + 8,496E-02MV - 2,997E-02Pc + 4,904E-02PSA + 0,364Ra$	0.81	0.77	0.05	0.75	0.53

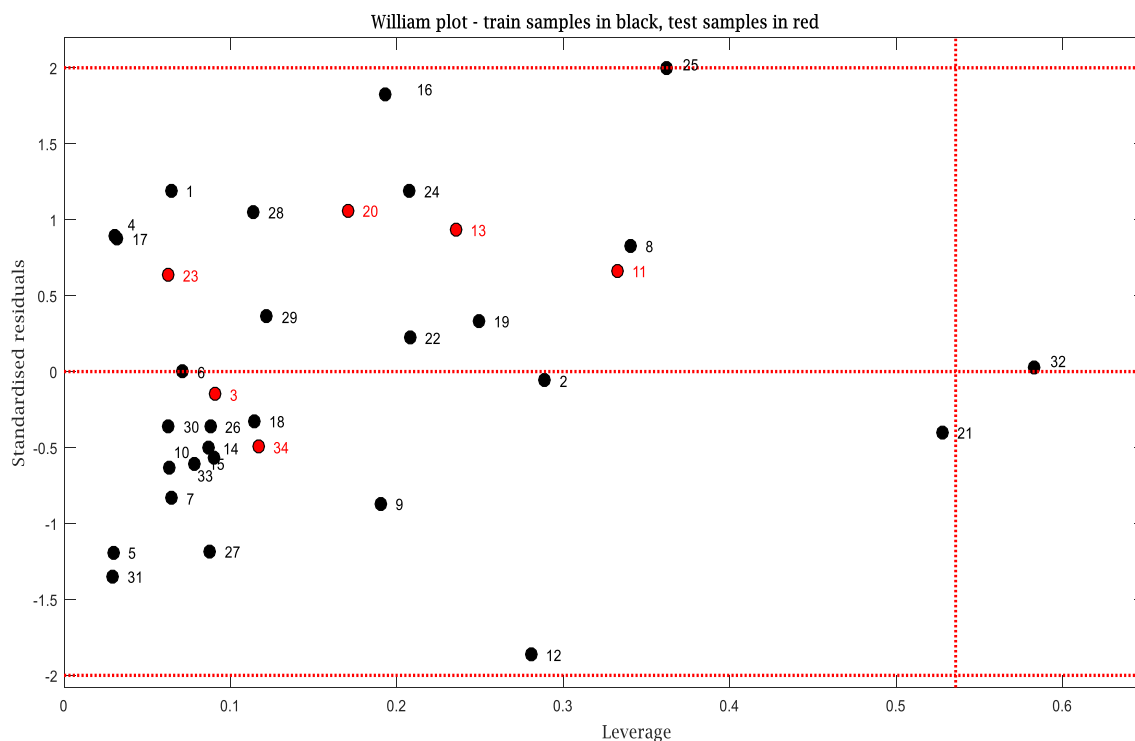


Fig. 2 Williams plot of standardized residual versus leverage for the MLR model (Train samples in black color and test samples in red color)

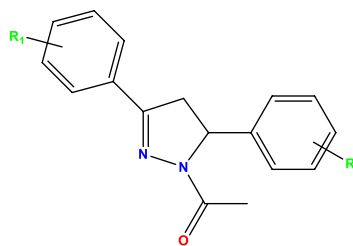
Table 3 Model 3 statistical metrics compared to Golbraikh and Tropsha's criteria

Parameter	Equation	Model score	Threshold
R^2	$R^2 = 1 - \frac{\sum (Y_{obs} - Y_{calc})^2}{\sum (Y_{obs} - \bar{Y}_{obs})^2}$	0.79	> 0.600
R^2_{adj}	$R^2_{adj} = \frac{(N-1)R^2 - p}{ N-p-1 }$	0.74	> 0.600
Q^2_{cv}	$Q^2_{cv} = 1 - \frac{\sum (Y_{calc} - Y_{obs})^2}{\sum (Y_{obs} - \bar{Y}_{obs})^2}$	0.64	> 0.500
MSE	$MSE = 1 - \frac{\sum (Y_{calc} - Y_{obs})}{N}$	0.05	A low value
R^2_{test}	$R^2_{test} = 1 - \frac{\sum (Y_{calc(test)} - Y_{obs(test)})^2}{\sum (Y_{obs(test)} - \bar{Y}_{obs(train)})^2}$	0.95	> 0.600
F_{test}	$F_{test} = \frac{\sum (Y_{calc} - \bar{Y}_{calc})^2}{\sum (Y_{obs} - \bar{Y}_{obs})^2} \times \frac{N-p-1}{p}$	16.856	A high value
R^2_{Rand}	Average of the 100 R^2_{Rand} (i)	0.185	< R ²
$Q^2_{cv LOO(Rand)}$	Average of the 100 $Q^2_{cv LOO}$ (Rand) (i)	-1.608	< Q^2_{cv}
cR^2_p	$cR^2_p = R \times \sqrt{(R^2 - (R^2 - (AverageR_{rand})^2))}$	0.702	> 0.500

Y_{obs} and Y_{calc} : refer to the observed and calculated/predicted response values.

\bar{Y}_{obs} and \bar{Y}_{calc} : refer to the mean of the observed and calculated/predicted response values.

N and p refer to the number of data points (compounds) and descriptors.

Table 4 Calculated anti-carbon anhydrase activity pIC50 and leverage effects (h) for the newly designed pyrazoline derivative

T	N	R ₁	R ₂	E _{HOMO-1}	Gap	Na	LogP	Ra	PIC50	h
1	1	4-OSO ₂ NH ₂	4-CH ₃	-6.60	4.31	45	2.61	7	7.65	0.07
	1a	4-OSO ₂ NH ₂	4-NH ₂	-8.75	7.47	44	0.98	7	9.17	0.79
	1b	4-OSO ₂ NH ₂	4-CCl ₃	-7.12	2.74	45	3.45	8	8.17	0.40
	1c	4-CCl ₃	4-CCl ₃	-7.17	3.93	42	6.18	7	6.95	0.02
	1d	4-NH ₂	4-OSO ₂ NH ₂	-6.46	2.16	44	0.98	7	8.27	0.70
	1e	4-CCl ₃	4-OSO ₂ NH ₂	-6.86	3.07	45	3.45	8	7.97	0.20
28	28	3,4Cl	4-OSO ₂ NH ₂	-6.8	4.29	42	3.41	7	7.88	0.19
	28a	3,4-NH ₂	3-OSO ₂ NH ₂	-5.86	1.66	46	0.17	7	7.98	0.40
	28b	3,4-CCl ₃	3-OSO ₂ NH ₂	-7.19	2.97	48	5.12	7	6.64	0.98
	28c	3,4-CCl ₃	3-CCl ₃	-0.26	0.14	45	7.85	7	6.52	0.6
	28d	3-OSO ₂ NH ₂	4-NH ₂	-5.89	2.09	44	0.98	7	7.93	0.7
	28e	3-OSO ₂ NH ₂	4-CCl ₃	-6.98	2.92	45	3.45	7	7.52	0.33
33	33	2,4Cl	3-OSO ₂ NH ₂	-6.80	4.31	42	3.41	7	7.51	0.10
	33a	4-NH ₂	3-OSO ₂ NH ₂	-5.90	2.02	44	0.98	7	7.94	0.72
	33b	4-CCl ₃	3-OSO ₂ NH ₂	-6.97	2.87	45	3.45	8	8.06	0.33
	33c	4-C(Cl) ₃	3-C(Cl) ₃	-7.07	4.25	42	6.18	7	6.86	0.99
	33d	3-OSO ₂ NH ₂	3,4-NH ₂	-5.90	1.99	46	0.17	7	7.97	0.95
	33e	3-OSO ₂ NH ₂	3,4-CCl ₃	-7.10	3.27	48	5.12	7	6.55	0.91

According to the result of Table 3, all of the parameters are better compared to the threshold.

Design of new compound

The model found is a very good model, so subsequently we can predict new molecules with good inhibitory activity by playing on the descriptors of the model. The importance of each descriptor in the built model is remarkable according to the absolute value of the *t*-test, which is an important value. These values for the five descriptors ($E_{\text{HOMO-1}}$, Gap, Na, LogP, Ra) are, respectively, -2.628 , -1.298 , -3.640 , -3.758 , and 5.735 . The influence of the descriptors on the activity differs from one descriptor to another according to the contribution coefficient of each descriptor shown in the model. According to the equation of the model, the positive sign means that the descriptor influences the activity proportionally, but the negative sign means the opposite. To increase the IC50 inhibitory activity against carbonic anhydrase, we must choose compounds with weak electronic effects to decrease the values of Gap and $E_{\text{HOMO-1}}$, as well as the number of atoms must be low. Nevertheless, the water solubility must be high for the LogP value to decrease. The suggested molecules must be large so that the radius Ra of the molecule is large. In the following paragraph, we will explain for each descriptor their influence on the activity of the compound by justifying the choice of the molecules proposed to have good inhibitory activity.

$E_{\text{HOMO-1}}$ The HOMO-1 energy has a relation with the ionization energy and designates the molecule's susceptibility to electrophilic attack. The variance of this descriptor is 24%; this value means that this descriptor has a remarkable influence on the IC50 activity. The HOMO-1 orbital can participate in the creation of a ligand–protein interaction bond. If this energy is high, it means that the HOMO-1 orbital loses the electron easily. We must therefore modify our compound to be nucleophilic by adding substituents that can decrease the nucleophilic character because this descriptor has a negative value in the equation of the model.

Gap It is the energy difference between the energy of the HOMO and LUMO orbital. In addition, this descriptor has an important contribution to IC50 considering its variance, which is 18.27%. A good inhibitory activity IC50 means a low gap value because a negative sign appears in the model equation for this descriptor; then the new compounds suggested for inhibition must contain groups that help to increase the gap.

Na This descriptor represents the number of atoms in the molecule; it has a very low variance value (0.18%), so the influence of this descriptor on the inhibitory activity of the molecule is very low.

LogP This descriptor has a negative sign in the model equation, which means that decreasing the value of this descriptor will increase the inhibitory activity IC50. The LogP represents the solubility of the molecule in water; by decreasing the LogP, the molecule becomes more soluble in water. The variance (8.93%) for this descriptor is modest, so the impact of this descriptor on the IC50 activity is small. We can then slightly modify the solubility of the molecule in water by adding substituents that facilitate the solubility of the compound in water.

Ra It is the radius of the molecule; a positive sign in the model means that this descriptor varies proportionally to the IC50 activity. We must therefore add bulky substituents to increase the radius of the molecule. The variance (32.88%) of this descriptor is very large and its influence on the IC50 activity is remarkable. These results show the decrease in molecule size and replacement of pyrazoline derivative with stronger electron-accepting groups (such as $-\text{NH}_2$ and CX_n). All the results obtained by the MLR model (3) are reliably indicating the performance of the model found, so we can subsequently design new compounds with better activity values compared to the studied compounds.

We made the relevant substitutions in light of the above results and estimated activities using the suggested model equation.

$$\begin{aligned} \text{pKi} = & 6,786 - 0,613E_{\text{HOMO-1}} \\ & - 0,101\text{Gap} - 0,132\text{Na} \\ & - 0,354\text{LogP} + 0,557\text{Ra} \end{aligned}$$

Therefore, the suggested approach will accelerate the process of synthesis of pyrazoline derivatives and the determination of their anti-carbonic anhydrase activity.

In the next part, we modified some structures (1, 28, and 33) with high pIC50 values and recalculated the theoretical pIC50 values by the chosen model as well as the value of *h*. The theoretical pIC50 values for the new proposed structures and their *h* are given in the following Table 4.

According to Table 4, we found very good results for the suggested structures because the majority of the theoretical pIC50 values for these structures are higher than those found experimentally as well as the *h* values are all lower than *h*.

Taking structure number 1, all the values found are higher than the pIC50 values found experimentally except molecule 1c; for example, molecule 1a has a very good pIC50 value (9.17). Moreover, according to molecule number 28, all the values found are close to or higher than the experimental values. For structure, 33 the results found are good; taking structure 33b, the value of pIC50 found for this structure is 8.06, so it is a very good value because it is higher than the experimental value.

Conclusion

To interpret the relationship between influenza virus inhibitor activity for 34 pyrazoline derivatives acting as anti-carbonic anhydrase and their structural descriptors obtained by the density functional theory calculation with Becke's three-parameter hybrid method and the Lee–Yang–Parr B3LYP functional employing 6-31G (d) basis set, the multi MLR approaches were used as a linear feature QSAR method. The model found in this work is very reliable because all the validation values are good, and from the descriptors of the model, we could suggest some molecules to be synthesized to be anti-carbonic anhydrase.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11356-023-28277-3>.

Author contribution This work is the result of collaboration among all authors. Imad Hmoudan: Conceptualization, investigation, writing—original draft. Mohammed Chafi: Formal analysis and validation.

Data availability The data sets used and/or analyzed during this study are available from the corresponding author upon reasonable request.

Declarations

Ethical approval This manuscript was prepared following ethical standards.

Consent to participate The authors have voluntarily agreed to participate in this research study.

Consent to publish The authors agree to publish the article in Environmental Science and Pollution Research.

Competing interests The authors declare no competing interests.

References

- Alafeefy AM, Abdel-Aziz HA, Carta F, Supuran CT, Pathak SK, Prasad O, Sinha L (2015) Exploring QSARs of some benzene-sulfonamides incorporating cyanoacrylamide moieties as a carbonic anhydrase inhibitors (specifically against tumor-associated isoforms IX and XII). *J Enzyme Inhib Med Chem* 30(4):519–523. <https://doi.org/10.3109/14756366.2014.948435>
- Eroğlu E (2019) DFT-based QSAR modelling of selectivity and inhibitory activity of coumarins and sulfocoumarins against tumor-associated carbonic anhydrase isoform IX. *Comput Biol Chem* 80:307–313. <https://doi.org/10.1016/j.compbiolchem.2019.04.011>
- Gaussian_09_ReferenceManual.pdf (2022) Retrieved May 19, 2022, from https://www.cwu.edu/chemistry/sites/cts.cwu.edu/chemistry/files/documents/Gaussian_09_ReferenceManual.pdf
- Gopinath P, Kathiravan MK (2022) Molecular field-based QSAR studies and docking analysis of mercaptoquinazolinone benzene sulfonamide derivatives against hCA XII. *Rasayan J Chem* 15(01):686–699. <https://doi.org/10.31788/RJC.2022.1516767>
- Hilvo M, Baranauskiene L, Salzano AM, Scaloni A, Matulis D, Innocenti A, Scozzafava A, Monti SM, Fiore AD, Simone GD, Lindfors M, Jänis J, Valjakka J, Pastoreková S, Pastorek J, Kulomaa MS, Nordlund HR, Supuran CT, Parkkila S (2008) Biochemical characterization of CA IX, one of the most active carbonic anhydrase isozymes *. *J Biol Chem* 283(41):27799–27809. <https://doi.org/10.1074/jbc.M800938200>
- Jaakkola P, Mole DR, Tian Y-M, Wilson MI, Gielbert J, Gaskell SJ, von Kriegsheim A, Hebestreit HF, Mukherji M, Schofield CJ, Maxwell PH, Pugh CW, Ratcliffe PJ (2001) Targeting of HIF- α to the von Hippel-Lindau ubiquitylation complex by O₂-regulated prolyl hydroxylation. *Science* 292(5516):468–472. <https://doi.org/10.1126/science.1059796>
- Josse J, Husson F, Pagès J (2009) Gestion des données manquantes en Analyse en Composantes Principales. *J De La Société Française De Statistique* 150(2):28–51. http://www.numdam.org/item/JSFS_2009__150_2_28_0/. Accessed 19 May 2022
- Kivelä A, Parkkila S, Saarnio J, Karttunen TJ, Kivelä J, Parkkila A-K, Waheed A, Sly WS, Grubb JH, Shah G, Türeci Ö, Rajaniemi H (2000) Expression of a novel transmembrane carbonic anhydrase isozyme XII in normal human gut and colorectal tumors. *Am J Pathol* 156(2):577–584. [https://doi.org/10.1016/S0002-9440\(10\)64762-1](https://doi.org/10.1016/S0002-9440(10)64762-1)
- Kumar V, Roy K (2020) Development of a simple, interpretable and easily transferable QSAR model for quick screening antiviral databases in search of novel 3C-like protease (3CLpro) enzyme inhibitors against SARS-CoV diseases. *SAR QSAR Environ Res* 31(7):511–526. <https://doi.org/10.1080/1062936X.2020.1776388>
- Lee M, Vecchio-Pagán B, Sharma N, Waheed A, Li X, Rarigh KS, Robbins S, Han ST, Franca AL, Pellicore MJ, Evans TA, Arcara KM, Nguyen H, Luan S, Belchis D, Hertecant J, Zabner J, Sly WS, Cutting GR (2016) Loss of carbonic anhydrase XII function in individuals with elevated sweat chloride concentration and pulmonary airway disease. *Hum Mol Genet* 25(10):1923–1933. <https://doi.org/10.1093/hmg/ddw065>
- Li Z, Wan H, Shi Y, Ouyang P (2004) Personal experience with four kinds of chemical structure drawing software: review on ChemDraw, ChemWindow, ISIS/Draw, and ChemSketch. *J Chem Inf Comput Sci* 44(5):1886–1890. <https://doi.org/10.1021/ci049794h>
- Matysiak J, Skrzypek A, Tarasiuk P, Mojzycz M (2017) QSAR study of pyrazolo[4,3-e][1,2,4]triazine sulfonamides against tumor-associated human carbonic anhydrase isoforms IX and XII. *Comput Biol Chem* 71:57–62. <https://doi.org/10.1016/j.compbiolchem.2017.09.006>
- Mendelsohn LD (2004) ChemDraw 8 Ultra, Windows and Macintosh Versions. *J Chem Inf Comput Sci* 44(6):2225–2226. <https://doi.org/10.1021/ci040123t>
- Mishra CB, Tiwari M, Supuran CT (2020) Progress in the development of human carbonic anhydrase inhibitors and their pharmacological applications: where are we today? *Med Res Rev* 40(6):2485–2565. <https://doi.org/10.1002/med.21713>
- Moi D, Nocentini A, Deplano A, Balboni G, Supuran CT, Onnis V (2019) Structure-activity relationship with pyrazoline-based aromatic sulfamates as carbonic anhydrase isoforms I, II, IX and XII inhibitors: synthesis and biological evaluation. *Eur J Med Chem* 182:111638. <https://doi.org/10.1016/j.ejmech.2019.111638>
- Rivero P, García-Suárez VM, Pereñíguez D, Utt K, Yang Y, Bellaïche L, Park K, Ferrer J, Barraza-Lopez S (2015) Systematic pseudopotentials from reference eigenvalue sets for DFT calculations. *Comput Mater Sci* 98:372–389. <https://doi.org/10.1016/j.commatsci.2014.11.026>
- Vidal NP, Manful CF, Pham TH, Stewart P, Keough D, Thomas RH (2020) The use of XLSTAT in conducting principal component analysis (PCA) when evaluating the relationships between sensory and quality attributes in grilled foods. *MethodsX* 7:100835. <https://doi.org/10.1016/j.mex.2020.100835>

- Wang Z-C, Qin Y-J, Wang P-F, Yang Y-A, Wen Q, Zhang X, Qiu H-Y, Duan Y-T, Wang Y-T, Sang Y-L, Zhu H-L (2013) Sulfonamides containing coumarin moieties selectively and potently inhibit carbonic anhydrases II and IX: design, synthesis, inhibitory activity and 3D-QSAR analysis. *Eur J Med Chem* 66:1–11. <https://doi.org/10.1016/j.ejmech.2013.04.035>
- Yorulmaz N, Eroğlu E (2020) DFT based QSARs for inhibitory activity of coumarins towards tumor-associated isoform (CA XII) of carbonic anhydrases. *J Mol Struct* 1208:127844. <https://doi.org/10.1016/j.molstruc.2020.127844>
- Zhang IY, Xu X (2021) Exploring the limits of the XYG3-type doubly hybrid approximations for the main-group chemistry: the xDH@B3LYP model. *J Phys Chem Lett* 12(10):2638–2644. <https://doi.org/10.1021/acs.jpcllett.1c00360>
- Zhou Z, Tang X, Dai W, Shi J, Chen H (2017) Nano-QSAR models for predicting cytotoxicity of metal oxide nanoparticles (MONPs) to *E. coli*. *Can J Chem* 95(8):863–866. <https://doi.org/10.1139/cjc-2017-0172>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.