ЖУРНАЛ СТРУКТУРНОЙ ХИМИИ

2017. Том 58, № 7 *Сентябрь – октябрь* *С. 1457 – 1465*

UDC 544.169

QSAR STUDY BY THE RASMS METHOD OF DABO DERIVATIVES AS HIV-1 REVERSE TRANSCRIPTASE NON-NUCLEOSIDE INHIBITORS

J.-B. Tong, M. Bai, X. Zhao

College of Chemistry and Chemical Engineering, Shaanxi University of Science and Technology, Xi'an, P.R. China
E-mail: jianbotong@aliyun.com

Received September, 10, 2015

The random sampling analysis on molecular surface (RASMS) method is applied to study quantitative structure activity relationships (QSARs) modeled by multiple linear regression for dihydroalkoxybenzyloxopyrimidines (DABO) derivatives as HIV-1 reverse transcriptase non-nucleoside inhibitors. The correlation coefficients (R^2_{cum}) and cross-validation correlation coefficients (Q^2_{CV}) obtained by the models are 0.852 and 0.755 for 74 DABO derivatives. Satisfactory results show that information about the biological activity can be systemically expressed by the RASMS method that can be a useful structural expression methodology for the study of QSAR of the DABO derivatives.

DOI: 10.26902/JSC20170720

K e y w o r d s: RASMS, dihydroalkoxybenzyloxopyrimidines derivatives, quantitative structure activity relationship (QSAR), MLR.

INTRODUCTION

Human immunodeficiency virus type-1 (HIV-1), the main etiological agent for the transmission of the acquired immunodeficiency syndrome (AIDS) is a retrovirus of the lentivirus family and contains a reverse transcriptase (RT) enzyme that makes a DNA copy of the viral RNA template [1, 2]. Looking at the anti-HIV drugs in the market, one can see that the most FDA approved anti-HIV drugs belong to one of the two classes of HIV-1 RT inhibitors: nucleoside reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs). AZT, 3TC, ddI, and ddC are typical examples of NRTIs and nevirapine, delaviridine, and efavirenz are NNRTIs approved by FDA. The non-nucleoside HIV-1 RT inhibitors under investigation are structurally different entities: dihydroalkoxybenzyloxopyrimidines (DABO), phenylethylthiazolylthiourea (PETT), tetrahydro-imidazo[4,5,1-jk] [1,4]-benzodiazepin-2(1H)-one and -thione (TIBO), 1-(2-hydroxyethoxymethyl)-6-(phenylthio)-thymine (HEPT), diarylpyrimidines (DAPY) and dipyridodiazepinone (with nevirapine in the market) derivatives [3, 4—11]. NNRTIs as one of the two kinds of inhibitors against the HIV-1 RT have attracted more attention due to their high specificity and low toxicity [12]. However, the rapid emergence of resistant HIV viral strains carrying mutation at residues surrounding the NNRTI binding pocket limits the usefulness of NNRTIs. Thus, the design and development of new and more potent mutation-resistant inhibitors is still an arduous task for the treatment of the HIV-1 infected patients [13, 14].

There is a large number of literature reports on the application of computational methods for describing the activity of biologically active compounds [15, 16—19]. Quantitative structure activity

relationship (QSAR) studies are the most extensively used methods in computational chemistry. An appropriate representation of the structural and physicochemical features of chemical agents is an essential key to the successful application of QSAR models [20, 21—24]. QSAR studies play a fundamental role in predicting the biological activity of new compounds and identifying ligand-receptor interactions. The first step in constructing the QSAR models is to find one or more molecular descriptors representing variation in the structural property of molecules by a number. Structural descriptors have been classified into different categories according to different approaches, including physiochemical, constitutional, geometrical, topological, and quantum chemical descriptors [25].

Based on the previous work, a 3D QSAR method, namely the random sampling analysis on molecular surface (RASMS) was proposed. The RASMS method is derived from atomic probe of protein (APP), pseudo-receptor accessible surface (PRAS), and approach to the aimed area by random sampling on molecular surface (ARSMS) and was used to express drug structures and biological activities of 74 DABO derivatives using the multiple linear regression (MLR) model. The proposed method was evaluated by predicting the activities of 74 DABO derivatives in this paper. The results indicate that it is a useful tool for the investigation of drug QSAR.

PRINCIPLE AND METHODOLOGY

Probe atoms. Proteins and active peptides have often been used as drug targets. There are eight different types of hybrid atoms from amino acids serving as probes. To characterize these probe atoms, the mean charge index (MCI), the van der Waals index (VWI), and the mean hydrophobic index (MHI) are used (Supplementary Table 1).

MCI. The average electrical charge of each atom in amino acid serves as MCI. Original molecular structures of twenty natural amino acids are primarily auto-produced using the HyperChem7.5 (Hypercube, Inc. Gainesville, FL 32601 USA) database and then refined by molecular dynamics. The obtained structures are further optimized using Gaussian 98W (Gaussian, Inc., Pittsburgh, PA, 1998) at the Hartree-Fock level. The analysis of oscillation frequencies of the obtained structures demonstrated that there was no virtual frequency. Ultimately, the amount of the net charge of all the atoms is calculated using the single point method within density functional theory (DFT).

VWI. Usually, van der Waals radii are the radii of isolated atoms from references. However, the hybrid state of actual atoms changes in different chemical microenvironments, so the van der Waals radius changes accordingly. In this experiment, the calibrated van der Waals radius was used as a probe atomic radius (i.e. VWI = $C_h \times R^*_{\mathrm{VDW}}$, with the calibration factor C_h of 1.00 in the case of sp^3 hybridization, 0.95 in the case of sp^2 hybridization, and 0.90 in the case of *sp* hybridization) [26]. Moreover, the standard van der Waals radii of all kinds of atoms were from the report of Bondi e*t al*. [27].

MHI. Similar to MCI, MHI was obtained from the average hydrophobic interaction of each probe atom from the natural amino acid. The atomic solvation parameter (ASP) defined by Pei *et al.* [28, 29] serves as a hydrophobic measurement.

PRAS. The concept of PRAS is proposed in this study. If atoms in biomolecules such as proteins, nucleic acids, and sugars, which were used as drug targets, reach the surface of the drug molecules, then the surface is defined as the pseudo-receptor accessible surface of molecule (PRASM). If the hydrogen atom in the eight probe atoms is the receptor probe, when a single hydrogen atom rolls on the van der Waals surface of the drug molecule, the curved surface to the center of which the hydrogen atom goes is defined as the hydrogen-pseudo-receptor accessible surface of molecule (H-PRASM). Similarly, the other seven kinds of pseudo-receptor probe atoms and their accessible surfaces can be calculated (Fig. 1). According to the above calculation method of PRASM, the isolated pseudo-receptor accessible surfaces of atoms (PRASA) can be defined. Obviously, PRASA is a spherical surface whose radius is the sum of the radii of drug atoms plus the radii of probe atoms (Fig. 2). As we can see, some parts of the PRASA of each drug atom may be involved in the formation of the PRASM of the drug molecule.

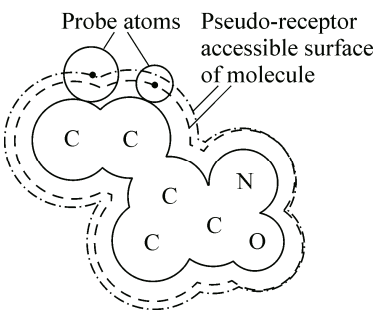


Fig. 1. Pseudo-receptor accessible surface of molecule

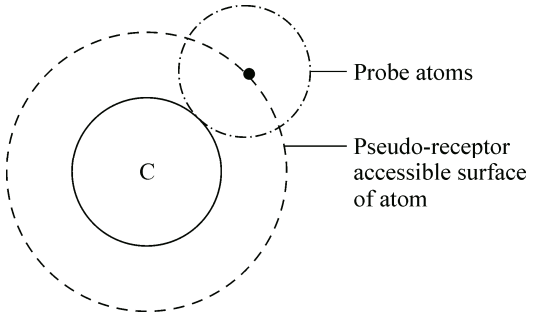


Fig. 2. Pseudo-receptor accessible surface of atom

Atomic types and interactions. The RASMS method was developed with three common non-bonding interactions of the biological activities, that is, electrostatic interaction, steric interaction, and hydrophobic interaction related with the atomic relative distance and atomic self-properties.

Electrostatic interaction. The electrostatic interaction field is an important non-bonded interaction expressed by the classical Coulomb theorem

$$E_p(E) = \sum_{i=1}^{n} \frac{e^2}{4\pi\varepsilon_0} \cdot \frac{\mathrm{MCI}_p \cdot Z_i}{r_{pi}} \qquad (1 \le p \le 8, \quad 1 \le i \le 10). \tag{1}$$

In the previous equation, n is the number of atoms in a drug molecule; r_{pi} is the Euclid distance from the probe to the i atom; e is the unit electrostatic charge of 1.6021892×10^{-19} C; ε_0 is the dielectric constant in vacuum with a value of 8.85418782×10^{-12} $C^2/J\cdot m$; Z is the atomic net charge. The electrostatic interaction components are calculated by this formula.

Steric interaction. The steric interaction described by the Lennard-Jones equation [30—32] is defined here as the interaction between the dipole and non-dipole field or the induced dipole interaction

$$E_p(S) = \sum_{i=1}^{n} \varepsilon_{pi} \left[\left(\frac{R_{pi}^*}{r_{pi}} \right)^{12} - 2 \cdot \left(\frac{R_{pi}^*}{r_{pi}} \right)^{6} \right] \qquad (1 \le p \le 8, \quad 1 \le i \le 10). \tag{2}$$

Here $\varepsilon_{pi} = (\varepsilon_{pp} \cdot \varepsilon_{ii})^{1/2}$ is the potential well of the probe atom and the receptor atom; $R_{pi}^* = (\mathrm{VWI}_p + C_h \cdot R_i^*)/2$ is the van der Waals radius with its calibration factor of 1.00 in the sp^3 hybridization state, 0.95 in sp^2 hybridization, and 0.90 in sp hybridization. R_{pi}^* is the calibration collision van der Waals radius between the probe and receptor atoms. Since the Lennard-Jones equation is extremely sensitive to changes in the distance, the lattice points close to atoms of compounds may lead to very large steric interactions (Supplementary Table 2).

Hydrophobic interaction. The hydrophobic interaction notably affects the binding interactions of drug molecules. Due to the entropy of systematic changes, such an interaction is difficult to be described. The HINT method is used here to express the hydrophobic interaction field. The formula of interatomic hydrophobic interactions in HINT is as follows:

$$E_p(H) = \sum_{i=1}^{n} S_p \cdot \mathrm{MCI}_p \cdot S_i \cdot a_i \cdot e^{-r_{pi}} \cdot T_{pi} \qquad (1 \le p \le 8, \quad 1 \le i \le 10). \tag{3}$$

In Eq. (3), S is the solvent accessible surface area of atom (SASA) (Supplementary Table 3); a is the hydrophobic constant expressed by ASP (Supplementary Table 4); T is the sign function, indicating changes in the entropy resulting from different types of atomic interactions [33—35] (Supplementary Table 5).

Implementation process of RASMS. The atoms of organic molecules include H, C, N, P, O, S, F, Cl, I, and Br which belong to IA, IVA, VA, VIA, and VIIA in the Periodic Table of Elements.

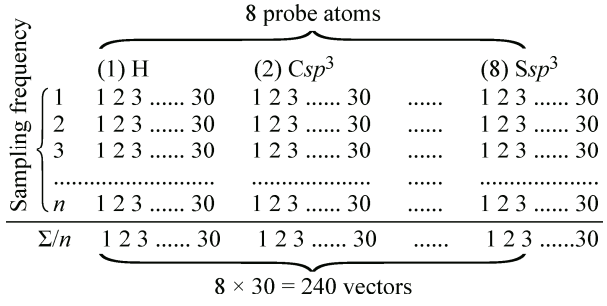


Fig. 3. The 240 types of interactions of the drug molecules with the RASMS method

Based on the point that the atoms of similar chemical properties belong to the same category and according to the hybridization states of the atoms, the atoms are furthermore subdivided into ten types (Supplementary Table 6) for a better expression of the microscopic structural features of the molecules. In this paper, the electrostatic, steric, and hydrophobic potential energies were involved in the formation of 240 interaction terms: 8×10×3 = 240 interaction items for organic compounds (Fig. 3).

Algorithm. Based on APP, PRAS, and ARSMS, 240 descriptors were produced with a self-made descriptor calculation software *Sampling-tool.EXE*, an applied program written in *C* language by the staff in the laboratory. The *Sampling-tool.EXE* was used to generate 240 descriptors for each molecule. The Cartesian coordinates and the Mulliken charges of the atoms need to be input into *Sampling-tool.EXE* after the geometry optimization when using the program. Then set the molecular surface sampling density and select the probe type, respectively.

Software used. Chemoffice 2004 (version 8.0.3, Cambridgesoft---CambridgeSoft Corporation CambridgePark Drive Cambridge, MA 02140 USA)) was employed to generate the structures of the drug molecules. Gaussian 98W (Gaussian, Inc., Pittsburgh, PA, 1998) was used to optimize the structures of these molecules. Stepwise multiple regression (SMR) was implemented using the SPSS 16.0 (IBM (New Orchard Road, Armonk, New York) 2007) statistical software. The application assembly (version 2.2.3, ICP/HN, 1997) software was employed to perform the MLR analysis.

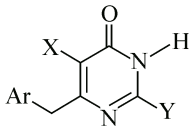
RESULTS AND DISCUSSION

Data selection and partition. The structures and experimental data of the 74 DABO derivatives (Table 1) were obtained from the literature [36]. IC_{50} is the drug concentration inhibiting 50 % of cellular growth followed by 1 h of drug exposure, and pIC_{50} (M) is used here for the calculation pIC_{50} = = $-\log IC_{50} = \log l/IC_{50}$.

MLR modeling and analysis. It is important to examine the estimation ability and prediction power of a QSAR model. In recent years, the statistical parameter correlation coefficient (Q^2_{CV}) and the leave-one-out cross-validation (LOOCV) coefficient have been used as means of indicating the predictive ability of a model. Generally, many researchers consider a high Q^2_{CV} value as an indicator or even as the ultimate proof of the high predictive power of a QSAR model. However, the recent study of Tropsha and his co-workers shows there is no evident relationship between the Q^2_{CV} value and the actual predictive power of a QSAR model, therefore, an external validation is required. Now, a novel method to further refine the predictive ability of the developed QSAR models was introduced by Roy *et al.* [37, 38—40] with an alternative group of metrics, r^2_m metrics, for the determination of the proximity between the observed and predicted activity. The r^2_m metrics are calculated based on the correlation of the observed and predicted response data with and without the intercept and also by interchanging the axes. Squared correlation coefficient values between the observed (Y axis) and predicted (X axis) values of the compounds with intercept (r^2) and without intercept (r^2_0) are calculated for the determination of r^2_m. A change in the axes gives the r'^2_0 value; the r'^2_m metric is calculated based on the r'^2_0 value. The k and k' parameters indicate the slopes in the former and later cases respectively. Presently, two different variants of this parameter, r^2_m and Δr^2_m, are calculated for both training (internal validation) and test (external validation) sets in addition to the total dataset (overall validation).

Table 1

Structures and pIC_{50} *activity of the* 74 DABO *derivatives*



No.	X	Ar	Y	pIC_{50}(M)	No.	X	Ar	Y	pIC_{50}(M)
1	Me	2-naphtyl	S-*sec*-Bu	4.23	38	H	2,6-di-F-Ph	S-Me	6.10
2	H	1-naphtyl	S-cyclopentyl	4.31	39	Me	2-Cl-Ph	S-*sec*-Bu	6.10
3	Me	1-naphtyl	S-cyclopentyl	4.35	40	Me	2-F-Ph	S-*sec*-Bu	6.10
4	Me	4-F-Ph	S-*sec*-Bu	4.59	41	Me	3-NO2-Ph	S-*sec*-Bu	6.10
5	Me	4-Cl-Ph	S-*sec*-Bu	4.77	42	H	2-F-Ph	S-*sec*-Bu	6.22
6	H	1-naphtyl	S-*sec*-Bu	4.79	43	H	3-NO_2-Ph	S-*sec*-Bu	6.22
7	H	2-naphtyl	S-*sec*-Bu	4.83	44	H	2,6-di-Cl-Ph	S-*tert*-Bu	6.22
8	H	4-F-Ph	S-*sec*-Bu	4.83	45	H	2,6-di-Cl-Ph	S-*n*-Bu	6.30
9	H	4-Cl-Ph	S-*sec*-Bu	5.02	46	H	2,6-di-Cl-Ph	S-cyclopentyl	6.40
10	H	Ph	S-*tert*-Bu	5.07	47	H	2,6-di-F-Ph	S-*n*-Bu	6.70
11	H	3-Me-Ph	S-*tert*-Bu	5.09	48	H	2,6-di-F-Ph	S-*tert*-Bu	6.70
12	Me	3-Me-Ph	S-*sec*-Bu	5.27	49	H	2,6-di-Cl-Ph	S-*sec*-Bu	6.70
13	Me	2,6-di-Cl-Ph	S-cyclohexyl	5.31	50	Me	2,6-di-Cl-Ph	S-*sec*-Bu	6.92
14	Me	Ph	S-Me	5.31	51	H	2,6-di-F-Ph	S-*sec*-Bu	7.00
15	Me	Ph	S-*sec*-Bu	5.32	52	Me	2,6-di-F-Ph	S-*sec*-Bu	7.00
16	Me	3-Me-Ph	S-*tert*-Bu	5.34	53	H	2,6-di-F-Ph	S-cyclohexyl	7.05
17	Me	Ph	S-cyclohexyl	5.37	54	Me	2,6-di-F-Ph	S-*tert*-Bu	7.05
18	H	3-Cl-Ph	S-*sec*-Bu	5.42	55	H	2,6-di-F-Ph	S-cyclopentyl	7.10
19	Me	4-NO_2-Ph	S-*sec*-Bu	5.44	56	Me	2,6-di-F-Ph	S-cyclopentyl	7.10
20	Me	3-Me-Ph	S-cyclopentyl	5.47	57	H	2,6-di-F-Ph	NH-cyclopentyl	7.15
21	H	2-Cl-Ph	S-*sec*-Bu	5.49	58	H	2,6-di-F-Ph	S-*iso*-Pr	7.30
22	Me	3-F-Ph	S-*sec*-Bu	5.52	59	Me	2,6-di-F-Ph	NH-cyclopentyl	7.52
23	H	2,6-di-Cl-Ph	S-Me	5.52	60[a]	Me	1-naphtyl	S-*sec*-Bu	4.35
24	H	Ph	S-cyclohexyl	5.52	61[a]	H	2-naphtyl	S-cyclohexyl	4.48
25	H	3-Me-Ph	S-*iso*-Pr	5.54	62[a]	H	Ph	S-*sec*-Bu	5.27
26	H	Ph	S-cyclopentyl	5.55	63[a]	Me	Ph	S-cyclopentyl	5.47
27	H	3-Me-Ph	S-cyclohexyl	5.59	64[a]	H	3-Me-Ph	S-cyclopentyl	5.59
28	Me	3-Me-Ph	S-Me	5.60	65[a]	Me	Ph	S-*iso*-Pr	5.60
29	Me	3-Me-Ph	S-*iso*-Pr	5.60	66[a]	H	3-Me-Ph	S-*sec*-Bu	5.62
30	H	4-NO_2-Ph	S-*sec*-Bu	5.62	67[a]	Me	3-Cl-Ph	S-*sec*-Bu	5.74
31	Me	3-Me-Ph	S-cyclohexyl	5.66	68[a]	H	3-F-Ph	S-*sec*-Bu	5.92
32	Me	Ph	S-*tert*-Bu	5.72	69[a]	H	2-NO_2-Ph	S-*sec*-Bu	6.22
33	Me	2,6-di-Cl-Ph	S-cyclopentyl	5.80	70[a]	H	2,6-di-Cl-Ph	S-cyclohexyl	6.40
34	H	2,6-di-Cl-Ph	S-*iso*-Pr	5.89	71[a]	Me	2,6-di-F-Ph	S-Me	6.70
35	Me	2,6-di-Cl-Ph	S-*iso*-Pr	5.94	72[a]	Me	2,6-di-F-Ph	S-*n*-Bu	7.05
36	Me	2,6-di-Cl-Ph	S-*n*-Bu	5.94	73[a]	Me	2,6-di-F-Ph	S-cyclohexyl	7.15
37	Me	2,6-di-Cl-Ph	S-*tert*-Bu	5.96	74[a]	Me	2,6-di-F-Ph	S-*iso*-Pr	7.30

[a] Chosen as the testing set.

The r^2, r_m^2, r'^2_m, r_0^2, r'^2_0, k, and k' values are calculated as the following equations:

$$r^2 = [\sum (Y_{\text{obs}} - \overline{Y_{\text{obs}}})(Y_{\text{pred}} - \overline{Y_{\text{pred}}})]^2 / [\sum (Y_{\text{pred}} - \overline{Y_{\text{pred}}})^2 \times \sum (Y_{\text{obs}} - \overline{Y_{\text{obs}}})^2] = \\ = [\sum (Y_{\text{pred}} - \overline{Y_{\text{pred}}})(Y_{\text{obs}} - \overline{Y_{\text{obs}}})]^2 / [\sum (Y_{\text{obs}} - \overline{Y_{\text{obs}}})^2 \times \sum (Y_{\text{pred}} - \overline{Y_{\text{pred}}})^2], \quad (4)$$

$$r_m^2 = r^2 \times (1 - \sqrt{r^2 - r_0^2}), \quad (5)$$

$$r'^2_m = r^2 \times (1 - \sqrt{r^2 - r'^2_0}), \quad (6)$$

$$r_0^2 = 1 - [\sum (Y_{\text{obs}} - k \times Y_{\text{pred}})^2 / \sum (Y_{\text{obs}} - \overline{Y_{\text{obs}}})^2], \quad (7)$$

$$r'^2_0 = 1 - [\sum (Y_{\text{pred}} - k' \times Y_{\text{obs}})^2 / \sum (Y_{\text{pred}} - \overline{Y_{\text{pred}}})^2], \quad (8)$$

$$k = \sum (Y_{\text{obs}} \times \overline{Y_{\text{pred}}}) / \sum (Y_{\text{pred}})^2, \quad (9)$$

$$k' = \sum (Y_{\text{obs}} \times \overline{Y_{\text{pred}}}) / \sum (Y_{\text{obs}})^2. \quad (10)$$

Here, r^2 and r'^2_0 are the squared correlation coefficients between the observed and predicted activity data, Y_{obs} and Y_{pred} are the observed and predicted response data while $\overline{Y_{\text{obs}}}$, and $\overline{Y_{\text{pred}}}$, refer to the mean values of the observed and predicted responses respectively. We may also note here the related tests for the QSAR model validation suggested by Tropsha, stating the following criteria for models to be considered acceptable:

$$r^2 > 0.6, \quad (11)$$

$$(r^2 - r_0^2)/r^2 < 0.1 \text{ or } (r^2 - r'^2_0)/r^2 < 0.1, \quad (12)$$

$$0.85 \le k \le 1.15 \text{ or } 0.85 \le k' \le 1.15, \quad (13)$$

$$|r_0^2 - r'^2_0| < 0.3. \quad (14)$$

In order to prove the validity and stability of the model, the whole dataset is systematically divided into two subsets. One was predicted using another subset as training. Samples 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, and 74 were chosen as the testing set and the others as the training set. Molecular steric structures of 74 DABO derivatives were firstly auto-constructed using *Chemoffice* 8.0 and then optimized at the AM1 level with MOPAC, the semi-empirical quantum chemistry software in *Chem3D*. Then the net electric charges of the atoms were calculated in the single point form by the Mulliken methods. After the aforementioned two items were input respectively into forms of the Cartesian coordinates and the net electric charge, 240 descriptors were produced using *Sampling-tool.EXE*. The ultimate vectors for 59 training compounds involve 240 items. The RASMS method may lead to some information overlap among these different descriptors. To solve the above problems, two approaches were adopted: SMR in the *SPSS* 16.0 software was employed to select the variables; MLR was applied to construct the model according to the values of Fisher prominent test with SMR. Twenty significant variables were selected from 240 items using the *SPSS* 16.0 software. The original variable matrix obtained by SMR was then subjected to MLR modeling and the optimal model was determined when the cross-validation correlative coefficient (Q_{CV}^2) in LOO-CV achieved the maximum value. Consequently, the MLR model was constructed for the training set with its fitting correlative coefficient R_{cum}^2=0.852 and cross-validation coefficient Q_{CV}^2 = 0.755. When 14 parameters were employed in the model, Q_{CV}^2 reached its maximum. The 14th stepwise regression equation is shown below

$$\text{pIC}_{50} = 12.150(+/-2.656) + 2.927(+/-1.234)E_{1-3} + 2.321(+/-1.018)S_{1-2} - 278.246(+/-63.303)E_{1-10} - \\ - 313.045(+/-72.510)E_{2-10} + 0.090(+/-0.188)H_{4-8} + 4.700(+/-1.817)H_{7-8} - 2.924(+/-1.589)H_{5-8} + \\ + 7.854(+/-2.626)E_{1-5} + 0.842(+/-0.412)S_{1-10} - 0.596(+/-0.271)S_{4-2} + 1.551(+/-1.358)E_{2-1} - \\ - 4.364(+/-1.551)S_{8-3} + 7.923(+/-2.986)S_{2-3} + 1.714(+/-0.860)H_{3-3} \quad (15)$$

Table 2

Comparison of various kinds of parameters of the first 15 *steps of* QSAR *models for* 74 DABO *derivatives*

No. variables	Internal parameters			External parameters			Tropsha parameters		Overall parameters
	r^2	Q^2_{CV}	$r^2_{m\,(LOO)}$	r^2	r^2_0	r^2_m	k	k'	r^2_m
1	0.620	0.599	0.589	0.904	0.891	0.800	1.036	0.963	0.650
2	0.657	0.623	0.608	0.929	0.925	0.869	1.039	0.961	0.673
3	0.685	0.639	0.620	0.889	0.889	0.880	1.030	0.969	0.673
4	0.720	0.642	0.615	0.811	0.803	0.739	1.021	0.975	0.640
5	0.738	0.656	0.618	0.833	0.833	0.829	1.025	0.972	0.658
6	0.768	0.697	0.667	0.841	0.836	0.780	1.017	0.980	0.691
7	0.780	0.704	0.672	0.824	0.820	0.773	1.025	0.972	0.689
8	0.789	0.708	0.676	0.852	0.851	0.821	1.026	0.971	0.702
9	0.804	0.721	0.687	0.884	0.883	0.853	1.028	0.971	0.716
10	0.813	0.729	0.696	0.860	0.859	0.831	1.036	0.963	0.713
11	0.822	0.731	0.692	0.842	0.839	0.799	1.036	0.962	0.731
12	0.826	0.730	0.690	0.841	0.839	0.802	1.034	0.964	0.703
13	0.839	0.743	0.700	0.808	0.801	0.739	1.036	0.962	0.696
14	0.852	0.755	0.711	0.793	0.789	0.748	1.036	0.961	0.705
15	0.858	0.753	0.708	0.768	0.768	0.751	1.040	0.957	0.698

In Eq. (15), *E*, *S*, and *H* represent the electrostatic interaction, the steric interaction, and the hydrophobic interaction respectively. $E_{1\text{-}3}$ represents the electrostatic interactions of the first kind of probes and the third kind of drug atoms, $S_{1\text{-}2}$ represents the steric interactions of the first kind of probes and the second kind of drug atoms, $H_{4\text{-}8}$ represents the hydrophobic interactions of the fourth kind of probes and the eighth kind of drug atoms, and so on. The first 15 stepwise regression results of SMR with R^2_{cum} and Q^2_{CV} of the MLR model of 74 DABO derivatives are shown in Table 2. Some parameters such as r^2_m, r^2_0, and k were calculated at the web http://aptsoftware.co.in/rmsquare/ and http://203.200.173.43:8080/rmsquare/. The input data required that the observed and predicted response values were either imported from a csv file (saved in *.csv format) or given manually for the calculation. The output data provide the values of $\overline{r^2_m}$ and Δr^2_m metrics for the respective set of compounds. The calculation of r^2_m metrics involves the determination of r^2, r^2_0, and k parameters together with information of the intercept of the regression line correlating the observed and predicted activity data. The results of $r^2_{m(LOO)}$, $r^2_{(test)}$, $r^2_{0(test)}$, $r^2_{m(test)}$, k, k', and $r^2_{m(all)}$ are 0.711, 0.793, 0.789, 0.748, 1.036, 0.961, and 0.705 respectively. Therefore, it is confirmed that the RASMS QSAR models are stable and generalized. Fig. 4 presents a plot of the observed values versus the calculated ones. It is shown that the results are relatively close to the predicted values.

CONCLUSIONS

In this paper, a RASMS method derived from APP, PRAS and ARSMS was proposed. 240 descriptors related to the method were generated by the calculation using the *Sampling-tool.EXE* soft ware. The obtained models, including classic electrostatic, steric, and hydrophobic interactions, have favorable stability and good predictive ability. It illustrates that the RASMS method is an effective description methodology for the characterization of complex interactions of drug molecules. It is suggested that the RASMS method behaves quite well in the represent-tation of both molecular structures and biological activities of DABO derivatives. Thus, it is suggested that the RASMS method has multiple advantages, such as plentiful structural information, good structural characterization ability, and

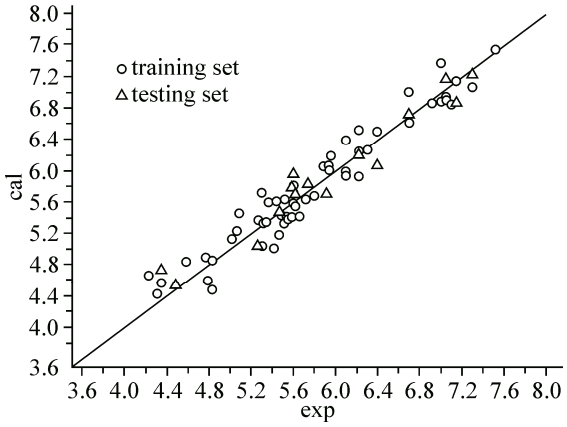


Fig. 4. The results of SMR with R^2_{cum} and Q^2_{CV} of the DABO derivatives with the RASMS method

the easy obtainment of the descriptors. It can be anticipated that the approach might hold a high potential to become a useful tool in the research of QSAR of DABO derivatives.

We gratefully acknowledge supports of this research by the National Natural Science Funds of China (21475081).

REFERENCES

1. *Fauci A.S.* // Science. – 1998. – **239**. – P. 617.
2. *Clercq E.D.* // J. Med. Chem. – 1995. – **38**. – P. 2491.
3. *Ahgren C., Backro K., Bell F.W., Cantrell A.S., Clemens M., Colacino J.M., Deeter J.B., Engelhardt J.A., Hogberg M., Jaskunas S.R., Johansson N.G.* // Antimicrob. Agents Chemother. – 1995. – **39**. – P. 1329.
4. *Pauwels R., Andries K., Desmyter J., Schols D., Kukla M.J., Breslin H.J., Raeymaeckers A., Gelder J.V., Woestenborghs R., Heykants J.* // Nature. – 1990. – **343**. – P. 470.
5. *Baba M., Tanakas H., Clercq E.D., Pauwels R., Balzarini J., Schols D., Nakashima H., Perno C.F., Walker R.T., Miyasaka T.* // Biochem. Biophys. Res. Commun. – 1989. – **165**. – P. 1375.
6. *Kleim J.P., Bender R., Billhardt U.M., Meichsner C., Riess G., Rosner M., Winkler I., Paessens A.* // Antimicrob. Agents Chemother. – 1993. – **37**. – P. 1659.
7. *Pauwels R., Andries K., Debyser Z., Daele P.V., Schols D., Vandamme A.M., Janssen C.G.M., Anne J., Desmyter G.J., Heykants J., Janssen M.A.C., Clercq E.D., Janssen P.A.J.* // Proc. Natl. Acad. Sci. U.S.A. – 1993. – **90**. – P. 1711.
8. *Balzarini J., Perez M.J., Felix A.S., Schols D., Perno D.C.F., Vandamme A., Camarasa M.J., Clercq E.D.* // Proc. Natl. Acad. Sci. U.S.A. – 1992. – **89**. – P. 4392.
9. *Romero D.L., Busso M., Tan C.K., Reusser F., Palmer J.R., Poppe S.M., Aristoff P.F., Downey K.M., Resnick A.G.L., Tarpley W.G.* // Proc. Nat. Acad. Sci. – 1991. – **88**. – P. 8806.
10. *Merluzzi V.J., Hargrave K.D., Labadia M., Grozinger K., Skoog M., Wu J.C., Shih C.K., Eckner K., Hattox S., Adams J.* // Science. – 1990. – **250**. – P. 1411.
11. *Razieh S., Afshin F., Behzad M.* // J. Mol. Graphics Modell. – 2009. – **28**. – P. 146.
12. *Sarafianos S.G., Das K., Hughes S.H., Arnold E.* // Curr. Opin. Struct. Biol. – 2004. – **14**. – P. 716.
13. *Das K., Lewi P.J., Hughes S.H., Arnold E.* // Prog. Biophys. Mol. Biol. – 2005. – **88**. – P. 209.
14. *Liang Y.H., Chen F.E.* // Eur. J. Med. Chem. – 2009. – **44**. – P. 625.
15. *Hemmateenejad B., Miri R., Akhond M., Shamsipur M.* // Chemom. Intell. Lab. Syst. – 2002. – **64**. – P. 91.
16. *Hemmateenejad B., Miri R., Akhond M., Shamsipur M.* // Arch. Pharm. Med. Chem. – 2002. – **10**. – P. 472.
17. *Hansch C., Hoekman D., Gao H.* // Chem. Rev. – 1996. – **96**. – P. 1045.
18. *Fassihi A., Sabet R.* // Int. J. Mol. Sci. – 2008. – **9**. – P. 1876.
19. *Fassihi A., Abedi D., Saghaie L., Sabet R., Fazeli H., Bostaki G.H., Deilami O., Sadinpour H.* // Eur. J. Med. Chem. – 2009. – **44**. – P. 2145.
20. *Nandy A., Kar S., Roy K.* // SAR QSAR Environ. Res. – 2013. – **44**. – P. 1062.
21. *Vyas B., Silakari O., Singh B.M., Singh B.* // SAR QSAR Environ. Res. – 2013. – **24**. – P. 733.
22. *Kar R.K., Suryadevara P., Sahoo B.R., Sahoo G.C., Dikhit M.R., Das P.* // SAR QSAR Environ. Res. – 2013. – **24**. – P. 215.
23. *Hemmateenejad B., Sanchooli M.* // J. Chemom. – 2007. – **21**. – P. 96.
24. *Kustrin S.A., Tucker I.G., Zecevic M., Ziva-novic L.J.* // Anal. Chem. Acta. – 2000. – **418**. – P. 181.

25. *Tong J.B., Che T., Liu S.L., Li Y.F., Wang P., Xu X.M., Chen Y.* // Arch. Pharm. Chem. Life Sci. – 2011. – **344**. – P. 719.
26. *Hahn M.* // J. Med. Chem. – 1995. – **38**. – P. 2080.
27. *Bondi A.* // J. Phys. Chem. – 1964. – **68**. – P. 441.
28. *Pei J.F., Wang Q., Zhou J.J., Lai L.H.* // Proteins. – 2004. – **57**. – P. 651.
29. *Zhou P., Tong J.B., Tian F.F., Li Z.L.* // Chin. Sci. Bull. – 2006. – **51**. – P. 1824.
30. *Xu L.N., Liang G.Z., Mei H., Zheng H.L., Zhou P., Wang J., Li Z.L.* // Asian J. Ecotoxicol. – 2008. – **1**. – P. 72.
31. *Tong J.B., Che T., Li Y.F., Wang P., Xu X.M., Chen Y.* // SAR QSAR Environ. Res. – 2011. – **22**. – P. 611.
32. *Tong J.B., Chen Y., Liu S.L., Xu X.M.* // Med. Chem. Res. – 2013. – **12**. – P. 4946.
33. *Tong J.B., Chen Y., Liu S.L., Che T., Xu X.M.* // J. Chemom. – 2012. – **26**. – P. 549.
34. *Sofie V.D., Patrick B.* // J. Mol. Struct: THEOCHEM. – 2010. – **943**. – P. 83.
35. *Tong J.B., Liu S.L., Zhou P., Wu B.L., Li Z.L.* // J. Theor. Biol. – 2008. – **253**. – P. 90.
36. *Monique A.B., Rodrigues C.R., Cirino J.J.V., Alencastro R.B., Castro H.C., Albuquerque M.G.* // J. Chem. Inf. Model. – 2008. – **48**. – P. 1706.
37. *Roy K., Mitra I.* // Comb. Chem. High Throughput Screening. – 2011. – **14**. – P. 450.
38. *Roy K., Mitra I., Kar S., Ojha P.K., Das R.N., Kabir H.* // J. Chem. Inf. Modell. – 2012. – **52**. – P. 396.
39. *Roy K., Chakraborty P., Mitra I., Ojha P.K., Kar S., Das R.N.* // J. Comput. Chem. – 2013. – **34**. – P. 1071.
40. *Mitra I., Saha A., Roy K.* // Mol. Simul. – 2010. – **13**. – P. 1067.