ЖУРНАЛ СТРУКТУРНОЙ ХИМИИ

*2015. Том 56, № 5* *Сентябрь – октябрь* *С. 914 – 921*

UDC 541.6:548.737

# QSAR STUDIES OF HEPT DERIVATIVES AS ANTI-HIV DRUGS USING THE RASMS METHOD

**J. Tong[1,2], X. Zhao[1,2], L. Zhong[1,2], J. Chang[1,2]**

[1]*College of Chemistry and Chemical Engineering, Shaanxi University of Science and Technology, Xi'an, P. R. China*

E-mail: jianbotong@sust.edu.cn

[2]*Key Laboratory of Auxiliary Chemistry and Technology for Chemical Industry, Ministry of Education, Shaanxi University of Science and Technology, Xi'an, P. R. China*

*Received January, 6, 2014*

Random sampling analysis on the molecular surface (RASMS) is used to describe the chemical structures of 35 HEPT derivatives as anti-HIV drugs. Here a quantitative structure activity relationship (QSAR) model is built by multiple linear regression (MLR). The estimation stability and prediction ability of the model are strictly analyzed by both internal and external validations. The correlation coefficients of the established MLR model, leave-one-out (LOO) cross-validation, and predicted values versus experimental ones of external samples were $r^2 = 0.851$, $Q^2_{CV} = 0.746$, and $r^2$(test) = 0.815 respectively. The satisfactory results show that RASMS can express the information related to the biological activity of HEPT derivatives.

DOI: 10.15372/JSC20150506

**K e y w o r d s:** random sampling analysis on molecular surface (RASMS), HEPT derivatives, quantitative structure activity relationship (QSAR), multiple linear regression (MLR).

## INTRODUCTION

The non-nucleoside HIV-1 RT inhibitors under investigation are structurally different entities: 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. Non-nucleoside reverse transcriptase inhibitors (NNRTIs) being one of the two kinds of inhibitors against the HIV-1 reverse transcriptase (HIV-1 RT) have attracted more attention due to their high specificity and low toxicity [ 1 ]. However, the rapid emergence of resistant HIV viral strains carrying mutation at residues that surround the NNRTIs′ 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 treatment of HIV-1 infected patients [ 2, 3 ].

There is a large number of literature reports on the application of computational methods for describing the activity of biologically active compounds [ 4—8 ]. 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 [ 9—12 ]. QSAR studies play a fundamental role in predicting the biological activity of new compounds and identifying ligand-receptor interactions [ 13—15 ].The first step in constructing the QSAR models is to find one or more molecular descriptors

that represent variation in the structural property of the molecules by a number [ 16 ]. Structural descriptors have been classified into different categories according to different approaches, including physiochemical, constitutional, geometrical, topological, and quantum chemical descriptors.

In this paper, a 3D QSAR method — random sampling analysis on molecular surface (RASMS) was proposed. The RASMS method derived from the atomic probe of protein (APP), the pseudo-receptor accessible surface (PRAS), and the approach to aimed area by random sampling on the molecular surface (ARSMS) was used to express the drug structures and biological activities with a multiple linear regression (MLR) model of 35 HEPT derivatives as anti-HIV drugs. The proposed method was evaluated by predicting the activities of the derivatives in this paper; the results indicated that it is a useful tool for the investigation of drug QSAR.

## PRINCIPLE AND METHODOLOGY

**Probe atoms.** Since drug targets are often protein and active peptides, there are eight different types of hybrid atoms from the amino acid 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.

**MCI.** Average electrical charges of each atom in the amino acid serve as MCIs. 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 using 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 imaginary frequency. Ultimately, the amount of the net charge of all the atoms is calculated using the single point method with the density functional theory (DFT).

**VWI**. Usually, the van der Waals radii are the radii of isolated atoms. 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 radii were used as probe atom radii (i.e. VWI = $C_h \times R_{\mathrm{VDW}}$*, with a 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 [ 17, 18 ]). Moreover, the standard van der Waals radii of all kinds of atoms were taken from the report of Bondi et al. [ 19 ].

**MHI.** Similarly to MCI, MHI was got from the average hydrophobic interaction of each probe atom from a natural amino acid. The atomic solvation parameter (ASP) defined by Pei et al. [ 20, 21 ] serves as a hydrophobic measurement.

**PRAS.** The concept of PRAS was proposed in this study. If atoms in the biomolecular systems 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 of the eight probe atoms (the receptor probe) 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 the molecule (H-PRASM). Similarly, the other seven kinds of atoms of the pseudo-receptor probe 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, the radius of which is the sum of the radii of drug atoms plus the radii of the 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.

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

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

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

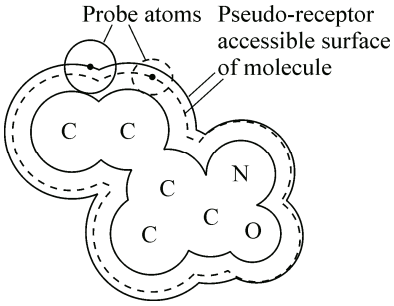



*Fig. 1.* Pseudo-receptor accessible surface of the molecule

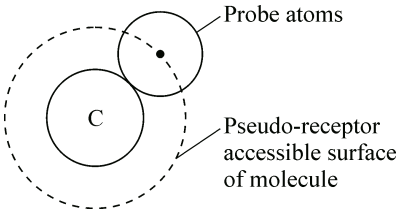



*Fig. 2.* Pseudo-receptor accessible surface of the atom

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 electrostatic charge unit of $1.6021892\times10^{-19}$ C; $\varepsilon_0$ is the dielectric constant in vacuum with a value of $8.85418782\times10^{-12}$ $C^2/J\cdot m$; $Z$ is the atomic net charges. The entire electrostatic interaction items are calculated with this formula.

***Steric interaction.*** The steric interaction, which is described by the Lennard—Jones equation [ 22 ] here, is defined as the interaction between the dipole and non-dipole fields or the induced dipole interaction.

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

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

***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 for the interatomic hydrophobic interactions in HINT is as follows:

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

In Eq. (3), $S$ is the solvent accessible surface area of the atom (SASA); $a$ is the hydrophobic constant expressed with the atomic solvation parameter (ASP); $T$ is the sign function, indicating the entropy changes resulting from different types of atomic interactions [ 23, 24 ].

***Implementation process of* RASMS.** The atoms of organic molecules include H, C, N, P, O, S, F, Cl, Br, and I which belong to IA, IVA, VA, VIA, and VIIA in the Periodic Table of Elements. 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 for a better expression of the microscopic structural features of the molecules. In this paper, 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 the *C* language by the staff of 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 *Sam-*

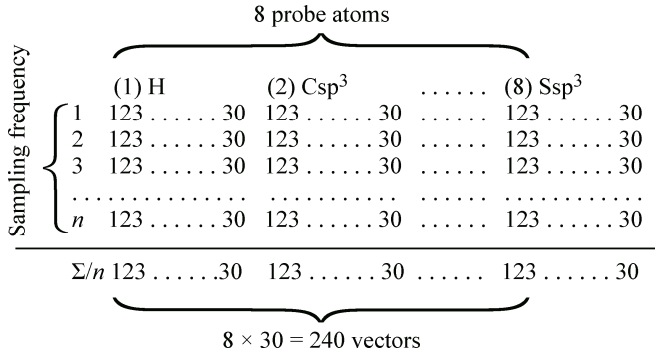



*Fig. 3.* 240 types of interactions of the drug molecules according to the RASMS method

*pling-tool.EXE* after the geometry optimization when using the program. Then the molecular surface sampling density is set and the probe type is selected.

## RESULTS AND DISCUSSION

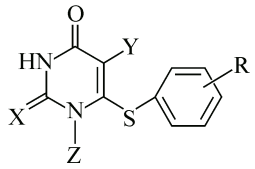
**Selection and partition of the data.** The data set implemented in this work consists of 35 HEPT derivatives. The structures and experimental data of these derivatives (Table 1) were obtained from the literature [ 25 ]. $EC_{50}$ is the effective concentration of the HEPT derivatives required to achieve 50 % protection of MT-4 cells against the cytopathic effect of the virus, and $pEC_{50}$ is used for the calculation in this paper ($pEC_{50} = -\log EC_{50} = \log 1/EC_{50}$). In order to prove the validity and stability of the model, the whole data set was divided into two groups: a training set and a testing set. The training set was used for establishing the QSAR model and the testing set was used to examine the validity of the developed model. Therefore, 7 out of 35 molecules, labeled with notation *a* in Table 1, were randomly selected as the testing set. To choose a suitable method for the geometry optimization in the QSAR study, 35 molecules were firstly auto-constructed using *Chemoffice* 8.0 and then optimized at the AM1 level with the MOPAC (the semi-empirical quantum chemical software in *Chem3D*). Then the net electric charges of the atoms were calculated in the single point form by the Mulliken methods.

***Descriptor generation.*** After the aforementioned two items were input correspondingly in forms of the Cartesian coordinates and the net electric charge, the *Sampling-tool.EXE* software was used to generate 240 descriptors for each HEPT derivative compound. The ultimate vectors for the 28 training compounds involved 240 items. The RASMS method may lead to some information overlap among these different descriptors. To solve the aforementioned 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 the Fisher prominent test with SMR. 22 significant variables were selected out of 240 items using the *SPSS* 16.0 software.

**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 [ 26 ]. 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, so an external validation is required. Recently, a novel method to further refine the predictive ability of the developed QSAR models was introduced by Roy et al. [ 27—30 ]. It is based on 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

T a b l e 1

*Structures and* $pEC_{50}$ *activity of* 35 HEPT *derivatives as* anti-HIV *drugs*




| No. | X | Y | Z | R | $pEC_{50}$ (exp) |
|---|---|---|---|---|---|
| 1 | O | Me | $CH_2OCH_2CH_2OMe$ | H | 5.06 |
| 2 | O | Me | $CH_2OMe$ | H | 5.68 |
| 3 | O | Me | $CH_2OMe$ | H | 6.48 |
| 4 | O | Me | $CH_2OC_3H_7$ | H | 5.44 |
| 5[a] | O | Me | $CH_2OC4H_9$ | H | 5.33 |
| 6 | O | Me | $CH_2OCH_2C_6H_5$ | H | 7.06 |
| 7 | S | $C_2H_5$ | $CH_2OC_2H_5$ | H | 7.59 |
| 8 | S | $C_2H_5$ | $CH_2OC_2H_5$ | 3,5-$Me_2$ | 8.36 |
| 9 | S | $C_2H_5$ | $CH_2OC_2H_5$ | 3,5-$Cl_2$ | 7.89 |
| 10[a] | S | $C_2H_5$ | $CH_2CHMe_2$ | H | 6.66 |
| 11 | S | $C_2H_5$ | $CH_2OC_6H_{11}$ | H | 5.80 |
| 12 | S | $C_2H_5$ | $CH_2OCH_2C_6H_{11}$ | H | 6.46 |
| 13 | S | $C_2H_5$ | $CH_2OCH_2C_6H_5$ | H | 8.11 |
| 14 | S | $C_2H_5$ | $CH_2OCH_2C_6H_5$ | 3,5-$Me_2$ | 8.16 |
| 15[a] | S | $C_2H_5$ | $CH_2OCH_2C_6H_4$-4-Me | H | 7.11 |
| 16 | S | $C_2H_5$ | $CH_2OCH_2C_6H_4$-4-Cl | H | 7.92 |
| 17 | S | $C_2H_5$ | $CH_2OCH_2CH_2C_6H_5$ | H | 7.04 |
| 18 | S | $CHMe_2$ | $CH_2OC_2H_5$ | H | 7.85 |
| 19 | S | $CHMe_2$ | $CH_2OCH_2C_6H_5$ | H | 8.17 |
| 20[a] | S | Cy-$C_3H_5$ | $CH_2OC_2H_5$ | H | 7.02 |
| 21 | O | $C_2H_5$ | $CH_2OC_2H_5$ | H | 7.72 |
| 22 | O | $C_2H_5$ | $CH_2OC_2H_5$ | 3,5-$Me_2$ | 8.27 |
| 23 | O | $C_2H_5$ | $CH_2OC_2H_5$ | 3,5-$Cl_2$ | 8.13 |
| 24 | O | $C_2H_5$ | $CH_2OCHMe_2$ | H | 6.47 |
| 25[a] | O | $C_2H_5$ | $CH_2OC_6H_{11}$ | H | 5.40 |
| 26 | O | $C_2H_5$ | $CH_2OCH_2C_6H_{11}$ | H | 6.35 |
| 27 | O | $C_2H_5$ | $CH_2OCH_2C_6H_5$ | H | 8.23 |
| 28 | O | $C_2H_5$ | $CH_2OCH_2C_6H_5$ | 3,5-$Me_2$ | 8.50 |
| 29 | O | $C_2H_5$ | $CH_2OCH_2CH_2C_6H_5$ | H | 7.02 |
| 30[a] | O | $CHMe_2$ | $CH_2OC_2H_5$ | H | 7.92 |
| 31 | O | $CHMe_2$ | $CH_2OCH_2C_6H_5$ | H | 8.57 |
| 32 | O | Cy-$C_3H_5$ | $CH_2OC_2H_5$ | H | 7.00 |
| 33 | O | Me | $C_2H_5$ | H | 5.66 |
| 34 | O | Me | $C_4H_9$ | H | 5.92 |
| 35[a] | O | Me | $CH_2OCH_2CH_2OH$ | H | 5.16 |

[a] Chosen as the testing set.

(*X* axis) values of the compounds with the intercept ($r^2$) and without the intercept ($r_0^2$) are calculated for the determination of $r_m^2$. 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 ($r_m^2$ and $\Delta r_m^2$) are calculated for both training (internal validation) and testing (external validation) sets in addition to the total dataset (overall validation). 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:

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

$$= \left[\sum (Y_{\text{pred}} - \overline{Y_{\text{pred}}})(Y_{\text{obs}} - \overline{Y_{\text{obs}}})\right]^2 \Big/ \left[\sum (Y_{\text{obs}} - \overline{Y_{\text{obs}}})(Y_{\text{pred}} - \overline{Y_{\text{pred}}})\right]^2,$$

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

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

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

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

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

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

Here, $r^2$ and $r'^2_0$ are the squared correlation coefficient values between the observed and predicted activity data; $Y_{\text{obs}}$ and $Y_{\text{pred}}$ are the observed and predicted response data, while $\overline{Y_{\text{pred}}}$ 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:

$$Q_{\text{CV}}^2 > 0.5, \quad (11)$$

$$r^2(\text{test}) > 0.6, \quad (12)$$

$$(r^2 - r_0^2) / r^2 < 0.1, \quad (13)$$

$$0.85 \le k \le 1.15 \text{ or } 0.85 \le k' \le 1.15. \quad (14)$$

When implementing the above method, the internal and external correlation coefficients should be considered together as a whole in order to achieve the good stability and fine predictability. The first 10 stepwise regression results of SMR with $r^2$ and $Q_{\text{CV}}^2$ of the MLR model of 35 HEPT derivatives were shown in Table 2. The first seven parameters were chosen to build the model; the stepwise multiple regression equation is shown below

$$\text{pEC}_{50} = 14.709 - 0.379H_{5\text{-}5} - 910.993E_{8\text{-}1} - 39.610E_{2\text{-}1} - 76.168H_{6\text{-}3} + \quad (15)$$
$$+ 0.016H_{8\text{-}2} + 2.773S_{5\text{-}3} - 5.063S_{7\text{-}5}.$$

In Eq.(15), *E*, *S*, and *H* represent the electrostatic, steric, and hydrophobic interactions respectively. $E_{8\text{-}1}$ represents the electrostatic interactions of the eighth kind of the probes and the first kind of the drug atoms, $S_{5\text{-}3}$ represents the steric interactions of the fifth kind of the probes and the third kind of the drug atoms, $H_{5\text{-}5}$ represents the hydrophobic interactions of the fifth kind of the probes and the fifth kind of the drug atoms, and so forth. Some parameters such as $r_m^2$, $r_0^2$, and *k* were calculated at the web http://aptsoftware.co.in/rmsquare/ and http://203.200.173.43:8080/rmsquare/. The input data require the observed and predicted response values either imported from a csv file (saved in *.csv format) or given manually for the calculation. The output data provide the values of $r_m^2$ and $\Delta r_m^2$ metrics

Table 2

*Comparison between various kinds of parameters of the first* 10 *steps of the* QSAR *models of* 35 HEPT *derivatives*

| No. of variables | Internal parameters | | | External parameters | | | Overall parameters | Tropsha parameters | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $r^2$ | $Q^2_{CV}$ | $r^2_m$(LOO) | $r^2$(test) | $r^2_0$ | $r^2_m$ | $r^2_m$ | $k$ | $k'$ | $(r^2 - r^2_0)/r^2$ |
| 1 | 0.531 | 0.466 | 0.447 | 0.761 | 0.546 | 0.408 | 0.375 | 0.858 | 1.150 | 0.283 |
| 2 | 0.644 | 0.585 | 0.484 | 0.459 | 0.614 | 0.457 | 0.393 | 0.851 | 1.164 | 0.005 |
| 3 | 0.636 | 0.525 | 0.495 | 0.780 | 0.641 | 0.489 | 0.391 | 0.841 | 1.179 | 1.179 |
| 4 | 0.662 | 0.532 | 0.494 | 0.787 | 0.653 | 0.499 | 0.410 | 0.848 | 1.169 | 0.847 |
| 5 | 0.776 | 0.676 | 0.643 | 0.661 | 0.657 | 0.617 | 0.499 | 0.840 | 1.181 | 0.063 |
| 6 | 0.830 | 0.743 | 0.710 | 0.780 | 0.771 | 0.706 | 0.606 | 0.851 | 1.169 | 0.012 |
| 7 | 0.851 | 0.746 | 0.706 | 0.815 | 0.776 | 0.654 | 0.640 | 0.866 | 1.148 | 0.048 |
| 8 | 0.867 | 0.751 | 0.713 | 0.822 | 0.775 | 0.642 | 0.634 | 0.860 | 1.156 | 0.058 |
| 9 | 0.876 | 0.740 | 0.692 | 0.652 | 0.622 | 0.539 | 0.564 | 0.846 | 1.171 | 0.046 |
| 10 | 0.889 | 0.749 | 0.694 | 0.574 | 0.562 | 0.509 | 0.557 | 0.846 | 1.169 | 0.022 |

Table 3

*Comparison between the* QSAR *models of the* RASMS *method and the literature*

| No. | Methods | Activity types | Model | No. samples | Outlier | No. descriptors | $r^2$ | $Q^2_{CV}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | Hansch [ 25 ] | $EC_{50}$ | MLR | 33 | 2 | 4 | 0.842 | 0.783 |
| 2 | RASMS | $EC_{50}$ | MLR | 35 | 0 | 7 | 0.851 | 0.746 |

for the respective set of compounds. The calculation of $r^2_m$ metrics involves the determination of values of the $r^2$, $r^2_0$, and $k$ parameters together with the information of the intercept of the regression line correlating the observed and predicted activity data. The calculated $r^2$, $Q^2_{CV}$, $r^2_m$(LOO), $r^2$(test), $r^2_0$, $r^2_m$ (external), $r^2_m$ (overall), $k$, $k'$, and $(r^2 - r_0{}^2)/r^2$ are 0.851, 0.746, 0.706, 0.815, 0.776, 0.654, 0.640, 0.866, 1.148, and 0.048, respectively. Therefore, it is confirmed that the RASMS QSAR models are stable and generalized. The comparison between the QSAR models of the RASMS method and the literature is shown in Table 3. It is shown that the result of the RASMS method is better than the literature one. 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.

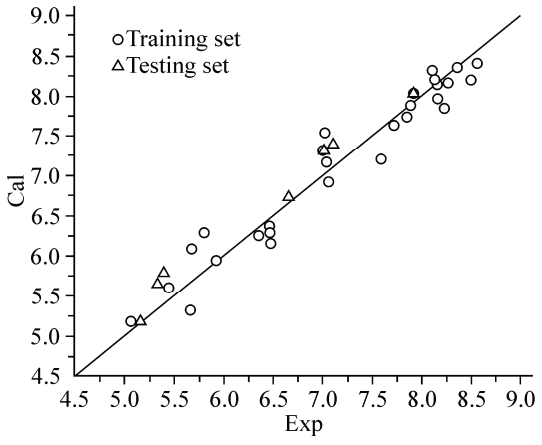



*Fig. 4.* Plots of the observed and calculated values of 35 HEPT derivatives with the RASMS method

## CONCLUSIONS

In this paper, all the descriptors involve classic electrostatic, steric, and hydrophobic interactions. The built model has a 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 representation of both molecular structures and

biological activities for the HEPT derivatives. It can be anticipated that the approach might hold a high potential to become a useful tool in the research of the QSAR of HEPT derivatives.

The authors appreciate the financial support from the National Natural Science Foundation of China (21275094) (21475081) and the Graduate Innovation Fund of Shaanxi University of Science and Technology.

*REFERENCES*

1. *Sarafianos S.G., Das K., Hughes S.H., Arnold E.* // Curr. Opin. Struct. Biol. – 2004. – **14**. – P. 716 – 730.
2. *Das K., Lewi P.J., Hughes S.H., Arnold E.* // Prog. Biophys. Mol. Biol. – 2005. – **88**. – P. 209 – 231.
3. *Liang Y.H., Chen F.E.* // Eur. J. Med. Chem. – 2009. – **44**. – P. 625 – 631.
4. *Hemmateenejad B., Miri R., Akhond M., Shamsipur M.* // Chemom. Intell. Lab. Syst. – 2002. – **64**. – P. 91 – 99.
5. *Hemmateenejad B., Miri R., Akhond M., Shamsipur M.* // Arch. Pharm. Med. Chem. – 2002. – **10**. – P. 472 – 480.
6. *Hansch C., Hoekman D., Gao H.* // Chem. Rev. – 1996. – **96**. – P. 1045 – 1075.
7. *Fassihi A., Sabet R.* // Int. J. Mol. Sci. – 2008. – **9**. – P. 1876 – 1892.
8. *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 – 2157.
9. *Hansch C., Fujita T.* // J. Amer. Chem. Soc. – 1964. – **86**. – P. 1616 – 1626.
10. *Wang J., Zhang L., Yang G., Zhan C.G.* // J. Chem. Inf. Comput. Sci. – 2004. – **44**. – P. 2099 – 2105.
11. *Hammet L.P.* // J. Amer. Chem. Soc. – 1937. – **59**. – P. 96 – 103.
12. *Hemmateenejad B., Sanchooli M.* // J. Chemom. – 2007. – **21**. – P. 96 – 107.
13. *Pasha F.A., Srivastava H.K., Singh H.K., Semiempirical P.P.* // Mol. Div. – 2005. – **9**. – P. 215 – 220.
14. *Cruz O.J.D., Uckun F.M.* // J. Antimicrob. Chemother. – 2006. – **57**. – P. 411 – 423.
15. *Gayen S., Debnath B., Samanta S., Jha T.* // Bioorg. Med. Chem. – 2004. – **12**. – P. 1493 – 1503.
16. *Kustrin S.A., Tucker I.G., Zecevic M., Ziva-novic L.J.* // Anal. Chem. Acta. – 2000. – **418**. – P. 181 – 195.
17. *Tong J.B., Li Y.F., Liu S.L., Meng Y.L.* // Chinese J. Struct. Chem. – 2010. – **29**. – P. 1893 – 1899.
18. *Hahn M.* // J. Med. Chem. – 1995. – **38**. – P. 2080 – 2090.
19. *Bondi A.* // J. Phys. Chem. – 1964. – **68**. – P. 441 – 451.
20. *Pei J.F., Wang Q., Zhou J.J., Lai L.H.* // Proteins-Structure Function, Bioinformatics. Proteins. – 2004. – **57**. – P. 651 – 664.
21. *Tong J.B., Chen Y., Liu S.L., Xu X.M.* // Med. Chem. Res. – 2013. – **22**. – P. 4946 – 4952.
22. *Tong J.B., Zhong L., Zhao X., Liu S.L., Wang P.* // Med. Chem. Res. – 2014. – **23**. – P. 1634 – 1642.
23. *Zhou P., Tong J.B., Tian F.F., Li Z.L.* // Chin. Sci. Bull. – 2006. – **51**. – P. 1824 – 1829.
24. *Tong J.B., Liu S.L.* QSAR & Combinat. Sci. – 2008. – **27**. – P. 330 – 337.
25. *Garg R., Gupta S.P., Gao H., Babu M.S., Debnath A.K., Hansch C.* // Chem. Rev. – 1999. – **99**. – P. 3525 – 3601.
26. *Tong J.B., Chen Y., Liu S.L., Xu X.M., Cheng F.L.* // J. Struct. Chem. – 2012. – **53**. – P. 1075 – 1080.
27. *Roy K., Chakraborty P., Mitra I., Ojha P.K., Kar S., Das R.N.* // J. Comput. Chem. – 2013. – **34**. – P. 1071 – 1082.
28. *Mitra I., Saha A., Roy K.* // Mol. Simul. – 2010. – **13**. – P. 1067 – 1079.
29. *Roy K., Mitra I., Kar S., Ojha P.K., Das R.N., Kabir H.* // J. Chem. Inf. Modell. – 2012. – **52**. – P. 396 – 408.
30. *Roy K., Mitra I.* // Comb. Chem. High Throughput Screening. – 2011. – **14**. – P. 450 – 474.