# QSAR Studies Regarding the Inhibition of the Carbonic Anhydrase by the Sulfonamides Containing a Picolinoyl Group

**TARKO LASZLO***

Romanian Academy, Center of Organic Chemistry, 202B Spl. Independenţei, 060023, Bucharest, Romania

*This paper presents two QSAR studies in which the dependent property is the inhibitor activity over carbonic anhydrase (hCAI isozyme) of 21 sulfonamides containing picolinoyl group. All statistical calculations have been done using PRECLAV program. In the first QSAR study, only the specific predictors of PRECLAV have been used. The correlation between the observed values and the calculated values of activity is good (s = 0.3431 , $r^2$ = 0.9135, F = 63.4 , $r^2_{CV}$ =0 .8588 ). The second QSAR study has used both specific descriptors of PRECLAV and descriptors calculated using the DRAGON program. The correlation between the observed values and the calculated values of activity is very good (s = 0.2199 , $r^2$ = 0.9645, F = 162.8 , $r^2_{CV}$ = 0 .9393). The QSAR equations obtained point to the fact that the inhibitor activity decreases, if a high percentage of hydrogen in the molecule is present and is favored by the presence of alkyl groups, a high internal topological diversity, and condensed aromatic rings.*

*Keywords: QSAR, sulfonamides, picolinoyl, carbonic anhydrase, PRECLAV, DRAGON*

Interconversion of carbon dioxide and water into carbonic acid is catalyzed by a family of metalloenzymes (isozymes) named "carbonic anhydrase". The active site of most isozymes contains a zinc ion. In the best studied isozyme (α-carbonic anhydrase), present in animals, the zinc ion is coordinated by some histidine residues. This isozyme interconvert carbon dioxide and bicarbonate, maintain acid-base balance in blood and help transport carbon dioxide out of tissues. A different isozyme, evolutionarily distinct, called β-carbonic anhydrase, is present in plants. This isozyme also uses a zinc ion in its active site and participates in the same reaction [19].

The sulfonamides are known for about 70 years as bactericides. In the last years, there have been many studies into their inhibitory effect over carbonic anhydrase isozymes. The inhibition of this enzymes allows the use of some sulfonamides as diuretics, in the treatment of some nervous ailments, in the treatment of glaucoma, etc.

Starting with 1999, a part of the research was directed towards a new class of sulfonamides, with a benefic effect in the treatment of intraocular pressure, molecules characterized by the presence of the picolinoyl fragment [1-5]. The presence of this fragment increases the water solubility and the inhibitory capacity of the enzyme. The QSAR methodology (Quantitative Structure Activity Relationship) has been applied [6, 18] in order to identify other molecular characteristics with a high influence over the inhibitory capacity.

*Methods and formulas*

In this paper we present the results of two QSAR studies performed without an external validation set, and where the calibration set contains the molecules from figure (R = 2 - picolinoyl), synthesized by Supuran and co-workers [1]. The two QSAR studies use two different sets of descriptors. The dependent property was the inhibition constant IC = log($K_i$), where $K_i$ is the concentration of inhibitor where the inhibitory effect on hCAI isozyme starts. The values for $K_i$ (nM), used in the computation, have been taken from literature [1]. Greater the value of IC, smaller the value of inhibitory activity.

The virtual construction of the molecules and the geometry optimization have been done using the molecular mechanics software PCModel [11]. The next step was the rigorous optimization of the geometry using the Mopac software for quantum mechanics calculations [12] using the key-words string "am1 pulay gnorm=0.01 shift=50 geo-ok camp-king mmok bonds vectors". The output file from Mopac was used as input file for PRECLAV program [7 - 9]. Using these data, PRECLAV calculated for each molecule the values for almost 400 whole molecule descriptors, specific for the program in question. No "grid" descriptors were used. Separately, for each molecule, the values for another 1400 descriptors, specific to this program have been calculated, using the DRAGON program [10]. The statistic calculations, including the QSAR equations, have been performed using the PRECLAV program.

The program has used only "significant" descriptors in calculating the QSAR equations, descriptors that fulfill criteria (1).

$$r^2 > 4 / N \quad (1)$$

where:

$r^2$ is the square of the Pearson linear correlation between the values of the analyzed descriptor and the values of the dependent property

N is the number of molecules in the calibration set (here N = 21)

The program combines successively sets with 2, 3, …, *k* significant descriptors (1 < k < 11). A set of descriptors contains only descriptors that are sufficiently low intercorrelated and fulfill criteria (2).

$$r_{ij}^2 < N^{-1/2} \quad (2)$$

where:

$r_{ij}^2$ is the square of Pearson linear correlation between the values of two descriptors present in the same set

Each set of descriptors has been used to calculate a multilinear QSAR equation of type (3).

$$IC = c_0 + \sum_{i=2}^{k} c_i \cdot D_i \quad (3)$$

where:

IC is the dependent property (inhibition constant)

**email: ltarko@cco.ro*

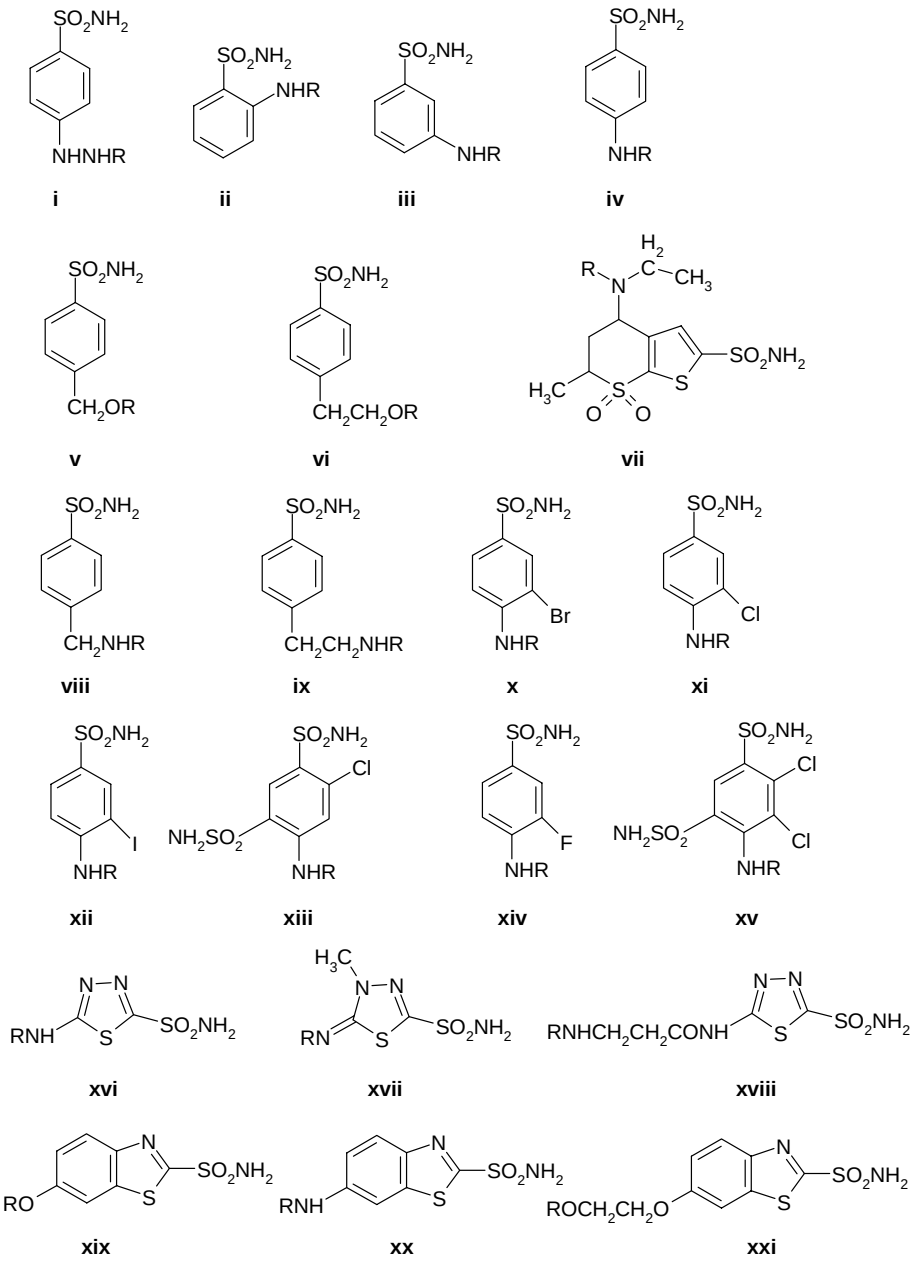


Fig. 1. The chemical structure of analysed sulfonamides

$c_0$ is the free term (intercept)

$c_i$ are the coefficients (weighting factors) of the descriptors

$D_i$ are some significant descriptors

*k* is the number of descriptors in the set

The $c_i$ coefficients have been calculated with OLSM (Ordinary Least Square Method). Tens of thousands of equations of type (3) have been calculated. With each of the type (3) equation, the values of IC have also been calculated. These computed values have been compared with the observed values. The concordance between the calculated/observed values has been calculated using the quality function Q.

$$Q = K_{CV} \cdot (N - k) / N \quad (4)$$

where:

$K_{CV}$ is Kendall cross-validated rank correlation between computed/observed values.

As the value of *k* increases, the quality Q of the equations increases, reaches a maximum, then decreases. For the prediction, the equation with the highest value of Q is used and the descriptors used in this equation are called "predictors".

For every predictor the "utility" U has been calculated.

$$U = (R^2 - r^2) / (1 - r^2) \quad (5)$$

where:

$R^2$ is the square of the Pearson linear correlation between the observed values of IC and the values calculated with the equation containing *k* predictors

$r^2$ is the square of the Pearson linear correlation between the observed values of IC and the values calculated with the equation containing *k-1* predictors, that is without the analyzed descriptor.

The values of U are then normalized (the highest becoming 1000). The "useful" predictors ( U > 400) are sufficiently well correlated with IC and not too well correlated with the other predictors. Every "useful" predictor describes (reasonably) well the variation in the values of IC, and, at the same time, describes a different aspect than the other predictors.

**Table 1**
OBSERVED / COMPUTED VALUES OF INHIBITION CONSTANT

| Sulfonamide | IC observed | IC computed | |
|---|---|---|---|
| | | by QSAR #1 | by QSAR #2 |
| i | 4.35 | 4.313 | 4.189 |
| ii | 4.33 | 4.643 | 4.006 |
| iii | 4.29 | 3.758 | 4.132 |
| iv | 4.17 | 3.588 | 4.060 |
| v | 3.32 | 3.447 | 3.534 |
| vi | 3.31 | 3.420 | 3.220 |
| vii | 3.20 | 3.187 | 3.009 |
| viii | 3.04 | 2.985 | 3.412 |
| ix | 3.04 | 2.960 | 3.258 |
| x | 2.79 | 2.740 | 2.841 |
| xi | 2.78 | 3.047 | 2.814 |
| xii | 2.78 | 2.633 | 2.891 |
| xiii | 2.78 | 2.920 | 3.051 |
| xiv | 2.72 | 2.916 | 2.796 |
| xv | 2.17 | 2.365 | 2.494 |
| xvi | 1.60 | 1.776 | 1.786 |
| xvii | 1.49 | 1.326 | 1.106 |
| xviii | 1.30 | 1.771 | 1.402 |
| xix | 1.04 | 1.569 | 0.995 |
| xx | 1.00 | 1.324 | 1.234 |
| xxi | 1.00 | 0.350 | 0.811 |

## Results and discussions

We present below the results of applying the PRECLAV algorithm in the two QSAR studies. The quality of the equations is given by the value of function Q, but also by the value of some usual statistical functions.

QSAR study #1

*Descriptors used:* specific descriptors of the PRECLAV program

*Number of significant descriptors:* 130

*QSAR of type (3) used for prediction:*

$c_0$ = -12.0954

$c_1$ = 9.2180

$D_1$ is Percent of hydrogen · Minimum net charge of H atoms product (U = 1000)

$c_2$ = 1.0845

$D_2$ is $E_{LUMO+1} - E_{HOMO}$ gap (U = 759)

$c_3$ = 4.6031

$D_3$ is Spherical shape index [13] (U = 756)

*Standard error s:* 0.3431

*Pearson square correlation* $r^2$: 0.9135

*Fisher function F:* 63.4

*Pearson cross-validated square correlation* $r^2_{CV}$: 0.8558

*Kendall rank correlation K:* 0.9143

*Kendall cross-validated rank correlation* $K_{CV}$: 0.8667

*Quality Q:* 0.7429

*Number of outliers:* 0

The lowest correlation with IC was computed for predictor $D_3$ ($r^2$ = 0.2496). The highest intercorrelation between predictors was computed for the pair $D_1$, $D_2$ ($r^2$ = 0.0195).

The predictor with the highest usability for describing the values of IC is in this case $D_1$. The value of IC is directly proportional with the value of $D_1$. The study of the files provided by Mopac shows the fact that a low value for the "Minimum net charge of H atoms" is determined by the presence of the alkyl groups in the molecule. On the other hand, in the molecules presented above, high percentage of hydrogen is exhibited by the ones with low molecular mass, a low degree of substitution of the aromatic rings and/or without condensed cycles. The fact that in QSAR #1 the product $D_1$ is a predictor can be interpreted as "the favorable effect on inhibitory activity of the presence of alkyl groups is diminished by a high percentage of hydrogen" or that "the favorable effect of the presence of alkyl groups is accentuated by a low percentage of hydrogen".

QSAR study #2

*Descriptors used:* descriptors of PRECLAV program and descriptors calculated with DRAGON program

*Number of significant descriptors:* 404
*QSAR of type (3) used for prediction:*

$c_0$ = 18.7491
$c_1$ = -17.9328
$D_1$ is Structural Information Content (neighborhood symmetry of 1- order) [14]
(U = 1000)
$c_2$ = -1.9352
$D_2$ is R-CR-X atom-centred fragment [15] (U = 965)
$c_3$ = -1.1106
$D_3$ is R autocorrelation of lag5 weighted by atomic Sanderson electronegativities
[16, 17] (U = 481)

*Standard error s:* 0.2199
*Pearson square correlation* $r^2$*:* 0.9645
*Fisher function F:* 162.8
*Pearson cross-validated square correlation* $r^2_{CV}$*:* 0.9393
*Kendall rank correlation K:* 0.8857
*Kendall cross-validated rank correlation* $K_{CV}$*:* 0.8762
*Quality Q:* 0.7510
*Number of outliers:* 0

We notice the fact that in QSAR #2 no PRECLAV descriptors are present.

Here, the lowest correlation with IC was computed for predictor $D_3$ ($r^2$ = 0.2014). The highest intercorrelation between predictors was computed for the pair $D_1$, $D_3$ ($r^2$ = 0.2480).

The most useful predictors for describing IC are in this case $D_1$ and $D_2$. The value of IC is inversely proportional with the values of $D_1$ and $D_2$. Considering the physical significance of predictor $D_1$ [14] we conclude that a high topological diversity of the molecular graph, from the point of view of the edge multiplicity and the vertex degree is favorable to the inhibitory activity. The presence of the $D_2$ predictor indicates the existence of condensed aromatic rings [15]. We can conclude that this structural feature is favorable to the inhibitory activity.

The table presents the observed and calculated values (rounded to three decimal points) of the inhibition constant IC.

## Conclusions

The QSAR studies of the inhibitory activity over carbonic anhydrase (hCAI isozyme) using 21 sulfonamides with picolinyl group lead to the following conclusions:

- the DRAGON descriptors seem to be more useful than the PRECLAV descriptors
- the inhibitory activity is disfavored by a high percentage of hydrogen in the molecule
- the inhibitory activity is favored by the presence of the alkyl groups, a high internal topological diversity and the presence of condensed aromatic rings.

## References


1.SUPURAN C.T., SCOZZAFAVA A., MENABUONI L., MINCIONE F., BRIGANTI F., MINCIONE G., Eur. J. Pharm. Sci., **8**, 1999, p. 317
2. CASINI A., SCOZZAFAVA A., MINCIONE F., MENABUONI L. SUPURAN C.T., J. Enzyme Inh. Med. Chem., **17,** 2002, p. 333
3.VULLO D., FRANCHI M., GALLORI E., ANTEL J., SCOZZAFAVA A., SUPURAN C.T., J. Med. Chem., **47,** 2004, p. 1272
4.ABBATE F., CASINI A., OWA T., SCOZZAFAVA A., SUPURAN C. T., Bioorg. Med. Chem. Lett., **14,** 2004, p. 217
5.NISHIMORI I., VULLO D., INNOCENTI A., SCOZZAFAVA A., MASTROLORENZO A., SUPURAN C.T., Bioorg. Med. Chem. Lett., **15,** 2005, p. 3828
6. SINGH S., SINGH J., INGLE M., MISHRA R., KHADIKAR P.V., Arkivoc, 2006, xvi, p. 1
7. TARKO, L. Rev. Chim. (Bucure°ti), **56**, 2005, p. 639
8. TARKO, L., LUPESCU, I., GROPOSILA-CONSTANTINESCU, D. Arkivoc, 2005**, *x***, p. 254
9. PRECLAV v.0609 program is available from Center of Organic Chemistry (CCO) - Bucharest, Romanian Academy; Director CCO E-mail address: pfilip@cco.ro
10. DRAGON v. 5.4 program is available from TODESCHINI, R., Talete srl., via V. Pisani, 13-20124, Milano, Italy
11. GAJEWSKI J.J., GILBERT K.E., PCModel v. 9.0 program is available from Serena Sofware, Box 3076, Bloomington, IN, USA
12. STEWART J.J.P., Mopac93, Fujitsu Ltd., Tokyo, Japan, 1993
13. TARKO L., CALAFETEANU S., Rev.Chim. (Bucure°ti), **49,** 1998, p. 169
14. MAGNUSON V.R., HARISS D.K., BASAK S.C., Studies in Physical and Theoretical Chemistry, Elsevier, Amsterdam, 1983, p. 178
15.VISWANADHAN V.N., GHOSE A.K., REVANKAR G.R., ROBINS R.K., J.Chem.Inf.Comput.Sci., **29,** 1989, p. 163
16. CONSONNI V., TODESCHINI R., PAVAN M., J. Chem. Inf. Comput. Sci., **42,** 2002, p. 682
17. CONSONNI V., TODESCHINI R., PAVAN M., GRAMATICA P., J. Chem. Inf. Comput. Sci., **42,** 2002, p. 693
18. WEBER A., BÖHM M., SUPURAN C. T., SCOZZAFAVA A., SOTRIFFER A., KLEBE G., J. Chem. Inf. Comput. Sci., **46,** 2006, p. 2737
19. Wikipedia encyclopedia http://en.wikipedia.org/wiki/Carbonic_anhydrase



Intrat în redac¡ie: 3.01.2007