Available online at www.sciencedirect.com

Bioorganic & Medicinal Chemistry 13 (2005) 1775–1781

Bioorganic & Medicinal Chemistry

Quantitative structure–activity relationship to predict toxicological properties of benzene derivative compounds

Maykel Pérez González,[a,b,*] Aliuska Morales Helguera[b,c] and Miguel Angel Cabrera[b]

[a] Unit of Service, Drug Design Department, Experimental Sugar Cane Station 'Villa Clara-Cienfuegos', Ranchuelo, Villa Clara 53100, Cuba
[b] Chemical Bioactive Center, Central University of Las Villas, Santa Clara, 54830 Villa Clara, Cuba
[c] Department of Chemistry, Faculty of Chemistry and Pharmacy, Central University of Las Villas, Santa Clara, 54830 Villa Clara, Cuba

Received 5 November 2004; revised 30 November 2004; accepted 30 November 2004

Abstract—TOPological Sub-structural MOlecular DEsign (TOPS-MODE) was used to assess acute aquatic toxicity of a series of 69 benzene derivatives. The obtained model was able to explain more than 88% of data variance, stressing the importance of molecule hydrophobicity and its dipolar moment, as well as the distance between their bonds to describe the property under study. On the other hand, this model was better than those obtained with Dragon software (Constitutional, Galvez topological charges indices and BCUT) using the same number of variables. This approach proved to be a very good method to assess acute aquatic toxicity of these king of compounds, which could be applied to other series of substances.

1. Introduction

Benzene derivative compounds have been used for many years in the chemical and pharmaceutical industry as herbicides, drugs, food additives and so on. The eventual release and accumulation of these compounds into the environment in both terrestrial and aquatic systems requires an assessment of their environmental risk. Since experimental measurements of aquatic toxicity are extremely time-consuming and expensive, it is imperative to develop robust quantitative structure–activity relationships (QSARs) that can predict their toxicity. The majority of QSARs correlate toxicity using the octanol–water partition coefficient (K_{ow}) and/or various physicochemical and quantum-chemical parameters.[1–6]

For instance, Karabunarliev and co-workers[7,8] used quantum-chemically derived acceptor superdelocalizabilities for the carbon of the polar group to which the electron transfer occurs (A_R), the bond order for the bridging single bond that acquires double bond character at the intermediate step of the reaction ($B_{\alpha\text{-R}}$) and K_{ow} to predict the acute aquatic toxicity for fathead minnows for a set of 18 Michael-type acceptors, mostly acrylates, with C=O as the polar group. Acceptor superdelocalizabilities described most of the variance, and the authors showed that quantum-chemical descriptors were compatible with the accepted mechanism.

On the other hand, Basak et al. developed a new approach to construct models for estimating toxicological properties of interest.[9] This approach uses increasingly more complex molecular descriptors in a graduated approach to model building. In this study, statistical and neural network methods were applied to the development of QSAR models for estimating the acute aquatic toxicity (LC_{50}) of 69 benzene derivatives to *Pimephales promelas* (*fathead minnow*), where Topostructural, topochemical, geometrical and quantum-chemical indices were used as the four levels of the hierarchical method.

In the context of in silico methods for modeling physicochemical and biological properties of chemicals, it has been introduced the TOPological Sub-structural MOlecular DEsign (TOPS-MODE) approach. The TOPS-MODE has been applied to describe physicochemical and biological properties of organic compounds.[10–16]

Keywords: QSTR; Aquatic toxicity; Topological indices.

* Corresponding author. Tel.: +53 42281473; fax: +53 42281130; e-mail: mpgonzalez76@yahoo.es

0968-0896/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2004.11.059

2. Aims and scope of the present work

The successful application of this theoretical approach to the modeling of toxicological and ecotoxicological properties[17] has also inspired us to perform a more exhaustive study, in order to test and/or validate TOPS-MODE applicability in assessing discovery of leads and environmental impact of chemicals. The selection here of a data set on acute aquatic toxicity to *Pimephales promelas* is not casual. Our research institute has been involved, over years, on the development of novel industrial additives with application as herbicides or drugs with direct environmental impact. Therefore, the development of predictive models, which may prevent the introduction of environmental pollutants before activity trials, is of major importance. The data we shall study here was previously studied by Toropov and Toropova considering a data set of 44 benzene derivatives.[18] These authors obtained accurate models for the prediction of this activity. However, even when it was a highly statistically significant model, the decision about the acceptance or rejection of a chemical due to predicted toxicity, must be based on different models, in order to ensure that estimation are as near as possible to reality. In this work a comparison with Constitutional,[19] Galvez topological charges indices and BCUT descriptors was carried out.[20]

3. The TOPS-MODE approach

The TOPS-MODE approach is based on the calculation of the spectral moments of the so-called bond matrix, whose theoretical basis has been described in previous reports.[16–22] Nevertheless, an overview of this approach will be given below.

The bond matrix is defined as a square and symmetric matrix whose entries are ones or zeros if the corresponding bonds are adjacent or not. The order of this matrix (m) is the number of bonds in the molecular graph, being two bonds adjacent if they are incident to a common atom. The spectral moments of the edge adjacency matrix are defined as the traces, that is, the sum of the main diagonal, of the different powers of such a matrix.

Table 1. Mains results of the *k*-means cluster analysis for the benzene derivate chemicals

Variance analysis				
Spectral moments	Between SS[a]	Within SS[b]	Fisher ratio (*F*)	*p*-Level[c]
μ_5^{dip}	40.02	13.19	137.13	0.000
$\mu_1{*}\mu_1^{\text{hid}}$	33.88	14.31	92.98	0.000
μ_1^{dist}	29.94	9.25	89.44	0.000
μ_2^{dist}	31.12	22.17	45.98	0.000

μ_5^{dip}: fifth spectral moment weights with dipole moment.
$\mu_1{*}\mu_1^{\text{hid}}$: squared of first spectral moment weights with hydrophobicity.
μ_1^{dist}: first spectral moment weights with atomic distance.
μ_2^{dist}: second spectral moment weights with atomic distance.
[a] Variability between groups.
[b] Variability within groups.
[c] Level of significance.

To apply the present approach to the structure–toxicity relationship, the following steps should be followed. First, to select an adequate training set with great structural diversity. Second, to draw the hydrogen depleted molecular graphs for each molecule of the training set. The third step is to differentiate the molecular bonds with appropriate weights. The fourth, to compute the spectral moments of the bond matrix for each molecule of the data set. Fifth, to find a quantitative structure–toxicity relationship by using a regression analysis:

$$P = a_0\mu_0 + a_1\mu_1 + a_2\mu_2 + a_3\mu_3 \ldots a_k\mu_k + b \qquad (1)$$

where P is the studied property, in our case, the $-\log(\mathrm{LC}_{50})$ partitioning, μ_k is the kth spectral moment and the a_ks are the coefficients obtained by the linear regression. Sixth, to test the predictive capacity of the regression model by cross-validation procedures and/or an external prediction set.

4. Material and methods

A data set of 69 compounds was selected, due to the toxicological action reported in the literature consulted.[18] The parameter of toxicological study was acute aquatic toxicity to *Pimephales promelas* ($-\log(\mathrm{LC}_{50})$). Compounds as well as the calculated and experimental values of aquatic toxicity are show in Tables 2 and 3.

TOPS-MODE computer software[22] was employed to calculate molecular descriptors. The standard bond distances, dipole moment and hydrophobicity were used as bond weight for making differentiations of heteroatoms. So, three sets of spectral moments were obtained, one using bond distances, another using dipole moment and the third using hydrophobicity. We also used multiplications of spectral moments as independent variables to describe toxicological characteristics. In this case we only multiplied μ_0 for the first three spectral moments obtaining four new variables. The total number of descriptors used in this model was 57 (45 spectral moments + 12 multiplications of moments). Three models were also carried out but using the computer software Dragon,[20] where the Constitutional, Galvez topological charges and BCUT descriptors were calculated. The statistical processing to obtain the QSAR models was carried out by using the Forward Stepwise regression methods,[27] 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.

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

One of the most important steps in computer-aided design is to have a representative and randomized training and predicting series. With this aim, we selected a data

Table 2. The observed, predicted and residual values of the 50 benzene derivatives used to derive the QSTR

Number	Compounds	Observed	Predicted	Deleted residuals
1	Bromobenzene	3.89	3.74	0.15
2	Phenol	3.51	3.34	0.18
3	1,2-Dichlorobenzene	4.4	4.30	0.10
4	3-Chlorotoluene	3.84	4.01	−0.18
5	1,3-Dihydroxybenzene	3.04	3.31	−0.28
6	3-Hydroxyanisole	3.21	3.23	−0.02
7	1,4-Dimethoxybenzene	3.07	3.14	−0.10
8	2,4-Dinitrophenol	4.04	4.42	−0.41
9	2-Methyl-5-nitroaniline	3.35	3.47	−0.13
10	4-Hydroxy-3-nitroaniline	3.65	3.69	−0.04
11	4-Methyl-3-nitroaniline	3.77	3.47	0.31
12	2,4-Dimethylphenol	3.86	3.84	0.01
13	2,6-Dimethylphenol	3.75	3.94	−0.20
14	3,4-Dimethylphenol	3.90	3.84	0.06
15	3-Nitrotoluene	3.63	3.46	0.17
16	2,6-Dinitrotoluene	3.99	4.16	−0.19
17	3-Methyl-2,4-dinitroaniline	4.26	4.54	−0.29
18	4-Methyl-2,6-dinitroaniline	4.21	4.45	−0.25
19	5-Methyl-2,6-dinitroaniline	4.18	4.53	−0.37
20	1,2,4-Trimethylbenzene	4.21	4.27	−0.08
21	5-Methyl-2,4-dinitroaniline	4.92	4.45	0.48
22	2,4-Dinitrotoluene	3.75	4.07	−0.35
23	1,2,4,5-Tetrachlorobenzene	5.85	5.26	0.64
24	4-Nitrophenol	3.36	3.50	−0.16
25	2-Methyl-4-nitroaniline	3.24	3.47	−0.25
26	4-Chlorotoluene	4.33	4.01	0.33
27	4-Nitrotoluene	3.76	3.46	0.31
28	2,4,6-Tribromophenol	4.70	4.63	0.07
29	2,4,6-Trichlorophenol	4.33	4.71	−0.41
30	Toluene	3.32	3.60	−0.32
31	2-Methyl-3-nitroaniline	3.48	3.56	−0.09
32	3-Methyl-6-nitroaniline	3.80	3.47	0.35
33	4-Methyl-2-nitroaniline	3.79	3.47	0.33
34	1,2,3-Trichlorobenzene	4.89	4.83	0.05
35	1,2,4-Trichlorobenzene	5.00	4.71	0.29
36	2,4-Dichlorophenol	4.30	4.20	0.09
37	3,4-Dichlorotoluene	4.74	4.56	0.18
38	2,4-Dichlorotoluene	4.54	4.53	−0.01
39	Chlorobenzene	3.77	3.77	0.00
40	1,3,5-Trinitrobenzene	5.29	5.55	−0.34
41	2-Methyl-3,5-dinitroaniline	4.12	4.46	−0.35
42	1,2,3,4-Tetrachlorobenzene	5.43	5.39	0.04
43	2,3,4,5,6-Pentachlorophenol	6.06	5.95	0.13
44	1,3-Dichlorobenzene	4.30	4.17	0.12
45	2-Chlorophenol	4.02	3.82	0.20
46	2-Methylphenol	3.77	3.61	0.16
47	3-Methylphenol	3.29	3.52	−0.24
48	1,3-Dinitrobenzene	4.38	4.16	0.24
49	3,4-Dinitrotoluene	5.08	5.00	0.08
50	2,3-Dinitrotoluene	5.01	4.89	0.12

set of 69 benzene derivate compounds with their aquatic toxicity coefficient. Later, *k*-Means Cluster Analysis (*k*-MCA) was performed for this set of compounds.

5. *k*-Means cluster analysis

The *k*-MCA may be used in training and predicting series design.[27] The idea consists of carrying out a partition of the set of compound under study in several statistically representative classes of chemicals. Thence, one may select from the members of all these classes the training and predicting series. This procedure ensures that any chemical class (as determined by the clusters derived from *k*-MCA) will be represented in both compounds series (training and predicting). It permits to design both, training and predicting series, which are representative of the entire 'experimental universe'. Figure 1 graphically illustrates the above-described procedure.

A *k*-MCA splits benzene derivate chemicals in four clusters with 22, 12, 14 and 21 members and standard deviations of 0.15, 0.43, 0.21 and 0.48, respectively. Selection

Table 3. The observed, predicted and residual values for the external prediction series of the TOPS-MODE model

Number	Compounds	Observed	Predicted	Deleted residuals
1	1,2-Dimethylbenzene	3.48	3.95	−0.48
2	1,4-Dicholobenzene	4.62	4.17	0.44
3	1,4-Dimethylbenzene	4.21	3.86	0.35
4	4-Chloro-3-methylphenol	4.27	4.02	0.25
5	2,3,4,5-Tetrachlorophenol	5.72	5.38	0.33
6	2,3,6-Trinitrotoluene	6.37	5.44	0.93
7	4-Methylphenol	3.58	3.52	0.06
8	4-Methyl-3,5-dinitroaniline	4.46	4.46	0.00
9	2-Methyl-6-nitroaniline	3.80	3.56	0.23
10	1,3,5-Trichlorobenzene	4.74	4.58	0.15
11	3,5-Dinitrotoluene	3.91	3.98	−0.07
12	Benzene	3.40	3.39	0.01
13	2-Nitrotoluene	3.57	3.55	0.02
14	1,4-Dinitrobenzene	5.22	5.43	0.21
15	2-Methyl-3,6-dinitroaniline	5.34	5.31	0.03
16	2,4,6-Trinitrotoluene	4.88	4.69	0.19
17	2-Methyl-4,6-dinitrophenol	5.00	4.98	0.02
18	2,5-Dinitrotolueno	5.15	5.00	0.15
19	1,2-Dinitrobenzene	5.45	5.51	−0.06

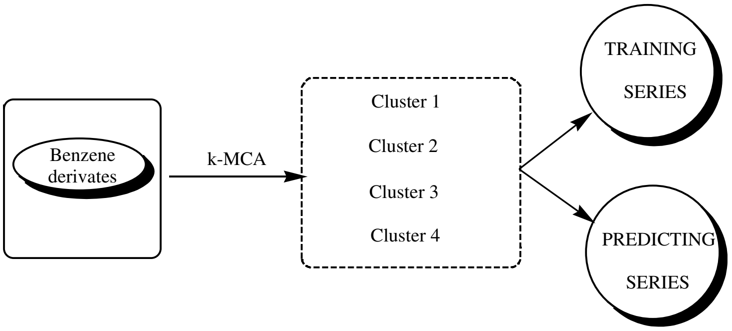

Figure 1. Training and predicting series design throughout *k*-MCA.

of the training and prediction set was carried out by taking, in a random way, compounds belonging to each cluster.

To ensure a statistically acceptable data partition into several clusters, we took into account the number of members in each cluster and the standard deviation of the variables in the cluster (as low as possible). We also made an inspection among and within clusters, where the respective Fisher ratio and their *p*-level of significance were considered to be lower than 0.05. All spectral moments (from μ_0 to μ_{15} of all the weights) were used in both analyses and all variables showed *p*-levels < 0.05 for Fisher test. The results are depicted in Table 1.

The main conclusion should be achieved from *k*-MCA: the structural diversity of several up-to-date known derivate benzene chemicals (as codified by TOPS-MODE descriptors) may be described at least by four statistically homogeneous clusters of chemicals. Anyhow, further conclusions about the mechanistic and molecular signification of these clusters seem to be speculative. Meanly, if it is considered that *k*-MCA based partitions of data, which consider not only 4 but also 5 or 6 clusters are statistically significant too. However, the use of the *k*-MCA analysis here points to a structurally representative distribution of chemicals into training and predicting series.

6. Results and discussion

QSTRs on benzene derivatives toxicity to *Pimephales promelas*: In first instance we developed a TOPS-MODE model, which accurately accounts for the toxicity of a benzene derivative series:

$$-\log(\mathrm{LC}_{50}) = (0.003 \pm 0.0004) \cdot \mu_5^{\mathrm{D}} + (0.037 \pm 0.004) \cdot \mu_1 \cdot \mu_1^{\mathrm{H}} - (0.215 \pm 0.056) \cdot \mu_1^{\mathrm{Dist}} + (4.860 \pm 0.23) \qquad (2)$$

$$N = 50 \quad R^2 = 0.888 \quad p < 0.00001$$
$$F = 105.73 \quad S = 0.246$$

The model establish a strong relationship between TOPS-MODE and toxicity characterized by a high squared regression coefficient (0.888), which explains around 89% of the variance of data. In our opinion this fact is a determining factor at the time of selecting the best model to be used later; besides it presents the greatest F of Fischer ($F = 105.73$) and the lowest standard deviation for the data ($S = 0.246$), which confirms the former selection.

A low susceptibility to over-fitting, due to excess in variable input, is ensured by a high adjusted squared regression coefficient (0.87), which decreases only 2.03% with respect to squared regression coefficient. Dispersion of predicted data with respect to the observed means was controlled by selecting only those models having an estimated Standard error (0.246) as low as possible.

There are, at least on a first view, three main facts involved in the variance explained by the model, which coincide with those classically recognized in QSAR/QSTR.[23] The first accounts for dipolar properties of benzene derivatives, codified here by μ_5 dipole. The introduction of μ_5 have a double significance because it highlights the importance of bonds dipole moment in biological activity, but remarking that it is not longer additive and it is conditioned to connectivity. The 5th order spectral moments of bond matrices codify information about molecular fragments containing five bonds.[21,24] The strong contribution of molecules' polarity may be related here with the necessary solubility in water of molecules to reach the fish environment. However, the second member of the equation confirms a classically reported behaviour of some membrane-crossing dependent biological properties, which increases with hydrophobicity until a cut off value.[25] Finally, the importance of geometry restriction to chemical-target interactions is characterized here by a negative influence of μ_1^{Dist} (the sum of all standard bond distances in the molecule). The standard errors of all coefficients in regression equations are not significant with respect to the absolute magnitude of the coefficient. Table 2 depicts the observed, predicted, and residual values of the 50 benzene derivatives used to derive the QSTR.

Leave-one-out (LOO) technique is one of the most successful procedures for models' validations.[26,27] The present model have a significant LOO squared regression coefficient ($Q^2 = 0.877$). This value represents a lack of only 1.23% in the variance, explained by the model in LOO with respect to training series. This fact ensures a strong stability of the model after data variation. On the other hand, the low value of root mean squared cross-validation (S_{cv}) indicates an acceptable dispersion of residuals in LOO.[28] The Figure 2 depicts the linear relation between observed and predicted aquatic toxicities of benzene derivatives for the Eq. 2.

In addition, the use of an external predicting series corroborates the predictive power of the model. Both, the regression coefficient $R^2_{\mathrm{pred}} = 0.9081$ and $S_{\mathrm{pred}} = 0.2826$ were considered in order to test either linearity or predictability. The Table 3 depicts the observed, predicted and residual values for the external prediction series of the TOPS-MODE model.

In order to carry out both, comparison of TOPS-MODE approach with other methodologies and to derive other useful QSTRs, which could be used to select chemicals with low environmental risk, we used the molecular indices calculated by DRAGON.[20] Three kinds of indices were, specifically, used: Constitutional indices, BCUTs and Galves's charge indices. The best models found were:

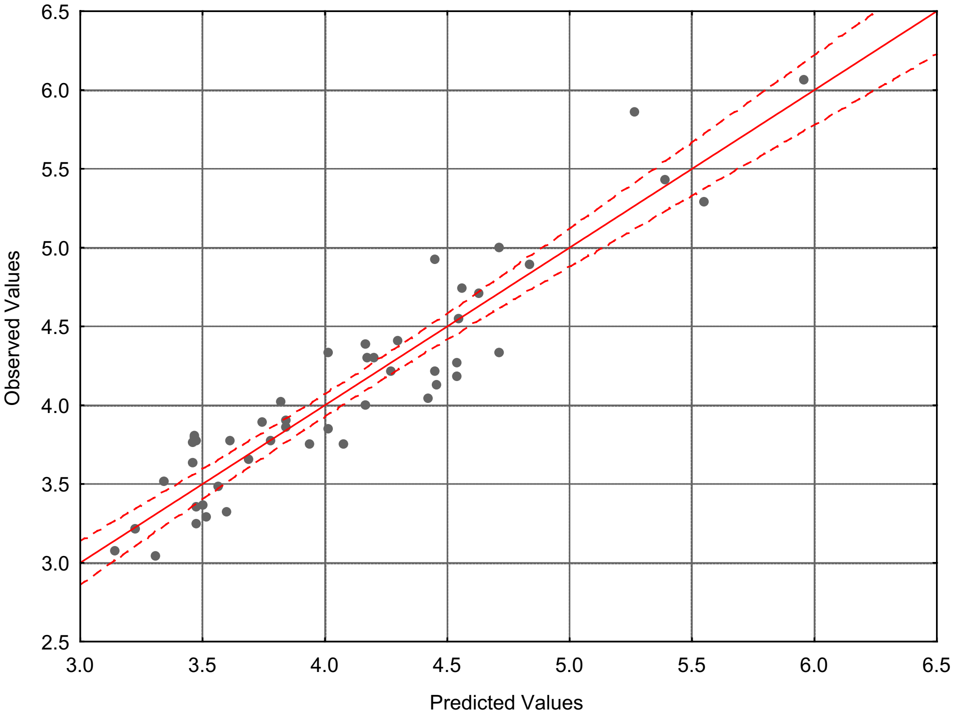

Figure 2. The linear relation between observed and predicted aquatic toxicities of benzene derivatives for the Eq. 2.

$$-\log(\mathrm{LC}_{50}) = 1.9202 + 0.52308 \cdot n\mathrm{X} + 0.1809 \cdot \mathrm{SCBO} - 6.1765 \cdot \mathrm{Me} \quad (3)$$

$$R^2 = 0.8095 \quad p < 0.0001 \quad N = 50 \quad F = 62.3449$$
$$S = 0.3079 \quad Q^2 = 0.7692 \quad S_{\mathrm{cv}} = 0.0478$$

$$-\log(\mathrm{LC}_{50}) = 0.0073 + 1.3085 \cdot \mathrm{BEHm4} - 1.2338 \cdot \mathrm{BEHp4} + 1.2418 \cdot \mathrm{BEHe3} \quad (4)$$

$$R^2 = 0.7481 \quad p < 0.0001 \quad N = 50 \quad F = 43.5804$$
$$S = 0.3541 \quad Q^2 = 0.6825 \quad S_{\mathrm{cv}} = 0.0561$$

$$-\log(\mathrm{LC}_{50}) = 2.5517 + 10.2125 \cdot \mathrm{JGI2} \quad (5)$$

$$R^2 = 0.3424 \quad p < 0.0001 \quad N = 50 \quad F = 23.9604$$
$$S = 0.5596 \quad Q^2 = 0.2739 \quad S_{\mathrm{cv}} = 0.0848$$

The three models are statistically significant, shown by their low p-level of Fishers' tests (all < 0.05). However, there is a gradation in the quality of models. The strongest linear regression, as measured by regression coefficient, appears for Constitutional indices > BCUTs > Galvez's charge indices, in this order. Constitutional indices model includes three variables: Me (mean atomic Sanderson's electronegativity), SCBO (sum of conventional bond orders, hydrogen depleted) and nX (number of halogen atoms). Sanderson's electronegativity is directly related with charge distribution,[20,29] while SCBO codifies geometric and electronic bonds global properties.[19] In essence, those are the aspects that TOPS-MODE based model codifies but do not seem to be a direct coincidence of all variables in both models. For instance, Constitutional indices model recognizes a special importance for the number of halogen atoms, which do not appear in any other model found here. On the other hand, BCUTs model also weights significantly the electronegativity and polarizability related features but appears here by first time the distribution of masses in the molecule. Finally, Galvez's charge indices based equation, although statistically significant, clearly suffers a lack of fit. This fact may be undoubtedly explained if we consider that these indices only take into account the charge distribution and connectivity, but they are not related with geometrical features, therein before identified as another important fact in the development of toxicity to *Pimephales promelas* by benzene derivatives. The total predictability of these models was tested by means of leave-one-out procedures, considering squared regression coefficient (Q^2) and root mean square error (S_{cv}). Both, BCUTs and Constitutional indices models have higher predictive power. Conversely, Galvez's charge indices predictability falls to a significant degree.

7. Concluding remark

A new application of the TOPS-MODE approach was used to predict acute toxicity of a series of 48 benzene derivatives. The theoretical model revealed that molecule's hydrophobicity and its dipolar moment as well as the distance between their bonds are very important factors when predicting aquatic toxicity of this type of compounds. Due to the easy calculation and interpretations of results are reliable, we are going to use this methodology to predict acute aquatic toxicity to other type of substances. For that reason, these results have shown that the TOPS-MODE approach not only explains the experimental data, but rather seems to be the best one in doing so.

Acknowledgements

The authors would like to express their gratitude to the Ministry of Science Technology and Environment (CITMA) of Villa Clara city for financial support and kindness.

References and notes

1. Konemann, H. *Toxicology* **1981**, *19*, 209.
2. Nendza, M.; Russom, C. L. *Xenobiotica* **1991**, *21*, 147.
3. Newsome, L. D.; Johnson, D. E.; Lipnick, R. L.; Broderius, S. J.; Russom, C. L. *Sci. Total Environ.* **1991**, *109/110*, 537.
4. Schultz, T. W.; Wilke, T. S.; Bryant, S. E.; Hosein, L. M. *Sci. Total. Environ.* **1991**, *109/110*, 581.
5. Karabunarliev, S.; Mekenyan, O. G.; Karcher, W.; Russom, C. L.; Bradbury, S. P. *Quant. Struct.–Act. Rel.* **1996**, *15*, 311.
6. Eldred, D. V.; Weikel, C. L.; Jurs, P. C.; Kaiser, K. L. E. *Chem. Res. Toxicol.* **1999**, *12*, 670.
7. Karabunarliev, S.; Mekenyan, O. G.; Karcher, W.; Russom, C. L.; Bradbury, S. P. *Quant. Struct.–Act. Rel.* **1996**, *15*, 302.
8. Karcher, W.; Karabunarliev, S. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 672.
9. Basak, S. C.; Grunwald, G. D.; Gute, B. D.; Balasubramanian, K.; Opitz, D. *J. Chem. Inf. Comput. Sci.* **1996**, *65*, 885.
10. Estrada, E. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 844.
11. Estrada, E. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 320.
12. Estrada, E. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 23.
13. González, M. P.; González, H. D.; Molina, R. R.; Cabrera, M. A.; Ramos, R. A. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1192.
14. González, M. P.; Morales, A. H.; González, D. H. *Polymer* **2004**, *45*, 2073.
15. Cabrera, M. A.; Bermejo, M.; González, M. P.; Ramos, R. *J. Pharm. Sci.* **2004**, *93*, 1701.
16. González, M. P.; Días, L. C.; Morales, A. H.; Rodríguez, Y. M.; Gonzaga de Oliveira, L.; Gómez, L. T.; González, H. D. *Bioorg. Med. Chem.* **2004**, *12*, 4467.
17. González, M. P.; González, H. D.; Cabrera, M. A.; Molina, R. R. *Bioorg. Med. Chem.* **2004**, *12*, 735.
18. Toropov, A.; Toropova, A. P. *J. Mol. Struct.* **2002**, *578*, 129.
19. Todeschini, R.; Consonni, V. Handbook of Molecular Descriptors; Wiley-VCH, Weinheim: Germany, 2000.

20. Todeschini, R.; Consonni, V.; Pavan, M. Dragon. Software version 2.1, 2002.
21. The method was originally introduced with the name TOSS-MODE that is now changed to TOPS-MODE: Estrada, E. On the topological sub-structural molecular design (TOSS-MODE) in QSPR/QSAR and drug design research. *SAR QSAR Environ. Res.* **2000**, *11*, 55–73.
22. Gutierrez, Y.; Estrada, E. TOSS-MODE (1997) (Topological Sub-Structural Molecular Design) for Windows Version 4.0, Universidad de Santiago de Compostela, Spain.
23. Kubinyi, H.; Taylor, J.; Ramdsen, C. Quantitative Drug Design In *Comprehensive Medicinal Chemistry*; Hansch, C., Ed.; Pergamon, 1990; Vol. 4, pp 589–643.
24. Estrada, E.; Molina, E.; Uriarte, E. *SAR QSAR Environ. Res.* **2001**, *12*, 309.
25. Hansch, C.; Leo, A. *Exploring QSARs Fundamentals and Applications in Chemical Biology*; ACS: Washington, DC, 1995; Chapter 6, p 349.
26. van der Waterbeemd, H. Discriminant Analysis for Activity Prediction. In *Method and Principles in Medicinal Chemistry*; Manhnhold, R., Krogsgaard-Larsen, H., Eds.; Chemometric Methods in Molecular Design; Van Waterbeemd, H. Timmerman, Ed.; VCH: Weinhiem, 1995; Vol. 2, pp 265–282.
27. Kowalski, R. B.; Wold, S. Pattern Recognition in Chemistry. In *Handbook of Statistics*; Krishnaiah, P. R., Kanal, L. N., Eds.; North Holland Publishing Company: Amsterdam, 1982; pp 673–697.
28. Draper, N. R.; Smith, H. Applied Regression Analysis. La Habana, E. R., 1980; pp 123–200.
29. Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219.