

QSAR Models of 2, 4-Disubstituted Imidazopyridines as Hemozoin Formation Inhibitors

Jieru Wang¹, Yuliang Li²

¹College of Computer Science & Technology of Qingdao University, Qingdao, 266071, China

²College of Computer Science & Technology of Qingdao University, Qingdao, 266071, China

Abstract: A new type of hemozoin formation inhibitors are the 2,4-disubstituted imidazopyridines, which can be used to design and develop anti-malarial drugs. Establishing a quantitative structure-activity relationship (QSAR) model can predict the physical and chemical properties of molecules. We used CODESSA software to find suitable molecule descriptors by heuristic method. Then a linear QSAR model with correlation coefficient (R²), square of standard error (S²) and the square of cross-validation coefficient (R²_{cv}) are 0.7, 0.54 and 0.25 respectively. In addition, we randomly divided the 60 compounds into 45 training sets and 15 test sets to establish the nonlinear model using gene expression programming (GEP), the R² and mean square error (MSE) of the training set are 0.92 and 0.17, and the test set are 0.85 and 0.10. It can be seen that the nonlinear results are significantly better than that of the linear results. It is hoped that this model will help in the design of hemozoin formation inhibitors.

Keywords: Hemozoin Formation Inhibitors, Quantitative structure-activity relationship (QSAR), Gene expression programming (GEP), Heuristic method

Received 15 Jan 2021, Revised 20 May 2021, Accepted 22 June 2021

1. Introduction

Malaria is one of the mosquito-borne infectious diseases and an important public health issue of global concern [1]. With the exception of Europe in 2019, malaria cases have been reported in other regions of the WHO. The number of reported cases of malaria in the world is on the rise. The number of reported malaria cases in Africa in each year accounts for more than 90% in the world, and there is an obvious trend of increasing year by year [2]. Although the global epidemic prevention and control has been effectively, we still need to pay attention to malaria [3]. Due to the increased substitution of the vicious parasite protozoa, we need to seek novel treatment drugs.

The formation of hemozoin is an effective target for most known existing antimalarials and is considered a suitable target for antimalarials [4]. Hemoglobin is a metabolic crystalline by-product of hemoglobin digested by parasites during red blood cell (RBC) infection. The formation of hemoglobin from heme residues is common in many blood-eating organisms that is not a phylogenetic relationship with Plasmodium species, such as schistosoma mansoni and rhodnius prolixus. During the red blood cell formation stage, the digestion of hemoglobin in the food vacuoles of the malaria parasite will release a large amount of redox active free heme as a by-product, which is toxic to the malaria parasite. Alpha-heme (iron protoporphyrin IX) is toxic to parasites, which is released during hemoglobin digestion. However, most likely as a protective strategy, the parasite converts α -hemin into hemozoin, a molecule with paramagnetic properties [5,6]. Therefore, hemozoin is the target of attracting new drug treatments. These new drugs may interfere with the biological

crystallization of hemoglobin and help fight diseases caused by these pathogens, especially malaria.

There are effective anti-malaria drugs such as quinine and artemisinin [7]. However, we should actively look for new drugs to fight malaria. According to the Crum-Brown equation, there is a close relationship between the structure, activity and properties of organic compounds. Therefore, we can analyze the structure of compounds to provide help for the development of new drugs. QSAR is the application of chemical theoretical calculation methods and various mathematical statistical analysis methods to quantitatively describe and study the relationship between the structure and their properties [8]. QSAR makes computer-aided drug design a perfect and emerging research field. It is a summary based on a large number of structures and the corresponding compound data [9]. Then it has good predictive ability. The establishment of a model steps are data collection, selection of descriptors, and model establishment.

The mathematical methods for constructing QSAR models usually include linear methods and nonlinear methods. In this study, a linear model was established by the HM. However, most problems in biology and chemistry are nonlinear, which requires the help of nonlinear algorithms. In this study, we used GEP to build the nonlinear QSAR model.

Gene expression programming is a new technology based on the gene expression law of biological genetics, which can evolve into a computer program coded by linear chromosomes with a fixed length. By simulating genetic evolution, the algorithm is continuously developed to find the best-fitting equation. A feature of the GEP method is that, because genetic operators work at the chromosome level, the creation of genetic

1246

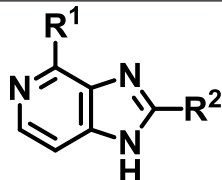
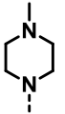
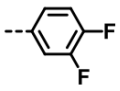
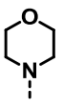
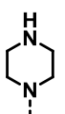
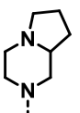
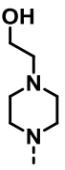
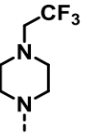
diversity is greatly simplified. Compared with other algorithms, the supervised machine learning model constructed by GEP has significantly improved data processing capabilities and generalization capabilities [10].

This study focused on 2,4-disubstituted imidazopyridines as inhibitors of hemozoin formation. In the QSAR model processes, the non-linear predictive ability is obviously better than the linear predictive ability. The GEP model has a better fitting effect.

2. Experiment

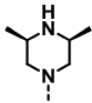
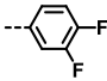
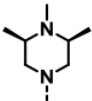

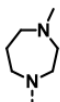
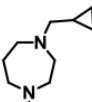

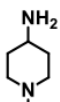
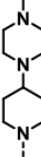
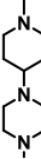
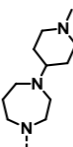
2.1. Data collection and its division

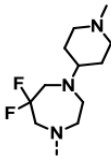
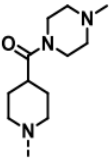
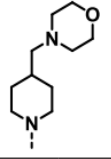
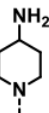
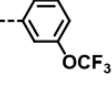
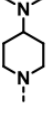
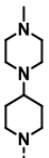
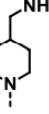
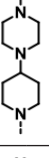
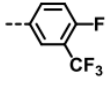
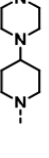
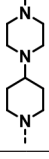
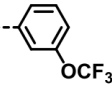
Table 1 The IC₅₀ values of compounds

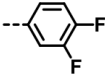
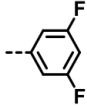
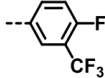

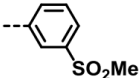
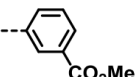
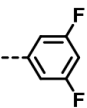
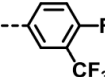
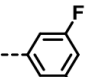
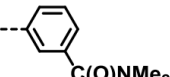
			
Number.	R ₁	R ₂	NF54 IC ₅₀ (μM) ^a
1 ^T			0.24
2			3.1
3			0.17
4			0.22
5 ^T			0.57
6 ^T			2.3

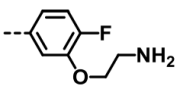

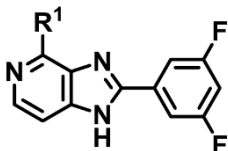
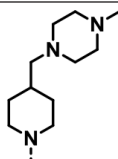
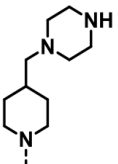
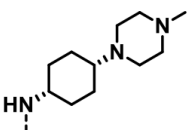
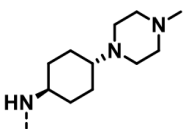
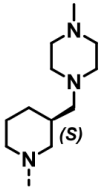
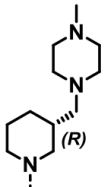
The first step of QSAR research is to select data and establish a data set. The reliability of the data is the key to the establishment of a model, and the model is meaningful.

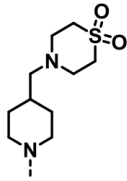
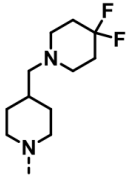
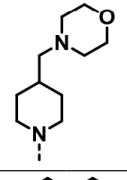
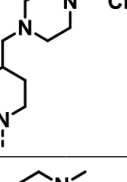
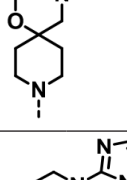
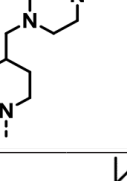
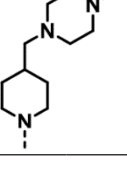
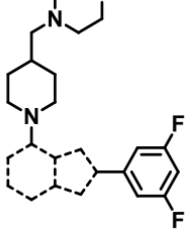
60 compounds and the IC₅₀ values were obtained from the literature [11] and listed in Table 1. In order to get good results by GEP method, we used system time as Random Seed to divide the data set into 45 compounds training set and 15 compounds test set. The training set was used to build, train and optimize the model, and the test set was used to evaluate the predictive ability of the model.

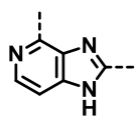
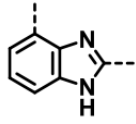
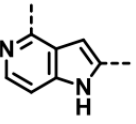
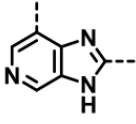
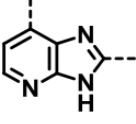
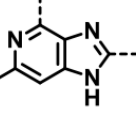
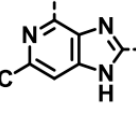
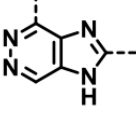
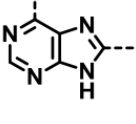
7			0.076
8			0.26
9			0.21
10 ^T			0.061
11			0.020
12			0.079
13			0.11
14 ^T			0.046
15			0.071
16			0.035

17			0.059
18 ^T			0.10
19			0.051
20			0.073
21 ^T			0.14
22			0.064
23			0.068
24			0.058
25			0.036
26 ^T			0.064

27			0.046
28			0.050
29			0.058
30			0.11
31			0.45
32			0.079
33			0.023
34			0.050
35			0.054
36			1.0

37			2.2
38			9.3
			
Number.	R ₁		NF54 IC ₅₀ (μM) ^a
39 ^T			0.023
40			0.038
41			0.16
42			0.13
43 ^T			0.028
44 ^T			0.073

45			1.4
46 ^T			0.20
47			0.29
48			0.075
49			0.18
50			0.029
51			0.019
			
Number.	Imidazopyridine Core		NF54 IC50 (μM) ^a

52 ^T		0.023
53		0.83
54 ^T		0.014
55		0.13
56		0.53
57		0.012
58 ^T		0.077
59		0.58
60		1.8

^T is test set for GEP.

2.2. Calculation of the descriptors

The calculation of molecular descriptors is the basis of QSAR research. The precise definition and rational using of molecular descriptors are very important. The so-called molecular descriptor refers to the measurement of a certain aspect of the properties of a molecule, which can be the physical and chemical properties of the molecule, or a numerical index derived through various algorithms based on the molecular structure. In this study, the chemical structure of the compound was drawn in the Chemdraw software, and then the molecular structure was imported into the Hyperchem [12] software for optimization. In order to achieve the lowest energy steady state of the compound, semi-empirical AM1 or PM3 methods [13] were used for precise optimization. The molecular structure was optimized by the Polak-Ribiere algorithm until the root mean square gradient reaches 0.01. Then 6 sets of descriptors were calculated.

2.3. Linear model by Heuristic method [14]

Feature selection is to reduce the number of descriptors and delete the descriptors that have less impact on the result. The remaining descriptors should represent the molecular structure and various properties as much as possible. The HM implemented in CODESSA software is used to calculate molecular descriptors and build linear models, and there is no software limitation on the size and speed of the data set. The detailed steps for HM to establish a linear model are: (1) Select of 1-parameter descriptor, use the R₂, F-test and t-test as standard selection, and delete the descriptors with lower correlation. (2) Select of 2-parameter descriptors, and pass R₂ and F-tests as standard selections. (3) Select the n parameter descriptor. After obtaining the two-parameter correlation coefficient with the best statistical characteristics, add descriptors that were not used in the previous selection process. Repeat this step until the established correlation equation contains the most parameters. The criteria for model evaluation are the R₂, the F-test value, the t-test value and the R_{2cv}. As a result, a linear model with 6 descriptors was developed by HM.

Therefore, linear models are not enough to find their correlation. Obviously, a non-linear model is needed.

2.4. Nonlinear model by GEP

GEP is a new adaptive evolutionary algorithm based on the structure and function of biological genes [15]. It was proposed by Ferreira at the end of the 20th century. GEP is essentially different from the individuals used by Genetic Algorithm (GA) and Genetic Programming (GP): in GA, individuals are linear strings (chromosomes) with a fixed length; in GP, individuals are non-linear entities with different lengths and shapes (split trees). In GEP, individuals are encoded as fixed-

length linear strings (genomes and chromosomes), and then expressed as non-linear entities of different lengths and shapes (simple graphical representations or expression trees). Due to the combination of the advantages of GA and GP, the efficiency of GEP to solve traditional problems is 2-4 orders of magnitude better than that of traditional GP methods.

In GEP, genes are composed of a head and a tail. The head is composed of terminal symbols and function symbols, and the tail can only be composed of terminal symbols. Assuming that the length of the head is h and the length of the tail is t , the relationship as follows:

$$t = h(n-1) + 1$$

Among them, n represents the number of parameters of the function that requires the most variables in the function symbol set. Obviously, the length of the gene can be calculated by the equation:

The $l = h + t = nh + 1$ main steps of the GEP algorithm are shown in Figure 1.

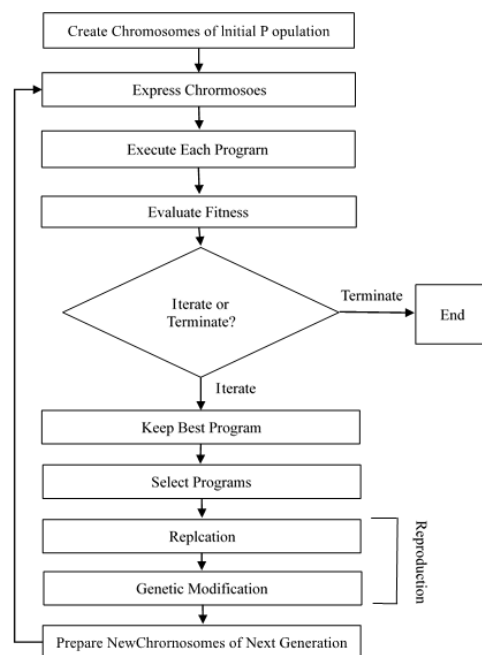


Figure 1. The flow chart of GEP.

The genetic manipulation of GEP begins with the formation of the initial population. First, defined the operator set F and the variable set P for the functional operation form. According to the coding method, first determine the value of h , and randomly selected data from F and P to generated a gene individual containing an expression code string and a numerical code string.

There are a variety of genetic operators in GEP, and different operators have different functions. Using these genetic operators can make individuals in the population

maintain diversity during evolution. At the same time, due to the unconstrained mapping relationship between the genotype and phenotype of GEP, it is very easy to implement these genetic operators.

In the genetic algorithm, fitness is an index that measures the adaptability of a species to the environment. According to the size of the fitness, the fitness function of the corresponding gene expression is given according to the principle of "survival of the fittest" of natural organisms, that is, the closer the gene expression is to the actual observation value, the higher the fitness. Ferreira proposed two evaluation models.

$$fitness = \sum_{i=1}^n (M - \left| \frac{y_i - \hat{y}_i}{y_i} \right|)$$

$$fitness = \sum_{i=1}^n (M - |y_i - \hat{y}_i|)$$

In fact, use relative error or absolute error for evaluation. Both error models have their inherent shortcomings. Therefore, we chose the MSE as the fitness function.

$$fitness = MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

In addition, we used the Automatic Problem Solver (APS) [16] software. The GEP algorithm had been well applied in this software, and a nonlinear model with good predictive ability had been established.

3. Results and discussion

3.1. Calculation results of HM

A total of 540 molecular descriptors of 60 compounds were calculated by CODESSA software. In order to find the best linear model, HM constructed linear models with 1-11 descriptors respectively. The R2, R2cv and S2 of these models are shown in Figure 2.

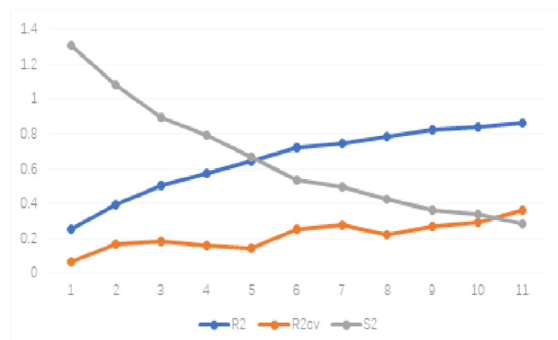


Figure 2. The influence of the number of descriptors on the correlation coefficient the cross-validation, correlation coefficient and the square of standard error of linear models.

By screening these molecular descriptors, 6 descriptors were obtained, which can more

comprehensively represent the structure and physical and chemical properties of the molecule. We can see that as the number of descriptors increases, R2 and R2cv will increase, while S2 will decrease. After the number of descriptors increased to 7, the increase rate of R2 and the decrease rate of S2 decreased slightly, and the generalization ability of the model with too many descriptors was poor, and finally 6 descriptors were selected as the best linear model (Figure 3).

$$IC50 = -4.1033 \times 10^2 + 3.9758 \times 10^3 MERA - 9.9646 \times 10 RNRB - 2.3382 \times 10 MREB + 1.7346 \times 10^{-1} MERB + 1.6076 \times 10 MNRB + 1.5752 \times 10 IOSE$$

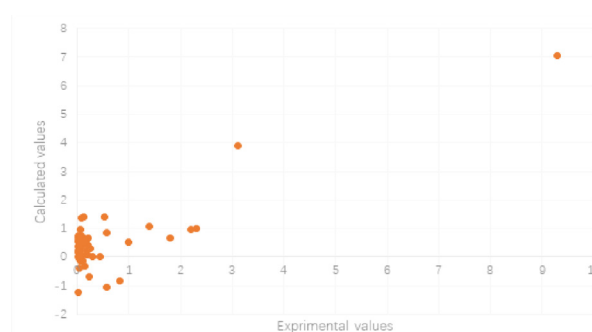


Figure 3. The plot of calculated IC50 versus experimental values for all compounds by HM.

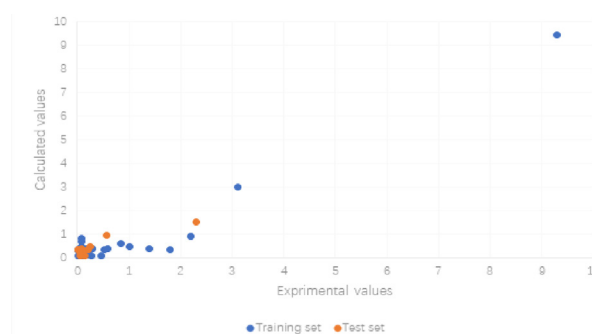


Figure 4. The relationship between the predicted and experimental values of GEP for 60 compounds

Table 2 lists the descriptors involved and their physical and chemical meanings. The R2, R2cv, F and S2 of this model are 0.7197, 0.2509, 22.6844 and 0.5367, respectively. Although R2 is satisfactory, R2cv and S2 are not very good. This shows that the generalization ability of the established linear model is not strong. Therefore, we imported the corresponding descriptor with the response variable into the APS software to develop the nonlinear model.

Table 2 The physical-chemical meaning, coefficient, DX and t-test values for the selected molecular descriptors of the best linear model by HM.

Descriptors	Physical-Chemical meaning	Coefficient	DX	t-test
Constant	Intercept	-410.33	121.89	-3.3664
MERA	Min electroph. react. index for a N atom	3975.8	904.15	4.3973
RNBR	Relative number of benzene rings	-99.646	14.147	-7.0437
MREB	Max resonance energy for a C-H bond	-23.382	3.1888	-7.3325
MERB	Min e-e repulsion for a C-N bond	0.17346	0.044176	3.9265
MNRB	Max n-n repulsion for a C-H bond	16.076	3.6684	4.3821
IOSE	Image of the Onsager-Kirkwood solvation energy	15.752	4.1462	3.7992

3.2. Results of GEP

In order to get a more accurate and generalized model, we introduced six descriptors in APS, and established a nonlinear model through the GEP feature set. We randomly selected 15 of the 60 compounds as

the test set, and the remaining 45 as the training set. After adjusting the function set, the most satisfactory result was obtained, where $R^2=0.92$, $MSE=0.17$ in the training set, $R^2=0.85$ and $MSE=0.10$ in the test set. The mathematical formula of the model is as follows:

$$IC50 = 4*x1 - \frac{x4}{\text{pow}(x0, x4) + \frac{x1}{x0}} + x2*x1 - x0*(x2*x3 + x3) + \frac{\text{pow}(\text{pow}(10, (x2*x4^2)), x2)}{x2}$$

The x_0 , x_1 , x_2 , x_3 and x_4 in the formula represent the descriptors MERA, RNBR, MREB, MERB, and IOSE respectively (Table 2). The function set used by

this model is {'+', '-', '*', '/', 'Pow', '10^x', 'Log', 'Inv'}; other definition parameters are shown in Table 3.

Table 3 The parameter setting for GEP

Parameter names	Values
Number of Chromosomes	100
Number of Genes	5
Function set	'+', '-', '*', '/', 'pow', '10^x', 'Log', 'Inv'
Gene head size	8
Linking function	'+'
Mutation rate	0.1
1-Point recombination rate	0.3
2-Point recombination rate	0.3
Gene recombination rate	0.1
IS transposition rate	0.1
RIS transposition rate	0.1
Gene transposition rate	0.1
Selection range	100
Precision	0.01

Through the experimental results of IC50 training set and test set in GEP, the nonlinear results are better than the linear results obtained by HM.

3.3. Discussion of selected descriptors

In order to better understand the factors that affect IC50, we discussed descriptors more deeply. The absolute value ($|t|$) of t-test is used to rank the selected six descriptors. The value indicates the importance of descriptors. It can be concluded that $MREB > RNBR > MERA > MNRB > MERB > LOSE$. It can be seen that MREB, RNBR, MERA and MNRB are more important descriptors.

MREB, MERA, MNRB, MERB and LOSE are Quantum chemical descriptors. The charge and local electron density in the molecule are very important for the physical and chemical properties of the compounds. The partial charge of an atom has been used as a static chemical reaction index, and the calculated σ and π electron densities on a specific atom also characterize the possible orientation of chemical interactions [17].

MREB is Max resonance energy for a C-H bond. Its definition is given as [18]:

$$E_{RAB} = \sum_{\mu \in A} \sum_{\nu \in B} P_{\mu\nu} \beta_{\mu\nu}$$

Here, A is given atomic species, B is another atomic species, $P_{\mu\nu}$ is density matrix elements over atomic basis $\{\mu\nu\}$, $\beta_{\mu\nu}$ is resonance integrals on atomic basis $\{\mu\nu\}$. The larger density matrix elements over atomic basis $\{\mu\nu\}$ and resonance integrals on atomic basis $\{\mu\nu\}$, the greater the value of MREB. Due to the negative correlation of MREB to IC50, lower value of this descriptor and a better IC50 value can be obtained.

MNRB is Max n-n repulsion for a C-H bond. It is calculated as follows[18,19]:

$$E_{nnAB} = \frac{Z_A Z_B}{R_{AB}}$$

A is given atomic species, B is another atomic species. Z_A is charge of atomic nucleus, Z_B is charge of atomic nucleus, R_{AB} is distance between the atomic nuclei, A and B

The smaller the charge of atomic nucleus of C and H and the greater the distance between the atomic nuclei, the lower MNRB can be obtained. Because the coefficients of MNRB are positively correlated, a lower MNRB can get a lower IC50.

MERA is Min electroph. react. index for a N atom, which reflects the electrophilic reactivity of the N atom on cinnamaldehyde compounds. For an atomic species A, it can be calculated as follows [20]:

ϵ LUMO is the energy of the lowest unoccupied molecular orbital (LUMO). Here, $CLUMO_i$ is the i th orbital coefficient of atom A on LUMO. Such

summation is conducted over all valence atomic orbitals i in atom A ($i = 1 \dots n_A$). Because MERA and IC50 are positively correlated, as MERA increases, IC50 decreases.

MERB is Min e-e repulsion for a C-N bond. It is calculated as follows [18]:

$$E_{ee}(AB) = \sum_{\mu, \nu \in A} \sum_{\lambda, \sigma \in B} P_{\mu\nu} P_{\lambda\sigma} \langle \mu\nu | \lambda\sigma \rangle$$

A is given atomic species. B is another atomic species.

$P_{\mu\nu}$ and $P_{\lambda\sigma}$ are density matrix elements over atomic basis.

$\langle \mu\nu | \lambda\sigma \rangle$ is electron repulsion integrals on atomic basis.

The smaller the density matrix elements and electron repulsion integrals of C and H atoms are, the lower the MERB will be. Because the coefficient of MERB is positively correlated with IC50, the lower value of IC50 will also decrease.

RNBR, constitutional descriptors, is independent from molecular connectivity and conformations. RNBR is relative number of benzene rings, the coefficient of RNBR is negative. This means that as the number of substituted benzene rings increases, the IC50 value will decrease. [21]

IOSE, is Image of the Onsager-Kirkwood solvation energy. The Onsager-Kirkwood solvation energy is given [22]:

$$E_{Onsager} = -\frac{(\epsilon - 1)\mu^2}{(2\epsilon + 1)a^3}$$

ϵ is the macroscopic relative permittivity of the solvent and μ is the total dipole moment of the solvent and a is the radius of the solvent cavity. As IOSE decreases, IC50 will also decrease.

Compounds with higher RNBR and MERB, lower MERA, MNRB, MERB and IOSE have better IC50 values, which means that these compounds have a better inhibitory effect on the formation of hydantoin. There is a possibility that the compounds with these concentrations will act as effective heme formation inhibitors.

4. Conclusion

In summary, 60 QSAR models of 2,4-disubstituted imidazopyridines as heme formation inhibitors have been established. After HM linear analysis and GEP nonlinear analysis, the nonlinear results are significantly better than the linear results. It can be seen that GEP has greater development potential and higher efficiency

in terms of data processing capabilities. In addition, through the analysis of the descriptor, it is obvious that the quantum chemistry plays a huge role in the change of IC50, and a similar design can be considered for the formation of hemozoin inhibitor. It is hoped that the above model can provide guidance for the development of related new drugs.

References

- [1] B.E. Barber, G.S. Rajahram, M.J. Grigg, T. William, N.M. Anstey, *BioMed Central*, 16 (2017) 1-2.
- [2] X. Hu, S. Wu, Y. Weng, H. Han, L. Tian, Y. Lv, R. Xu, P. Zhou, H. Wu, *EpidemDisease Surveillance*, 36 (2021) 2-3.
- [3] P. Alonso, A.M. Noor, *Lancet*, 2532 (2017) 1-3.
- [4] M. Navarro, W. Castro, C. Biot, *Organometallics*, 31 (2012) 9-11.
- [5] S. Kumar, M. Guha, V. Choubey, P. Maity, U. Bandyopadhyay, *Life Sciences*, 80 (2007) 3-4.
- [6] L.M. Coronado, C.T. Nadovich, C. Spadafora, *BBA - General Subjects*, 1840 (2014) 2-6.
- [7] E.A. Ashley, A.P. Phyto, *Drugs*, 78 (2018) 1-2.
- [8] A. Cherkasov, E.N. Muratov, D. Fourches, A. Varnek, I.I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y.C. Martin, R. Todeschini, V. Consonni, V.E. Kuz'min, R. Cramer, R. Benigni, C. Yang, Jn, L. Terfloth, J. Gasteiger, A. Richard, A. Tropsha, *Journal of medicinal chemistry*, 57 (14) 8-18.
- [9] O.A. Santos-Filho, A.J. Hopfinger, *Journal of Computer-Aided Molecular Design*, 15 (2001) 1-12.
- [10] C. Ferreira, Springer London, (2002) 1-2.
- [11] A. Horatscheck, A. Andrijevic, A. T. Nchinda, C. L. Manach, T. Paquet, L. P. Khonde, J. Dam, K. Pawar, D. Taylor, N. Lawrence, C. Brunschwig, L. Gibhard, M. Njoroge, J. Reader, M. van der Watt, K. Wicht, A.C.C. de Sousa, J. Okombo, K. Maepa, T.J. Egan, L.M. Birkholtz, G.S. Basarab, S. Wittlin, P.V. Fish, L.J. Street, J. Duffy, K. Chibale, *Journal of Medicinal Chemistry*, 63 (2020) 3-8.
- [12] HyperChem. 4.0, Hypercube, 1994.
- [13] P. P. Stewart. University of Indiana, Bloomington, (1989).
- [14] R. Katritzky, V. S. Lobanov, M. Karelson, Gainesville, Florida, (1994).
- [15] Ferreira, *Complex System*, 13 (2001) 87-129.
- [16] <http://www.gepsoft.com/gepsoft>.
- [17] M. Karelson, V. S. Lobanov, A. R. Katritzky, *Chemical reviews*, 96 (1996) 2-7.
- [18] E. Clementi, SPRINGER-VERLAG 19 (1980) 167.
- [19] M.A. Kirpichenok, N.S. Zefirov, *J. Org. Chem. USSR (Engl. Transl.)(United States)*, 23 (1987) 673-703.
- [20] H. Wang, M. Jiang, S. Li, C.Y. Hse, C. Jin, F. Sun, Z. Li, *Royal Society Open Science*, 5 (2017) 8.
- [21] <http://www.codessa-pro.com/descriptors/const.htm>.
- [22] L. Mu, R. S. Drago, D. E. Richardson, *Journal of the Chemical Society*, 1 (1998) 3.