Mongolian Journal of Biological Sciences 2008 Vol. 6(1-2): 45-49

Calculation of Quantitative Structure-Activity Relationship Descriptors of Artemisinin Derivatives

Jambalsuren Bayarmaa[1] and Vladimir Frecer[2]

[1]Department of Biochemistry and Bioorganic Chemistry, Faculty of Biology, National University of Mongolia E-mail: bayarmaa_j@biology.num.edu.mn
[2]Area of Pure and Applied Chemistry, International Center for Sciense and High Technology-UNIDO, Trieste, Italy E-mail: vladimir.frecer@ics.trieste.it

Abstract

Quantitative structure-activity relationships are based on the construction of predictive models using a set of known molecules and associated activity value. This accurate methodology, developed with adequate mathematical and computational tools, leads to a faster, cheaper and more comprehensive design of new products, reducing the experimental synthesis and testing on animals. Preparation of the QSAR models of artemisinin derivatives was carried out by the genetic function algorithm (GFA) method for 91 molecules. The results show some relationships to the observed antimalarial activities of the artemisinin derivatives. The most statistically significant regression equation obtained from the final GFA relates to two molecular descriptors.

Key words: molecular model, descriptor, QSAR technique, artemisinin derivatives

Introduction

Malaria is one of the most serious parasitic diseases in the world. There are millions of acute cases of malaria each year globally, resulting in more than one million deaths. The clinical disease caused by *Plasmodium falciparum* is widely distributed and entrenched in areas of the world where climates are suitable for its transmission. The present level of annual global malaria incident consisted of about 500 million, 350 million in the African region and 125 million in the South East Asian region. Deaths due to malaria are occurring in increasing numbers because of frequent failure of the conventional treatments using drugs such as chloroquine, mefloquine and sulfadoxine-pyrimethamine, against *P. falciparum*, the populations of which have developed a high degree of resistance.

Combination therapies with formulations containing an artemisinin compound have emerged as a more reliable treatment option. Artemisinin is naturally formed in *Artemisia annua* L (sweet wormwood). This herb of the Asteraceae family has been used for treatment of fever and malaria in China over many centuries. Additional names found in China for artemisinin are qinghaosu and huanghuahaosu. Artemisinin and its derivatives are only the group of compounds that is still effective against drug-resistant *P. falciparum* strains, and has the ability to quickly reduce parasite level (Krishna, 2004). From the other side, it has short plasma half-life, limited bioavailability, poor solubility in oil and water, and the low yield from natural sources (Woodrow, 2005). These reasons prompted to develop new chemotherapeutic artemisinin derivatives. By the chemical modification of artemisinin at the position of C10 semisynthetic derivatives as artesunate, artemether, artelinic acid and dihydroartemisinin were produced, which have more antimalarial activities *in vitro* than artemisinin itself (Pinheiro, 2003). Many derivatives have been synthesized using dihydroartemisinin by adding different radicals on its rings. But, the development of a new drug is a very long and expensive process. To overcome these difficulties scientists start to use a variety of computational methods to identify novel compounds, design compounds with increased selectivity, efficacy and safety and develop compounds into clinical trial candidates. One of these methods is the Quantitative Structure-Activity Relationship (QSAR) technique. The present work is related to the design of a new antimalarial active artemisinin analog.

Methods

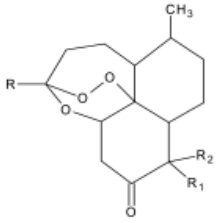
The physical properties of drugs dictate their biological activity. Use of descriptors of physicochemical properties allow for the application of mathematical models to analyze and predict drug activity. The physicochemical properties, which include parameters to account for atomic charges, topology, electronic properties and steric effects are determined by computational methods.

Results of calculation

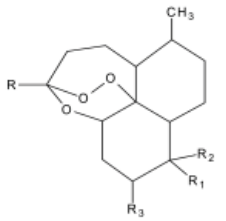
A. Preparation of the QSAR models

In this field we use ArgusLab software

Table 1. Artemisinin analogs.



Structure	R	R_1	R_2	Log RA
1	CH_3	CH_3	H	0.00
2	C_4H_8Ph	H	H	0.45
3	CH_3	H	Z-crotyl	-1.10
4	CH_3	H	H	0.79
5	CH_3	H	CH_3	-0.17
6	CH_3	H	E-crotyl	-0.60
7	CH_3	Allyl	H	-0.10
8	CH_3	C_4H_9	H	0.17
9	C_4H_8Ph	C_4H_9	H	-0.32
10	C_3H_6(*p*-ClPh)	C_4H_9	H	-0.28
11	$C_2H_4CO_2Et$	C_4H_9	H	1.36
12	C_4H_9	C_4H_9	H	-0.48
13	CH_3	C_2H_5	H	1.40
14	CH_3	C_6H_{13}	H	0.86
15	CH_3	*i*-C_4H_9	H	-0.55
16	CH_3	*i*-C_6H_{13}	H	-0.04
17	CH_3	*i*-C_3H_7	H	-0.04
18	CH_3	*i*-C_5H_{11}	H	0.07
19	C_3H_6(*p*-ClPh)	H	H	0.10
20	C_2H_4Ph	H	H	-2.00
21	C_4H_9	H	H	-0.74
22	$C_2H_4CO_2Et$	H	H	0.37
23	C_2H_5	H	H	0.05
24	H	H	H	-2.76
25	*i*- C_4H_9	H	H	-0.35
26	C_3H_7	H	H	0.83
27	CH_3	C_3H_6(*p*-ClPh)	H	1.37
28	CH_3	Br	CH_2Br	-.1.64
29	CH_3	OH	CH_2OH	-4.00
30	CH_3	β-OCH_2-	-	-2.09
31	CH_3	α-OCH_2-	-	-4.00
32	CH_3	$=CH_2$	-	-0.89
33	CH_3	-N=NC_2H_4-	-	-2.70
34	CH_3	$=CHCH_3$	-	-0.36
35	CH_3	-C_2H_4-	-	-0.94
36	CH_3	=O	-	-2.47
37	CH_3	C_5H_{11}	H	1.02
38	CH_3	C_4H_8Ph	H	0.63
39	CH_3	C_2H_4Ph	H	0.12
40	CH_3	C_3H_6Ph	H	0.78
41	CH_3	C_3H_7	H	1.13



Structure	R	R_1	R_2	R_3	Log RA
42	CH_3	CH_3	H	H	0.75
43	CH_3	CH_3	H	OH	0.55
44	CH_3	CH_3	H	Oet	0.34
45	CH_3	C_4H_9	H	OH	0.96
46	CH_3	H	H	H	0.28
47	CH_3	CH_3	Br	NH-2-(1,3-thriazole)	0.66
48	CH_3	CH_3	Br	m-F-aniline	0.79
49	CH_3	CH_3	Br	Aniline	0.18
50	CH_3	CH_3	Br	NH-2-pyridine	-0.09
51	CH_3	CH_3	Br	NH-2-pyrimidine	-0.77
52	CH_3	CH_3	H	OMe	0.28
53	CH_3	CH_3	H	α-OEt	0.32
54	CH_3	C_4H_9	H	H	1.32
55	CH_3	C_2H_5	H	H	0.67
56	CH_3	C_3H_7	H	OEt	-0.04
57	CH_3	H	H	OEt	0.43
58	CH_3	C_2H_5	H	OEt	0.50
59	CH_3	CH_3	H	C_3H_6OH	0.78
60	CH_3	CH_3	H	C_4H_9	0.06
61	CH_3	CH_3	H	OCH_2CO_2Et	0.52
62	CH_3	CH_3	H	$OC_2H_4CO_2Me$	0.10
63	CH_3	CH_3	H	OCH_2(p-$PhCO_2Me$)	-0.07
64	CH_3	CH_3	H	$OC_3H_6CO_2Me$	-0.03
65	CH_3	CH_3	H	(R)-$OCH_2CH(CH_3)CO_2Me$	1.79
66	CH_3	CH_3	H	(S)-$OCH_2CH(CH_3)CO_2Me$	2.25
67	CH_3	CH_3	H	(R)-$OCH(CH_3)CH_2CO_2Me$	0.87
68	CH_3	CH_3	H	(S)-$OCH(CH_3)CH_2CO_2Me$	1.70
69	CH_3	CH_3	H	OCH_2-adamantyl	0.28
70	$C_2H_4CO_2Et$	H	H	H	0.70
71	C_3H_6(*p*-ClPh)	H	H	H	-0.55
72	C_4H_9	H	H	H	0.75
73	C_2H_5	H	H	H	-1.00
74	*i*-C_4H_9	H	H	H	0.40
75	C_3H_7	H	H	H	0.84
76	C_4H_8Ph	H	H	H	0.58
77	C_2H_4Ph	H	H	H	-1.7
78	CH_3	β-OCH_2-	OOH	OOH	-0.62
79	CH_3	α-CH_2O-	OOH	OOH	-0.57
80	CH_3	$=CH_2$	OH	OH	-2.37
81	CH_3	CH_3	α-OH	α-OH	-0.89
82	CH_3	C_5H_{11}	H	H	0.16
83	CH_3	C_3H_6Ph	H	H	1.40
84	CH_3	C_3H_7	H	H	0.74
85	CH_3	CH_3	CH_2CHF	CH_2CHF_2	0.11
86	CH_3	CH_3	$CH_2CF_2CH_3$	$CH_2CF_2CH_3$	-0.17
87	CH_3	CH_3	OCH_2CF_3	OCH_2CF_3	0.29
88	CH_3	OH	OCH_2CF_3	OCH_2CF_3	-0.70
89	CH_3	CH_3	OEt	OEt	-0.44
90	CH_3	OH	OEt	OEt	-1.13
91	CH_3	CH_3	OOt-C_4H_9	OOt-C_4H_9	0.92

(www.arguslab.com) to build molecular models, display their structure and optimize geometry of artemisinin and its analogs (Table1). Data set of analogs was obtained from the published results (*Journal of Medical Chemistry,* Supporting info, 2001). Crystal structure of artemisinin from Cambridge Crystallographic Database (*Journal of Medical Chemistry*, Supporting info, 2001) was used as the template for creation of the analogs.

B. Calculation of the QSAR descriptors

Preparation of the QSAR models was carried out by the genetic function algorithm (GFA) method of Cerius[2] (Accelrys) software for 91 molecules listed in the Table 1. This method generates regression equations containing the pre-calculated descriptors and optimizes the combinations of descriptors using genetic crossing to maximize the relationship with the experimentally observed biological activities of the artemisinin derivatives.

In this approach, an initial set of 1000 equations was evolved throughout 50 000 iterations (crossing over) with the aim to establish a statistically significant regression model involving over 280 molecular and submolecular descriptors for the 91 molecular models of artemisinin and its derivatives (Fig. 1).

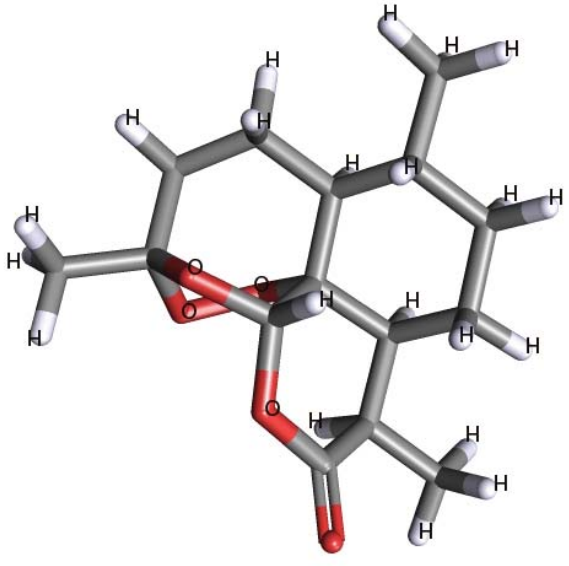


Figure 1. Molecular model of artemisinin.

From the preliminary results we could see that among the all used molecular descriptors the net atomic charges of carbons bearing substituents show some relationships to the observed antimalarial activities of the artemisinin derivatives. Therefore, we used Gaussian software to calculate atomic charges of the artemisinin and its analogs by a single point Hartree-Fock calculation in a double-zeta basis set (6-31G). Atomic charges were then transferred to Cerius[2] – software package of Accelrys, which is specifically devoted to QSAR analyses of large sets of molecules using the genetic function algorithm method.

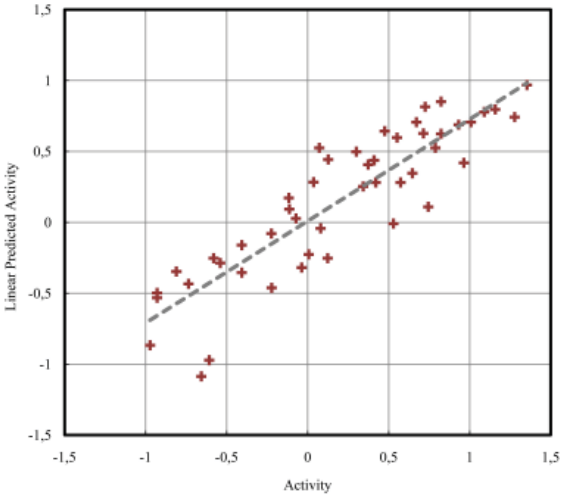


Figure 2. Linear predicted activity of artemisinin analogs.

The most statistically significant regression equation obtained from the final GFA runs relates the observed antimalarial activity of artemisinin derivatives to two molecular descriptors: Area and Zagreb.

$$\log R = \frac{IC_{50}[deriv]}{IC_{50}[artemis] \cdot M_w[deriv] / M_w[artemis]}$$

Area is the molecular van der Waals surface area, a spatial descriptor which describes the dimensions of the molecular surface exposed for binding with receptors. Zagreb is a 2D topological index (sum of squares of vertex valencies), which describes the complexity and extent of the molecular structure (size, branching, composition, flexibility, shape) in a simplified manner and reflects the presence and type of substituents in the studied artemisinin derivatives. We have obtained the following regression equation by removing a larger number of outliers from the original set of data containing 91 molecules:

$$\log R = 0.5159 + 0.0216 \cdot \text{Area} - 0.0618 \cdot \text{Zagreb} \quad (1)$$

($n = 52$, $R^2 = 0.768$, $R_{xv}^2 = 0.736$, F-test = 81.295, $\alpha > 95\%$)

Where: n - the number of molecules considered,

R^2 - squared correlation coefficient of the regression,

R_{xv}^2 - leave-one-out cross validated correlation coefficient,

F-test - Fisher test,

α - Statistical significance of the correlation.

The obtained regression is shown in Fig. 2. Based on the regression equation (1) it is in principle possible to predict the antimalarial activity of new artemisinin derivatives. The QSAR model can thus serve for designing new more active artemisinin derivatives.

Acknowledgements

We would like to express our sincere thanks to Prof. Stanislav Miertus, ICS AREA Pure and Applied Chemistry for his permanent orientation and advice that have helped us successfully finish this work.

References

Journal of Medical Chemistry, American Chemical Society, Supporting info. 2001. p. 1-2, 5-6.

Krishna, S., Uhlemann, A. & Haynes, R. K. 2004. Artemisinins: mechanisms of action and potential for resistance. *Drug Resistance Updates*, 7: 233-244.

Pinheiro, J. C., Kiralj, R. & Ferriera. M. M. C. 2003. Artemisinin derivatives with antimalarial activity against *Plasmodium falciparum* with the aid of quantum chemical and partial least squares methods. *QSAR and Combinatorial Science*, 22: 830-844.

Woodrow, C. J., Haynes, R. K. & Krishna, S. 2005. Artemisinins. *Postgraduate Medical Journal*, 81: 71-78.

Хураангуй

Бүтэц, идэвхийн хамаарлын судалгаа нь тодорхой молекулын орон зайн бүтэц болон түүний биологийн идэвхийн хоорондын хамаарлыг математик, компьютерийн арга, хэрэглэгдэхүүнүүдээр тооцоолоход үндэслэгддэг ба богино хугацаанд амьтад дээр турших, шинээр нийлэгжүүлэх үе шатуудыг алгасах замаар бага зардлаар шинэ бүтээгдэхүүнийг загварчлан гаргахад оршдог. Артемизинины бүтэц, идэвхийн хамаарлыг GFA (genetic function algorithm)-н аргаар артемизинины уламжлалт 91 молекулын хувьд судалж хумхаагийн эсрэг тодорхой хэмжээний идэвхи үзүүлсэн үр дүнг гаргасан бөгөөд хоёр уламжлалын хувьд статистикийн хувьд хамгийн магадлал өндөртэй регрессийн тэгшитгэлийг тооцоолон гаргав.

Received: 04 February 2008
Accepted: 23 March 2009