QSAR Study on the anti-tumor activity of levofloxacin-thiadiazole HDACi conjugates

Ziqiang Tang1,2, Hui Feng1, Yan Chen1, Wei Yue1, Changjun Feng1*

1School of Chemistry and Chemical Engineering, Xuzhou University of Technology, Xuzhou, 221111;
2Xuzhou Technical Institute, Xuzhou 221151
*corresponding author’s e-mail address: fengecj@xzit.edu.cn

ABSTRACT. A molecular electronegativity distance vector (Mt) based on 13 atomic types is used to describe the structures of 19 conjugates (LHCc) of levofloxacin-thiadiazole HDAC inhibitor (HDACi) and related to the anti-tumor activity (Mt and Pc) of LHCc against MCF-7 and PC-3. The quantitative structure-activity relationships (QSAR) was established by using leaps-and-bounds regression analysis for the anti-tumor activities (Mt and Pc) of 19 above compounds to MCF-7 and PC-3 along with the Mt. The correlation coefficients (R2) and the leave-one-out (LOO) cross validation Rcv2 for the Mt and Pc models were 0.792 and 0.679; 0.773 and 0.565, respectively. The QSAR models have favorable correlation, as well as robustness and good prediction capability by R2, F, Rcv2, AIC, FIT, VIF tests. The results indicate that the molecular structural units: −CHg− (g=1, 2), −NH2, −NH−, −OH, O=, −O−, −S− and −X are main factors which can affect the anti-tumor activity Mt and Pc bioactivities of these compounds directly.

1. Introduction
Experimental determination of biological activity for all compounds requires large amounts of manpower, materials and financial resources, while quantitative structure-activity relationship (QSAR) is an effective way to solve this problem. QSAR[1-4] studied the mathematical relationship between the structure of several compounds and their biological activity by theoretical calculations and statistical analysis tools. To estimate and predict the relative properties (such as toxicity, mutagenicity, carcinogenicity, etc.) of other compounds, and to explore the possible mechanism of the microbial structure on the biological activity of the compounds at the molecular level.

The key to QSAR research is the establishment of descriptors for molecular structures. Scientists have done a great deal of meaningful work, such as the establishment of two-dimensional (2D) descriptors. Topological indexes are 2D descriptors which are encoded in numerical form information about molecular structure, such as molecular size, shape, branching, presence of hetero-atoms, and multiple bonds. They 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. Molecular connectivity indexes put forward by Randic[5] and developed by Kier and Hall[6] are one of the most broadly applied topological indexes. Recently, a novel molecular electronegativity distance vector (Mt) [7-9] based on 2D-topologies and 13 atomic types has been reported to successfully establish QSAR models between multiple (Mt) and organic compounds.
Histone deacetylases (HDACs) is a new target of anti-tumor research in recent years, and HDACi has become a hot research topic of anti-tumor drugs both at home and abroad. Eighteen kinds of new, efficient and low toxicity of levofloxacin-thiadiazole HDACi conjugates (referred to as LHCc) were designed and synthesized from levofloxacin by Li Hui and other [10]. The results showed that these new conjugates displayed potent anti-tumor activity against Michigan Cancer Foundation-7 (MCF-7, human breast adenocarcinoma cell) and non-androgen-dependent PC-3 cell (PC-3). Specifically, conjugate 10 exhibited the most activities, which was more potent than SAHA (Vorinostat). Therefore, it is important to study the levofloxacin-HDACi conjugate by QSAR method.

Based on the electronegativity distance vector \( M_t \) of Liu Shushen et al [7-9], the robust QSAR model of LHCc biological activity was established by leaps-and-Bounds regression, to estimate and predict the anti-tumor activity of LHCc and reveal the microstructure that affected its antitumor activity at the molecular level. It provides theoretical reference a reasonable design and screening of novel and highly effective lead compounds for LHCc.

2. Material and methods

2.1 Studied compounds and their biological activity data

The compounds studied herein are a series of HDACs inhibitors with anti-tumor activity of MCF-7 and PC-3. The matrix structure [10] of these compounds is shown in Fig.1. Their bioactivity data for MCF-7 and PC-3 were \( IC_{50} \), and drug concentrations resulted in 50% inhibition of MCF-7 and PC-3, in units of \( \mu \text{mol} \cdot \text{dm}^{-3} \). Their MCF-7 and PC-3 inhibitory activity \( (IC_{50}) \) were expressed as \( M_F \) and \( P_C \), respectively. The biological activity of 18 HDACs inhibitors and SAHA are listed in Table1 [10].

| No. | R   | n | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|-----|-----|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|
| 1   | OH  | 3 | 5.8440 | 0 | 10.117 | 10.3300 | -0.5457 | 11.7 | 10.96 | 15.3 | 14.15 |
| 2   | OH  | 4 | 7.1312 | 0 | 9.2323 | 9.6117 | -0.5048 | 9.8 | 9.83 | 12.7 | 12.48 |
| 3   | OH  | 5 | 8.9586 | 0 | 9.7850 | 9.1302 | -0.4780 | 8.2 | 9.25 | 9.6 | 11.58 |
| 4   | OH  | 6 | 11.1760 | 0 | 9.6841 | 8.7907 | -0.4599 | 6.5 | 9.01 | 7.2 | 11.14 |
| 5   | OH  | 7 | 13.6760 | 0 | 9.6089 | 8.5421 | -0.4473 | 9.3 | 8.99 | 11.5 | 10.99 |
| 6   | OH  | 8 | 16.3840 | 0 | 9.5516 | 8.3543 | -0.4384 | 10.1 | 9.13 | 13.9 | 11.05 |
| 7   | NHOH | 3 | 5.9526 | 0 | 12.5630 | 6.8814 | -0.5457 | 6.2 | 4.42 | 9.6 | 7.42 |
| 8   | NHOH | 4 | 7.2979 | 0 | 12.4840 | 6.2001 | -0.5058 | 3.5 | 3.37 | 7.5 | 5.51 |
| 9   | NHOH | 5 | 9.1654 | 0 | 12.4220 | 5.7401 | -0.4795 | 1.7 | 2.84 | 3.7 | 4.43 |
| 10  | NHOH | 6 | 11.4100 | 0 | 12.3740 | 5.4139 | -0.4615 | 0.8 | 2.63 | 1.5 | 3.85 |
| 11  | NHOH | 7 | 13.9280 | 0 | 12.3370 | 5.1735 | -0.4489 | 2.9 | 2.63 | 2.4 | 3.61 |
| 12  | NHOH | 8 | 16.6490 | 0 | 12.3100 | 4.9910 | -0.4400 | 5.3 | 2.78 | 5.3 | 3.58 |
| 13  | —   | 3 | 24.8850 | 0.0512 | 12.2380 | 3.2291 | -0.5700 | 9.2 | 7.36 | 13.3 | 12.67 |
| 14  | —   | 4 | 26.4540 | 0.0461 | 12.0620 | 2.6528 | -0.5282 | 6.5 | 5.90 | 11.6 | 10.95 |
| 15  | —   | 5 | 28.4850 | 0.0416 | 11.9360 | 2.2713 | -0.5000 | 4.7 | 5.90 | 9.8 | 9.96 |
| 16  | —   | 6 | 30.8590 | 0.0378 | 11.8450 | 2.0049 | -0.4804 | 1.6 | 4.44 | 6.2 | 9.43 |
| 17  | —   | 7 | 33.4760 | 0.0345 | 11.7770 | 1.8110 | -0.4664 | 2.9 | 4.14 | 7.1 | 9.20 |
| 18  | —   | 8 | 36.2750 | 0.0316 | 11.7260 | 1.6652 | -0.4561 | 5.3 | 4.01 | 12.4 | 9.17 |
| 19  | SAHA | 32.7280 | 0 | 12.2740 | 4.7459 | 4.0315 | 12.2740 | 4.4 | 3.89 | 1.7 |

2.2 Molecular descriptors
The molecular structure information of the compounds is a prerequisite for establishing a good structure-activity relationship. At present, the two-dimensional descriptors used in the molecular structure characterization of QSAR are molecular holographic, topological index and so on. They have been studied in environmental science, life sciences, drug design and so has been widely used. In this paper, the molecular electronegativity distance vector descriptor($M_f$)\textsuperscript[7-9] based on 13 atomic types is used to characterize the molecular structure of different classes of organic compounds. The specific calculations see the literature\textsuperscript[7-9].

In this paper, there are only 1, 2, 3, 5, 6, 7, 9, 10 and 13 species of atoms in the molecules. They interact with each other (including the self-interactions). Theoretically, 45 kinds of electrical distance vectors can be formed. Since some of the atomic types do not exist, some interaction types do not exist, so there are only 41 electrical distance vectors that are not all zero. Part of the electrical distance vectors can be formed. Since some of the atomic types do not exist, some interaction types do not exist, the molecular structure of the differences have a unique characterization, showing good structural selectivity.

2.3 Statistical regression analysis

For QSAR derivation, $IC_{50}$ values act as the dependent variables, and electronegativity distance vector descriptors act as the independent variables. The regression analyses are carried out by using multiple linear regression(MLR), partial least squares(PLS), leaps-and-bounds regression (LBR) program. The correlation between variables in model was estimated by the variance inflation factor ($VIF$). The statistic significance of the model was validated by $t$-tests. If the absolute values of $t$ for all the variables in the validated model are larger than the standard $t$-value (at $\alpha/2$) at one confidence level $\alpha$, it will suggest that the model passes $t$-tests and has obvious statistic significance. We also applied the Akaike’s information criterion($AIC$; $Eq.2$) 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 is considered potentially the most useful) and Kubinyi function($F_{IT}$; $Eq.3$; the best model will present the highest value of this function)\textsuperscript[13,14] to determine if a variable should be included in the model.

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

in which $\beta$ 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 “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 $R_{cv^2}>0.50$\textsuperscript[12] are considered reasonable. Where parameter $R_{cv^2}>0.50$ is used as a criterion of both robustness and predictive ability of the model.

The statistic significance of the model was validated by $t$-tests. If the absolute values of $t$ for all the variables in the validated model are larger than the standard $t$-value (at $\alpha/2$) at one confidence level $\alpha$, it will suggest that the model passes $t$-tests and has obvious statistic significance. We also applied the Akaike’s information criterion($AIC$; $Eq.2$) 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.

$$AIC = RSS \times \frac{f + b}{(f - b)\hat{\sigma}^2}, \hspace{1cm} F_{IT} = \frac{R^2 (f - b - 1)}{(f + b)(1 - R^2)}$$ \hspace{1cm} (2)

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 anti-tumor activity

The electronegativity distance vector($M_f$) and anti-tumor activities($M_r$, $P_C$)\textsuperscript[10] of HDACs inhibitors 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 shown in Table 2, 3. Where $R^2$ is the square of the correlation coefficient, $R_{adj^2}$ is the square of the adjusted
correlation coefficient, $R_{CV}^2$ is the LOO cross-validation correlation coefficient, $S_D$ is the standard deviation of the regression, $S_D$ is the standard deviation of the regression of LOO, and $F$ is the Fisher ratio. $AIC$ is the Akaike’s information criterion and $F_{IT}$ is the Kubinyi function. Table 2 shows the QSAR models for $M_F$ and $M_t$.

**Table 2** Results of electronegativity distance vector descriptors and anti-tumor activities $M_F$ with leaps-and-bounds regression

| No. | $R^2$ | $R_{adj}^2$ | $R_{cv}^2$ | $AIC$ | $FIT$ | $S_T$ | $S_{CV}$ | $F$ | Variables |
|-----|-------|-------------|------------|-------|-------|-------|----------|-----|-----------|
| 1   | 0.370 | 0.333       | 0.238      | 8.712 | 0.499 | 2.620 | 2.881    | 9.993| $M_{17}$  |
| 2   | 0.700 | 0.662       | 0.580      | 5.655 | 1.623 | 1.864 | 2.205    | 18.661| $M_{77}, M_{55}$ |
| 3   | 0.792 | 0.750       | 0.679      | 5.114 | 2.040 | 1.604 | 1.992    | 19.010| $M_{78}, M_{44}, M_{55}$ |
| 4   | 0.811 | 0.757       | 0.536      | 8.781 | 1.716 | 1.582 | 2.477    | 15.014| $M_{77}, M_{44}, M_{55}, M_{14}$ |

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

$$M_F = -9.749(\pm 2.639) + 0.183(\pm 0.071)M_{14} + 1.902(\pm 0.071)M_{77} + 125.464(\pm 31.319)M_{55}$$

(3)

When $f = 19$, $R^2 = 0.792$, $R_{adj}^2 = 0.750$, $F = 19.010$, $S_D = 1.604$ (Modeling);

$$R_{cv}^2 = 0.679, S_{cv} = 1.992$$

(LOO Cross-validation)

Table 3 shows the QSAR models for anti-tumor activities $P_C$ and $M_t$.

**Table 3** Results of $M_t$ and $P_C$ with leaps-and-bounds regression

| No. | $R^2$ | $R_{adj}^2$ | $R_{cv}^2$ | $AIC$ | $FIT$ | $S_T$ | $S_{CV}$ | $F$ | Variables |
|-----|-------|-------------|------------|-------|-------|-------|----------|-----|-----------|
| 1   | 0.251 | 0.207       | 0.073      | 31.083| 0.285 | 3.181 | 5.442    | 5.703| $M_{55}$  |
| 2   | 0.657 | 0.615       | 0.400      | 14.368| 1.332 | 2.657 | 3.515    | 15.354| $M_{77}, M_{55}$ |
| 3   | 0.773 | 0.728       | 0.565      | 12.311| 1.824 | 2.233 | 3.090    | 17.038| $M_{44}, M_{55}, M_{14}$ |
| 4   | 0.810 | 0.756       | 0.536      | 402.133| 1.705 | 2.112 | 16.763   | 14.967| $M_{77}, M_{44}, M_{55}, M_{14}$ |

Table 3 shows that with the increase in the number of independent variables in the model, in addition to $R^2$, $R_{adj}^2$, $S_D$, the rest of the statistical indicators in the ternary equation has a turning point, which $R_{cv}^2$, $F_{IT}$, $F$ has the maximum, $AIC$, $S_{CV}$ has a minimum. So choose the best ternary QSAR model:

$$P_C = 9.150(\pm 3.325) + 0.155(\pm 0.056)M_{14} - 2.757(\pm 0.478)M_{77} - 58.603(\pm 8.212)M_{78}$$

(4)

When $f = 19$, $R^2 = 0.773$, $R_{adj}^2 = 0.728$, $F = 17.038$, $S_D = 2.233$ (Modeling);

$$R_{cv}^2 = 0.565, S_{cv} = 3.090$$

(LOO Cross-validation)

Using the QSAR equations(3) and (4), we can predict the anti-tumor activities $M_F$ and $P_C$, respectively, and their predicted values are $M_F$ cal. and $P_C$ cal. in table 1.

3.2 Validation of the QSAR equation

The predictivity 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(3) and (4) are 0.679 and 0.565, respectively; which are well above 0.5, indicating that the model has good robustness and prediction ability.

Where the values in Equation(3), (4) after a symbol “±” refer to the standard deviation corresponding to the regression coefficient. All standard deviations were less than 1/2 of regression coefficients, indicating that the model is stable. The standard regression coefficients($S_T$) and $t$-value of three independent variables in equation(3) are listed in Table 4. When confidence level is 95%, the standard $t$ value ($t_{a/2}$) of the model is 2.2281. From Table 4, we can see that the absolute value of $t$ of each independent variable in model(3) is bigger than the standard $t_{a/2}$ value, which proves the credibility of the model. At the meantime, the absolute values of $t$ of $M_{77}, M_{55}, M_{14}$ decrease in turn, which is consistent with the law of $S_T$, which indicates that $M_{77}$ has the strongest effect on $M_F$. For
models (4), a similar rule is obtained from Table 4: $M_{78}$ is the strongest factor affecting antitumor activity. As shown in Table 4, the $V_{IF}$ values of the variables in model (3), (4) are less than 5.0, indicating that all models are statistically significant and have good stability.

| Table 4 | $S_R$ and $t$ values of independent variables in equation (3), (4) |
|---------|---------------------------------------------------------------|
| model.3 | $M_{77}$ | 1.751 | 7.022 | 4.292 |
|         | $M_{14}$ | 0.600 | 2.572 | 3.921 |
|         | $M_{55}$ | 0.771 | 4.006 | 2.667 |
| model.4 | $M_{77}$ | 1.751 | 7.022 | 4.292 |
|         | $M_{14}$ | 0.600 | 2.572 | 3.921 |
|         | $M_{59}$ | 0.381 | 2.765 | 1.257 |

3.3 Analysis of the QSAR equation

According to the theory of molecular electronegativity distance vector, it can be seen that the electrical distance vector $M_{14}$ in the models reflects the interaction of the second carbon atom ($-\text{CH}_g-$, $g=1, 2$) with the second carbon atom, $M_{55}$ reflects the interaction of the thirteen type of halogen atom ($-\text{X}$) with the second carbon atom, $M_{59}$ reflects the interaction between the ninth kind of oxygen atom ($-\text{O}$) and the sixth group of nitrogen atom ($-\text{NH}_2$), and $M_{77}$ reflects the interaction between the ninth kind of oxygen atom and the tenth kind of oxygen atom ($-\text{O}$) in ether or sulfur atom ($-\text{S}$) in thioether. The five electronegativity distance vectors respectively implied the structure information of eight non-hydrogen atoms. Wherein $-\text{CH}_g-$ is a non-polar group, having hydrophobicity; the remaining seven classes are highly electronegative polar groups that are capable of forming hydrogen bonds, and can form the coordination compounds formed with metal ions in enzyme.

In addition, determination coefficient $R^2$ is also called the reduction error ratio. $R^2 = 0.792$ in model(3), indicating that $M_{14}$, $M_{77}$, $M_{55}$ and constant items together show 79.2% of the factors affecting anti-tumor activities ($M_F$) of HDACs inhibitors to MCF-7, and only 21.8% are random factors; $R^2 = 0.773$ in model (4), indicating that $M_{14}$, $M_{59}$, $M_{78}$ and constant items together show 77.3% of the factors affecting anti-tumor activities ($P_C$) of HDACs inhibitors to PC-3, and only 22.7% are random factors. which further prove the correctness of the models.

4. Conclusion

(1) The molecular electronegativity distance vector can characterize the abstract molecular structure to achieve numerical representation of 19 HDACs inhibitors, and show good structural selectivity.

(2) The optimal ternary QSAR model of anti-tumor activity $M_F$ and $P_C$ of HDACi to MCF-7 and PC-3 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.

(3) Drug activity is usually the result of synergistic action by multiple structural factors, rather than by a single parameter. Therefore, according to the electrical distance vector entering the three models, the main molecular structural units that affect their anti-tumor activity $M_F$ and $P_C$ are $-\text{CH}_g-$ ($g=1, 2$), $-\text{NH}_2$, $-\text{NH}$, $-\text{OH}$, $-\text{O}$, $-\text{S}$ and $-\text{X}$.

In summary, this study provides theoretical guidance for the further design of novel and efficient HDACs inhibitors.

Acknowledgments

This work was financially supported by the the National Natural Science Foundation of China (21272095, 21075138) and special fund of State Key Laboratory of Structure Chemistry (20160028).
References

[1] C J Feng, W H Yang. Linear QSAR regression models for the prediction of bioconcentration factors of chloroanilines in fish by density functional theory. Chin. J. Struct. Chem., 2014, 33(6):830-834.

[2] C J Feng. 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. 2012, 70, 512–518.

[3] B G Liu, J W Liu, J Q Li, et al. 3D-QSAR and interaction mechanism of flavonoids as P-glycoprotein inhibitors. Chem.J.Chinese Universities, 2017, 38, 41-46

[4] M A Abdelrahman, I Salama, M S Gomaa, et al. Design, synthesis and 2D QSAR study of novel pyridine and quinolone hydrazone derivatives as potential antimicrobial and antitubercular agents[J]. European Journal of Medicinal Chemistry, 2017, 138(9): 698-714

[5] M Randic. On charactermization of molecular branching. J. Amer. Chem. Soc., 1975, 97: 6609-6615

[6] L B Kier, L H Hall. Molecular Connectivity in Structure-Activity Analysis. England: Researchh Studies Press, 1986: 82.

[7] S S Liu, C S Yin, Z L Li, et al. QSAR Study of Steroid Benchmark and Dipeptides Based on MEDV-13. J. Chem. Inf. Comput. Sci., 2001, 41: 321-329.

[8] S S Liu, C S Yin, L S Wang. Combined MEDV-GA-MLR Method for QSAR of Three Panels of Steroids, Dipeptides, and COX-2 Inhibitors. J. Chem. Inf. Comput. Sci., 2002, 42: 749-756.

[9] Y H Zhang, Z N Xia, L T Qin, et al. Prediction of blood–brain partitioning: A model based on molecular electronegativity distance vector descriptors. J. Molec. Grap. Model., 2010, 29: 214-220.

[10] H Li, X Han, D W Li. Synthesis and anti-tumor activity of levofloxacin-thiadiazole histone deacetylase inhibitor conjugates. Acta Pharmaceutica Sinica, 2017, 52: 582 –591.

[11] D B Wei, A Q Zhang, C D Wu, et al. Progressve study and robustness test of QSAR model based on quantum chemical parameters for predicting BCF of selected polychlorinated organic compounds. Chemosphere, 2001, 44, 1421-1428.

[12] S Xia, Y Feng, J G Cheng, et al. QAAR exploration on pesticides with high solubility: An investigation on sulfonylurea herbicide dimers formed through p–p stacking interactions[J]. Chin. Chem. Let., 2014, 25: 973–977

[13] L S Urra, M P Gonza'lez, M Teijeira. 2D-autocorrelation descriptors for predicting cytotoxicity of naphthoquinone ester derivatives against oral human epidermoid carcinoma. Bioorg. Med. Chem., 2007, 15, 3565-3571.

[14] L S Urra, M P Gonza'lez, M Teijeira. 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., 2006, 14, 7347-7358