Available online at www.sciencedirect.com
SCIENCE DIRECT®
Polymer 45 (2004) 2045–2050
polymer
www.elsevier.com/locate/polymer

TOPS-MODE approach to predict mutagenicity in dental monomers

Aliuska Morales Helguera[a,b], Maykel Pérez González[a,c,*], Jaques Rieumont Briones[d]

[a]Department of Drug Design, Chemical Bioactives Center, Central University of Las Villas, Santa Clara, Villa Clara, Cuba
[b]Department of Chemistry, Faculty of Chemistry and Pharmacy, Central University of Las Villas, Santa Clara, Villa Clara, Cuba
[c]Service Unit, Department of Drug Design, Experimental Sugar Cane Station 'Villa Clara-Cienfuegos', Ranchuelo, Cuba
[d]Department of Physical Chemistry, Faculty of Chemistry, University of Havana, Havana City, Cuba

Received 11 December 2003; received in revised form 14 January 2004; accepted 14 January 2004

Abstract

The TOPological Sub-Strutural Molecular Design (TOPS-MODE) approach has been introduced for the study of mutagenic properties. The mutagenicity of 23 dental monomers was studied with this approach obtaining a good quantitative structure–toxicity model. For the comparison were involved four different weights in the diagonal entries of the bond matrix for selecting the best TOPS-MODE model. TOPS-MODE was used to derive the contribution of different fragments to the toxicity of studied compounds.

Keywords: Monomers; QSAR; Mutagenicity

1. Introduction

Currently, much effort is expended toward improving the quality and durability of polymer based dental restoratives [1]. These materials when prepared in situ may risk leaching of chemicals should the polymerization not be ideal [2]. Because substance such as monomers, initiators, and polymerization sensitizers must be reactive, there is also risk of interaction with genetic bio-molecules and therefore the possibility for inducing genotoxicity [3].

Since experimental determination of mutagenicity is difficult and expensive, to simplify the laboratory process, it would be desirable to use a model based on a dataset of published results to predict mutagenicity, of previously untested chemicals. Such a result could then be used to predict a suitable, more selective, dose range over which to perform the Ames assay, reducing the necessary experiments to obtain an approximate local minimum for the linear range.

QSAR models can be used to illuminate the modes of action of genotoxic agents aiding in initial laboratory investigations. Such information can be achieved from the descriptors that are identified as contributing to genotoxicity. Thence, chemical knowledge can then be combined with the specific descriptors to develop design hypotheses leading to new dental resins [2].

In the context of novel in silico methods for modeling physicochemical and biological properties of chemicals, the topological sub-structural molecular design (TOPS-MODE) approach has been introduced. The TOPS-MODE has been applied to the description of physicochemical and biological properties of organic compounds [4–10].

The successful application of this theoretical approach to the modeling of toxicological and neurotoxicological properties [6,11] has inspired us to perform an exhaustive study in order to test and/or validate TOPS MODE applicability in assessing discoveries about the human mutagenic impact.

We will show here how TOPS-MODE is able to produce good QSAR models that permit easy structural interpretation of the results in terms of group contributions to mutagenicity.

2. Materials and methods

2.1. TOPS MODE approach

Here, we use the TOPS-MODE approach to obtain

* Corresponding author. Address: Drug Design Department, Estacion Territorial de Investigaciones de la Caña de Azucar, Autopista Nacional Km 246, 53100 Ranchuelo, Cuba. Tel.: +53-42281473; fax: +53-42281230.
E-mail address: mpgonzalez76@yahoo.es (M.P. González).

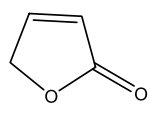
0032-3861/$ - see front matter © 2004 Published by Elsevier Ltd.
doi:10.1016/j.polymer.2004.01.022

molecular descriptors through which we developed the QSAR function. The mathematical details of the method have been largely reported [12–14], thus we will outline only the fundamental remarks.

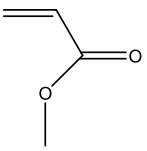
Briefly, this method codifies the molecular structure by means of the edge adjacency matrix **E** (likewise called bond adjacency matrix **B**). The **E** or **B** matrix is a square table of order m (the number of chemical bonds in the molecule) [15]. The elements of such a matrix (e_{ij}) are equal to 1 if the bonds i and j are adjacent (it means that an atom exist, which participates either in the bond *i* or in the bond *j*) or 0 otherwise. In order to codify information related to heteroatoms, the TOPS-MODE approach use $\mathbf{B}(\mathbf{w}_{ij})$ weighted matrices instead of **B**. The weights ($\mathbf{w}_{ij}$) are chemically meaningful numbers such as bond distances, bond dipole, bond polarizabilities, or even mathematical expressions involving atomic weightings such as hydrophobicity or Van der Waals radii [10,11]. These weightings are introduced in the main-diagonal of the matrix $\mathbf{B}(\mathbf{w}_{ij})$. Afterwards, the spectral moments of this matrix may be used as molecular fingerprints in QSAR studies in order to codify molecular structure. By definition, the expression 'spectral moments' must be understood as the sum of the elements in the natural powers of $\mathbf{B}(\mathbf{w}_{ij})$. It means that the spectral moment of order *k* (μ_k) is the sum of the main diagonal elements (e_{ii}) of the matrix $\mathbf{B}(\mathbf{w}_{ij})^k$. In the present work, the $\mathbf{B}(\mathbf{w}_{ij})$ matrix was weighted in the main diagonal with the bond distance, bond dipole, bond molar refractivity and atomic radius of Van Der Waals. Such a parameter equals μ_1 to sum the atomic radii of Van Der Waals, sum the atomic molar refractivity, sum the bonds dipole, or sum the bond distances in the molecule according to selected case. The calculation of the μ_k was carried out by means of the software package ModesLab 1.0 b® [16].

2.2. Data set and computational strategies

A series of 23 methacrylates and their cyclic analogs were reported by Yourtee et al. [2] and Tuparainen [17]. This set was used in the present work. As these authors report, many fur-2-enones could be used because their structures were similar to those of the methacrylates compounds [2].



Fur-2-enones



Methacrylates

Also the fur-2-enones derivatives has been selected for obtaining a training set in a reasonable amount of compounds because of the details of mutagenicity of this dental monomers are not found easily.

TOPS-MODE computer software [18] was employed to calculate molecular descriptors. The standard dipole moments, standard bond distance, molar refractivity and atomic radius of Van Der Waals were used as bond weightings for making differentiation of heteroatoms [13]. The selection of only these types of descriptors from the whole pool of 10 types included in TOPS-MODE methodology was carried out on the sake of simplicity and on the belief that polarity and steric parameters influence the mutagenicity of many compounds [2,3].

So, four sets of spectral moments were obtained, one for each used bond weightings. A brief descriptions of these schemes used in the current work are given in Table 1.

In general, 15 spectral moments were calculated for each of the studied schemes, which make a total number of 60 descriptors. We also used multiplication of spectral moments as independent variables for describing mutagenicity on these monomers. In this case, we multiplied μ_0 and μ_1 for the first 11 spectral moments obtaining 24 new variables.

However, we develop the QSAR models with each independent scheme and not with all the calculated descriptors at a time. The statistical processing to obtain the QSAR models was carried out by using the Forward stepwise regression methods [19], where the independent variables are individually added or deleted from the model at each step of the regression depending on the Fisher ratio values selected to enter and to remove until the 'best' model is obtained.

Examining the regression coefficient, cross validation leave-one-out and the proportion between the cases and variables in the equation determined the quality of the model.

3. Results and discussion

3.1. Quantitative structure association constant relations

In this work, the model selection was subjected to the principle of parsimony. Then, we choose a function with high statistical signification but having so few parameters (b_k) as possible.

Table 1
Definition of the different weighting bonds used in the current work

Weighting bonds	Definition[a]
Distance	Standard bonds distances
Dipole	Bond parameters computed with the relative electronegativity
Molar refractivity	Bond parameters computed with the molecular refractivity
Radius of Van Der Waals	Bond parameters computed with the atomic radius

[a] Consulting Ref. [28] for more complete definition of bond parameters.

Statistical parameters of the linear regression models obtained by using TOPS-MODE to describe mutagenicity are given in Table 2. The order of spectral moments that are included in such models varies from one model to another. It is due to the fact that the structural information encoded by the different weighting schemes used here is different and they have different influences on the description of the variable studied. The four models used for modeling mutagenicity are show in Table 2.

As can be seen, these models are statistically significant because of their $p < 0.05$ [20]. This confirms that all variables conforming the models are significant and essentially all of them could be used for predicting the studied property of this set of compounds. Furthermore all models have the same number of significant variables and in all of them the same training set was used which was formed by 23 compounds as it is shown in Table 3.

However, there are remarkable differences concerning the explanation of the experimental variants which give the same (R^2) where it can be seen that models obtained using the weighting spectral moments with standard dipole moment explain more than 91% of mutagenicity data variability. Thus, in our opinion, it is in fact a determining factor at the time of selecting the best model to be used later (Table 3), besides it present a greater F of Fischer ($F = 50.16$) and minor standard deviation of data ($S = 1.30$) which confirms the former selection. Predicted, observed and residual values for all compounds in training series appear in Table 3. Fig. 1 of predicted values against the observed ones mutagenicity can be seen immediately.

A cross-validation of leave-one-out type was done where it was possible to confirm that model obtained using standard dipole moments for the bond weightings had a greater coefficient of correlation (q^2) and showed a minor standard deviation (S_{cv}) for this test.

Equation of the model obtained by this bond weight is as follows:

$$\ln TA100 = -3.24 + 0.72\mu_1\mu_1 - 0.11\mu_1\mu_3 + 0.13\mu_4 - 1.01 \times 10^{-5}\mu_0\mu_{11} \quad (1)$$

In this equation, $\mu_1\mu_1$ is the square of the sum of dipole moments in the molecule, the μ_4 is the four order spectral moment, $\mu_1\mu_3$ and $\mu_0\mu_{11}$ is the multiplication of the respective spectral moment in the molecule according to selected case.

Predicted, observed and residual values for all compounds in training set, using this model, appear in Table 3. The interpretation of these models is given in a following section. However, before making this interpretation, we need to orthogonalize the molecular descriptors included in the model obtained by Yourtee et al. (Eq. (2)) to eliminate the intercorrelation existing among of them.

3.2. The orthogonalization of molecular descriptors

In order to avoid collinearity, Randić's orthogonalization procedure was carried out [21–24]. The main philosophy of this approach is to avoid the exclusion of descriptors on the basis of its collinearity with other variables previously included in the model. It is known that the interrelatedness among the different descriptors can result in highly unstable regression coefficients, which makes impossible to know the relative importance of an index and underestimates the utility of the regression coefficients in a model.

The Randić method of orthogonalization has been described in details in several publications [21–25]. Thus, we will give a general overview here.

In this sense, in the model 3 we used $H_v = {}^1O(H_v)$ as the first orthogonal variable. Afterwards, the successive residuals of the step-by-step regressions between each variable selected in the model and the others in order of statistical significance were calculated [26]. All these residuals were used as the remnant orthogonal variables in the model 3 [26]. In this analysis the least squares method selected all orthogonal analogs of collinear variables. It ensured us that, in spite of variables collinearity, each variable carry an amount of information do not encoded in the others [23,26,27].

3.3. Interpretation of QSPR models

As can be appreciated in Eq. (1), the variable $\mu_1\mu_1$ contributes to increase the property under study, indicating that an increase in the total monomer polarity can induce an increase in the mutagenic potential of this molecule. This hypothesis is not in contradiction with the behavior of the $\mu_1\mu_3$ variable, which negatively contributes to the mutagenicity of this type of structure. In this case, the third order spectral moment [29] negatively contributes too, thus it is necessary carry out a group analysis to determine which sub-structures negatively contribute to this property. In a previous study with these same monomers Yourtee et al. [2] have tried to model the mutagenicity using

Table 2
Statistical parameters of the lineal regression models for mutagenicity obtained for the four kinds of descriptors

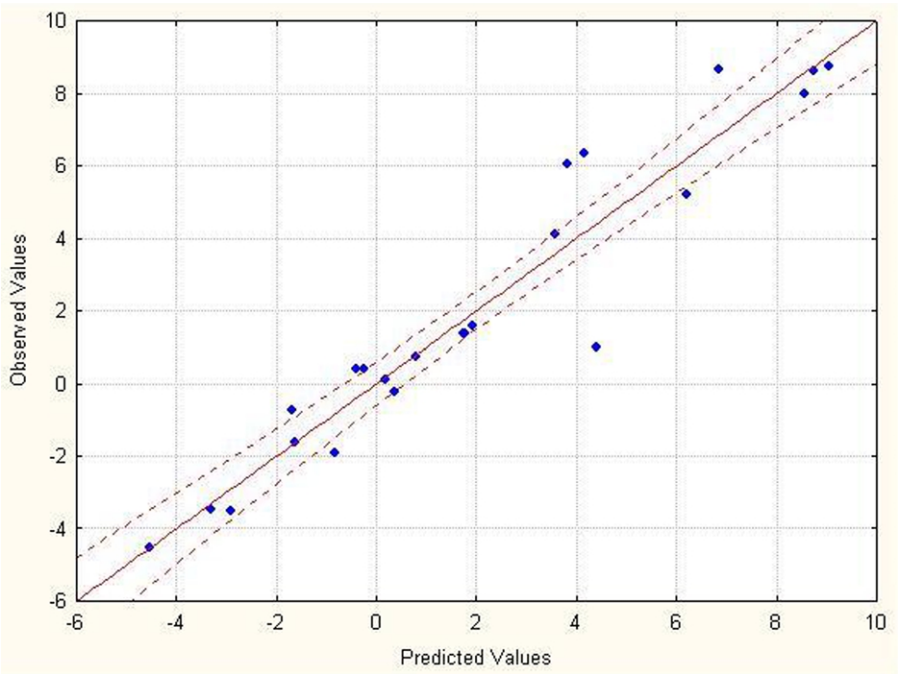
Weighting bonds	Spectral moments	N	S	R^2	F	p	q^2	S_{cv}
Distance	μ_2, μ_{11}, $\mu_1\mu_{11}$, μ_5	23	0.22	0.757	41.30	0.00	0.72	0.50
Dipole	$\mu_1\mu_1$, $\mu_1\mu_3$, μ_4, $\mu_0\mu_{11}$	23	0.13	0.918	50.16	0.00	0.89	0.28
Molar refractivity	μ_7, μ_{13}, μ_{15}, $\mu_1\mu_{13}$	23	0.26	0.490	36.58	0.00	0.40	0.68
Radius of Van Der Waals	μ_0, μ_3, $\mu_2\mu_{11}$, $\mu_1\mu_9$	23	0.15	0.830	45.25	0.00	0.79	0.42

Table 3
Predictions, residuals and names of the monomeric compounds in the training set

Name	Observed (ln TA100)	Predicted (ln TA100)	Residual (ln TA100)
Urethane dimethacrylate	−3.47	−3.29	−0.17
Glicidyl methacrylate	−1.92	−0.81	−1.10
Bisphenol A dimethacrylate	−4.51	−4.54	0.03
Glicidyl acrylate	−0.75	−1.67	0.92
2-Chloro-3-dichloromethyl-4-hydroxyfur-2-enone	8.75	9.05	−0.30
2-Chloro-3-dichloromethyl-4-methoxyfur-2-enone	8.65	6.84	1.80
2-Chloro-3-dibromomethyl-4-hydroxyfur-2-enone	8.61	8.74	−0.13
2-Bromo-3-dibromomethyl-4-hydroxyfur-2-enone	7.97	8.56	−0.59
2-Chloro-3-chloromethyl-4-hydroxyfur-2-enone	6.36	4.15	2.20
2-Bromo-3-bromomethyl-4-hydroxyfur-2-enone	6.04	3.82	2.21
2-Chloro-3-dibromomethylfur-2-enone	5.20	6.21	−1.01
2,3-Dichloro-4-hydroxyfur-2-enone	4.09	3.58	0.50
2-Chloro-3-chloromethylfur-2-enone	1.59	1.91	−0.32
2-Chloro-3-bromomethylfur-2-enone	1.37	1.74	−0.37
2-Bromo-3-chloromethylfur-2-enone	1.37	1.77	−0.40
2,3-Dichloro-4-methoxyfur-2-enone	0.99	4.41	−3.42
2-Chloro-3-methyl-4-ethoxyfur-2-enone	0.74	0.79	−0.05
2-Chloro-3-methyl-4-hydroxyfur-2-enone	0.41	−0.25	0.66
2-Bromo-3-methyl-4-hydroxyfur-2-enone	0.41	−0.39	0.80
2,3-Dichlorofur-2-enone	0.11	0.17	−0.06
3-Chloro-4-ethoxyfur-2-enone	−0.22	0.36	−0.58
2-Chloro-4-hydroxyfur-2-enone	−1.60	−1.62	0.02
3-Methyl-4-hydroxyfur-2-enone	−3.51	−2.91	−0.59

quantum-mechanical descriptors. In this study, only 22 compounds were used for the training set, whereas the whole set was of 23 compounds. An explanation for letting aside this compound was not offered. In our model this compound was not included leading to a correlation coefficient of 0.9518. However, from a practical point of view this decision is not quite good because the restricted number of compounds considered. On the other hand, these authors do not take into account the possible collinearity between the variables. This fact can lead to mistaken hypothesis as has been proved [26].

The equation that describe the mutagenicity of ‘dental’



Graphic 1. The linear relation between observed and predicted mutagenicity in dental monomers for the Eq. (1).

monomers according to Yourtee et al. for this family of compounds is the following:

$$\ln TA100 = -1.98 + 0.06H_v - 4.91F_2 - 24.39Q_{R-} - 0.79Q_{R+} \quad (2)$$

In this equation, H_v is the vibrational enthalpy, F_2 is the fractional partial positively charged surface area, Q_{R-} is the relative negative charge of the most negatively charged atom and Q_{R+} is the relative surface area of the most positively charged atom.

In order to show the existing collinearity between independent variables was carried out a correlation between them. For F_2 and Q_{R+} was found $R^2 = 0.5302$, and for H_v and Q_{R-} was obtained $R^2 = 0.3751$ indicating collinearity in some extent between these independent variables of the model. It means that these variables contain information of each other. The weightings of these variables in the model was determined using the method of forward stepwise and for the orthogonalization of them was used the method of Randić as we already explain in the previous section.

After the orthogonalization of the variables of Eq. (1) is obtained an 'orthogonalized model'. Independent variables of this model do not show statistical dependence between them.

Orthogonalization of such variables leads to the Eq. (3):

$$\ln TA100 = -2.32 + 3.22OHv + 1.98OF_2 - 0.67OQ_{R-} - 0.49OQ_{R+} \quad (3)$$

Differences in the sign of the coefficients can be appreciated between both equations. This result leads us to think that a chemical explanation for the contribution of these descriptors is not realiable if a process of orthogonalization is not carried out.

To complete our mutagenicity study of this set of monomers is absolutely necessary to know which are the possible contributions of some chemical groups that form part of the training set under study.

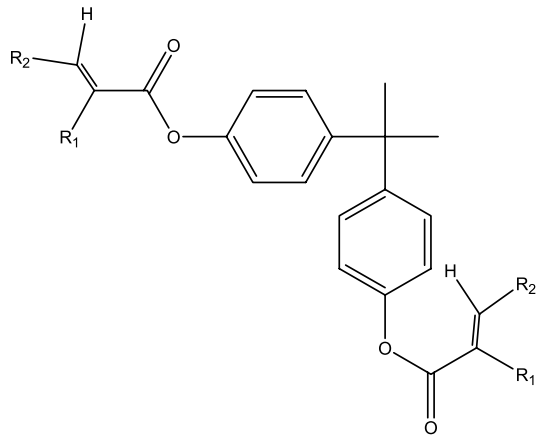
3.4. Study of group's contribution to mutagenic property

The study of group contribution to mutagenic property is very important, because it is a guide for the development and synthesis of new molecules with low mutagenic potenciality. The compound 3 of the training set, Bisphenol A dimethacrylate, was taken as base structure for calculating the contributions of groups, so that it shows the most lower mutagenic character of selected data. As can be seen in Table 4 the contributions to mutagenic power of different groups in two zones of the former mentioned compound are shown.

Some comments about the contribution to the mutagenicity of each group according to our model must be pointed out. Contribution of chloride in position I (8.27) is greater than bromide (8.14) due to the higher electronegativity of the former leading to a higher dipolar moment in this part of the molecule giving a higher mutagenicity as was explained above. Something similar occurs with the substituted methyl groups, which are mono, di and trichlorides respect to the bromides. Inside this set the property increases too in the order $CX_3 > CHX_2 > CH_2X$ for both halides. This observation was already reported [2]. The same trend was observed for the position II but the values are lesser that in position I except for the tri-halides. By other hand only the groups that contribute to decrease the mutagenicity are $CHBrCH_3$ and $CHClCH_3$ even in comparison with the ethyl group. It seems to indicate that the lesser polarization of the electronic cloud leads to a lesser mutagenicity because inductive effects of halide and methyl groups are counterbalanced.

It is to be noted that all monomers substituted in position I (R_1), can be in principle homopolymerized by radical initiators when $R_2 = H$. These monomers are 1,1 disubstituted olefins. It could be consider some transference or decrease in molecular weight can occur but not interfering

Table 4
Contributions to mutagenic power of different groups in the two zones under study



Group	Contribution in position	
	R_1	R_2
–Acetate	6.40	6.42
–Benzene	8.40	8.60
–Br	8.14	8.06
$-CBr_3$	11.54	11.68
$-CCl_3$	11.76	11.93
$-CH_2Br$	4.20	4.15
$-CH_2Cl$	4.23	4.17
$-CHBr_2$	6.96	6.87
$-CHBrCH_3$	−0.35	−0.45
$-CHCl_2$	7.10	7.02
$-CHClCH_3$	−0.45	−0.56
–Cl	8.27	8.18
–COOH	12.87	12.77
$-CH_2CH_3$	2.59	2.86
–OH	6.37	6.33

polymerization. There is an exception when the molecule holds the OH group because is known that vinyl alcohols are unstable. For the methacrylates set substituted in position II the homopolymerization is not possible because are 1,2 disubstituted olefins that are prone to polymerization. For this set the only way of reaction is the copolymerization with other biocompatible polymers.

4. Concluding remarks

We have shown that TOPS-MODE approach is able to describe mutagenicity of dental monomers with an appropriate degree and robustness. In fact we have developed a model for predicting mutagenicity of a data set of 23 dental monomers. This model explains more than 91% of the variance in the experimental mutagenicities with appropriated predictive power.

On the other hand, the main advantage of using TOPS-MODE approach in QSAR has been confirmed again in this work. This approach is able to derive group contributions and gives simultaneously a valuable capability of interpretation contributing to drug discovery [30].

Acknowledgements

Thanks are given to the owner of the software Modeslab 1.0 for the donation of this valuable tool by Maykel Pérez González. On the other hand, the authors would like to express they more sincerely gratitude to unknown referees and the editor Professor J.E. Mark by useful comments and kind attention.

References

[1] Chan D, Titus H, Chung K, Dixon H, Wellinghoff S, Rawls H. Dent Mater 1999;3:219–22.
[2] Yourtee D, Holder A, Smith R, Morrill J, Kostoryz E, Brockmann W, Glaros A, Chappelow C, Eick D. J Biomater Sci Polym Edn 2001;1:89–105.
[3] Atkinson A, Roy D. Biochem Biophys Res Commun 1995;210:424–8.
[4] Estrada E, Gutiérrez Y. J Chromatogr A 1999;858:187–99.
[5] Estrada E, Peña A. Bioorg Med Chem 2000;8:2755–70.
[6] Estrada E, Uriarte E. SAR QSAR Environ Res 2001;12:309–24.
[7] Estrada E. J Chem Soc Faraday Trans 1998;94:1407–11.
[8] Estrada E, Gutiérrez Y, González H. J Chem Inf Comput Sci 2000;40:1386–99.
[9] Estrada E, Peña A, García-Domenech R. J Comput-Aided Mol Des 1998;12:583–95.
[10] Estrada E, Uriarte E, Montero A, Tejeira M, Santana L, De Clercq E. J Med Chem 2000;43:1975–85.
[11] Estrada E, Molina E, Uriarte E. SAR QSAR Environ Res 2001;12:445–59.
[12] Estrada E. J Chem Inf Comput Sci 1996;36:844–9.
[13] Estrada E. J Chem Inf Comput Sci 1997;37:320–8.
[14] Estrada E. J Chem Inf Comput Sci 1998;38:23–7.
[15] Estrada E. J Chem Inf Comput Sci 1995;35:31–3.
[16] Gutierrez Y, Estrada E. Modes Lab®, version 1.0 b; 2002.
[17] Tuparainen K. Chemosphere 1999;38:3015–30.
[18] Gutierrez Y, Estrada E. TOSS-MODE (Topological Sub-strutural Molecular Desing) for windows, Version 4.0, Universidad de Santiago de Compostela, Spain; 1997.
[19] Manhnhold R, Krogsgaard-Larsen H, Timmerman. Discriminant analysis for activity prediction. In: Van der Waterbeemd H, editors. Chemometric methods in molecular design. Method and principles in medicinal chemistry, vol. 2. Weinhiem: VCH; 1995. p. 265–82.
[20] Kowalski RB, Wold S. Pattern recognition in chemistry. In: Krishnaiah PR, Kanal LN, editors. Handbook of statistics. Amsterdam: North Holland; 1982. p. 673–97.
[21] Randić M. J Chem Inf Comput Sci 1991;31:311–20.
[22] Randić M. New J Chem 1991;15:517–25.
[23] Randić M. J Mol Struct (Theochem) 1991;233:45–59.
[24] Lučić B, Nikolić S, Trinajstić N, Jurić D. J Chem Inf Comput Sci 1995;35:532–8.
[25] Klein DJ, Randić M, Babić D, Lučić B, Nikolić S, Trinajstić N. Int J Quantum Chem 1997;63:215–22.
[26] Estrada E, Perdomo I, Torres-Labandeira J. J Chem Inf Comput Sci 2001;41:1561–8.
[27] Randić M. J Comput Chem 1993;4:363–70.
[28] Ivanciuc O, Ivanciuc T, Balaban AT. Vertex and edge-weighted molecular graphs and derived structural descriptors. In: Topological indices and related descriptors. (Devillers J, Balaban AT, Eds.) Gordon & Breach, The Netherlands. p. 169–220.
[29] Estrada E, Molina E, Uriarte E. SAR QSAR Environ Res 2001;12:445–59.
[30] Devillers J. Curr Opin Drug Discov Dev 2000;3:275–9.