Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 12 (2004) 735–744
Bioorganic & Medicinal Chemistry

A novel approach to predict a toxicological property of aromatic compounds in the *Tetrahymena pyriformis*

Maykel Pérez González,[a,b,c,*] Humberto González Díaz,[b,c] Miguel Angel Cabrera[b] and Reinaldo Molina Ruiz[b]

[a]*Experimental Station of Sugar Cane Research 'Villa Clara-Cienfuegos', Ranchuelo, 53100, Cuba*
[b]*Department of Drug Design, Chemical Bioactives Center, Central University of 'Las Villas', Santa Clara 54830, Cuba*
[c]*Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela, 15706, Spain*

Received 22 April 2003; accepted 14 November 2003

Abstract—The TOPological Substructural MOlecular DEsign (TOPS-MODE) has been successfully used in order to explain the toxicity in the *Tetrahymena pyriformis* on a large data set. The obtained models for the training set had good statistical parameters ($R^2 = 0.72–0.81$, $p < 0.05$) an also the prediction power of the models found was adequate ($Q^2 = 0.70–0.80$). A detailed study of the influence of variable numbers in the equation and the statistical outliers was carried out; leading to a good final model with a better physicochemical interpretation than the rest of the published models. Only two molecular descriptors codifying dipolar and hydrophobic features were introduced. Finally, the fragment contributions to the toxicity prediction evidenced the powerful of this topological approach.

1. Introduction

Quantitative Structure–Toxicity Relationships (QSTRs) are predictive tools for a preliminary evaluation of the hazard of chemical compounds by using computer-aided models.[1–7] These theoretical models represent an alternative to the 'real' world of assaying chemical compounds for determining their toxicological properties on living organisms in the laboratory. This 'virtual' world consists of datasets of compounds, their computational analysis, hypothesis generation and toxicity prediction, made and stored on a computer. By this means, the expensive, time-consuming and in many cases animal-aggressive bioassays are made only after exploring the initial concepts with computational models.[8–10]

Turner et al. published an excellent review about the use and abuse of QSAR techniques in Toxicology and Ecotoxicology during the 1980s[11] and, at the same time, a monograph from the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) appeared to serve as guide for this kind of study.[12] Today, we can still use this rules outlined by ECETOC in 1986. Also in the International Workshop on In vitro Methods for Assessing Acute Systemic Toxicity was recognized the complementary uses of quantitative structure–activity relationships (QSAR) as part of an in vitro strategy for acute systemic toxicity testing.[13]

Both the ECETOC and Turner publications recognize three main steps in any QSAR study: (i) collecting data on the presence or magnitude of a specified biological activity for a series of chemicals, (ii) listing certain structural elements and (iii) applying techniques to seek a relationship between molecular structure and biological activity (QSAR).[11,12]

Among the statistical methods developed to found a relation between structure and toxicological activity, the Multivariate Regression Techniques, the Factor Analysis and the Neural Networks appear to be the best choice.[14–21] On the other hand, there is a wide range of molecular indexes that we can use to codify the necessary information to develop the second step, being the topological, quantum-chemical, informational, graph theoretical, constitutional and molecular mechanics indices the most popular molecular descriptors.[22]

Keywords: QSAR; *Tetrahymena pyriformis*; Topological indices; TOPS-MODE.
* Corresponding author. Tel.: +1-534-228-1473; fax: +1-534-228-1130; e-mail: mpgonzalez76@yahoo.es

0968-0896/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2003.11.028

In the context of novel *in silico* methods for modeling physicochemical and biological properties of chemicals, the TOPS-MODE approach has been introduced. This methodology has been successfully applied to the description of physicochemical and biological properties of organic compounds.[23–29]

The application of this theoretical approach to the modeling of toxicological properties[30] has inspired us to perform a more exhaustive study in order to test and/or validate the application of the TOPS-MODE approach in the assessing of the chemical toxicological impact. For this reason, the selection of a large data set on toxicity in the *Tetrahymena pyriformis* is not casual; this property was previously studied with the TOPS-MODE approach but considering only a more reduced data set.[31]

Other research papers about the theoretical prediction in the *Tetrahymena pyriformis* have been publish by Cronin et al.[32] In this sense could be interesting to test the TOPS-MODE potentialities through a comparison with the Cronin' data set,[32] and to evaluate the contribution of different molecular fragments to the prediction of the toxicity in the *Tetrahymena pyriformis*. Thence, it may result very interesting to test the potentialities of TOPS-MODE approach with this extended data set.

2. The TOPS-MODE approach

The TOPS-MODE approach is based on the computation of the spectral moments of the bond adjacency matrix, and its mathematical basis has been described in previous reports.[23–29] In addition, a methodological explanation about the use of this approach as well as a software description[33] have been ecently published.[34]

The application of TOPS-MODE approach to the study of quantitative structure—toxicity relationships can be resumed in the following set of steps:

1. To draw the hydrogen-depleted molecular graphs for each molecule of the data set,
2. To use appropriated bond weights in order to differentiate the molecular bonds, for example, bond distance, bond dipoles, bond polarizabilities, etc.,
3. To compute the spectral moments of the bond matrix with the appropriated weights for each molecule in the data set generating a table in which rows correspond to the compounds and columns correspond to the spectral moments of the bond matrix. Spectral moments are defined as the trace of the different powers of the bond matrix,
4. To find a quantitative structure-toxicity relationship (QSTR) by using any appropriated linear or non-linear multivariate statistical technique, such as multi-linear regression analysis (MRA), etc.:

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

where, P is the measurement of toxicity, μ_k is the kth spectral moment, and the a_k's are the coefficients obtained by the MRA,

5. To test the predictive capacity of the QSTR model by using cross-validation techniques.
6. To compute the contribution of the different fragments of interest in order to determine their quantitative contribution to the toxicity of the molecules under study.

The computation of fragment contributions to the toxicological property under study is probably the most important advance of the TOPS-MODE approach to the study of toxicological variables compared to the traditional QSAR and QSTR methods. This procedure can be useful for the identification of possible toxicophores that can be further studied by using different theoretical and experimental techniques. The procedure consists of calculating the spectral moment for all the fragments contained in a given substructure, and by difference of these moments we obtain the contribution of the substructure. The general algorithm for this computational approach is as follows:

First, we select the substructure whose contribution to the moments we would like to determine. Then, we generate all the fragments, which are contained in the corresponding substructure, and calculate the spectral moments for both, the substructure and all their fragments. The contribution of the substructure to the spectral moments is finally obtained as the difference between the spectral moments of the substructure and all those from their fragments. Once, the contributions of the different structural fragments are obtained, we only need to substitute these contributions into the quantitative model developed to describe the property studied.

3. Tops-mode QSTR chemical toxicity in the *Tetrahymena pyriformis*

In this study we selected a data set of 202 aromatic compounds (with nitro and cyano groups) for which the toxicity in the *Tetrahymena pyriformis* was reported.[32] The toxicological parameter studied was the 50% inhibitory growth concentration of aromatic compounds in the *Tetrahymena pyriformis* (-logIGC_{50}). The experimental values of this property are given in Tables 1 and 2.

The Modeslab 1.0 computer software was employed to calculate molecular descriptors.[34] The standard dipole moments and hydrophobicity were used as weights in the main diagonal of the bond adjacency matrix. In general, a total set of 115 descriptors were calculated. The forward stepwise was used as the strategy for the variable selection.[35,36]

A first theoretical model, having 5 variables and acceptable statistical parameters was obtained. However, a step-by-step outlier extraction procedure led to different

Table 1. Experimental and predicted values of toxicity to *Tetrahymena pyriformis* of compounds in the training set

No	Compounds	Obs.	Pred.	Residual.
1	4-Cyanopyridine	−0.82	−0.18	−0.634
2	2-Cyanopyridine	−0.79	−0.21	−0.577
3	3-Cyanopyridine	−0.74	−0.19	−0.546
4	3-Cyano-4,6-dimethyl-2-hydroxypyridine	−0.70	0.15	−0.856
5	2-Cyanoaniline	−0.50	−0.15	−0.348
6	3-Cyanoaniline	−0.47	−0.12	−0.342
7	2,3-Dicyanohydroquinone	−0.44	0.31	−0.750
8	2-Amino-5-nitropyrimidine	−0.43	0.01	−0.439
9	4-Acetylbenzonitrile	−0.37	−0.01	−0.356
10	1,2-Dicyanobenzene	−0.34	−0.004	−0.335
11	2-Cyanobenzamide	−0.32	−0.100	−0.219
12	4-Fluorobenzonitrile	−0.26	0.179	−0.439
13	2- Tolunitrile	−0.24	−0.071	−0.168
14	3-(Hydroxymethyl)nitrobenzene	−0.22	0.252	−0.472
15	3-Nitrobenzamide	−0.19	0.378	−0.568
16	2-Hydroxy-4-methyl-5-nitropyridine	−0.17	0.511	−0.681
17	4-Tolunitrile	−0.10	−0.059	−0.040
18	3-Chlorobenzonitrile	−0.06	0.266	−0.326
19	Methyl-4-cyanobenzoate	−0.06	0.042	−0.102
20	3-Cyanophenol	−0.06	0.038	−0.098
21	2-Amino-3-nitropyridine	−0.01	0.245	−0.255
22	2-Metoxy-5-nitropyridine	0.00	0.296	−0.306
23	4-Chlorobenzonitrile	0.00	0.266	−0.266
24	3-Nitroaniline	0.03	0.279	−0.249
25	4-Cyanobenzaldehyde	0.04	−0.016	0.056
26	3-Methoxybenzonitrile	0.05	−0.132	0.182
27	2-Methylnitrobencene	0.05	0.343	−0.293
28	3-Methylnitrobencene	0.05	0.371	−0.321
29	4-Methoxybenzonitrile	0.10	−0.13	0.232
30	4-Nitrobenzyl alcohol	0.12	0.25	−0.132
31	4-Nitrophenylacetonitrile	0.13	0.19	−0.064
32	3-Nitrobenzaldehyde	0.14	0.42	−0.288
33	2-Nitrobenzaldehyde	0.17	0.39	−0.228
34	4-Methylnitrobenzene	0.17	0.37	−0.201
35	4-Nitrobenzamide	0.18	0.37	−0.198
36	4-Nitrobenzaldehyde	0.20	0.42	−0.228
37	2-Amino-5-nitropyridine	0.22	0.26	−0.041
38	1-Fluoro-2-nitrobenzene	0.23	0.58	−0.350
39	4-Cyanoaniline	0.24	−0.12	0.367
40	4-Fluoronitrobenzene	0.25	0.60	−0.358
41	3-Hydroxy-4-nitrobenzaldehyde	0.27	0.58	−0.319
42	2-Chloro-4-methyl-3-nitropyridine	0.29	0.57	−0.287
43	4-Bromobenzonitrile	0.29	0.31	−0.021
44	2,6-Dimethylnitrobenzene	0.30	0.44	−0.146
45	5-Hydroxy-2-nitrobenzaldehyde	0.33	0.60	−0.275
46	2-Fluoro-4-nitrotoluene	0.33	0.72	−0.397
47	4-Methyl-2-nitroaniline	0.37	0.12	0.249
48	Ethyl-4-cyanobenzoate	0.37	−0.08	0.382
49	4-Nitroanisole	0.38	0.30	0.071
50	4-Bromonitrobenzene	0.38	0.75	−0.375
51	5-Nitroquinoline	0.39	0.49	−0.105
52	3-Hydroxy-6-methyl-2-nitropyridine	0.39	0.53	−0.140
53	4-Nitropyridine	0.41	0.17	0.239
54	2-Chloro-4-methyl-5-nitropyridine	0.42	0.60	−0.181
55	4-Ethylnitrobenzene	0.43	0.19	0.231
56	4-Chloronitrobenzene	0.43	0.70	−0.275
57	3-Nitrobenzonitrile	0.45	0.27	0.170
58	4,5-Dimethyl-2-nitroaniline	0.45	0.49	−0.045
59	6-Nitroquinoline	0.46	0.53	−0.073
60	2-Amino-4-nitrophenol	0.47	0.41	0.057
61	3-Nitrophenol	0.51	0.45	0.056
62	4-Amino-3,5-dinitrobenzamide	0.51	1.18	−0.676
63	4-Cyanophenol	0.52	0.03	0.481
64	4-Nitrophenylene-1,2,diamine	0.52	0.23	0.289
65	2,4-Dinitroaniline	0.53	0.93	−0.406
66	2,6-Dinitrophenol	0.54	1.08	−0.541
67	2,3-Dimethylnitrobenzene	0.56	0.46	0.098
68	4-Nitrobenzonitrile	0.57	0.45	0.114
69	1,2-Dimethyl-4-nitrobenzene	0.59	0.49	0.098
70	2-Chloro-5-nitrobenzaldehyde	0.60	0.92	−0.323
71	4-Hydroxy-3-nitrobenzaldehyde	0.61	0.58	0.020

(*continued on next page*)

Table 1 (*continued*)

No	Compounds	Obs.	Pred.	Residual.
72	2-Nitroresorcinol	0.66	0.54	0.110
73	2-Nitrophenol	0.67	0.41	0.257
74	3-Metoxynitrobenzene	0.67	0.30	0.361
75	4-Nitrobenzaldoxime	0.68	0.59	0.089
76	2-Chloronitrobenzene	0.68	0.67	0.001
77	3-Chloro-2-methylnitrobenzene	0.68	0.80	−0.124
78	1-Cyanonapthalene	0.69	0.63	0.052
79	Ethyl-4-nitrobenzoate	0.71	0.52	0.185
80	4-Methyl-3-nitrophenol	0.74	0.54	0.196
81	4,5-Difluoro-2-nitroaniline	0.75	0.96	−0.214
82	2-Chloromethyl-4-nitrophenol	0.75	0.39	0.352
83	4-Ethoxy-2-nitroaniline	0.76	0.21	0.543
84	3-Chloro-4-fluoronitrobenzene	0.80	1.07	−0.270
85	Methyl-4-chloro-2-nitrobenzoate	0.82	1.11	−0.297
86	5-Chloro-2-methylnitrobenzene	0.82	0.59	0.225
87	4-Nitrophenetole	0.83	0.23	0.591
88	3-Chloronitrobenzene	0.84	0.70	0.134
89	2-Bromonitrobenzene	0.86	0.72	0.131
90	2,4,6-Trimethylnitrobenzene	0.86	0.63	0.226
91	6-Methyl-1,3-dinitrobenzene	0.87	1.11	−0.247
92	3-Hydroxy-2-nitropyridine	0.87	0.45	0.415
93	1,3-Dinitrobenzene	0.89	0.71	0.174
94	5-Amino-6-nitroquinoline	0.92	0.71	0.205
95	3-Fluoro-4-nitrophenol	0.93	0.77	0.151
96	3,5-Dinitroaniline	0.94	0.99	−0.050
97	4-Amino-2-nitrophenol	0.98	0.41	0.567
98	2,4-Dichloronitrobenzene	0.99	0.93	0.052
99	1-Nitronaphthalene	1.00	1.17	−0.176
100	2,3-Dichloronitrobenzene	1.07	0.92	0.147
101	2-Bromo-5-nitropyridine	1.07	0.56	0.504
102	1-Fluoro-3-iodo-5-nitrobenzene	1.09	1.28	−0.191
103	3,4-Dinitrobenzyl alcohol	1.09	0.97	0.114
104	4-Nitro-1-naphthylamine	1.12	0.98	0.131
105	5-Fluoro-2-nitrophenol	1.12	0.76	0.356
106	2,5-Dichloronitrobenzene	1.13	1.16	−0.032
107	3,4-Dichloronitrobenzene	1.16	1.17	−0.018
108	2-Bromo-5-nitrotoluene	1.16	0.88	0.270
109	2-Amino-4-chloro-5-nitrophenol	1.18	0.88	0.298
110	3,5-Dinitrobenzonitrile	1.22	0.80	0.413
111	2-Chloro-4,6-dinitroaniline	1.22	1.50	−0.288
112	3-Bromonitrobenzene	1.22	0.75	0.464
113	2-Bromo-4,6-dinitroaniline	1.24	1.57	−0.337
114	4-Biphenylcarbonitrile	1.24	0.84	0.397
115	1,2-Dinitrobenzene	1.25	0.89	0.356
116	4-Chloro-3-nitrophenol	1.27	0.88	0.383
117	1,4-Dinitrobenzene	1.30	0.94	0.356
118	2-Phenylnitrobenzene	1.30	1.17	0.124
119	2,6-Dibromo-4-nitrophenol	1.36	1.50	0.148
120	2-Nitro-1-naphthol	1.36	1.41	−0.056
121	2,5-Dibromonitrobenzene	1.37	1.27	0.097
122	3,5-Dinitrophenol	1.39	1.18	0.201
123	2,4,6-Trichloronitrobenzene	1.43	1.69	−0.260
124	2,3,5,6-Tetrachloronitrobenzene	1.47	2.06	−0.589
125	2,3,4-Trichloronitrobenzene	1.51	1.69	−0.181
126	3,4-Dinitrotoluene	1.52	0.86	0.651
127	3-Phenylnitrobenzene	1.57	1.21	0.351
128	2-Chloro-4-nitrophenol	1.59	0.88	0.703
129	2,4-Dibromo-6-nitroaniline	1.62	1.30	0.318
130	4-Chloro-3-methyl-6-nitrophenol	1.63	1.03	0.592
131	4,5-Dichloro-2-nitroaniline	1.66	1.19	0.460
132	4-Chloro-2-nitrophenol	1.67	0.87	0.798
133	2,4-Dinitro-5-fluoroaniline	1.69	1.37	0.319
134	2,4-Dinitrofluorobenzene	1.71	1.35	0.359
135	2-Methyl-4,6-dinitrophenol	1.73	1.34	0.385
136	2,4-Dichloro-6-nitrophenol	1.75	1.37	0.377
137	2,3,4,5-Tetrachloronitrobenzene	1.78	2.30	−0.528
138	4-Terbutyl-2,6-dinitrophenol	1.80	2.66	−0.867
139	2,3,4,6-Tetrafluoronitrobenzene	1.87	1.77	0.092
140	1,2,3-Trifluoro-4-nitrobenzene	1.89	1.33	0.551
141	4-Nitrodiphenylamine	1.89	1.07	0.813
142	2,4-Dinitronaphth-1-ol	1.89	2.54	−0.653

(*continued on next page*)

Table 1 (*continued*)

No	Compounds	Obs.	Pred.	Residual.
143	1,5-Difluoro-2,4-dinitrobenzene	2.03	0.91	1.114
144	4,6-Dichloro-5-nitropyrimidine	2.08	1.80	0.272
145	4-Iodo-1,3-dinitrobenzene	2.12	1.70	0.412
146	2,4,6-Trichloro-1,3-dinitrobenzene	2.19	2.48	−0.294
147	4-Bromo-1,3-dinitrobenzene	2.31	1.54	0.767
148	3,5-Dichloro-1,2-dinitrobenzene	2.42	1.83	0.588
149	Pentafluoronitrobenzene	2.43	2.26	0.163
150	1,3-Dinitro-2,4,5-trichlorobenzene	2.60	2.50	0.094
151	2,3,5,6-Tetrachloro-1,4-dinitrobenzene	2.74	3.46	−0.724
	Outliers			
	Nitrophenols			
	3-Methyl-4-nitrophenol			
	2,6-Dichloro-4-nitrophenol			
	2,3-Dinitrophenol			
	3,4-Dinitrophenol			
	4-Nitrophenol			
	Nitrogen containing rings			
	4,6-Dichloro-5-nitropyrimidine			
	2-Chloro-3,5-dinitropyridine			
	2-Hydroxy-4-methyl-3-nitropyridine			
	Others			
	2-Cyanonitrobenzene			
	2-Nitrobenzamide			
	4-Nitroaniline			
	2,5-Difluoronitrobenzene			

Table 2. Experimental and predicted values of toxicity in the *Tetrahymena pyriformis* of compounds in the predicting set

No	Compounds	Observed	Predicted	Residual
1	Benzonitrile	−0.520	−0.662	−0.142
2	4-Cyanobenzamide	−0.380	−0.662	−0.282
3	3-Tolunitrile	−0.250	−0.635	−0.385
4	2-(Hydroxymethyl)nitrobenzene	−0.160	−0.353	−0.193
5	3-Cyanobenzaldehyde	−0.020	−0.591	−0.571
6	2-Cyanophenol	0.040	−0.561	−0.601
7	8-Nitroquinoline	0.080	0.132	0.052
8	Nitrobenzene	0.140	−0.281	−0.421
9	1-Fluoro-3-nitrobenzene	0.200	0.033	−0.167
10	4-Fluoro-2-nitrotolueno	0.250	0.136	−0.114
11	3-Nitroacetophenone	0.320	−0.333	−0.653
12	2-Amino-4-methyl-5-nitropyridine	0.370	−0.245	−0.615
13	Methyl-4-nitrobenzoate	0.400	−0.028	−0.428
14	2-Amino-5-chlorobenzonitrile	0.440	−0.334	−0.774
15	2-Methyl-4-nitroaniline	0.490	−0.215	−0.705
16	3,5-Dinitrobenzyl alcohol	0.530	0.456	−0.074
17	4-Methyl-2-nitrophenol	0.570	−0.048	−0.618
18	2-Methyl-5-nitrophenol	0.660	−0.258	−0.918
19	2-Nitroaniline	0.680	−0.337	−1.017
20	2-Chloro-4-nitroaniline	0.750	0.128	−0.622
21	2-Chloro-5-nitropyridine	0.800	−0.059	−0.859
22	2,6-Dinitroaniline	0.840	0.306	−0.534
23	2-Chloro-3-nitropyridine	0.870	−0.081	−0.951
24	2,5-Dinitrophenol	0.950	0.558	−0.392
25	2-Methyl-1-nitronaphthalene	1.040	0.854	−0.186
26	2,4-Dinitrophenol	1.100	0.558	−0.542
27	3,5-Dichloronitrobenzene	1.130	0.616	−0.514
28	4-Nitrobenzyl chloride	1.180	−0.135	−1.315
29	2,6-Dinitromethylphenol	1.230	0.515	−0.715
30	2,4-Dichloro-6-nitroaniline	1.260	0.606	−0.654
31	2-Chloro-6-methoxy-3-nitropyridine	1.360	0.153	−1.207
32	4-Butoxynitrobenzene	1.420	−0.515	−1.935
33	2,4,5-Trichloronitrobenzene	1.530	1.130	−0.400
34	3-Trifluoromethyl-4-nitrophenol	1.650	1.637	−0.013
35	4-Chloro-3-nitrobenzonitrile	1.710	0.359	−1.351
36	2,6-Diiodo-4-nitrophenol	1.810	1.230	−0.580
37	4-Chloro-1,3-dinitrobenzene	1.980	0.901	−1.079
38	1,2-Dichloro-4,5-dinitrobenzene	2.210	1.274	−0.936
39	4-Chloro-3,5-dinitrobenzonitrile	2.660	1.284	−1.376

Table 3. Step by step outlier deleting study

Models	Variables						Statistical parameters				
	μ_0*$\mu_{3(H)}$	$\mu_{2\ (Dh)}$	μ_0*$\mu_{7(H)}$	μ_0*$\mu_{4(Dh)}$	$\mu_{8(G)}$	b_0	N	R^2	S	Scv	Q^2
1	7.61 e-3	−8.2 e-2	−4.7 e-5	2.01 e-4	7.5 e-5	1.87	202	0.72	0.409	0.029	0.70
2	7.65 e-3	−7.87 e-2	−4.6 e-5	1.89 e-4	7.2 e-5	1.71	198	0.76	0.370	0.027	0.75
3	7.63 e-3	−7.85 e-2	−4.9 e-5	1.88 e-4	9.6 e-5	1.69	194	0.78	0.353	0.026	0.77
4	7.80 e-3	−7.54 e-2	−4.8 e-5	1.77 e-4	7.8 e-5	1.57	190	0.81	0.325	0.024	0.80

models with better statistical profile. In this study the outliers' number were continuously extracted from 0 to 12, considering that a number of outliers lower than 10% of the general data are classically accepted in the literature as threshold limit value for outliers' extraction.[36] In our case the higher extracted outlier number represented a 5.94% of the whole data. All the models together are depicted in Table 3.

A deep analysis of the outlier compounds shown that its can be grouped into two main families of chemicals, one the nitro-phenols and the others aromatic rings contained nitrogen atoms such as pyridines and pyrimidines (only three compounds in the data set). Particularly, 4-nitrophenols (as well as 4-nitroaniline) are chemicals that undergo *abiotic* transformation.[37] It is also possible that 4-substituted nitro-aromatics may be metabolized to quinone-type compounds.[38] Furthermore, Schüürmann et al.[39] postulated that other metabolic and toxicokinetic factors could result in poor prediction for 4-substitued nitrobenzenes. There were detected other outliers without any apparent structural pattern. One of these compounds was the 2-cyanonitrobenzene which presents both the CN and NO_2 groups in the chemical structure. These high electron-withdrawing groups appear separated in the rest of the analyzed chemicals.

Both 2,3-dinitrophenol and 2-hydroxy-4-methyl-3-nitropyridine seemed to be outliers due to intramolecular hydrogen bonding, which affect the reactivity of the nitro group.[40] On the other hand, the 2-chloro-3,5-dinitropyridine was an outlier because it presents an unique mechanism of action.[37] It is remarkable, that the outlier compounds detected in this study do not fully coincide with reported by Cronin et al.,[32] but they match in many cases.

Once the outliers selection was carry out, the number of variables in QSTR equation was studied. By manual fixing of the number of steps in the forward stepwise analysis, the number of molecular descriptors in the model was diminished from 5 to 2 as shown in Table 4. All models were statistically significant, which was evidenced by a p-level <0.05 in the Fisher test.[35,36] The overall accuracy and predictability of these models were measured throughout R^2 and Q^2 values, respectively.

On the other hand, the design and use of both training and predicting series were carried out in order to develop a depth validation of the model by means of an external prediction series. The chemicals in the prediction series were never used to fit the new model. The decision about including a chemical either in the training or predicting set was taken at random.[32] For this study the model with lower number of variables was selected in order to avoid any over-fitting problem after predicting cases leave out. The model per se is presented as follows:

$$-\log(IGC_{50}) = 0.5742 + 3.24 * 10^{-4}\mu_0 * \mu_3(H) - 0.025\mu_2(Dh) \quad (2)$$

$$N = 151 \;\; R^2 = 0.750 \;\; R^2adj = 0.746 \;\; F = 221.574 \;\; p = 0.000 \;\; S = 0.376$$

As can be seen above, the extraction of 39 cases does not affect the magnitude of the regression coefficient. Otherwise, the adjusted R^2 (R^2_{adj}) is only 0.45% lower than the R^2, this fact may be consider as an indicator that the model is not over-fitted by an excess in the variable's numbers.[35,36] The predicted, observed and residual values for all compounds in training series appear in Table 1. Additionally, Figure 1 graphically illustrates the relation between observed and predicted values in the training series.

The regression analysis of predicting series shows a fairly good predictability power for this model ($N = 39$, $R^2_{pred} = 0.66$, $F = 71.72$, $p = 0.0000$, $S = 0.3624$). The predicted, observed and residual values for all compounds in predicting series appear in Table 2. Alternatively, Figure 2 graphically depicts the relation between observed and predicted values in the predicting series.

In the original Cronin' paper[32] only one model had a similar value of the determination coefficient ($R^2 = 0.75$) to our reported model (eq 2). Nevertheless, a comparison between both models is not possible due to in our case a lineal regression techniques was used, while in the Cronin paper a PLS (with two principal component) was developed. This fact permits us to give a simple and more direct interpretation of the model in physicochemical terms.

As can be noted in eq 2, the −log (IGC_{50}) and the toxicity in the *Tetrahymena pyriformis* increase in the same way with high values of μ_0*$\mu_{3\ (H)}$. This fact may be easily explained if we consider that $\mu_{3(H)}$ codify information about topology but mainly molecular hydrophobicity.

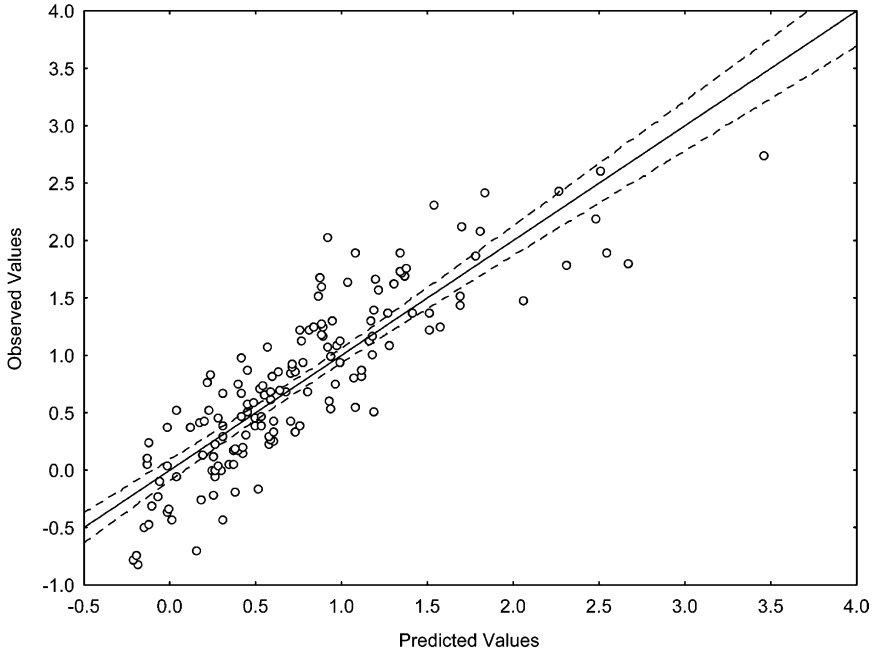

Figure 1. The linear relation between observed and predicted aquatic toxicities for the aromatic compounds of the training set.

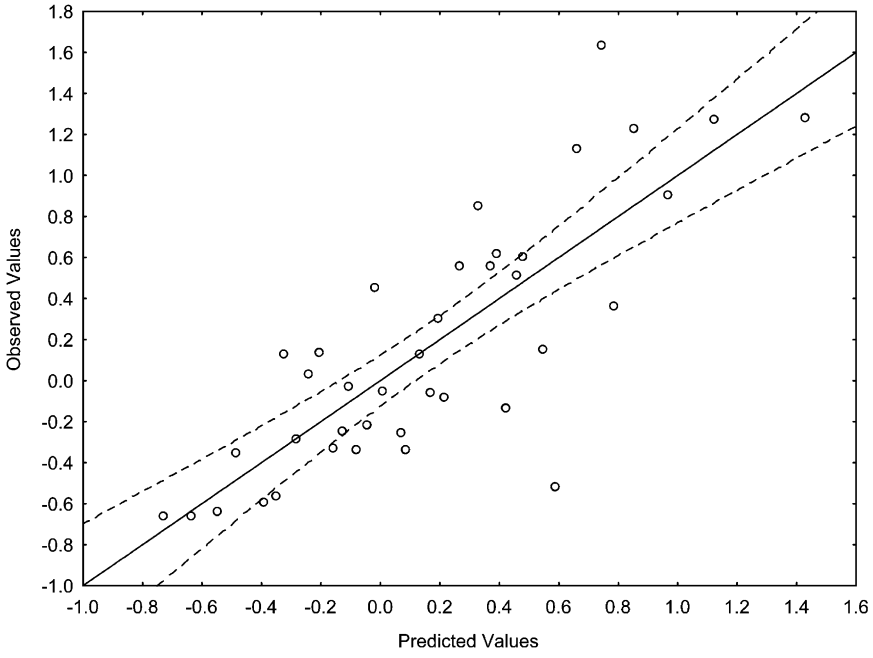

Figure 2. The linear relation between observed and predicted aquatic toxicities for the aromatic compounds of the prediction set.

Besides, it is fully remarkable that this effect is directly modulated by the number of atoms in the molecule (μ_0). This term brings us some information about the influence of the molecular size in the mechanism of penetration to this organism.

On the other hand, an inverse influence of $\mu_{2(Dh)}$ over toxicity seems to be clearly related with the possibility of a stronger affinity of the molecules by water than by the organism tissues. However, this is probable but in some degree speculative interpretation of the model reported.

4. Computation of fragment contributions

As we previously explain, the TOPS-MODE approach is able to compute the contribution of any structural fragment to the toxicological property studied. In the present case, we can find the positive and negative contributions of such fragments to the development of the toxicology activity.

In Figure 3, we show the structures of a series of fragments selected from our database. The contributions to

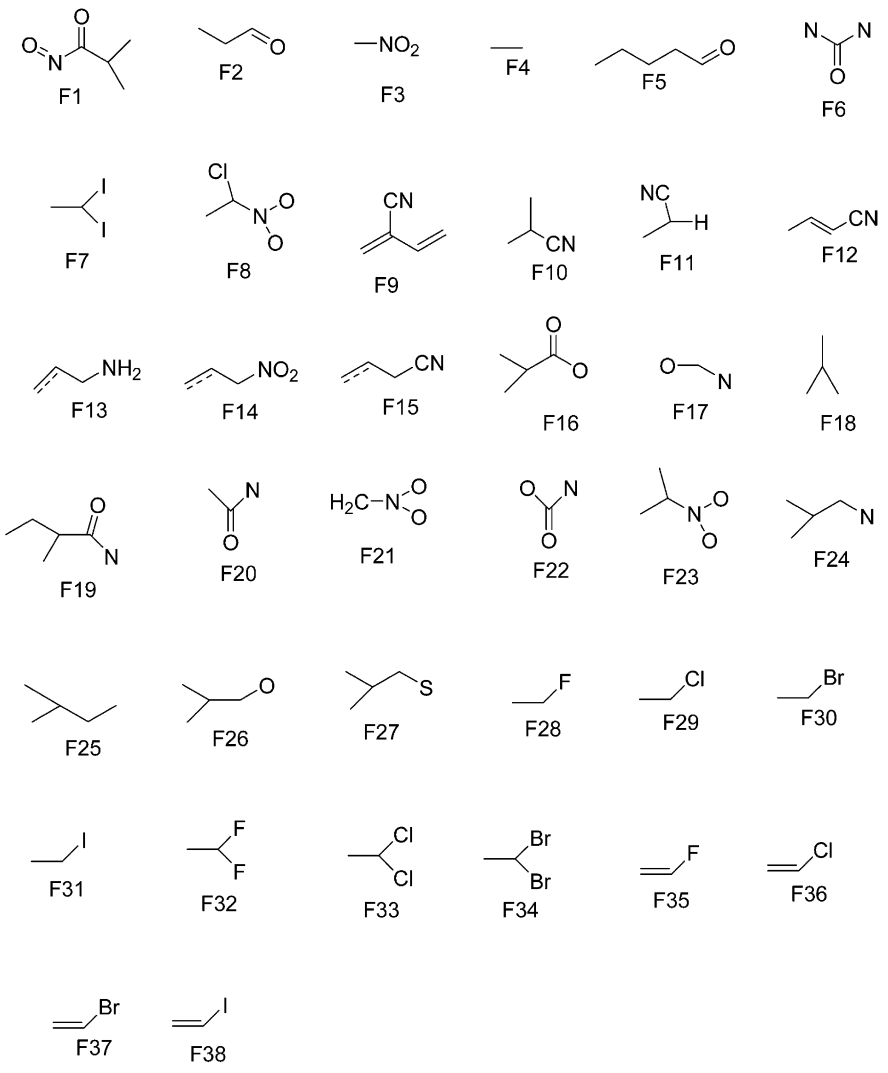

Figure 3. Structures of selected fragments for which their contributions to the toxicological activity are evaluated.

Table 4. Forward stepwise step-by-step analysis

Models	Variables						Statistical parameters				
	$\mu_0*\mu_{3(H)}$	$\mu_{2\ (Dh)}$	$\mu_0*\mu_{7(H)}$	$\mu_0*\mu_{4(Dh)}$	$\mu_{8(G)}$	b_0	N	R^2	S	Scv	Q^2
5	7.80 e-3	−7.54 e-2	−4.8 e-5	1.77 e-4	7.8 e-5	1.57	190	0.81	0.325	0.024	0.80
6	7.21 e-3	−7.14 e-2	−3.3 e-5	1.61 e-4		1.65	190	0.80	0.33	0.025	0.79
7	6.87 e-3	−2.66 e-2	−3.0 e-5			0.26	190	0.79	0.34	0.026	0.78
8	3.21 e-3	−2.32 e-2				0.51	190	0.73	0.38	0.029	0.72

the herbicide activity of these fragments were compute by using the eq 2. These quantitative contributions are given in Table 5.

As can be seen in Figure 3 and Table 5 the set of fragment from F28 to F31 show and interesting behaviour than can be explained if we consider the tendency to increase the fragment contribution value the property under study when increases the hydrophobic character of the carbon–halogen bond. This contribution is justified by our model; where the high toxicity can be explain with the hydrophobicity increase for this serie, contributing to increase in more extension the toxicity of the compounds in the *Tetrahymena pyriformis*. This is easily explicable because these halogens in these type of fragments are very little reactive and therefore their influence is relatively small in the toxicity of the compounds. However, a different contribution can be observed in the set of the fragment F35–F38 where a decreases of the dipolar character of the carbon-halogen bond, increase the contribution to the toxicity in this type of fragments. In the case of the fragment F36, the decreases in the toxicity variation is very small and it does not deserve to be commented as an important difference in the behavior of the set. Finally, the contribution of this set can be explained by the halogen reactivity in this position, which is higher than the set F28–F31. Several other interesting relations can be

Table 5. Contribution of some selected fragments to the toxicological activity

Fragment	Contribution	Fragment	Contribution	Fragment	Contribution
F1	−0.778	F14	−0.587	F27	−0.857
F2	−0.361	F15	−0.751	F28	−0.091
F3	0.025	F16	−0.672	F29	−0.095
F4	−0.036	F17	0.251	F30	−0.087
F5	−0.976	F18	−0.644	F31	−0.073
F6	0.275	F19	−0.963	F32	−0.136
F7	−0.094	F20	−0.052	F33	−0.117
F8	−0.240	F21	0.265	F34	−0.124
F9	−0.414	F22	0.261	F35	0.219
F10	−0.741	F23	−0.491	F36	0.217
F11	−0.443	F24	−0.645	F37	0.223
F12	−0.461	F25	−0.949	F38	0.238
F13	−0.505	F26	−0.660		

obtained by analyzing the contributions of fragments in this table or by computing them for other fragments.

5. Concluding remarks

There is an increasing necessity of QSAR/QSTR models to assess drug discovery and chemicals environmental risk.[41,42] However, most of recently published about QSTR studies have used the artificial neural network techniques with the consequent in the model interpretation.[14–16,18] The TOPS-MODE has been largely probed here to generate fairly good predictive linear models in order to account for *Tetrahymena pyriformis* chemical toxicity. Additionally, TOPS-MODE models reported here are very simple and easily interpreted in physicochemical terms. Then, we have now another strong reason to continue with TOPS-MODE methodology validation in larger data sets of compound.

Acknowledgements

The authors would like to express their gratitude to the Cuban Ministry of Sugar and Ministry of science technology and environment of Villa Clara city (project 0124) for financial support and kindness. Additionally Maykel Pérez thanks the collaboration of the the owner of the software Modeslab 1.0 for the donation of this valuable tool for the realization of this work.

References and notes

1. Lewis, D. F. V. Computer-assisted methods in the evaluation of chemical toxicity. In: *Reviews in Computational Chemistry Vol. III*, Lipkowitz, K. B. and Boyd, D. B., Eds. VCH Publishers, Inc., New York, 1992, p 173.
2. Cronin, M. T. D.; Dearden, J. C. *Quant. Struct.-Act. Relat* **1995**, *14*, 1.
3. Cronin, M. T. D.; Dearden, J. C. *Quant. Struct.-Act. Relat* **1995**, *14*, 117.
4. Cronin, M. T. D.; Dearden, J. C. *Quant. Struct.-Act. Relat* **1995**, *14*, 329.
5. Katritzky, R. A.; Douglas, B. T. J. *Chem. Inf. Comput. Sci.* **2001**, *41*, 1141.
6. Martin, M. T.; Young, M. D. *Chem. Res. Toxicol.* **2001**, *14*, 1378.
7. Cronin, M. T. D.; Dearden, J. C. *Quant. Struct.-Act. Relat* **1995**, *14*, 518.
8. Cronin, M. T. D. *Pharm. Pharmacol. Commun.* **1998**, *4*, 157.
9. Estrada, E.; Uriarte, E.; Montero, A.; Tejeira, M.; Santana, L.; De Clercq, E. *J. Med. Chem.* **2000**, *43*, 1975.
10. Barratt, M.; Richard, A.; Benigni, R. Predicting Toxicity from chemical structure In: *Progress in the Reduction and Refinement and Replacement of Animal Experimentation*. Balls, M., Van Zeller, A.-M., Halder, M.E. (Eds), Elselvier, 2000, p 445.
11. Turner, L.; Choplin, F.; Dugard, P.; Hermens, J.; Jaeckh, R.; Marsman, M.; Roberts, D. *Toxic. In Vitro* **1986**, *3*, 143.
12. Turner, L.; Choplin, F.; Dugard, P.; Hermens, J.; Jaeckh, R.; Marsman, M.; Roberts, D. *ECETOC Monogarph, Brussels* **1986**, *8*, 1.
13. Report of the International Workshop on In vitro Methods for Assessing Acute Systemic Toxicity Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM). National Toxicology Program (NTP). Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). National Toxicology Program Research Triangle Park, NC 27709. NIH Publication No. 01-4499. National Institute of Environmental Health Sciences. National Institutes of Health. US Public Health Service. Department of Health and Human Services. 2001, p 92.
14. Nicolescu, S. P.; Kaiser, K. L. E.; Schults, T. W. *Arch. Environ. Toxicol.* **2000**, *39*, 289 –298.
15. Nicolescu, S. P.; Kaiser, K. L. E. *Chemosphere* **1999**, *38*, 3237.
16. Kaiser, K. L. E.; Nicolescu, S. P. *Environmental Toxicology and Chemistry*. *Setac Press* **2001**, *20*, 420.
17. Cronin, D. T. M.; Schultz, W. T. *Chem. Res. Toxicol.* **2001**, *14*, 1284.
18. Kaiser, K. L. E.; Niculescu, P. S. *Chemosphere* **1999**, *38*, 3237.
19. Roberts, D. W. *Chem. Res. Toxicol.* **1995**, *8*, 545 –551.
20. Cronin, M. T. D.; Manga, N.; Seward, J. R.; Sinks, G. D.; Schultz, W. T. *Chem. Res. Toxicol.* **2001**, *14*, 1498.
21. Barratt M.; Langowski J. Validation and development of the DEREK skin sensitization rule base by analysis of the BgVV list of contact allergens. In *Progress in the Reduction and Refinement and Replacement of Animal Experimentation*. Balls, M., Van Zeller, A.-M., Halder, M. E. (Eds.), Elselvier. 2000, p 445.
22. Todeschini, R.; Consonni, V. *Handbook of molecular descriptors*; Wiley-VCH Weinheim: Germany, 2000.
23. Estrada, E. J. *Chem. Inf. Comput. Sci.* **1996**, *36*, 844.
24. Estrada, E. J. *Chem. Inf. Comput. Sci.* **1997**, *37*, 320.

25. Estrada, E. J. *Chem. Inf. Comput. Sci.* **1998**, *38*, 23.
26. Estrada, E. J. *Chem. Soc. Faraday Trans.* **1998**, *94*, 1407.
27. Estrada, E.; Gutiérrez, Y. J. *Chromatogr. A* **1999**, *858*, 187.
28. Estrada, E.; Gutiérrez, Y.; González, H. J. *Chem. Inf. Comput. Sci.* **2000**, *40*, 1386.
29. Estrada, E.; Peña, A. *Bioorg. Med. Chem.* **2000**, *8*, 2755.
30. Estrada, E.; Uriarte, E. *SAR and QSAR in Environ. Research* **2001**, *12*, 309.
31. Cronin, M. T. D.; Gregory, B. W.; Schultz, T. W. *Chem. Res. Toxicol.* **1998**, *11*, 902.
32. Cronin, T. D. M.; Schultz, T. W. *Chem. Res. Toxicol.* **2001**, *14*, 1284.
33. The method was originally introduced with the name TOSS-MODE that is now changed to TOPS-MODE: Estrada, E. *Environ. Res.* **2000**, *11*, 55–73.
34. Gutierrez, Y. and Estrada, Modeslab 1.0 b, 2002.
35. STATISTICA for Windows, Statsoft Inc. **2002,** version 6.0.
36. Draper, N. R.; Smith, H. Applied Regression Analysis. E. R.; (Ed.), La Habana. 1980.
37. Seward, J. R.; Cronin, M. T. D.; Schultz, T. W. *SAR QSAR Environ. Res.* **2001**, *11*, 489.
38. Schultz, T. W.; Lin, D. T.; Arnold, L. M. *Sci. Total Environ.* **1991**, *109/110*, 569.
39. Schüürmann, G.; Flemming, B.; Dearden, J. C., Cronin, M. T. D.; Schultz, T. W. CoMFA study of acute toxicity of nitrobenzenes to *Tetrahymena pyriformis*. In *Quantitative Structure-Activity Relationships in Enviromental Sciences-VII*, Fredenslund, F. C. and Schüürmann, G., (Eds.), 1997, p 315.
40. Schultz, T. W.; Wesley, S. W.; Lin, D. T. *Qual. Assur.: Good Pract., Law* **1992**, *1*, 132.
41. Devillers, J. *Curr. Op. Drug Discov. Dev.* **2000**, *3*, 275.
42. Kulkarni, S. A.; Hopfinger, A. J. *Pharm. Res.* **1999**, *16*, 8 1245.