

PAPER • OPEN ACCESS

QSAR Study on the Relative Activity of Thiazide diuretics in Doping

To cite this article: Li-Lan Zhu 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **218** 012137

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

QSAR Study on the Relative Activity of Thiazide diuretics in Doping

Li-Lan ZHU*

Guangdong Industry Polytechnic, Guangzhou 510300, China

*Corresponding author's e-mail: zhulilan_8983@sina.com

Abstract. Electrotopological state indices (E_K) of atom types is used to describe the structures of 10 thiazide diuretics in doping and related to the relative activity(R_a). The optimal three-parameter(E_{16} , E_{23} , E_9) quantitative structure-activity relationships (QSAR) model is established by using leaps-and-bounds regression analysis for the relative activity(R_a) of thiazide diuretics along with the E_K . The correlation coefficients (R^2) and the leave-one-out (LOO) cross validation R_{cv}^2 for the R_a was 0.975 and 0.936, respectively. The model has favorable correlation, as well as robustness and good prediction capability by F , R_{cv}^2 , A_{IC} , F_{IT} , V_{IF} tests. The results indicate that the molecular structural units: =O(in sulfuryl), =N- and =C<(in aromatic ring) are main factors which can affect the relative activity(R_a) bioactivities of these compounds directly.

1. Introduction

Doping(the stimulant has been used in horse racing since 16th century is called doping.) is the general name of all the athletic banned substance in international sports, including all the banned drugs and methods in sports prescribed by international sports organizations. There are five categories of dopings: irritant, anesthetic, anabolic agents, diuretic and peptide and glycoprotein hormone, according to Prohibited Classes of Substances and Prohibited Methods by International Olympics Committee. Thousands of compounds have been involved, including diuretics, a type of stimulant banned since the 1988 Seoul Olympics^[1]. Diuretics can increase the amount of urine excretion, and promote the accumulation of excess water in the subcutaneous and abdominal cavity in a short period of time. As a result, sports competition is unfair, fair and banned. At present, the diuretics in the list of prohibited substances of IOC are divided into four categories: loop diuretics, thiazide diuretics, potassium diuretics, carbonic anhydrase inhibitors^[2].

Quantitative structure-activity relationship (QSAR)^[3-6] can establish the quantitative relationship between the molecular structure and activity of organic compounds, which shows practical significance for predicting the biological activity of new compounds, initial screening and evaluation of activity compounds, and reducing high cost of the activity evaluation. Topological indexes could directly derive from chemical molecular structure independent of the experimental mensuration. Therefore, the appearance of topological indexes in general QSAR models was popular for the development of reliable QSAR models. The electrotopological state (E-state) indexes(E_K) put forward by Kier and Hall^[7-9] are one of the most broadly applied topological indexes.

Based on the electrotopological state (E-state) indexes^[7-9], the robust QSAR model of relative activity(R_a)^[10]for 10 thiazide diuretics in doping was established by leaps-and-bounds regression, to estimate and predict R_a of thiazide diuretics and reveal the microstructure that affected its relative activity at the molecular level. It provides a theoretical reference for the effective detection of thiazide diuretics and plays an important role in anti-doping.



2. Material and methods

2.1 Studied compounds and their biological activity data

The compounds studied in this paper are 10 thiazide diuretics^[10]. The parent structure is shown in Fig.1, where R₁, R₂, R₃, X, Y are the substituent groups. Their substituents and their relative activity data(Ra)^[10] are shown in Table 1.

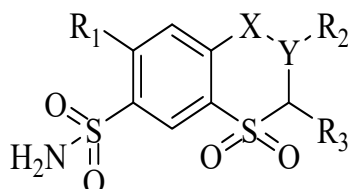


Figure1. Basic structure of thiazide diuretics

Table 1. The molecule structures and Ra of thiazide diuretics

No.	R ₁	R ₂	R ₃	X	Y	Ra ^[10]
1	Cl	H	H	-N=	>C=	0.80
2	Cl	SCH ₂ C ₆ H ₅	H	-N=	>C=	1.30
3	Cl	H	H	-NH-	>CH-	1.40
4	Cl	CHCl ₂	H	-NH-	>CH-	1.70
5	Cl	CHCl ₂	CH ₃	-NH-	>CH-	1.80
6	Cl	CH ₂ SCH ₂ CF ₃	CH ₃	-NH-	>CH-	2.00
7	CF ₃	CH ₂ Cl	H	-NH-	>CH-	1.30
8	CF ₃	CH ₂ C ₆ H ₅	H	-NH-	>CH-	1.80
9	Cl	CH ₂ Cl	CH ₃	-NH-	>CH-	1.80
10	CF ₃	H	H	-NH-	>CH-	1.30

2.2 Molecular descriptors

In this paper, electrotopological state indices (E_K)^[7-9] is used to characterize the molecular structure of different classes of organic compounds. Hall *et al.* have investigated the localization of several famous topologies, and provided E_K which can show the topological, geometrical and electrical characters comprehensively. The E_K of all atoms in the compound molecular skeleton was calculated following the method given by Hall *et al.*, and the details can be found in references^[7-9].

2.3 Statistical regression analysis

For QSAR derivation, Ra values^[10] act as the dependent variables, and electrotopological state indices descriptors act as the independent variables. The regression analyses are carried out by using multiple linear regression(MLR), leaps-and-bounds regression (LBR) program. By using Fischer statistics (F-tests) to eliminate insignificant descriptor when entering a new descriptor, stepwise regression can identify the most important descriptors contributing to the relative activity (Ra) of thiazide diuretics.

The correlation between variables in model was estimated by the variance inflation factor (V_{IF})^[11] value. V_{IF} is defined as follows:

$$V_{IF} = 1/(1 - \beta^2) \quad (1)$$

in which β is the correlation coefficient of multiple regressions between one variable and the others in the equation. $V_{IF} = 1.0$ suggests no self-correlation among each variable; if V_{IF} ranges from 1.0 to 5.0, indicating that there is no obvious autocorrelation between variables, the model is stable; when V_{IF} is larger than 5.0, the regression equation is unstable and recheck of variables correlation coefficient is necessary.

The quality of each regression model was evaluated, using a squared correlation coefficient(R^2), cross validation squared correlation coefficient(R_{cv}^2). The “leave-one-out”(LOO) cross-validation

coefficient R_{cv}^2 was considered as an indicator of the predictive performance and stability of a QSAR model. As a rule of thumb, the equations with regression coefficients $R^2 > 0.80$ ^[12] and $R_{cv}^2 > 0.50$ ^[13] are considered reasonable. Where parameter $R_{cv}^2 > 0.50$ is used as a criterion of both robustness and predictive ability of the model, whereas the conventional correlation coefficient R^2 defines the goodness-of-fit.

Another indicator used to evaluate the quality of model is the standard deviation (S_D). The model is good and the prediction accuracy is acceptable when the ratio of the standard deviation to the value range (the samples between the maximum and minimum values) is less than 10%^[14].

Akaike's information criterion (A_{IC} ; Eq.2; the model that produces the minimum AIC value is considered potentially the most useful) and Kubinyi function (F_{IT} ; Eq.3; the best model will present the highest value of this function)^[15,16] are applied to determine if a variable should be included in the model. That is to say, if the Akaike's information criterion decreases in value when adding an additional variable and the Kubinyi function increases in value, then, the introduction of this new variable is justified.

$$A_{IC} = RSS \times \frac{f+b}{(f-b)^2} \quad (2)$$

$$F_{IT} = \frac{R^2(f-b-1)}{(f+b^2)(1-R^2)} \quad (3)$$

Where RSS is the residual sum of squares, f is the number of compounds included in the model, b is the number of variables included in the model, R^2 is the square of the correlation coefficient.

3. Results and discussion

3.1 QSAR equation of the R_a

The electrotopological state indices and relative activity (R_a) of thiazide diuretics were input into MINITAB14.0 statistical analysis software, and the leaps-and-bounds regression was used to select the best variable combinations, to establish the best QSAR models, the results are shown in Table 2.

Table 2. Results of electrotopological state indices descriptors and inhibitory activities R_a with leaps-and-bounds regression

No.	R	R_{cv}^2	A_{IC}	F_{IT}	S_D	F	Variables
1	0.558	0.264	0.118	0.918	0.255	10.119	E_{16}
2	0.914	0.872	0.028	5.314	0.120	37.423	E_{16}, E_{23}
3	0.975	0.936	0.020	12.316	0.070	77.254	E_{16}, E_{23}, E_9
4	0.979	0.847	0.070	8.965	0.070	59.598	E_{16}, E_{23}, E_9, E_2

Table 2 shows that with the increase in the number of independent variables in the model, in addition to R^2 , S_D , the rest of the statistical indicators in the ternary equation has a turning point, which R_{cv}^2 , F_{IT} , F has the maximum, A_{IC} has a minimum. So choose the best three-variable QSAR model:

$$Ra = -0.929(\pm 1.154) + 0.070(\pm 0.018)E_9 + 0.230(\pm 0.024)E_{16} - 0.157(\pm 0.016)E_{23} \quad (4)$$

$$f = 10, R^2 = 0.975, R_{cv}^2 = 0.936, R_{adj}^2 = 0.962, F = 77.254, S_D = 0.070$$

The QSAR equation (4) can be used to predict R_a , and the predicted value ($R_{a, cal.}$) is basically consistent with the experimental data in Table 3.

Table 3. The correlativity between R_a and E_k of thiazide diuretics

No	E_2	E_9	E_{16}	E_{23}	$R_{a, exp.}^7$	$R_{a, cal.}$	$R_{a, err.}$
1	0.0000	-0.8650	45.545	3.7373	0.80	0.79	0.01
2	0.5144	0.1280	47.827	4.2093	1.30	1.31	-0.01
3	0.0143	-0.5125	45.684	0.0000	1.40	1.43	-0.03
4	0.0000	-0.9296	46.740	0.0000	1.70	1.65	0.05
5	0.0000	-0.8591	47.652	0.0000	1.80	1.86	-0.06

6	-1.3053	-1.2354	48.054	0.0000	2.00	1.93	0.07
7	-0.2530	-3.9609	46.589	0.0000	1.30	1.40	-0.10
8	0.1645	-3.0636	47.981	0.0000	1.80	1.78	0.02
9	0.0176	-0.6305	47.417	0.0000	1.80	1.82	-0.02
10	-0.3042	-3.7725	45.769	0.0000	1.30	1.23	0.07

3.2 Validation of the QSAR equation

The predictive capability of a QSAR model should be tested through model validation. Cross-validation is one of the most often used model validation methods. The R_{cv}^2 values of models(4) is 0.936, which is well above 0.5, indicating that the model has good robustness and prediction ability.

The V_{IF} values of the variables(E_9 , E_{16} , E_{23}) in model (4) are 1.319, 1.069, 1.291, respectively, which is less than 5.0. Indicating that model is statistically significant and have good stability.

The relative activities(R_a) range of the sample between the maximum and minimum values is 1.20($2.00 - 0.80 = 1.20$). The ratio of the standard deviation ($S_D=0.070$) to 1.20 is 5.83%. This is less than 10%, indicating that the model has acceptable predictive accuracy.

3.3 Analysis of the QSAR equation(4)

According to the theory of electrotopological state indices, it can be seen that E_9 in the model (4) reflects carbon atom($=C<$) in aromatic ring, E_{16} in the model (4) reflects oxygen atom($=O$) in sulfuryl, E_{23} in the model (4) reflects nitrogen atom($=N-$). Wherein $=O$ and $=N-$ are highly electronegative polar groups that are capable of forming hydrogen bonds. $=C<$ is nonpolar or weak polar groups that is hydrophobic interaction of forming. The diuretic contains $=C<$, $=O$ and $=N-$, which is hydrogen bond and hydrophobic with the enzyme in the human body.

In addition, determination coefficient R^2 is also called the reduction error ratio. $R^2 = 0.975$ in model (4), indicating that E_{16} , E_{23} and E_9 constant items together show 97.5% of the factors affecting relative activities(R_a) of the diuretic, and only 2.5% are random factors, which further prove the correctness of the model.

4. Conclusions

(1) The optimal three-variable QSAR model of relative activity for thiazide diuretics was constructed by using the leaps-and-bounds regression method. The QSAR models have showed good correlation, as well as robustness and prediction ability by statistical indicators: R^2 , R_{cv}^2 , S_R , V_{IF} , F_{IT} and A_{IC} tests.

(2) According to the electrotopological state indices entering the model (4), the main molecular structural units that affect their relative activity are $=O$ (in sulfuryl), $=N-$ and $=C<$ (in aromatic ring).

(3) According to the main molecular structure units that affect relative activity, the interactions between thiazide diuretics molecules and enzyme in the human body are mainly hydrophobic, hydrogen bond.

All in all, this study provides theoretical guidance for diuretic detection, and also has certain guiding effect on anti-doping.

References

- [1] International Olympic Committee. (1990) Medical Commission. International Olympic charter against doping in sport. Lausanne:IOC,
- [2] Qin Y., Zhu S.T., Zhang C.J. (2001) Determination of Diuretics in Doping, Journal of Instrumental Analysis, 20: 82-83.
- [3] Feng C.J., Yang W.H., (2014) Linear QSAR regression models for the prediction of bioconcentration factors of chloroanilines in fish by density functional theory, Chin. J. Struct. Chem., 33: 830-834.
- [4] Feng C. J. (2012) Theoretical studies on quantitative structure-activity relationship and structural modification for 3-substituted sulfur-5-(2-hydroxyphenyl)-4H-1,2,4-triazole compounds, Acta Chim. Sin., 70: 512-518.

- [5] Liu B.G., LIU Jiang W., Li J.Q., Geng S., Mo H.Z., Liang G.Z., (2017)3D-QSAR and interaction mechanism of flavonoids as P-glycoprotein inhibitors. *Chem. J. Chinese Universities*, 38: 41-46
- [6] Liao L.M., Li J.F., Lei G.D. (2016)Structural characterization and Kovats retention indices(RI) prediction for alkylbenzene compounds. *J.Struct.Chem.*, 57: 1627-1634.
- [7] Hall L.H., Kier L.B., Brown B.B. (1995)Molecular similarity based on novel atom-type electrotopological state indices. *J.Chem.Inf.Comput.Sci.*, 35: 1074-1080.
- [8] Hall L.H., Kier L.B. (1995)Electrotopological state indices for atom types: A novel combination of electronic, topological, and valence state information. *J. Chem. Inf. Comput. Sci.*, 35:1039-1045.
- [9] Hall L.H., Mohny B., Kier L.B. (1991)The electrotopological state: structure information at the atomic level for molecular graphs. *J.Chem.Inf.Comput.Sci.*, 31: 76-82.
- [10] Bai D.L., Chen K.X. (2011)Advanced medicinal chemistry. The Beijing.
- [11] Wei D.B., Zhang A.Q., Wu C.D., Han S.K., Wang L.S. (2001)Progressive study and robustness test of QSAR model based on quantum chemical parameters for predicting BCF of selected polychlorinated organic compounds. *Chemosphere*, 44: 1421-1428.
- [12] Yao S.W., Lopes V.H., Fernández F., García-Mera X., Morales M., Rodríguez-Borges J. E., Cordeiro D. S. (2003)Synthesis and QSAR study of the anticancer activity of some novel indane carbocyclic nucleosides. *Bioorg. Med. Chem.*, 11: 4999–5006.
- [13] Xu X., Li X. G. (2012) QSAR for predicting biodegradation rates of polycyclic aromatic hydrocarbons in aqueous systems. *Chin. J. Struct. Chem.*, 31: 1212–1221.
- [14] Sung-Sau S., Kaprlus M., (1999)A comparative study of ligand-receptor complex binding affinity prediction methods based on glycogen phosphorylase inhibitors. *Journal of Computer - Aided Molecular Design*, 13: 243-258.
- [15] Urra L.S., Gonza'lez M.P., Teijeira M. (2007)2D-autocorrelation descriptors for predicting cytotoxicity of naphthoquinone ester derivatives against oral human epidermoid carcinoma. *Bioorg. Med. Chem.*, 15: 3565-3571.
- [16] Urra L.S., Gonza'lez M.P., Teijeira M. (2006)QSAR studies about cytotoxicity of benzophenazines with dual inhibition toward both topoisomerases I and II: 3D-MoRSE descriptors and statistical considerations about variable selection. *Bioorg. Med. Chem.*, 14: 7347-7358