Original article

Quantitative structure−activity relationship studies of a series of non-benzodiazepine structural ligands binding to benzodiazepine receptor

Binbin Xia [a], Weiping Ma [a], Bo Zheng [a], Xiaoyun Zhang [a,*], Botao Fan [b]

[a] *Department of Chemistry, Lanzhou University, 222 Tianshui South Road, Lanzhou 730000, Gansu, PR China*
[b] *Université Paris 7-Denis Diderot, ITODYS 1, Rue Guy de la Brosse, 75005 Paris, France*

Abstract

Heuristic method (HM) and radial basis function neural network (RBFNN) methods were proposed to generate QSAR models for a set of non-benzodiazepine ligands at the benzodiazepine receptor (BzR). Descriptors calculated from the molecular structures alone were used to represent the characteristics of the compounds. The six molecular descriptors selected by HM in CODESSA were used as inputs for RBFNN. Compared with the results of HM, more accurate prediction could be obtained from RBFNN. The correlation coefficients (R) of the nonlinear RBFNN model were 0.9113 and 0.9030 for the training and test sets, respectively. This paper proposed an effective method to design new ligands of BzR based on QSAR.
© 2007 Elsevier Masson SAS. All rights reserved.

1. Introduction

γ-Aminobutyric acid (GABA), the major inhibitory neurotransmitter system in the central nervous system (CNS), operates through three different classes of receptors consisting of the ionotropic $GABA_A$ and $GABA_C$ receptors and the metabotropic $GABA_B$ receptors [1]. The $GABA_A$ receptors, as one of the most important classes, have been associated with certain neurological and psychiatric disorders and become therapeutic targets in certain diseases. The benzodiazepine receptor (BzR) is located on the $GABA_A$ receptor channel, and it is capable of binding with high affinity to benzodiazepine and other chemical families [2]. Researchers in the fields of medicine, pharmacology, and chemistry have paid more attention to $GABA_A$/BzR over the last several years.

Various ligands bind to GABA/BzR and then enhance, diminish or block the chloride flux through the neuronal membranes. Therefore, the ligands are classified as agonists, inverse agonists or antagonists (blockers), respectively [3]. The inter-relationships of the three classes of BzR ligands can be interpreted on the basis of changes in the conformation of the receptor from its unoccupied resting state [4]. For an accurate study, it is very important to distinguish between the three classes of the ligands. Unluckily, the experimental conditions do not allow the investigator to make such a differentiation among the examined ligands.

Benzodiazepines, a well-known class of ligands, have been used as anxiolytics, as tranquilizers, and as anticonvulsants. In recent years, various structurally different non-benzodiazepine molecules have been identified as endogenous ligands for BzR. Meanwhile, some quantitative structure−activity relationship (QSAR) models also have been developed to have a better understanding on the interactions between non-benzodiazepine structures and BzR, and thus to reach a more precisely outlined target for the design of new active compounds [5−10]. Mostly, biological activity is expressed by IC_{50}, which refers to the concentration of the test compound causing 50% inhibition of the

* Corresponding author. Tel.: +86 931 891 2578; fax: +86 931 891 2582.
  E-mail address: xyzhang@lzu.edu.cn (X. Zhang).

specific binding of [3H]-diazepam or [3H]-flunitrazepam to BzR from rat brain preparation [11]. It therefore reflects the binding affinities between BzR and these ligands. However, there has been much difficulty to obtain the experimental data because of the complex biological process in organisms.

QSAR as one of the most important areas in chemometrics gives information that is useful for drug design and medicinal chemistry. As we know, chemical and biological effects are closely related to molecular properties, which can be calculated or predicted by various kinds of methods from their structure. Once a reliable model is established, we can predict the activities of compounds and know which structural factors play an important role in biological process. The advances in QSAR studies have widened the scope of rational drug design and the search for the mechanisms of drug actions [12,13]. On the other hand, the advantage of this approach over other methods lies in the fact that it requires only the knowledge of chemical structure and is not dependent on the experimental process.

Regarding the studies, we have developed a QSAR method for the prediction of pIC_{50} of some non-benzodiazepine compounds. It can also identify and describe important structural features of the molecules that are relevant to variations in molecular activities. The QSAR models are useful because they rationalize a large number of experimental observations and allow saving of time and money in the drug design process.

Many different technologies can be applied to the QSAR development, such as multiple linear regression (MLR), partial least squares (PLS), heuristic method (HM), and different types of artificial neural networks (ANN). Recently, ANN has gained great popularity in QSAR research due to their flexibility in modeling nonlinear problems [14]. As one type of neural networks (NN), the radial basis function neural network (RBFNN) has advantages of short training time and are guaranteed to reach the global minimum of error surface during training process. Moreover, their parameters can be adjusted by fast linear methods [15]. In the past decade, RBFNN has been used to solve many problems in chemistry and chemical engineering: classification [16,17], QSPR [18,19], and QSAR [20].

RBFNN, as a novel machine learning technique, for the first time, was used for the prediction of the biological activities of 58 non-benzodiazepine structures based on the large and diverse dataset using the descriptors calculated from the molecular structure alone by the software CODESSA. Six descriptors were selected as inputs by heuristic method (HM). The aim of this study was to explore the active mechanisms of binding of these ligands to the $GABA_A$/BzR receptor and to establish a new and accurate quantitative structure−activity relationship model and at the same time, to seek the structural factors affecting the binding affinity.

2. Experimental section

2.1. Dataset

The dataset was extracted from a recent work reported by Cuadrado et al. [21]. The 58 non-benzodiazepine $GABA_A$/BzR ligands were three classes of compounds, namely,

dihyroindolo-β-carboline structures (**1−33**), dihydro-pyrazolo-quinolinone compounds (**34−45**), and β-carboline ligands (**46−58**). The biological activity (pIC_{50}) was expressed by inhibition of the [3H]-diazepan-specific which ranges from 3.6 to 9.51. The values of these compounds are listed in Table 1. The dataset was randomly divided into a training set of 45 compounds and a test set of 13 compounds. The training set was used to adjust the parameters and construct the QSAR models and the test set was used to evaluate their prediction ability.

The structures are summarized in Table 2.

2.2. Descriptor generation

The structures of the molecules were drawn with the ISIS Draw 2.3 program [22]. All molecules were preoptimized using molecular mechanics force field in the HyperChem program [23]. A more precise optimization is done with semiempirical PM3 method in the MOPAC 6.0 program [24]. The output files exported from MOPAC were transferred into the software CODESSA, developed by Katritzky et al. [25,26]. In the present work, five classes of structural descriptors were obtained and about 618 descriptors were provided. Topological descriptors include valence and non-valence molecular connectivity indices calculated from the hydrogen-suppressed formula of the molecule, encoding information about the size, composition and the degree of branching of a molecule. Geometrical descriptors are calculated from 3D atomic coordinates of the molecule. These descriptors comprise moments of inertia, shadow indices, molecular volume, molecular surface area, and gravitation indices. Electrostatic descriptors reflect characteristics of the charge distribution of the molecule. The quantum-chemical descriptors include information about binding and formation energies, partial atom charge, dipole moment, and molecular orbital energy levels.

3. Methodology

3.1. Heuristic method

HM in CODESSA was used to select descriptors and build the linear model. After calculating a large number of descriptors, a feature selection step was carried out to reduce the large set of descriptors to a suitable number without losing any important information [27]. The HM of the descriptor selection proceeds with a pre-selection by eliminating descriptors that (i) are not available for each structure; (ii) have a small variation in magnitude for all structures; (iii) have a Fisher F-criterion greater than 1 unit; (iv) have a student's t-criterion less than that defined (by default 0.1), etc. This procedure ordered the descriptors by the decreasing correlation coefficient when used in one-parameter correlations. As a next step, the program calculated the pair correlation matrix of descriptors and further reduced the descriptor pool by eliminating highly correlated descriptors. After the pre-selection of descriptors, multilinear regression models were developed in a stepwise procedure [28]. Thus, descriptors and correlations were ranked according to the values of the F-test and the correlation

Table 1
Lab and predicted activities (using HM and RBFNN) for the whole set

No.	Lab values (pIC$_{50}$)	Cal. HM	Residual	Cal. RBFNN	Residual
1[a]	8.40	7.06	−1.34	7.31	−1.09
2	8.40	7.54	−0.86	7.48	−0.92
3	5.92	6.83	0.91	6.45	0.53
4[a]	6.90	6.97	0.07	6.60	−0.30
5[a]	7.36	7.42	0.06	7.01	−0.35
6[a]	7.01	7.49	0.48	6.13	−0.88
7[a]	8.22	7.95	−0.27	7.67	−0.55
8	6.51	6.89	0.38	6.38	−0.13
9	8.22	7.67	−0.55	7.70	−0.52
10	5.67	7.01	1.34	6.76	1.09
11	6.12	8.00	1.88	7.52	1.40
12	6.65	6.75	0.10	7.24	0.59
13	6.05	7.15	1.10	6.35	0.30
14	5.80	5.70	−0.10	5.71	−0.09
15	6.94	7.54	0.60	7.49	0.55
16	7.92	7.59	−0.33	7.60	−0.32
17	7.10	6.80	−0.30	6.99	−0.11
18	7.55	6.91	−0.64	7.04	−0.51
19	7.08	6.51	−0.57	7.22	0.14
20	8.15	7.92	−0.23	7.54	−0.61
21[a]	8.00	7.53	−0.47	6.79	−1.21
22	7.72	8.17	0.45	7.58	−0.14
23	8.10	7.00	−1.10	7.59	−0.51
24	8.10	7.46	−0.64	6.86	−1.24
25	8.22	7.55	−0.67	7.57	−0.65
26	6.15	6.78	0.63	7.14	0.99
27[a]	6.60	6.22	−0.38	6.16	−0.44
28	6.24	6.98	0.74	7.15	0.91
29	6.60	6.48	−0.12	6.64	0.04
30[a]	5.94	6.26	0.32	7.17	1.23
31	6.80	6.16	−0.64	7.06	0.26
32	5.71	5.36	−0.35	5.63	−0.08
33	5.72	5.27	−0.45	5.22	−0.50
34[a]	9.51	9.69	0.18	8.94	−0.57
35	8.94	9.34	0.40	8.93	−0.01
36	9.35	8.71	−0.64	8.99	−0.36
37	8.99	9.44	0.45	8.80	−0.19
38	8.99	8.63	−0.36	9.13	0.14
39[a]	8.96	9.14	0.18	8.19	−0.77
40[a]	9.06	8.66	−0.40	8.52	−0.54
41	9.05	9.34	0.29	9.37	0.32
42	9.29	9.12	−0.17	9.44	0.15
43	8.87	8.51	−0.36	8.45	−0.42
44	9.27	8.36	−0.91	8.64	−0.63
45	6.52	7.75	1.23	7.72	1.20
46	5.79	5.86	0.07	5.50	−0.29
47	3.60	4.58	0.98	3.19	−0.41
48	4.91	5.38	0.47	5.42	0.51
49	7.35	6.80	−0.55	6.97	−0.38
50	6.90	6.75	−0.15	6.27	−0.63
51[a]	5.40	6.75	1.35	6.15	0.75
52	6.91	6.68	−0.23	7.23	0.32
53	8.30	7.35	−0.95	8.29	−0.01
54[a]	4.59	6.14	1.55	4.59	0.00
55	5.30	5.67	0.37	5.26	−0.04
56	5.04	4.80	−0.24	5.13	0.09
57[a]	6.17	5.84	−0.33	5.92	−0.25
58[a]	5.77	5.11	−0.66	5.15	−0.62

[a] The sample of the test set, the remaining are the samples of the training set.

coefficient. Starting with the top descriptor from the list, two-parameter correlations are calculated.

In the following steps new descriptors were added one by one until the pre-selected number of descriptors in the model is achieved. The final result is a list of the 10 best models according to the values of the F-test and correlation coefficient. The goodness of the correlation is tested by the coefficient regression (R^2), the F-test (F), and the standard deviation (s^2). The Heuristic method usually produces correlations 2–5 times faster than other methods, with comparable quality [26]. The rapidity of calculations from the heuristic method renders it the first method of choice in practical research. Thus, we used this method for our calculations.

3.2. Radial basis function neural networks

As a type of neural networks, RBFNN has been widely used for modeling and classification. The theory of RBFNN has been adequately described elsewhere [29,30]. Here, only a brief description of the RBFNN principle is given. As presented schematically in Fig. 1, the RBFNN consists of three layers: input layer, hidden layer and output layer. The input layer does not process the information; it only distributes the input vectors to the hidden layer. The hidden layer of RBFNN consists of a number of RBF units (n_h) and bias (b_k). Each hidden layer unit represents a single radial basis function, with associated center position and width. Each neuron on the hidden layer employs a radial basis function as nonlinear transfer function to operate on the input data. The more often used RBF is a Gaussian function that is characterized by a center (c_j) and width (r_j). The RBF functions by measuring the Euclidean distance between input vector ($\mathbf{x}$) and the radial basis function center (c_j) and performs the nonlinear transformation with RBF in the hidden layer as given below:

$$\mathbf{h}_j(\mathbf{x}) = \exp\left(-\left\|\mathbf{x} - \mathbf{c}_j\right\|^2 / r_j^2\right) \tag{1}$$

in which, $\mathbf{h}_j$ is the notation for the output of the jth RBF unit. For the jth RBF, $\mathbf{c}_j$ and r_j are the center and width, respectively. The operation of the output layer is linear, which is given in Eq. (2):

$$\mathbf{y}_k(\mathbf{x}) = \sum_{j=1}^{n_h} \mathbf{w}_{kj} \mathbf{h}_j(\mathbf{x}) + b_k \tag{2}$$

where $\mathbf{y}_k$ is the kth output unit for the input vector $\mathbf{x}$, $\mathbf{w}_{kj}$ is the weight connection between the kth output unit and the jth hidden layer unit and b_k is the bias.

From Eqs. (1) and (2), one can see that designing RBFNN involves selecting centers, number of hidden layer units, width and weights. In general, the number of hidden layer units is equal to the number of centers. There are various ways for selecting the centers, such as random subset selection, K-means clustering, orthogonal least squares learning algorithm, RBF–PLS, etc. In the present paper, a forward subset selection routine was used to select the centers from training set samples. The widths of the radial basis function can either be chosen

Table 2
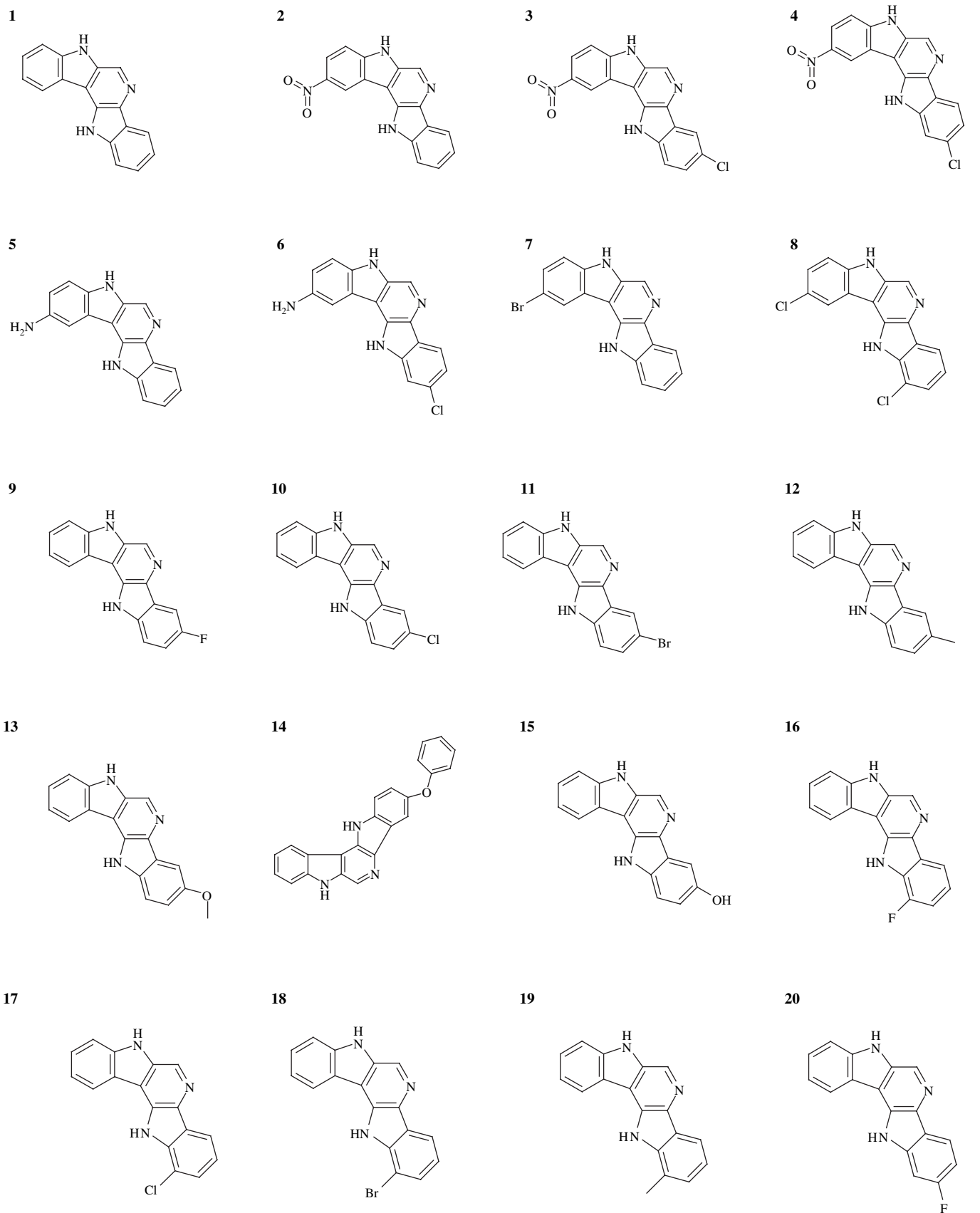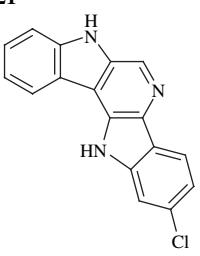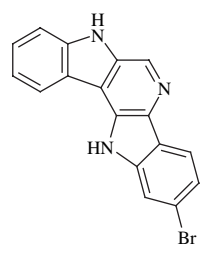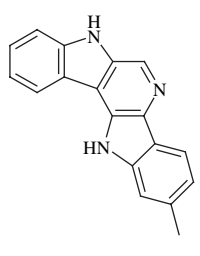Structures of the 58 non-benzodiazepine ligands

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20
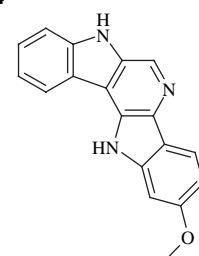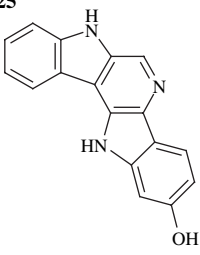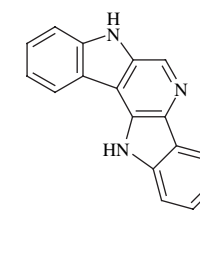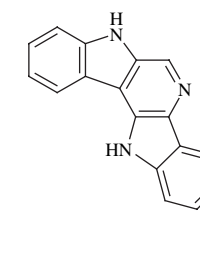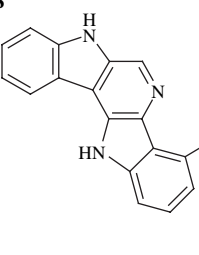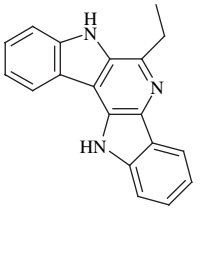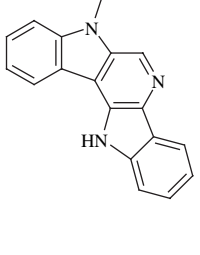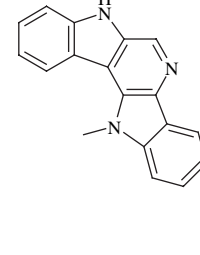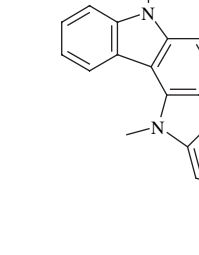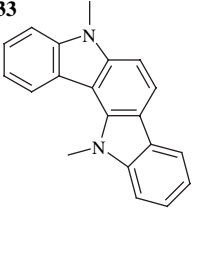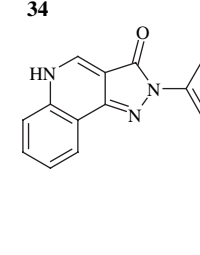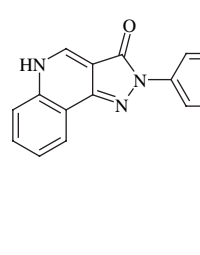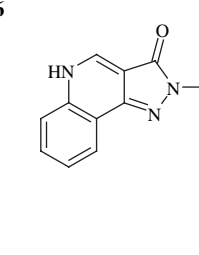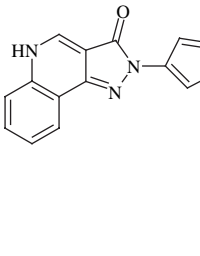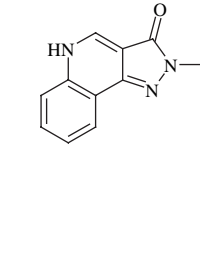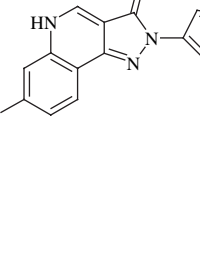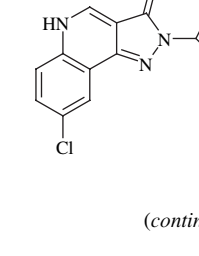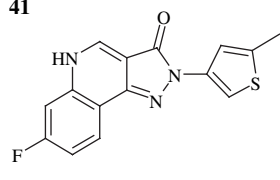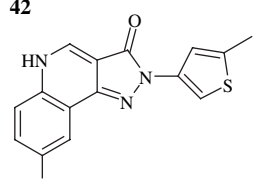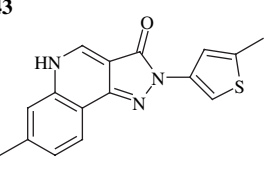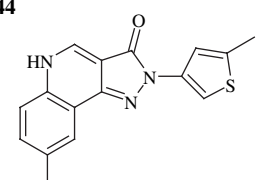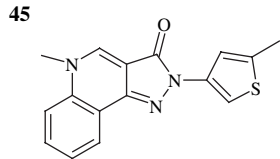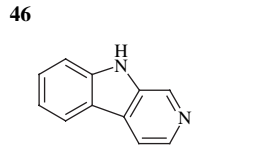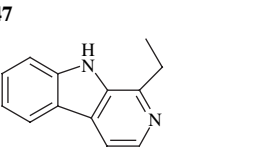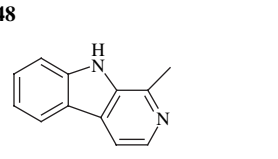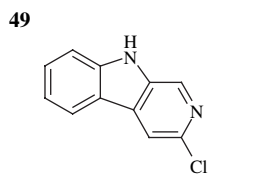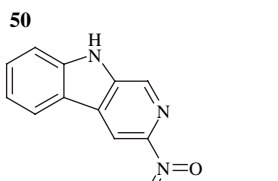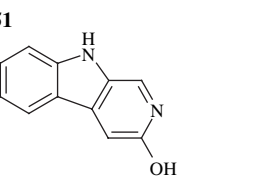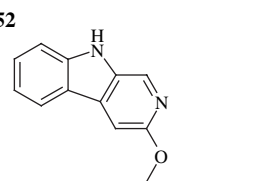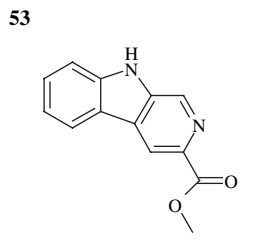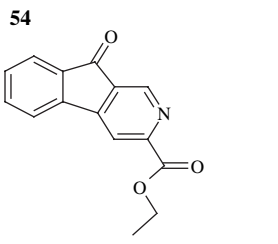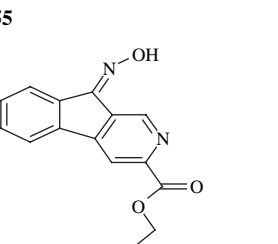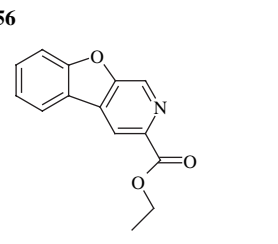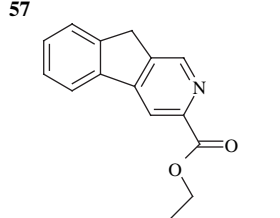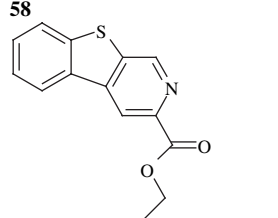
Table 2 (*continued*)

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

Table 2 (*continued*)

the same for all the units or can be chosen different for each unit. In this paper, considerations were limited to the Gaussian functions with a constant width, which was the same for all units. The adjustment of the connection weight between the hidden layer and output layer is performed using a least squares solution after the selection of centers and width of radial basis functions. The overall performance of RBFNN is evaluated in terms of root mean squared error (*RMS*) according to the equation below:

$$RMS = \sqrt{\frac{\sum_{i=1}^{n_s}(\mathbf{y}_k - \widehat{\mathbf{y}}_k)^2}{n_s}} \qquad (3)$$

where $\mathbf{y}_k$ is the desired output, $\widehat{\mathbf{y}}_k$ is the actual output of the network, and n_s is the number of compounds in analyzed set.

3.3. Algorithm implementation and computation environment

All calculation programs implementing RBFNN were written in M-file based on a basis MATLAB script for RBFNN. The scripts were run on a Pentium IV PC with 256M RAM.

4. Results and discussion

4.1. Results of HM

HM was used to develop the linear model for the prediction of binding activity using calculated structural descriptors. After the heuristic reduction, the pool of descriptors was reduced from 618 to 287. To determine the optimum number of

Fig. 1. The architecture of radial basis function neural network.

Fig. 2. R^2 and R_{cv}^2 vs. the number of descriptors.

descriptors, a variety of subset sizes were investigated. When adding another descriptor did not improve significantly the statistics of a model, it was determined that the optimum subset size had been achieved. To avoid 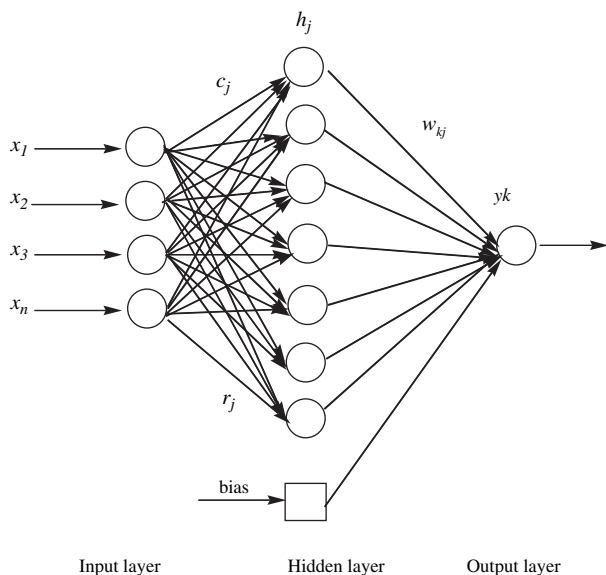the 'over-parameterization' of the model, an increase of the R^2 value of less than 0.02 was chosen as the breakpoint criterion. The influences of the number of the descriptors on the correlation coefficient (R^2) and cross-validation (R_{cv}^2) are shown in Fig. 2. From Fig. 2, it can be seen that six descriptors appear to be sufficient for a successful regression model. The multilinear analysis of the biological activity values for the 45 compounds of the training set resulted in the six-parameter model summarized in Table 3.

The six-parameter correlation model are listed as follows:

$$pIC_{50} = 1.483 \times MTIcc + 2.947 \times MnnRch - 15.363 \\ \times XY\ Shadow + 0.972 \times HOMO - 1.867 \times Balaban \\ - 0.020 \times (WPSA\text{-}1) - 174.470 \quad (4)$$

The obtained correlation coefficient (R) was 0.8698 and the LOO cross-validated coefficient (R_{cv}^2) was 0.6922. The RMS error was 0.6040. The prediction results of the test set were obtained with R of 0.8662 and the RMS error of 0.6554. The model produced R of 0.8688 and RMS error of 0.6060 for the whole dataset. The predicted activity values based on heuristic regression are listed in Table 1, and Fig. 3 shows the predicted pIC_{50} vs. lab values for all of the 58 compounds.

4.2. Results of RBFNN

From Table 1 and Fig. 3, it can be seen that the model of the heuristic method was not sufficiently accurate. The factors affecting the activities of these compounds were complex and none of them had linear correlation with the biological activity. Therefore, RBFNN was performed to develop a nonlinear model of biological activities based on the same subset of
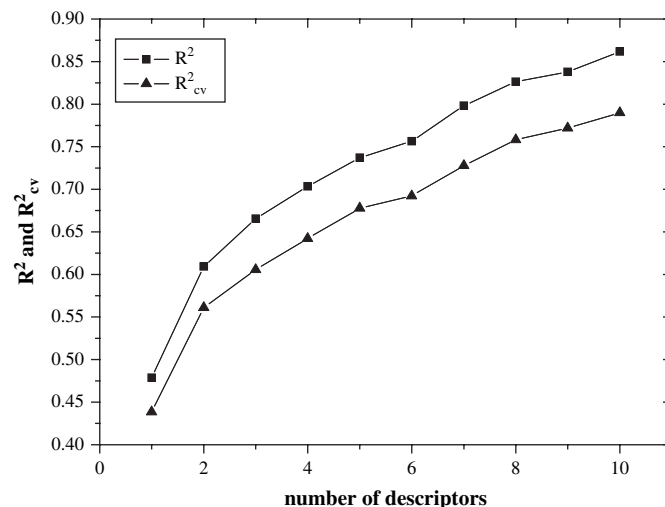
descriptors. Such an RBFNN can be designed as 6-n_k-1 net to indicate the number of units in the input, the hidden layer and output layer, respectively. To obtain better results, the parameters that influence the performance of RBFNN were optimized. The selection of the optimal width value for RBFNN was performed by systemically changing its value in the training step. The RMS error on LOO cross-validation of the training set was plotted against the width, as shown in Fig. 4. The width of 1.5 with the minimum RMS error was chosen as the optimal condition for the RBFNN model of active values. In this case, the number of hidden layer units (n_k) was selected as 13. The predicted results of the nonlinear models are shown in Table 1 and Fig. 5. The obtained model had a correlation coefficient R of 0.9113 and RMS error of 0.5700 for the training set. The statistical parameters of test set were $R = 0.9030$ and RMS $= 0.7242$. The RMS error in prediction for overall dataset was 0.6136.

To demonstrate the absence of chance correlations, we used the internal validation method. The full set of 58 structures was sorted in ascending order according to the pIC_{50} value and then divided evenly into four subsets A, B, C and D: structures **1**, **5**, **9**, etc. formed group A, structures **2**, **6**, **10**, etc.

Table 3
The six-parameter linear model of structure and activity ($R^2 = 0.7436$, $R_{cv}^2 = 0.6922$, $F = 17.4040$, $s^2 = 0.5860$)

Descriptor	Chemical meaning	Coefficient	Error	t-Test value
Constant	Intercept	−114.4700	69.0600	−1.6575
MTIcc	Maximum total interaction for a C−C bond	1.4827	0.2181	6.7972
MnnRch	Minimum n−n repulsion for a C−H bond	2.9467	1.6204	1.8184
XY Shadow	XY Shadow/XY Rectangle	−15.3630	4.0984	−3.7484
HOMO	HOMO energy	0.9723	0.5483	1.7733
Balaban	Balaban index	−1.8670	1.1314	−1.6502
WPSA-1	WPSA-1 weighted PPSA (PPSA-1 × TMSA/1000) [quantum-chemical PC]	−0.0197	0.0071	−2.7735

Fig. 3. Scatter plot of the predicted values by HM vs. the lab values for the two sets.
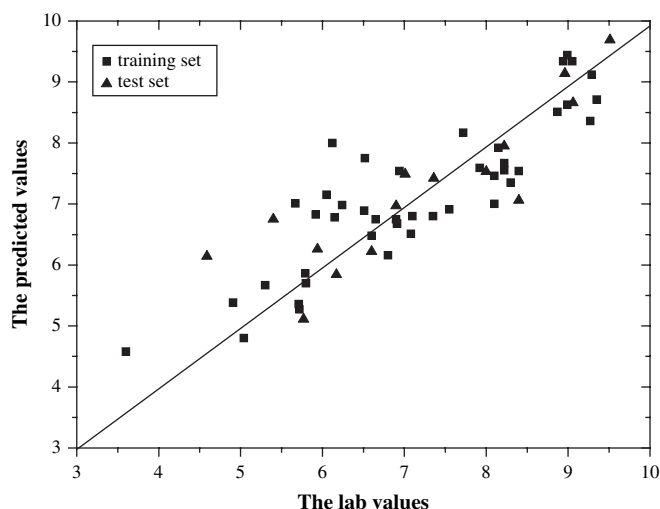

Fig. 5. Scatter plot of the predicted values by RBFNN vs. the lab values for the two sets.

formed group B, structures **3**, **7**, **11**, etc. formed group C, and structures **4**, **8**, **12**, etc. formed group D. Each subset was predicted by using the other three subsets as the training set. For each training set, the correlation equation was derived with the same descriptors, and the equation obtained was used to predict the activity of interest values for the compounds from the corresponding test set. The results shown in Table 4 disclose an average training quality of *R* 0.9032 and an average predicting quality of *R* 0.8983, which demonstrate that the proposed model had a satisfactory statistical stability and validity. The results of the validation showing the summary of all four predictions are given in Table 4. At last, the best model (Table 4, bold characters) was selected for predicting the pIC_{50} of the 58 non-benzodiazepines.

Analysis of the result obtained indicates that the models we proposed can correctly represent the relationship between these compounds and the molecular descriptors calculated solely from molecular structures. Comparing the correlation
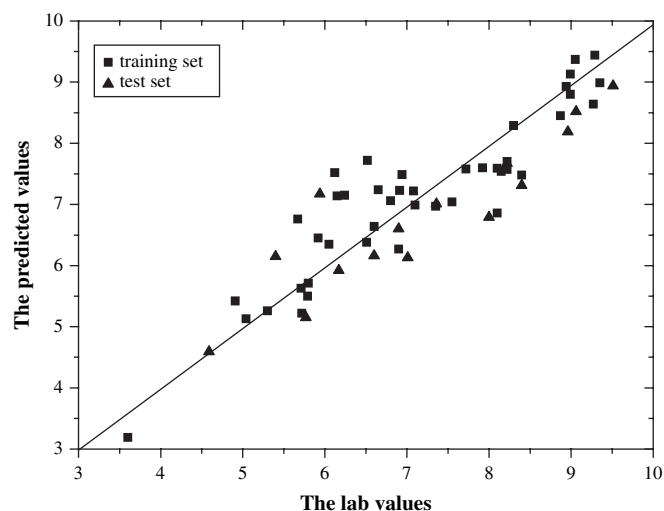
models obtained by HM and RBFNN, it can be seen that the performance of RBFNN is better than that of HM, which indicates that the nonlinear model can simulate the relationship between the structure descriptors and bioactivity, and is more accurate than the linear model.

4.3. Discussions of the input parameters

By interpreting the descriptors in the regression model, it is possible to gain some insight into factors affecting the affinity ability of ligands and understand which interaction plays an important role during the binding process. In the linear model, six descriptors were found to be important for these compounds.

Balaban index (Balaban) is a topological descriptor and encodes information about the size, composition and the degree of branching of a molecule. The Balaban index is the average distance sum connectivity which is based on the graph-theoretical distance matrix and represents extended connectivity [31]. The more the average distance sum connectivity increases, the higher is the size of the compounds. From the negative coefficients in the linear model, we can see that increasing the descriptor will decrease the affinity ability of ligands. This can be explained that the larger molecule has a larger stereospecific blockade when the binding process arose between the BzR ligands and receptors. So, the calculated activity values (pIC_{50}) decrease with an increase in the Balaban index.
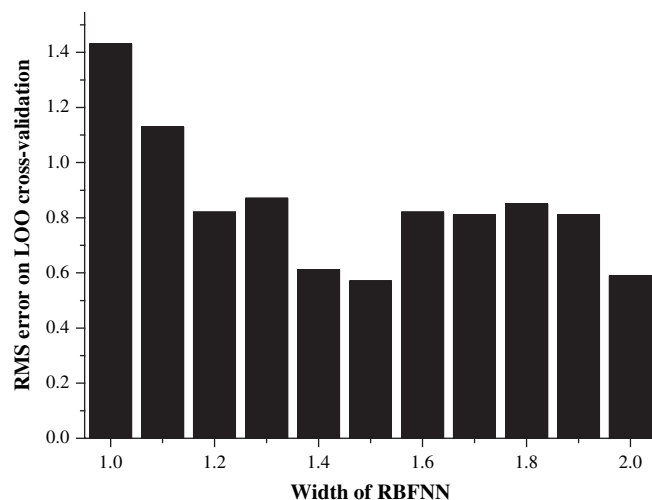
Fig. 4. The width of RBFNN vs. RMS error on LOO cross-validation.

Table 4
Validation of the nonlinear model

Training set	*R*	RMS	Test set	*R*	RMS
B + C + D	0.9006	0.5699	A	0.9027	0.7538
A + C + D	**0.9113**	**0.5700**	**B**	**0.9030**	**0.7242**
A + B + D	0.9004	0.6339	C	0.8804	0.6107
A + B + C	0.9006	0.5969	D	0.9069	0.6693
Average	0.9032	0.5927	Average	0.8983	0.6895

The geometrical descriptors describe the size of the molecules and are derived from the three-dimensional coordinates of the atomic nuclei, the atomic masses, and the atomic radii in the molecule. XY Shadow/XY Rectangle (XY Shadow) reflects the size (natural shadow indices) and geometrical shape (normalized shadow indices) of the molecule. It received negative coefficients in the regression which indicates that the binding ability increases with the decrease of XY Shadow. In other words, the larger the molecule, the more difficult is the binding of BzR to the ligands. The reason can be concluded that the binding ability is dependent on the special groups in the molecules.

Highest occupied molecular orbital (HOMO) energy is a quantum-chemical descriptor. Molecular orbital (MO) surfaces visually represent the various stable electron distributions of a molecule. According to frontier orbital theory, the shapes and symmetries of the highest occupied and lowest unoccupied molecular orbitals (HOMO and LUMO) are crucial in predicting the reactivity of a species and the stereochemical and regiochemical outcome of a chemical reaction [32]. Molecules with high HOMO values can donate their electrons more easily compared to molecules with low HOMO energy values, and hence are easier to bond and should be readily bioactivated. If these non-benzodiazepine compounds have larger HOMO values, the binding stability will increase correspondingly. Generally speaking, the HOMO level is affected by electron-donating groups, such as the alkyl group.

Surface weighted charged partial positive charged surface area (WPSA-1) is an electrostatic descriptor. It can be defined as follows:

$$WPSA\text{-}1 = \frac{PPSA\text{-}1 \times TMSA}{1000} \qquad (5)$$

In Eq. (5), partial positive surface area (PPSA-1) indicates the polar surface areas formed by positive charge distribution in the molecule and it encodes information about polar interaction. And TMSA is the total molecular surface area. This descriptor can roundly reflect how charge is distributed in the molecule. The regression coefficient of this descriptor indicates that it has a negative effect on the affinity ability of these compounds. The reason may be that the polarity has a negative influence on the hydrophobicity of the molecules, so the binding becomes more difficult.

Maximum total interaction for a C–C bond (MTIcc) and minimum n–n repulsion for a C–H bond (MnnRch) are both quantum-chemical descriptors. They represent the stability of the molecules and other properties that correspond to the bond such as chemical reactivity. The positive coefficient of MTIcc suggests that probably the binding of the ligands to the receptor site involves a certain degree of electron transfer. The higher the descriptor is, the higher are the pIC$_{50}$ values. MnnRch describes the nuclear repulsion driven processes in the molecule and may be related to the conformational (rotational, inversional) changes or atomic reactivity in the molecule [33]. If the descriptor is lower, the repulsion between the ligands and receptor will become smaller. so the binding tends to become more stable.

Thus, the two descriptors may again be related to the conformational changes or atomic reactivity in the molecule.

From the above discussion, it can be seen that all the descriptors involved in the model have explicit physical meaning, and these descriptors can account for the structural features responsible for binding of the BzR ligands. According to the analysis of the *t*-test values (Table 3), MTIcc, XY shadow and WPSA-1 have larger values which indicate that polar, hydrophobic, and spatial conformation descriptors play very significant roles in predicting the binding affinity. Furthermore, choosing the most important descriptors that influence the construction of the compounds and offer suggestion to synthesize the new drug with high binding affinities or high inhibition activity is valuable in QSAR analysis. One can adjust the shape of the molecules, the polar interactions between molecules and the conformation of the compounds to increase the binding affinities of these non-benzodiazepines.

5. Conclusion

The present study demonstrated that both linear and nonlinear QSAR models can be used for the prediction of binding affinities of a set of non-benzodiazepines to BzR based on descriptors calculated from molecular structures alone. Satisfactory results were obtained with the two proposed methods. The proposed linear model could identify and provide some insight into what structural features were related to binding affinities of those compounds. Additionally, nonlinear RBFNN model showed better predictive ability. The advantages of this approach lie in the fact that it requires only the knowledge of chemical structure and is not dependent on any experimental properties once the model is built.

Furthermore, the proposed approaches confirmed that the interaction between the receptors and ligands are dependent on the properties of the special groups of ligands, not the whole ligand structure. The polarity, size, hydrophobicity and conformation of the groups play key roles in binding. Again, the method can be extended to other QSAR investigations.

References

[1] A.J. Lu, J.J. Zhou, J. Chem. Inf. Comput. Sci. 44 (2004) 1130–1136.
[2] J. Karolak-Wojciechowska, J. Lange, W. Ksiazek, M. Gniewose, S. Rump, Il Farmaco 53 (1998) 579–585.
[3] Y. Chumakov, A. Terletskaya, A. Dimoglo, S. Andronati, Quant. Struct.-Act. Relat. 19 (2000) 443–447.
[4] D. Hadjipavlou-Litina, R. Garg, C. Hansch, Chem. Rev. 104 (2004) 3751–3793.
[5] H. Verli, M.G. Albuquerque, R.B. Alencastro, E.J. Barreiro, Eur. J. Med. Chem. 37 (2002) 219–229.

[6] L. Teuber, F. Watjen, L.H. Jensen, Curr. Pharm. Des. 5 (1999) 317−343.

[7] G. Palazzino, L. Cecchi, F. Melani, V. Colotta, G. Fillachione, C. Martini, A. Lucacchini, J. Med. Chem. 30 (1987) 1737−1742.

[8] Q. Huang, R.Y. Liu, P.W. Zhang, X.H. He, R. McKernan, T. Gan, D.W. Bennett, J.M. Cook, J. Med. Chem. 41 (1998) 4130−4142.

[9] S.P. Gupta, A. Paleti, Quant. Struct.-Act Relat. 15 (1996) 12−16.

[10] A. Paleti, S.P. Gupta, Quant. Struct.-Act Relat. 16 (1997) 367−371.

[11] S.P. Gupta, A. Paleti, Bioorg. Med. Chem. 6 (1998) 2213−2218.

[12] S. Deswal, N. Roy, Eur. J. Med. Chem. 42 (2007) 463−470.

[13] W.M. Shi, Q. Shen, W. Kong, B.X. Ye, Eur. J. Med. Chem. 42 (2007) 81−86.

[14] K. Hasegawa, T. Deushi, O. Yaegashi, Y. Miyashita, S. Sasaki, Eur. J. Med. Chem. 30 (1995) 569−574.

[15] X.J. Yao, Y.W. Wang, X.Y. Zhang, R.S. Zhang, M.C. Liu, Z.D. Hu, B.T. Fan, Chemom. Intell. Lab. Syst. 62 (2002) 217−225.

[16] A. Pulido, I. Ruisanchez, F.X. Rius, Anal. Chim. Acta 388 (1999) 273−281.

[17] T. Stubbings, H. Hutter, Chemom. Intell. Lab. Syst. 49 (1999) 163−172.

[18] H. Lohniger, J. Chem. Inf. Comput. Sci. 33 (1993) 736−744.

[19] X.J. Yao, X.Y. Zhang, R.S. Zhang, M.C. Liu, Z.D. Hu, B.T. Fan, Comput. Chem. 25 (2001) 475−482.

[20] F. Luan, W.P. Ma, X.Y. Zhang, H.X. Zhang, M.C. Liu, H.D. Hu, B.T. Fan, Chemosphere 63 (2006) 1142−1153.

[21] M.U. Cuadrado, I.L. Ruiz, M.A. Gómez-Nieto, J. Chem. Inf. Model. 46 (2006) 2022−2029.

[22] ISIS Draw 2.3, MDL Information Systems, Inc., 1990−2000.

[23] HyperChem 6.01, Hypercube, Inc., 2000.

[24] J.P.P. Stewart, MOPAC 6.0, Quantum Chemistry Program Exchange, QCPE, No. 455, Indiana University, Bloomington, IN, 1989.

[25] A.R. Katritzky, V.S. Lobanov, M. Karelson, CODESSA: Training Manual, University of Florida, Gainesville, FL, 1995.

[26] A.R. Katritzky, V.S. Lobanov, M. Karelson, CODESSA: Reference Manual, University of Florida, Gainesville, FL, 1994.

[27] A.R. Katritzky, R. Petrukhin, R. Jain, M. Karelson, J. Chem. Inf. Comput. Sci. 41 (2001) 1521−1530.

[28] J. Shi, F. Luan, H.X. Zhang, M.C. Liu, Q.X. Guo, Z.D. Hu, B.T. Fan, QSAR Comb. Sci. 25 (2006) 147−155.

[29] X.J. Yao, A. Panaye, P. Doucet, R.S. Zhang, H.F. Chen, M.C. Liu, Z.D. Hu, B.T. Fan, J. Chem. Inf. Comput. Sci. 44 (2004) 1257−1266.

[30] F. Luan, X.Y. Zhang, H.X. Zhang, R.S. Zhang, M.C. Liu, Z.D. Hu, B.T. Fan, Comput. Mater. Sci. 37 (2006) 454−461.

[31] A.T. Balaban, Chem. Phys. Lett. 89 (1992) 399−404.

[32] A. Afantitis, G. Melagraki, H. Sarimveis, P.A. Koutentis, J. Markopoulos, O. Igglessi-Markopoulou, Polymer 47 (2006) 3240−3248.

[33] A.T. Balaban, S.C. Basak, A. Beteringhe, D. Mills, C.T. Supuran, Mol. Divers. 8 (2004) 401−412.