Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 26 (2008) 1057–1065
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Modeling calcium channel antagonistic activity of dihydropyridine derivatives using QTMS indices analyzed by GA-PLS and PC-GA-PLS

Afshan Mohajeri [a], Bahram Hemmateenejad [a,b,*], Ahmadreza Mehdipour [b], Ramin Miri [b]

[a] Department of Chemistry, Shiraz University, Shiraz, Iran
[b] Medicinal & Natural Products Chemistry Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

Received 22 April 2007; accepted 8 September 2007
Available online 14 September 2007

Abstract

The usefulness of a novel type of electronic descriptors called quantum topological molecular similarity (QTMS) indices for describing the quantitative effects of molecular electronic environments on the antagonistic activity of some dihydropyridine (DHP) derivatives has been evaluated. QTMS theory produces a matrix of descriptors, including bond (or structure) information in one dimension and electronic effects in another dimension, for each molecule. Some different modeling tools such as multiple linear regression (MLR), principal component analysis (PCA), partial least squares (PLS) and genetic algorithms (GA) were employed to find some appropriate models for noted biological activity. GA was used in order to select the proper variables and also PCA was used for data compression. Then modeling was performed by MLR and PLS. The model performances were accessed by both cross-validation and external prediction set. The results showed that the proposed models could explain above 90% of variances in the biological activity. The significant effects of chemical bonds on the antagonistic activity were identified by calculating variable important in projection (VIP). It was obtained that those belonging to the substituted 4-phenyl ring represent high influence on the biological activity which, confirms their importance in mechanism of action of DHP derivatives.

Keywords: Quantum topological molecular similarity (QTMS); Calcium channel; DHP; QSAR; PLS

1. Introduction

Quantitative structure–activity relationship (QSAR) research field has been widely developed because of its powerful ability to predict drug activity [1,2]. QSAR models generally are mathematical equations relating chemical structure to their biological activity. The official birth date of QSAR is considered to be 1962, when Hansch et al. published a paper which showed a correlation between biological activity and octanol–water partition coefficient [3]. QSAR models have another ability which is obtaining a deeper knowledge about the mechanism of biological activity. This can be helpful for finding active site of action and also finding a basic structure for drug design.

A major step in constructing the QSAR models is finding appropriate molecular descriptors that represent variations in structural property of molecules quantitatively. Nowadays, a wide range of descriptors has been used in QSAR modeling [4]. These descriptors have been classified into different categories according to Karelson approach including constitutional, geometrical, topological, quantum chemical and etc. [5]. Among these categories, quantum chemical descriptors have attracted many researchers because of recent progress in computational hardware and also development of efficient algorithms.

Recently, a topological approach according to the theory of ''atoms in molecules'' (AIM) [6,7] was introduced by Popelier and his research group called quantum topological molecular similarity (QTMS) indices which aim to represent a molecules based on the topology of its electron density [8]. For first time, these indices were successfully employed to predict Hammett σ values of *para*- and *meta*-benzoic acids [9]. One of QTMS's main abilities is to localize the active centre which is the major goal of descriptive QSAR modeling. QTMS is also able to rank bonds which are important in biological activity/chemical property of interest [10]. It has been demonstrated that QTMS offers a reliable alternative to electronic parameters so it can be applied to QSAR/QSPR modeling [7–13].

* Corresponding author at: Department of Chemistry, Shiraz University, Shiraz, Iran. Tel.: +98 711 228 4822; fax: +98 711 228 6008.
E-mail address: hemmatb@sums.ac.ir (B. Hemmateenejad).

1093-3263/$ – see front matter
doi:10.1016/j.jmgm.2007.09.002

The 1,4-dihydropyridine (DHP) derivatives, known as calcium channel antagonists, are used for treatment of cardiovascular diseases such as hypertension and angina pectoris [14]. The DHP family, in which nifedipine is the prototype, have been the aim of many SAR [15–17] and QSAR [18–23] studies. Lately, our research group parallel to synthesis of new derivatives of DHP [24–26], has performed an extensive investigation on DHP's QSAR modeling by different classes of molecular descriptors including topological, geometrical and especially quantum chemical parameters [27,28] and also applied various types of modeling like GA-MLR [29], GA-PLS [29], GA-ANN [28] and GA-PC-ANN [27].

In this article, we applied the QTMS method to drive parameters for QSAR study of 41 DHP derivatives whose calcium channel antagonist activity was reported by Coburn et al. [18]. MLR, GA-PLS and GA-PC-PLS were operated to model structure–activity relationship aiming to find an active centre for calcium channel blocker activity of the DHP derivatives.

2. Methodology

2.1. Biological activity data

The biological data used in this study are the calcium channel antagonist activity in guinea pig Ileal, $-\log(IC_{50})$, of a set of 41 DHP derivatives. The biological activity of these compounds was previously reported by Coburn et al. [18]. This data set has been used by the others for different QSAR studies [18–23]. In Table 1 the structural features and the biological activity of the DHP derivatives used in this study are represented.

2.2. Computational procedure

The theoretical bases of QTMS indices have been extensively described by Popelier and coworkers [7–13,30,31] and herein we described them briefly. The theory of ''atoms in molecules'' retrieves chemical insight from electronic wave functions [8]. Bond critical points (BCPs) are defined as points in real 3D space where the gradient of ρ vanishes ($\bigtriangledown\rho = 0$). A BCP is characterized by sign of its principal curvature of ρ-one positive and two negative curvatures. Another useful quantity to characterize a bond is the Hessian of the charge density. The Hessian is a matrix describing all possible second derivatives of ρ with respect to coordinates. Diagonalization of the Hessian yields three eigen-values, λ_1, λ_2 and λ_3. The Laplacian of density, $\bigtriangledown^2\rho$, at the BCP is the sum of its three principal curvatures at each point in space: $\bigtriangledown^2\rho = \lambda_1 + \lambda_2 + \lambda_3$ that provides a measure of the extent to which the charge density is locally compressed or expanded in a bond.

Another quantity derived from the Hessian eigen-values is the ellipticity of a bond (ε) at the BCP and is defined as: $\varepsilon = \lambda_1/\lambda_2$. Ellipticity provides the measure of the extent to which charge is accumulated in a given plane and can be used as a quantitative index of the π-character of a bond [31]. Bonds can be further characterized by evaluating of two types of kinetic energy densities denoted by Lagrangian kinetic energy $G(r)$ and Hamiltonian kinetic energy $K(r)$:

$$G(r) = \frac{1}{2N}\int \mathrm{d}\tau\prime\nabla\psi^*\nabla\psi$$

$$K(r) = -\frac{1}{4N}\int \mathrm{d}\tau\prime\nabla\psi[\psi^*\nabla^2\psi + \psi\nabla^2\psi^*]$$

where N is the number of electrons and $N\int \mathrm{d}\tau'$ summarizes the one electron integration mode. The quantities introduced above can be used together as chemical descriptors for a bond.

Calculation of QTMS indices is preceded in two steps. In the first step a geometry optimization is performed for each molecule to obtain structural parameters and wave functions at a level of theory. Here, the molecules optimized at the semi-

Table 1
Experimental and predicted values of $\log(1/IC_{50})$

No.	X	$\log(1/IC_{50})_{Exp}$	$\log(1/IC_{50})_{Pred}$ [a]	$\log(1/IC_{50})_{Pred}$ [b]
1	H	7.88	7.13	7.51
2	2-CF_3	8.82	8.80	8.82
3	2-CN	7.80	7.68	7.94
4	2-Et	7.72	7.60	7.59
5	2-F	7.37	7.38	7.77
6	2-Vinyl	8.35	7.80	8.31
7	2-NO_2	8.29	8.39	8.32
8	2-Cl	8.16	8.40	8.02
9	2-Oet	7.33	7.56	7.20
10	2-NH_2	4.40	5.01	5.59
11	2-Me	8.22	8.07	8.12
12	2-Br	8.12	8.03	7.92
13	2-Ome	7.24	6.92	7.55
14	3-Br	8.89	8.84	8.96
15	3-Cl	7.80	7.44	7.55
16	3-F	7.76	7.77	8.00
17	3-Me	6.96	7.16	6.83
18	3-NO_2	8.40	8.07	7.51
19	3-Ome	6.72	6.92	6.17
20	3-OH	6.00	5.47	5.39
21	3-Oac	5.22	5.71	6.42
22	3-NMe_2	6.05	6.11	5.65
23	3-CF_3	7.13	7.12	7.28
24	3-CN	7.46	7.60	8.01
25	3-NH_2	5.70	5.67	5.69
26	4-Cl	5.09	5.32	5.26
27	4-CN	5.46	5.60	5.45
28	4-NMe_2	4.00	4.17	4.05
29	4-NO_2	5.50	5.73	5.27
30	4-F	6.89	6.85	7.29
31	4-Br	5.40	5.00	4.81
32	2,6-Cl_2	8.72	8.84	8.70
33	F_5	8.36	8.20	8.57
34	2-F-6Cl	8.12	7.89	7.90
35	2,3-Cl_2	7.72	8.46	7.92
36	2-Cl-5-NO_2	7.52	6.98	7.50
37	3,5-Cl_2	7.03	6.87	7.21
38	2-OH-5NO_2	7.00	7.85	6.74
39	2,4,5-Ome_3	3.00	2.70	2.95
40	2,5-Me_2	7.00	7.43	6.99
41	2,4-Cl_2	6.40	6.36	6.18

[a] Predicted by GA-PLS.
[b] Predicted by GA-PC5-PLS.

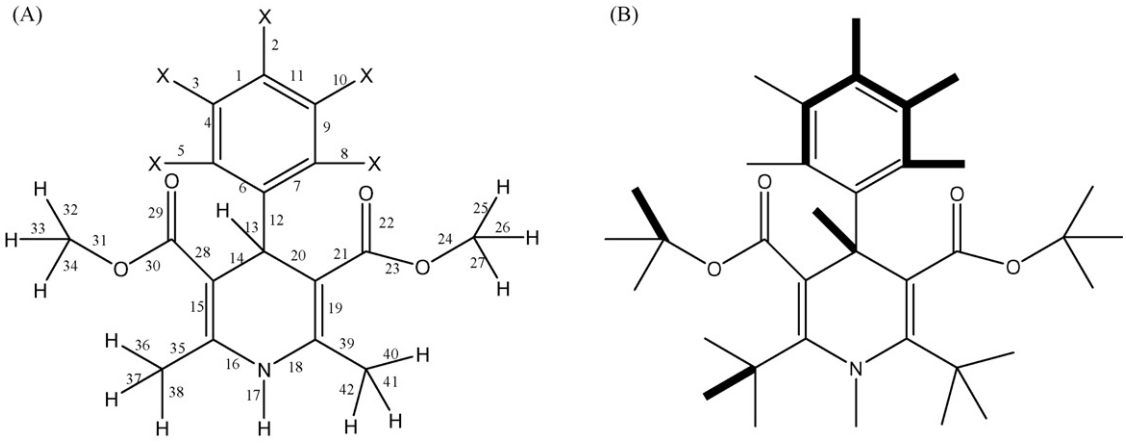

Fig. 1. Structural backbone of the DHP derivatives used in this study: (A) bond numbering and (B) highlighted highly influential bonds.

empirical PM3 method which followed by a density functional B3LYP/6-311++G** calculation. All calculations were carried out using GAUSSIAN 98 program [32]. In the second step, wave functions calculated at the B3LYP/6-311++G** were used for the topological analyses of the electron densities. The AIM2000 program [33] was employed for calculating the bond critical points and visualizing the bond paths. In this step the BCPs are located for each individual bond in the molecule. Analysis of the electron densities produces eight properties (including $G(r)$, $K(r)$, ρ, $\bigtriangledown^2\rho$, λ_1, λ_2, λ_3 and ε) for each BCP.

2.3. Model development

As noted previously, a total of 41 DHP derivatives were used in this study (Table 1), among which 10 molecules were randomly selected as external test set and the rest was used in model development step.

Each QTMS property can be used as a descriptor variable. Thus, for each molecule a matrix of QTMS descriptor with dimension of $(8 \times n)$ was obtained, where n is the number of chemical bonds in the basic molecular skeleton. The structural backbone of the DHP derivatives (Fig. 1) contains 42 chemical bonds hence the QTMS data matrix of each molecule has a dimension of (8×42), and the total number of calculated descriptors for each molecule is equal to 336. Therefore, the descriptor data matrix of entire set of molecules has a dimension of (41×336).

Data analysis was followed by MLR and PLS [34,35] methods. MLR was used to account the separate effects of each bond and each QTMS indices on the calcium channel antagonist activity of the DHP derivatives. To do so, for each chemical bond and each QTMS index separate QSAR models was developed. In the case of chemical bonds, the descriptor data matrix was the set of QTMS descriptors belonging to that bond and hence had a dimension of (41×8). In the case of each QTMS index, the descriptor data matrix was composed of that property for all chemical bonds in the structural backbone of the DHPs and therefore had a dimension of (41×42) for entire set of molecules.

The input (or predictor variables) of PLS model was the original QTMS indices unfolded to a row vector for each molecules. Therefore, the dimension of descriptor data matrix for PLS analysis was (41×336) for entire set of molecules. In another trial, the matrix of QTMS indices of an individual molecule was subjected to principal component analysis (PCA) and its most significant principal components (or eigen-vectors) were considered as the input variables of that molecule for PLS analysis. In both methods, PLS regression was achieved using entire set of input variables as well as using the subset of variables selected by genetic algorithm (GA). The model refinement procedure was the use of leave-one-out cross-validation method to select the optimum number of PLS latent variables; also a set of samples was put out as an external test set for checking the prediction ability of resulted model.

The GA used here was similar to those we used in previous publications [36–38]. The presence or absence of a descriptor in a chromosome was coded by 1 and 0, respectively. So, in the present study, each string was composed of 336 genes. The characteristics of different GA runs were as below: the population size was between 50 and 200, the probability of generating 0 for a gene was set greater than for 1 (at least 70%), cross-validation process was used as evaluation function of models. The strings with the least number of selected variables and the highest fitness were marked as informative strings. These strings were kept for further processing (i.e. mutation and cross over) in the next generations. The runs were terminated when 90% of the strings reached to same fitness.

The subroutines for doing PCA and PLS was written in MATLAB (Mathwork Inc., version 7). For variable selection by GA, the PLS-toolbox developed by eigen-vector company (Eigenvector Research, Inc.) was employed. A Pentium IV personal computer with windows XP operating system was used throughout.

3. Results and discussion

QTMS theory provided a matrix of descriptors for each molecule whose number of rows & columns is equal to number

of bond and number of QTMS, respectively. As it is observed from Fig. 1, the structural backbone of DHP derivatives contains 42 bonds (i.e. 15 C–C bonds, 13 C–H bonds, 5 C–X bonds, 6 C–O bonds, 2 C–N bonds and 1 N–H bond). It should be noted that for the sake of simplicity, in Fig. 1 numbering was used for chemical bonds instead of atoms. It should be note that the chemical bonds will be denoted by Bi through the rest of manuscript, so that bond number 1 of Fig. 1 will be addressed as B1 and so on. Since eight QTMS parameters were calculated for each bond, matrix of descriptors for each molecule had a dimension of (8×42) and the total number of calculated descriptors for each molecule was equal to 336. Since the modeling methods employed in this work (i.e., MLR and PLS) need two dimension array of data, it was necessary to convert the three-dimensional array to a two-dimensional one by unfolding of QTMS data of each molecule to a row vector.

Collinearity between the variables can introduce instability into QSAR models based on MLR analysis. In addition, if a high correlation exists between the descriptors, it is impossible to delimitate the effects of each descriptor in MLR analysis. Therefore, MLR models should be carried out only using uncorrelated descriptors. Meanwhile, this analysis reduces the dimension of the data set before applying PLS analysis. To do so, correlation between each one of descriptors and with calcium channel antagonist activity data was calculated. The co-linearity threshold was set to $R^2 > 0.9$. Among the detected collinear descriptor, one of them representing higher correlation with biological activity was retained and the rest was removed. Therefore, PLS and MLR analyses were applied to non-collinear descriptors.

To handle QTMS indices for QSAR study of the calcium channel antagonist activity of the studied DHP derivatives, some different approaches were employed, whose results will be given in the following sections.

3.1. QSAR models for chemical bonds

According to the suggestions made by Popelier and coworkers [7–13,30], the biological activity of a set of molecules can be attributed to the change in electronic features of selected bonds in molecule and none all of bonds have significant contribution to the biological activity of interest. Therefore, we firstly attempted to find which chemical bonds have significant contribution in the calcium channel blocker activity of the studied DHP derivatives. To do so, separate descriptors data matrices were prepared for each bond and therefore separate QSAR model were obtained that bond. The descriptor data matrix of each bond is composed of a set of eight QTMS indices calculated for that bond. MLR with stepwise selection and elimination of variables was used to find the most convenient QSAR model for each bond. The chemical bonds, for which significant QSAR models have been obtained, are listed in Table 2. The models were evaluated by leave-one-out cross-validation as well as using of an external test set. As is observed, the QSAR models obtained from some chemical bonds could explain more than 40% of variances in the calcium channel antagonist activity of DHPs. Five bond including B1, B4, B11, B14 and B17 represented significant QSAR models with moderate statistical quality. Some other bonds like B2, B5, B16, B18 and B20 showed poor quality models, and the rest exhibited models with very poor quality or even no significant models. This finding suggests that the calcium channel antagonist activity of the studied molecules cannot be significantly affected by a single bond in the molecule; rather a combination of chemical bonds should be used to describe this biological activity.

Table 2
Summary of some QSAR equations obtained from analysis of chemical bonds

No.	Bond	Selected QTMS indices	R^2	Q^2	R^2_P
1	B1	λ_2, λ_3	0.511	0.466	0.57
2	B2	λ_2	0.247	0.178	0.212
3	B4	λ_2, λ_3	0.453	0.351	0.210
4	B5	$\bigtriangledown^2\rho$, $K(r)$	0.333	0.238	0.171
5	B11	λ_2, ε	0.458	0.351	0.600
6	B13	ε	0.224	0.147	0.640
7	B14	$\bigtriangledown^2\rho$, λ_1	0.395	0.282	0.302
8	B16	$G(r)$	0.239	0.147	0.382
9	B17	λ_2, ε, $G(r)$	0.437	0.111	0.598
10	B18	ρ	0.278	0.192	0.439
11	B19	ρ	0.156	0.100	0.660
12	B20	$K(r)$	0.327	0.259	0.536
13	B21	λ_3	0.214	0.125	0.168
14	B32	$G(r)$	0.272	0.181	0.356

Among the chemical bonds represented much higher impact on the calcium channel antagonist activity, three of them (i.e., B1, B4 and B11) belongs to the 4-substituted phenyl ring, and the two others are, i.e. B14 and B17, are the own of DHP ring. Table 2 shows that among the QTMS descriptors selected for the chemical bonds, λ_2 has the major popularity, especially for the more significant bonds. This describes the significance of this QTMS descriptor in calcium channel antagonist activity.

3.2. QSAR models for QTMS descriptors

In the next step, we tried to find a significant model for each QTMS parameter considering its values for 42 different chemical bonds of DHP's structural backbone. Hence, for each QTMS index, the descriptor data matrix had a dimension of (41×42). The strategy for data analysis was the same as that used in the previous section. The resulted QSAR models for all QTMS indices are summarized in Table 3. This table shows correlation coefficients for calibration, prediction and cross-validation (R^2, R^2_p and Q^2, respectively), and also the bonds selected used by each QSAR model. As it is observed, significant QSAR models have been obtained for all QTMS descriptors. Whilst, the QSAR models obtained for λ_2, λ_3 and $K(r)$ represents high statistical quality, explaining more than 70% of variances in calcium channel antagonist activity of the studied DHPs, the QSAR models of the other QTMS descriptors represent moderate statistical quality. This supports the significance of almost all of QTMS descriptors for describing calcium channel antagonist activity of DHP derivatives. The results obtained in this section and those found from chemical bonds in the previous section indicate that

Table 3
Summary of some QSAR equations obtained from analysis of QTMS indices

No.	QTMS	Selected bonds	R^2	Q^2	R^2_p
1	ρ	B11, B2	0.482	0.393	0.469
2	$\bigtriangledown^2\rho$	B10, B13	0.316	0.187	0.344
3	λ_1	B17, B2, B36	0.586	0.443	0.553
4	λ_2	B1, B21, B2, B36	0.714	0.56	0.351
5	λ_3	B1, B4, B12, B18, B2, B36	0.739	0.565	0.56
6	ε	B4, B20	0.453	0.364	0.367
7	$G(r)$	B12, B20, B31	0.521	0.452	0.542
8	$K(r)$	B18, B19, B2, B8, B27, B40	0.774	0.420	0.390

neither individual bond nor a specific QTMS descriptor can be used to explain the total variances in the calcium channel antagonist activity. Instead, a combination of bond and QTMS information should be used.

3.3. PLS-based QSAR models

In the third approach for analyzing the QTMS descriptors, both bond information and descriptor information were used to generate the QSPR models. For this purpose, the matrix of the calculated descriptors of each molecule was unfolded to a row vector by collecting the QTMS descriptors of each bond beside each others. In this case, all calculated descriptors are used in the QSAR model development. According to the discussion given in the previous section, by using both bond and QTMS information higher variances in the calcium channel antagonist activity data will be described. Since the number of calculated descriptors (336 descriptors) is much higher than the number of molecules (41 molecules), PLS regression was employed instead of MLR analysis.

In the first step for PLS modeling, it was performed using all 336 calculated descriptors without variable selection. As accepted procedure of refinement process in selecting the optimum number of PLS latent variables (LVs), leave-one-out cross-validation (LOO-CV) was used [35]. The model predictivity was investigated by using an external test set composed of 10 molecules. The resultant statistical parameters are listed in Table 4. As it is observed, this model has high calibration statistical quality (i.e. $R^2 = 0.948$) whereas it has low prediction ability measured by cross-validation and using a separate test set (i.e. Q^2 and R^2_p are equal to 0.483 and 0.501, respectively). A large difference between the prediction and calibration statistics (high calibration and low prediction statistics) indicates the presence of over-fitting problem in the resulted model [39]. Two main sources of over-fitting in the factor analysis-based regression method are the use of higher number of latent variables and higher number of predictor variables than they are necessary [39]. Here, the number of latent variables was optimized by cross-validation procedure, which currently is a standard method for optimizing model complexity. The use of calibration statistics in optimizing the PLS latent variables generally leads to over-fitted model. As an example, the variation of the root mean square of errors (R.M.S.E.) from calibration data or from cross-validation as a function of the PLS latent variables has been plotted in Fig. 2. The plots show the gradual decreasing of R.M.S.E. of calibration while number of latent variables is increased. In the other hand the plot of R.M.S.E. of cross-validation as a function of latent variables is passed through a distinct minimum at number of latent variables of 5. Thus, getting over-fitted model cannot be attributed to the using higher number of latent variables. Accordingly, it can be concluded that the major source of over-fitting in the mentioned above PLS model is using higher numbers of predictor variables than they are necessary. As it was discussed in the previous sections, none of

Table 4
Statistical parameters for different PLS models

Model	LVs	R^2	Q^2	R^2_p
PLS	5	0.948	0.483	0.500
GA-PLS	7	0.952	0.819	0.921
PC5-PLS	9	0.971	0.300	0.468
PC1-PLS	6	0.857	0.336	0.238
GA-PC1-PLS	9	0.833	0.723	0.514
GA-PC5-PLS	7	0.923	0.828	0.918

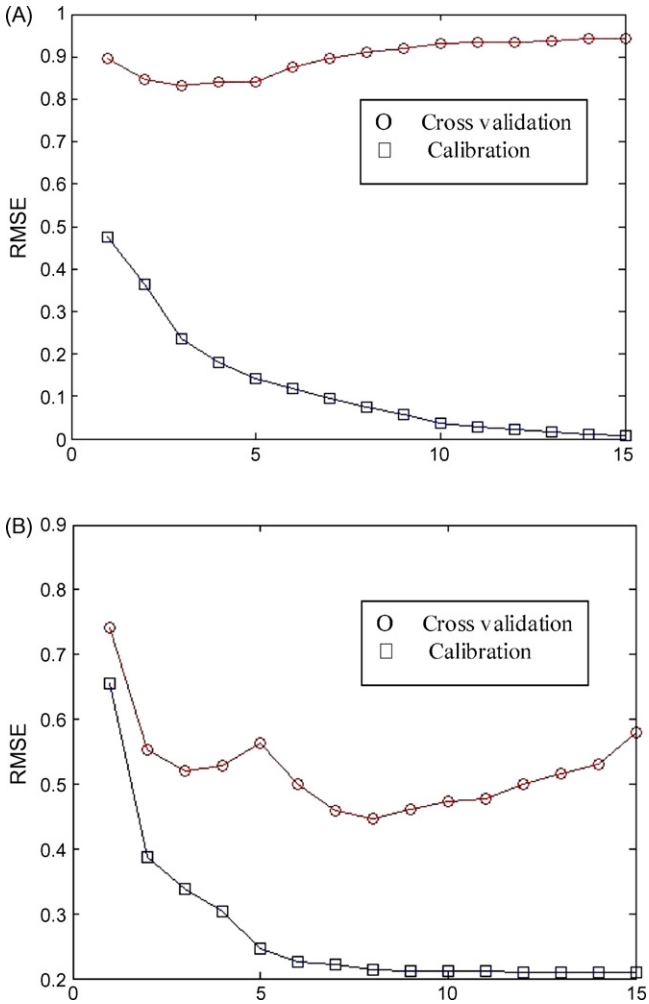

Fig. 2. Variation of R.M.S.E. as the function of PLS-latent variables for (A) PLS model and (B) GA-PLS model.

entire set of calculated descriptors represented significant effects on the calcium channel antagonist activity of DHP derivatives. Therefore, a subset of descriptors, describing structure–activity relationships of DHPs in a good manner, should be selected. Detecting these variables will help us to obtain a more deep insight about the significance of QTSM properties and molecular bonds in the biological process under study.

Nowadays, many different methods are used for variable selection in QSAR studies in order to find suitable parameters for modeling the biological process [40–42]. Among these methods, genetics algorithm was chosen due to its popularity in QSAR studies [36–38]. Genetic algorithm is a problem solving method which uses genetic rules and evolution theory to build pseudo organisms for variable selection on the basis of a fitness criterion to survive and pass information on to the next generation. The statistical quantity of the most convenient GA-PLS model that produced the best cross-validation and prediction statistics is represented in Table 4. GA selected 25 variables for PLS modeling. According to cross-validation examination the optimum number of factors was set to 8 (Fig. 2B). It is clearly observed from Table 4 that GA-PLS has similar calibration statistics with respect to the conventional PLS; however, the former has improved cross-validation and prediction statistics. The respective squared correlation coefficients for cross-validation and prediction are 0.819 and 0.921, which indicates the ability of the model to reproduce about 90% of variances in biological data.

For measuring the importance of selected QTMS descriptors, variable importance in projection (VIP) was calculated for each descriptor [43]. According to Erikson et al., predictor variables could be classified to their relevance in explaining Y (dependent variable) [44] so that variables with VIP values more than 1 indicated as highly influence while those with VIP values less than 0.8 considered as the least influential. VIP values lying between 0.8 and 1 mean moderate influences. The calculated VIP values for the descriptors selected by GA-PLS method are represented in Fig. 3. Obviously, eight parameters have VIP values greater than 1 which can be considered as highly effective parameters in antagonistic activity of DHP derivatives. These parameters in the order of their decreasing

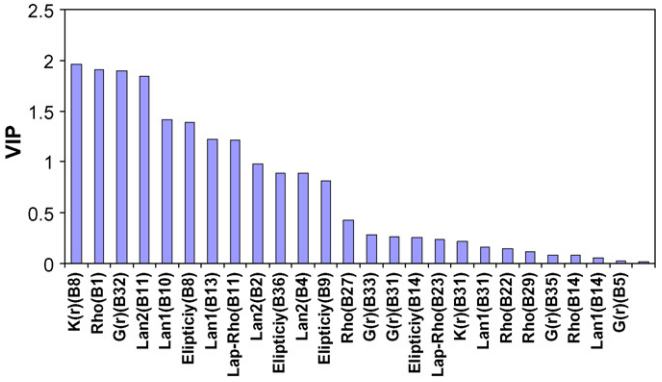

Fig. 3. Plot of variable important in projection (VIP) for the QTMS descriptors of the chemical bonds given by GA-PLS model.

importance are $K(r)$ (B8), ρ (B1), $G(r)$ (B32), λ_2 (B11), λ_1 (B10), ε (B8), λ_1 (B13) and $\bigtriangledown^2\rho$ (B11). Also, four descriptors including, λ_2 (B1), ε (B36), λ_2 (B4) and ε (B9), represent moderate effects. Refereing to Fig. 1 reveals that the majority of the influential chemical bonds belong to the 4-substituted phenyl ring. The B8 bond (the chemical bond jointing substituents to C-2 position of phenyl ring) was selected as the most influential one, which is in direct agreement with findings of previous authors explaining the important role of the electronic features of 2-substitution on the calcium channel antagonist activity of DHPs. B1, B14, B9, B10, B11 are the other bonds of phenyl ring identified as influential parameters. Interestingly, three C–H bonds, i.e., B13, B32 and B36 are also represented significant effect on the calcium channel antagonist activity. The QTMS indices of the influential bonds are different from one bond to the other bonds, however, among them λ_2 and λ_1 have the major popularity.

In another approach, Popelier et al. have used principal component analysis (PCA) [34] as a tool for reduction of size of data matrix of QTMS descriptors for each molecule into a principal component (PC) [8–10,30]. They used the PCs of the chemical bonds as the input of regression methods such as MLR and PLS, by which they could obtain a more deep insight about the role of chemical bonds in the system under study. The results of application of PCA to the QTMS descriptors data

Table 5
Results of application of PCA on the bond descriptor data matrices of some representative DHP derivatives

Derivative	Eigen-value					Cumulative percent of variance				
	PC1	PC2	PC3	PC4	PC5	PC1	PC2	PC3	PC4	PC5
H	11.52	1.32	0.13	0.04	0.010	88.0	98.5	99.5	99.8	99.96
2,5-Me_2	9.92	2.25	0.80	0.14	0.030	75.5	92.6	98.5	99.6	99.86
2,4,5-Ome_3	11.40	1.58	1.15	0.11	0.300	79.6	90.7	98.7	99.5	99.76
2-Me	11.48	1.27	0.13	0.03	0.010	88.7	98.6	99.6	99.9	99.94
2-Cl	10.05	1.34	0.24	0.12	0.040	85.1	96.5	98.5	99.5	99.78
2-Ome	11.70	1.34	0.15	0.02	0.005	88.4	98.6	99.7	99.9	99.99
3-CF_3	11.58	1.26	0.13	0.04	0.010	88.8	98.6	99.5	99.8	99.96
3-CN	10.60	0.90	0.16	0.08	0.020	89.9	97.6	99.0	99.7	99.90
3-NH_2	11.57	1.35	0.18	0.04	0.006	88.0	98.0	99.6	99.9	99.99
4-F	11.80	1.40	0.15	0.03	0.010	87.8	98.6	99.7	99.9	99.97
4-Br	11.46	1.30	0.14	0.03	0.006	88.5	98.6	99.7	99.9	99.99

matrix of some selected DHP derivatives are listed in Table 5. In this table the eigen-values associated with each PC and the corresponding percent of variances explained by that PC are represented. In the case of each molecule, the data are shown for the first five PCs. Obviously, in all most all case, the first PC could explain 88% of variances in the QTMS data and about 99.99% of variances could be explained by first five PCs of the QTMS data matrix of each molecule. The extracted PCs of the chemical bonds on each molecule was used as input of PLS regression to make a connection between structure and activity.

In the first trail for modeling relationships between PCs and calcium channel antagonist activity, only the first PC of each bond was used as input variables. Therefore, the predictor variable data matrix had a dimension of (41 × 42) for entire set of molecules; each column is the PC of one chemical bond in the molecules. PLS modeling was followed by entire set of predictors (PC1-PLS) as well as those selected by GA (GA-PC1-PLS). The statistical parameters of the resulted models are listed in Table 4. As it is observed, the model with entire set of variables is highly over-fitted for high calibration statistics (i.e., $R^2 = 0.857$) and very poor predictivity (i.e., $Q^2 = 0.336$ and $R^2_p = 0.238$). In addition, the models obtained by selected variables (i.e., GA-PC1-PLS) has not statistical quality as high as those of GA-PLS model (i.e., $Q^2 = 0.723$ and $R^2_p = 0.514$). This can be attributed to the fact that the first PCs have not sufficient information from the original QTMS indices. Moreover, we and other authors notified previously that the magnitude of eigen-value of PCs cannot be essentially the measure of their significant in QSAR analysis [26,36–38 and references therein].

To consider more extra information from the QTMS data matrices, PLS-based QSAR models was obtained by using first five PCs of each bonds so the predictor data matrix composed of 210 PCs for each molecule (i.e. 42 bonds each with 5 PCs). The results, obtained by PLS analysis of entire set of 210 PCs without variable selection, are given in Table 4 (PC5-PLS). Obviously, the PC5-PLS model is highly over-fitted because of its high statistical quality for calibration set ($R^2 = 0.971$) and bad results for cross-validation and prediction set ($R^2_{CV} = 0.3$, $R^2_P = 0.468$). In the other hand, the model obtained after selecting significant variables by GA (GA-PC5-PLS model) represented very good performances, close to those found by GA-PLS model. This model, which used 23 PCs as predictor variable could explain and reproduce in average about 90% of variances in the calcium channel antagonist activity data

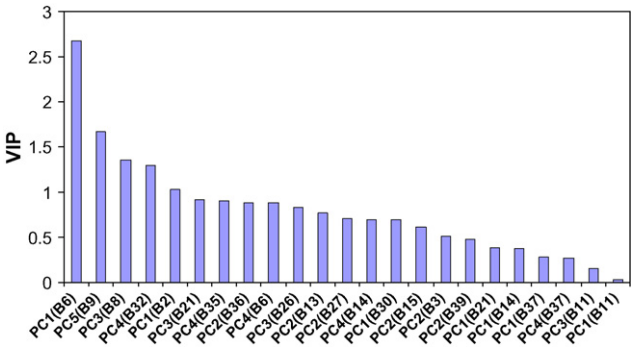

Fig. 4. Plot of variable important in projection (VIP) for the principal components of the bonds given by GA-PC5-PLS model.

employing seven numbers of PLS latent variables. The calculated VIP values for the PCs used in the GA-PC5-PLS (Fig. 4) indicates that among the 32 selected PCs, five of them, including PC1 (B6), PC5 (B9), PC3 (B8), PC4 (B32) and PC1 (B2), have been identified as highly influential parameters. Interestingly, the chemical bonds whose PC have been selected as highly influential, are similar to those obtained by GA-PLS. The most of these bonds belong to the 4-substituted phenyl ring. Clearly, only for B6 and B2 the first PC are in the list of highly significant variables and for other bonds including B9, B8 and B32, the selected PCs are not the first ones. This confirms the statement that says the magnitude of eigen-value of PCs cannot be essentially the measure of their significant in QSAR analysis [24,34–36 and references therein].

According to the results obtained in this and previous sections, among 42 chemical bonds presented in the structural backbone of DHP derivatives, the variations in the electronic properties of only some bonds can affect the calcium channel antagonist activity of the DHP derivatives. These chemical bonds are highlighted in Fig. 1B. Obviously, the majority of these bonds belong to the phenyl ring. The binding mode of DHP derivatives in the calcium channel is considered to be a sandwich-like interaction, where the DHP ring entered in the calcium channel and substitutions on the C-3, C-4 and C-5 positions of the DHP ring affect such interaction. The results obtained in this work suggest that the C–H bond of methyl group attached to C-2 position of the DHP ring (i.e. B36) can also affect the calcium channel antagonist activity, and hence changing substituent effect may lead to new calcium channel blockers. In our research group we synthesized DHP derivatives

Table 6
Statistical data for comparing the efficiency of the proposed models with pervious studies

	Regression method	Descriptors used	R^2_c	R^2_{cv}	R^2_p	Ref.
1	MLR analysis	Hansch and Hammett constants	0.81	NR	NR	[18]
2	MLR analysis	Quantum chemical and Hansch and Hammett constants	0.90	NR[a]	NR	[20]
3	Least squares support vector machines	A wide variety of descriptors	0.87	0.817	NR	[22]
4	Gene expression programming	A wide variety of descriptors	0.91	0.86	NR	[23]
5	GA-MLR	*ab initio*-derived quantum chemical descriptors	0.87	NR	NR	[24]
6	GA-PLS (this work)	QTMS data	0.952	0.819	0.921	–

[a] Not reported.

with a phenyl ring on the C-2 position, and some derivatives represented comparable activity with respect to nifedipine [25].

As it was mentioned in the introduction, there are some literature reports on QSAR modeling of the calcium channel antagonist activity of DHP derivatives [18–23]. The results for some linear QSAR models are summarized in Table 6. Since in this study MLR and PLS as linear modeling methods were used, the results of nonlinear QSAR models are not included in this table. Obviously, the QSAR model obtained in this work (GA-PLS model) has higher calibration statistics and comparable cross-validation correlation coefficient. Interestingly, no external prediction results have been reported by the previous works. The model presented in this article showed very good prediction ability. In addition to the predictivity of the models obtained in this article, they represented interesting descriptive results, detecting part of molecular skeleton involved in the calcium channel antagonist activity of the DHP derivatives.

4. Conclusion

Different modeling tools including MLR, PCA, PLS and GA were used in order to find some reliable relationship equations between antagonistic activity of DHP derivatives and a novel group of chemical descriptors called quantum topological molecular similarity (QTMS) indices. As is seen in the result of GA-PLS (Fig. 1) most of bonds which had high values of VIP were bonds of phenyl ring (i.e. B8, B1, B10, and B2) which may be as indicator of importance of phenyl ring in the activity of DHP derivatives as antagonists of Ca channel. Previously this was reported by Mahmoudian and Richards [19]. They showed that conformation of phenyl ring had a significant role; so our findings were in agreement with that study and could be an additive document for importance phenyl ring in mechanism of action these derivatives.

Acknowledgements

Financial supports of this project by research councils of Shiraz University and Shiraz University of Medical Sciences are acknowledged.

References

[1] C. Hansch, D. Hoekman, H. Gao, Comparative QSAR: toward a deeper understanding of chemicobiological interactions, Chem. Rev. 96 (1996) 1045–1076.
[2] C. Hansch, A. Leo, Exploring QSAR: Fundamentals and Applications in Chemistry and Biology, ACS Publishers, Washington, DC, 1995.
[3] C. Hansch, P.P. Maloney, T. Fujita, R.M. Muir, Nature 194 (1962) 178–180.
[4] R. Todeschini, V. Consonni, Handbook of Molecular Descriptors, Wiley-VCH, Weinheim, 2000.
[5] M. Karleson, Molecular Descriptors in QSAR/QSPR, Wiley-Interscience, New York, 2000.
[6] R.F.W. Bader, Atoms in Molecules: A Quantum Theory, Oxford University Press, Oxford, 1990.
[7] M. Rafat, M. Shaik, P.L.A. Popelier, Transferability of quantum topological atoms in terms of electrostatic interaction energy, J. Phys. Chem. A 110 (2006) 13578–13583.
[8] S.E. O'Brein, P.L.A. Popelier, Quantum topological molecular similarity. Part 4. A QSAR study of cell growth inhibitory properties of substituted (E)-1-phenylbut-1-en-3-ones, J. Chem. Soc., Perkin Trans. (2002) 478–483.
[9] P.L.A. Popelier, Quantum molecular similarity. 1. BCP space, J. Phys. Chem. A 103 (1999) 2883–2890.
[10] P.L.A. Popelier, U.A. Chaudry, P.J.J. Smith, Quantum topological molecular similarity. Part 5. Further development with an application to the toxicity of polychlorinated dibenzo-p-dioxins (PCDDs), J. Chem. Soc., Perkin Trans. (2002) 1231–1237.
[11] R.J. Loader, N. Singh, P.J. O'Malley, P.L.A. Popelier, The cytotoxicity of ortho alkyl substituted 4-X-phenols: a QSAR based on theoretical bond lengths and electron densities, Bioorg. Med. Chem. Lett. 16 (2006) 1249–1254.
[12] P.L.A. Popelier, P.J. Smith, QSAR models based on quantum topological molecular similarity, Eur. J. Med. Chem. 41 (2006) 862–873.
[13] N. Singh, R.J. Loader, P.J. O'Malley, P.L.A. Popelier, Computation of relative bond dissociation enthalpies (DeltaBDE) of phenolic antioxidants from quantum topological molecular similarity (QTMS), J. Phys. Chem. A 110 (2006) 6498–6503.
[14] M.W. Wolowyk, E.E. Knaus, S. in Abraham, G. Amital (Eds.), Calcium Channel Modulator in Heart and Smooth Muscle: Basic Mechanism and Pharmacological Aspect, VCH, 1990.
[15] D.J. Triggle, Calcium-channel drugs: structure–function relationships and selectivity of action, J. Cardiovasc. Pharmacol. 18 (1991) S1–S6.
[16] D.A. Langs, Y.W. Kwon, P.D. Strong, D.J. Triggle, Molecular level model for the agonist/antagonist selectivity of the 1,4-dihydropyridine calcium channel receptor, J. Comp. Aided Mol. Des. 5 (1991) 95–106.
[17] N.R. Natale, M.E. Rogers, R. Staples, D.J. Triggle, A. Rutledge, Lipophilic 4-isoxazolyl-1,4-dihydropyridines: synthesis and structure–activity relationships, J. Med. Chem. 42 (1999) 3087–3093.
[18] R.A. Coburn, M. Wierzba, M.J. Suto, A.J. Solo, A.M. Triggle, D.J. Triggle, 1,4-Dihydropyridine antagonist activities at the calcium channel: a quantitative structure–activity relationship approach, J. Med. Chem. 31 (1988) 2103–2107.
[19] M. Mahmoudian, W.G. Richards, QSAR of binding of dihydropyridine-type calcium antagonists to their receptor on ileal smooth muscle preparations, J. Pharm. Pharmacol. 38 (1986) 272–276.
[20] A.C. Gaudio, A. Korolkovas, Y. Takahata, Quantitative structure–activity relationships for 1,4-dihydropyridine calcium channel antagonists (nifedipine analogues): a quantum chemical/classical approach, J. Pharm. Sci. 83 (1994) 1110–1115.
[21] V.N. Viswanadhan, G.A. Mueller, S.C. Basak, J.N. Weinstein, Comparison of a neural net-based QSAR algorithm (PCANN) with Hologram- and multiple linear regression-based QSAR approaches: application to 1,4-dihydropyridine-based calcium channel antagonists, J. Chem. Inf. Comput. Sci. 41 (2001) 505–511.
[22] X. Yao, H. Liu, R. Zhang, M. Liu, Z. Hu, A. Panaye, J.P. Doucet, B. Fan, QSAR and classification study of 1,4-dihydropyridine calcium channel antagonists based on least squares support vector machines, Mol. Pharm. 2 (2005) 348–356.
[23] H.Z. Si, T. Wang, K.J. Zhang, Z.D. Hu, B.T. Fan, QSAR study of 1,4-dihydropyridine calcium channel antagonists based on gene expression programming, Bioorg. Med. Chem. 14 (2006) 4834–4841.
[24] R. Miri, C.A. McEwen, E.E. Knaus, Synthesis and calcium channel modulating effects of modified Hantzsch nitrooxyalkyl 1,4-dihydro-2,6-dimethyl-3-nitro-4-(pyridinyl or 2-trifluoromethylphenyl)-5-pyridinecarboxylates, Drug Dev. Res. 51 (2000) 225–232.
[25] R. Miri, H. Niknahad, G. Vesal, A. Shafiee, Synthesis and calcium channel antagonist activities of 3-nitrooxyalkyl, 5-alkyl 1,4-dihydro-2,6-dimethyl-4-(1-methyl-5-nitro-2-imidazolyl)-3,5-pyridinedicarboxylates, Farmaco 57 (2002) 123–128.
[26] R. Miri, K. Javidnia, H. Sarkarzadeh, B. Hemmateenejad, Synthesis, study of 3D structures, and pharmacological activities of lipophilic nitroimidazolyl-1,4-dihydropyridines as calcium channel antagonist, Bioorg. Med. Chem. 14 (2006) 4842–4849.
[27] B. Hemmateenejad, M.A. Safarpour, R. Miri, F. Taghavi, Application of ab initio theory to QSAR study of 1,4-dihydropyridine-based calcium

channel blockers using GA-MLR and PC-GA-ANN procedures, J. Comput. Chem. 25 (2004) 1495–1503.
[28] M.A. Safarpour, B. Hemmateenejad, R. Miri, M. Jamali, Quantum chemical-QSAR study of some newly synthesized 1,4-dihydropyridine calcium channel blockers, QSAR Comb. Sci. 22 (2003) 993–1005.
[29] B. Hemmateenejad, R. Miri, M. Akhond, M. Shamsipur, Quantitative structure–activity relationship study of recently synthesized 1,4-dihydropyridine calcium channel antagonists. Application of the Hansch analysis method, Arch. Pharm. 335 (2002) 472–480.
[30] S.E. O'Brien, P.L.A. Popelier, Quantum molecular similarity. 3. QTMS descriptors, J. Chem. Inf. Comput. Sci. 41 (2001) 764–775.
[31] P.L.A. Popelier, Atom in Molecules. An Introduction, Pearson Education, London, 2000.
[32] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A.J. Montgomery, R.E. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, A.G. Baboul, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, W.M. Wong, J.L. Andres, C. Gonzalez, M. Head–Gordon, E.S. Replogle, J.A. Pople, Gaussian 98, Revision A. 7, Gaussian, Inc., Pittsburgh, PA, 1998.
[33] AIM2000, Version 2.0, 2002.
[34] E.R. Malinowski, Factor Analysis in Chemistry, John Wiley & Sons, New York, 2002.
[35] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemometr. Intell. Lab. Syst. 58 (2001) 109–130.
[36] B. Hemmateenejad, M. Akhond, R. Miri, M. Shamsipur, Genetic algorithm applied to the selection of factors in principal component-artificial neural networks: application to QSAR study of calcium channel antagonist activity of 1,4-dihydropyridines (nifedipine analogous), J. Chem. Inf. Comput. Sci. 43 (2003) 1328–1334.
[37] B. Hemmateenejad, Optimal QSAR analysis of the carcinogenic activity of drugs by correlation ranking and genetic algorithm-based PCR, J. Chemometr. 18 (2004) 475–485.
[38] B. Hemmateenejad, Correlation ranking procedure for factor selection in PC-ANN modeling and application to ADMETox evaluation, Chemometr. Intell. Lab. Syst. 75 (2005) 231–245.
[39] D.M. Hawkins, The problem of overfitting, J. Chem. Inf. Comput. Sci. 44 (2004) 1–12.
[40] D.M. Livingstone, D.W. Salt, Variable selection—spoilt for choice? Rev. Comput. Chem. 21 (2005) 287–348.
[41] Y. Liu, A comparative study on feature selection methods for drug discovery, J. Chem. Inf. Comput. Sci. 44 (2004) 1823–1828.
[42] S.P. Niculescu, Artificial neural networks and genetic algorithms in QSAR, J. Mol. Struct. (Theochem) 622 (2003) 71–83.
[43] M. Olah, C. Bologa, T.I. Oprea, An automated PLS search for biologically relevant QSAR descriptors, J. Comput. Aided Mol. Des. 18 (2004) 437–449.
[44] L. Erikson, E. Johansson, N. Kettaneh-Wold, S. Wold, Multi- and Mega-Variate Data Analysis. Principle and Applications, Umetrics Academy, Umea, 2001.