Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 13 (2005) 2477–2488
Bioorganic & Medicinal Chemistry

A topological substructural approach applied to the computational prediction of rodent carcinogenicity

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

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

Received 29 November 2004; revised 20 January 2005; accepted 21 January 2005

Abstract—The carcinogenic activity has been investigated by using a topological substructural molecular design approach (TOPS-MODE). A discriminant model was developed to predict the carcinogenic and noncarcinogenic activity on a data set of 189 compounds. The percentage of correct classification was 76.32%. The predictive power of the model was validated by three test: an external test set (compounds not used in the develop of the model, with a 72.97% of good classification), a *leave-group-out* cross-validation procedure (4-fold full cross-validation, removing 20% of compounds in each cycle, with a good prediction of 76.31%) and two external prediction sets (the first and second exercises of the National Toxicology Program). This methodology evidenced that the hydrophobicity increase the carcinogenic activity and the dipole moment of the molecule decrease it; suggesting the capacity of the TOPS-MODE descriptors to estimate this property for new drug candidates. Finally, the positive and negative fragment contributions to the carcinogenic activity were identified (structural alerts) and their potentialities in the lead generation process and in the design of 'safer' chemicals were evaluated.

1. Introduction

A main objective of toxicological research has always been to establish a relationship between the structure of chemical and their harmful biological effects.[4] Among the toxicological endpoints, chemical carcinogenicity has been the target of some attempts to create alternative predictive models because of the experimental determination, for a single chemical in a 2-year standard rodent assay, involves a lot of time and resources. Also the increasing societal and economic pressures to reduce the use of animal testing and the existence of relatively large databases on rodent carcinogenicity in relation to other whole-animal toxicity appear as important reasons to develop 'in silico' carcinogenicity models.[8,48] Nevertheless, among the main problems of poor 'in silico' prediction of the carcinogenicity for any drug or drug-like appears the following: the biological complexity of the biochemical mechanism involved in chemical carcinogenicity to be modeled by machine learning or statistical techniques; the descriptors for chemical structures and properties are inadequate to predict carcinogenicity; the structure–activity relationship (SAR) models ignore the importance of the biological variables and the rodent carcinogenicity classifications are too inaccurate to learn accurate models from them.[48,55]

Most of the general quantitative structure–activity relationship (QSAR) modeling efforts have focused on the challenge of predicting rodent carcinogenicity[18,49] and its has been applied to congeneric and noncongeneric chemicals.[7,4,3,9,15,31,46] There have been a number of approaches for the prediction of carcinogenesis, being these predictive methods split into qualitative and quantitative techniques. In the qualitative (or rule based) assessment of carcinogenicity, a large number of structural features have been associated with this endpoint.

Keywords: Rodent carcinogenicity; Topological approach; In silico; Topological descriptors; TOPS-MODE approach.
* Corresponding author. Tel.: +53 42 281473/281192; fax: +53 42 281130; e-mail addresses: migue@gammu.com; macabrera@uclv.edu.cu

0968-0896/$ - see front matter © 2005 Published by Elsevier Ltd.
doi:10.1016/j.bmc.2005.01.035

This information has been frequently utilized in a number of 'knowledge-based' expert system approaches such as: DEREK, HazardExpert, and OncoLogic.[50,53,60] Other 'correlative' expert system approaches to prediction of carcinogenicity have been developed, including those based on the TOPKAT methodology and MULTI-CASE technology.[17,39,40] Application of MULTI-CASE and the similar 'graph' theory methodologies have attempted to discern structural features with carcinogenicity.[52]

From the topological representation the graph–theoretical methods have become one of the most important tools for quantifying molecular structure. The TOPS-MODE approach[23] based on the calculation of the spectral moments of the bond matrix of molecular graph has been used to generate graph–theoretical descriptors, expressing physical and biological properties in terms of substructural features of molecules. This approach has been successfully applied to different QSPR and QSAR studies.[13,14,21,22,24,27,29,42–44]

Taking into consideration the above mentioned, the aims of the present paper were: to use the TOPS-MODE approach in the generation of discriminant functions by a linear discriminant analysis (LDA) that permits the classification of drugs in carcinogenic and noncarcinogenic and to demonstrate the validity of the 'in silico' models by the use of different validation tests. Finally, to evaluate the positive and/or negative fragment contribution to the carcinogenicity, in order to identify carcinogenic alerts, as an important tool in the drug design field.

2. Materials and methods

2.1. The TOPS-MODE approach

The TOPS-MODE approach is based on the calculation of the spectral moments of the so-called bond matrix,[20] whose theoretical basis has been widely described in previous reports.[21,22,25,26,29] 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.

In order to apply the present approach to the structure–property relationship, the following steps should be followed. First, to select an adequate training set with great structural diversity. Second, to draw the 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 qualitative structure–property relationship by using a discriminant analysis:

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

where P is the studied property, in our case, the carcinogenic activity (C_{act}), μ_k is the kth spectral moment, and the a_k's are the coefficients obtained by the discriminant equation. Sixth, to test the predictive capacity of the model by cross-validation procedures and an external prediction set. Finally, to compute the contribution of the different fragments in order to determine their quantitative contribution to the carcinogenic activity of the molecules studied.

2.2. Selection of bond weights and calculation of molecular descriptors

The current selection of bond weight for calculating TOPS-MODE descriptors was carried out through accounting for hydrophobic/polarity, electronic, and steric features of molecules. Thus, atomic contributions for partition coefficient (H),[57] polar surface area (PSA),[19] polarizability (Pol),[41] are transformed into bond contributions. A similar approach is used to transform Gasteiger–Marsilli atomic charge (Ch),[32] van der Waals atomic radii (vdW),[11] molar refraction (MR),[33] and atomic mass (AM) into bond weights. The way in which these atomic contributions were transformed into bond contributions have been described by Estrada et al.:[28]

$$w(i,j) = \frac{w_i}{\delta_i} + \frac{w_j}{\delta_j} \qquad (2)$$

where w_i and δ_i are the atomic weight and vertex degree of the atom i. Also the standard bond dipole moments (Dp and Dp2) were used as bond weight.[30,45] The molecular weight (MW) of each compound was considered as other variable. The calculation of the TOPS-MODE descriptors was carried out with the computer software Modeslab 1.0.[38] The input of the chemical structures into the Modeslab software was as simplified molecular input line entry specification (SMILES).[58] We calculated the first 15 spectral moments (μ_1–μ_{15}) for each bond weight and the number of bonds in the molecules (μ_0). Also considering the nonlinearity of the biological process studied (carcinogenic activity) were evaluated the interactions between μ_0 and μ_1 with all variables (descriptors). The spectral moment were calculated considering the molecules with (H) and without hydrogen.

2.3. Biological data

The carcinogenic data used was collected from the Carcinogenic Potency Database (CPDB) established by Gold and Zeiger[35] available at (http://potency.berkeley.edu/cpdb.html). The CPDB is a single standardized resource of many years of chronic, long-term carcinogenesis bioassays. It contains a larger diversity of chemical structures (more than 1300 tested substance), and includes tumor data reproduced from all of the NCI/NTP rodent bioassay *Technical Reports* as well as additional data extracted from over 1200 literature sources subjected to extensive review.[34]

2.4. Development of a discrimination function for carcinogenic activity

In order to obtain the discriminant functions, different carcinogenic, and noncarcinogenic criterions were followed:

Criterion 1: A chemical is categorized as a carcinogen if it causes tumors, by statistically significant measure, at a single site in any one of four rodent species/sex groups.[47] The noncarcinogenic compounds were selected when the reported experiments were negative or negative/cero.

Criterion 2: The previous concept was followed but considering only the rat specie, it has been demonstrated that rat bioassay results are considerably more reproducible than mouse bioassay.[37] The noncarcinogenic compounds were selected when all the experiments had negative values.

Criterion 3: A chemical is categorized as a carcinogen only if it causes tumors at multiple organ sites or in multiple rodent species.[16] The noncarcinogenic compounds were selected when the reported experiments were negative or negative/cero, in multiple rodent species.

The data for the first model (first criterion) was composed by 830 compounds (533 carcinogenic and 297 noncarcinogenic). The second data (second criterion) had 283 compounds (108 carcinogenic and 175 noncarcinogenic) and the last data set (third criterion) was composed by 189 compounds (119 carcinogenic and 70 noncarcinogenic).

The data set for each case was randomly divided in two subsets, the training and the test set (20% of the general data set). The compounds belonging to the test set were never used in the development of the discriminant function and were reserved to validate the discriminant model.

The discriminant function was obtained by using the stepwise LDA as implemented in the STATISTICA version 6.0.[54] The default parameters of this program were used in the development of the model. The variables to be included in the equation were selected using a forward stepwise procedure as a variable selection strategy.

The quality of the model was determined by examining the Wilks λ, the Mahalanobis distance (D^2), the Fisher ratio (F), and the corresponding p-level ($p(F)$) as well as the percentage of good classification and the proportion between the cases and variables in the equation. The Wilks λ statistic establish a perfect discrimination for $\lambda = 0$ and not discrimination when $\lambda = 1$. The Mahalanobis distance indicates the separation of the respective groups, showing whether the model possesses an appropriate discriminatory power for differentiating between the two respective groups.

In developing the classification function the values of 1 and -1 were assigned to active and inactive compounds. We will use the posteriori probabilities in order to classify the compounds as carcinogenic and noncarcniogenic. This is probable that the respective case belongs to a particular group (carcinogenic or noncarcinogenic). It is proportional to the Mahalanobis distance from that group centroid.

Compounds were considered unclassified by the model (NC) when the differences in the percentage of classification between groups do not differ in more than 5%. However, we have used the posteriori probabilities in order to classify the compounds as carcinogenic/noncarcinogenic. This is the probability that the respective case belongs to a particular group (carcinogenic or noncarcinogenic). It is proportional to the Mahalanobis distance from that group centroid.

Also a *full* cross-validation test for the final model was investigated. From the general data set, 20% of the data were randomly selected four times. Each group was left out (*leave-group-out*, LGO) and it was predicted by the model developed from the remaining observations. In this way, every observation was left out once and its value predicted. The significant criterion for assessing model quality was the classification percentage, the squared Mahalanobis distance (D^2) and the Wilks statistic (λ).

Finally, the predictive power of all discriminant models was assessed, in retrospective form, with two external data set reported in the literature.[12] These data sets belong to the first and second exercises of the National Toxicology Program on the prediction of rodent carcinogenicity.

2.5. Computation of fragment contributions

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 QSPR methods. This procedure can be useful for the identification of structural alerts in the study of carcinogenic prediction. The computation of fragment contribution has been described in more details in previous reports.[23,22,24,27,29] The procedure consists of calculating the spectral moments for all the fragments contained in a given substructure, and by the difference of these moments we obtain the contribution of the substructure. The general algorithm followed is 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 (subgraphs), 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. Having the contributions of the different structural fragments of interest, we only need to substitute these contributions into the quantitative model developed to describe the carcinogenic activity (Eq. 5), and we obtain the quantitative contribution of the different fragments to carcinogenic activity.

3. Results

The functions to discriminate carcinogenic from noncarcinogenic compounds in the training set (C_{act}), are given below:

Model 1 (first criterion)

$$C_{act} = -0.005\mu_4^{Dp} + 0.001\mu_5^{Dp} - 0.038\mu_3^{ChH} + 1.41\mu_1^{Ch} + 0.009\mu_4^{HH} - 3.02 \times 10^{-9}\mu_{13}^{HH} - 0.037\mu_1\mu_1^{HH} + 1.07 \times 10^{-4}\mu_1\mu_4^{HH} - 0.554\mu_1^{H} - 0.089\mu_1\mu_1^{H} + 0.03\mu_1\mu_3^{H} - 0.004\mu_1\mu_4^{H} + 1.626 \quad (3)$$

Model 2 (second criterion)

$$C_{act} = -0.529\mu_0^{Dp} - 0.494\mu_1^{Dp2H} - 0.001\mu_0\mu_0^{Dp2H} + 0.022\mu_1\mu_1^{Dp2H} - 0.012\mu_1\mu_1^{H} + 2.73 \times 10^{-5}\mu_1\mu_3^{MR} - 3.9 \times 10^{-5}\mu_7^{Pol} + 3.66 \times 10^{-9}\mu_1\mu_{10}^{Pol} + 0.183\mu_2^{vdW} - 0.026MW + 0.081 \quad (4)$$

Model 3 (third criterion)

$$C_{act} = 0.324\mu_1^{DpH} + 0.048\mu_2^{Dp2} - 0.080\mu_3^{Dp2} + 3.4 \times 10^{-10}\mu_1\mu_{14}^{Dp2} - 6.58 \times 10^{-10}\mu_0\mu_{12}^{HH} + 2.15 \times 10^{-11}\mu_1\mu_{15}^{HH} + 1.554 \quad (5)$$

The statistical parameters obtained for each model are given in Table 1.

As can be seen, the values reported of Wilks' λ decrease from model 1 to model 3, while the squared Mahalanobis distance (D^2) and the percentage of total classification (% Class) increase in this direction, evidencing that the last model (Eq. 5) possess the best discriminatory power of the three models. The number of compounds per variables in these models always was higher than the traditionally accepted values of 5. Taking into consideration the above mentioned, in the present paper, we will only report and discuss the results achieved by the third criterion of carcinogenicity (model 3). The numerical values of the spectral moments used in the model 3 (Eq. 5) are given as Supporting information.

Table 1. Statistical parameters of the linear discriminant analysis for the three carcinogenicity models

Models	N	λ	D^2	F_{exp}	N_v	% C_{act}	% n-C_{act}	% Class
1	663	0.86	0.72	9.04	12	68.55	59.07	63.81
2	226	0.75	1.42	7.24	10	69.77	69.29	69.53
3	152	0.65	2.24	12.79	6	78.13	73.21	76.32

N: number of compounds in the training set; λ: Wilks' statistic; D^2 squared of Mahalanobis distance; F_{exp}: experimental Fisher ratio; N_v: number of variable in the equation; % C_{act}: percentage of good classification for carcinogenic compounds in the training set; % n-C_{act}: percentage of good classification for noncarcinogenic compounds in the training set; % Class: percentage of total good classification in the training set.

The classification results for the training set, using model 3 (Eq. 5), are given in Table 2.

The model classified correctly the 78.13% of chemicals with carcinogenic activity in the training set and the 73.21% of noncarcinogenic compounds, for a global good classification of 76.32% (see Table 1). If the unclassified compounds are considered the percentages of good classification will be 81.11% and 76.92% for carcinogenic and noncarcinogenic chemicals, respectively. The percentages of *false positives* and *false negatives* in the training set were 11.18% (17/152) and 7.89% (12/152), respectively. *False positives* are those compounds without carcinogenic activity that are classified as active, and the *false negatives* are those compounds with carcinogenic activity and the model classified it as inactive (see Table 2). From a practical point of view, in the development of the classification model, is considered more important avoiding *false positives* because those are compound that will be rejected for their wrong predicted property and therefore they will never be evaluated experimentally, and their true carcinogenic activity would never be discovered. On the contrary, the *false negatives* compounds eventually will be detected.

The most important criterion for the quality of the discriminant model is based on the statistics for the external test set. Equation 5 classified correctly the 73.91% and 71.45% of carcinogenic and noncarcinogenic compounds, respectively. The global classification was 72.97%. The percentage of *false negative* and *false positive* compounds was 16.21% (6/37) and 10.81% (4/37), respectively. In Table 3 the classification of compounds in the external test set is presented.

Also a cross-validation *leave-group-out* procedure (removing 20% of the complete data set) was carried out. The range of good classification was between 74.56% and 78.07%. The global classification of the model was 76.32% (see Table 4).

Another criterion to judge the performance of the discriminant model (Eq. 5) was the prediction of compounds belonging to the first and second exercise of the National Toxicology Program (NTP), respectively.[5,10] Although, it was not a prospective exercise, it was considered as such, due to the predictions that were obtained once validating the theoretical model. The good global classification was 55% for the first exercise and 63% for the second one (see Tables 5 and 6). Nevertheless, when the third carcinogenic criterion was considered, the global classification was 58% for both exercises.

The results achieved with the present retrospective approach were compared with other prospective theoretical predictions reported in the literature for both data sets.[5,10] In Tables 7 and 8 are depicted the comparative results. The results, with the TOPS-MODE approach, were similar to other achieved by the best prediction sys-

Table 2. Results of the classification of the compounds in the training set, according to the third carcinogenic criterion

No.	Compounds	Prob	Class	No.	Compounds	Prob	Class
Carcinogenic compounds							
1	1-(2-Hydroxyethyl)-1-nitrosourea	64.65	+	**49**	Diallylnitrosamine	94.25	+
2	1,1-Dimethylhydrazine	77.90	+	**50**	Dibromodulcitol	76.10	+
3	1,2-Dibromoethane	89.10	+	**51**	Dibromomannitol	76.10	+
4	1,2-Dimethylhydrazine·2HCl	86.98	+	**52**	Dieldrin	95.51	+
5	1′-Hydroxysafrole	82.35	+	**53**	Diethylstilbestrol	92.16	+
6	1-Nitroso-3,4,5-trimethylpiperazine	66.48	+	**54**	Dihydrosafrole	84.88	+
7	2,2,2-Trifluoro-*N*-[4-(5-nitro-2-furyl)-2-thiazolyl]acetamide	41.13	–	**55**	*dl*-Ethionine	86.54	+
8	2,3,7,8-Tetrachlorodibenzo-*p*-dioxin	14.55	–	**56**	Ethylene thiourea	69.98	+
9	2,4,5-Trimethylaniline·HCl	81.12	+	**57**	Formaldehyde	83.16	+
10	2,4,6-Trimethylaniline·HCl	81.17	+	**58**	Glu-P-1	77.02	+
11	2,4-Diaminotoluene·2HCl	79.22	+	**59**	Glu-P-2	79.92	+
12	2,4-Dinitrotoluene	90.06	+	**60**	Hexachlorobenzene	17.87	–
13	2,5-Dimethoxy-4′-aminostilbene	88.49	+	**61**	Hydrazine	84.40	+
14	2,5-Xylidine·HCl	83.47	+	**62**	Hydrazine sulfate	55.90	+
15	2-Amino-4-(5-nitro-2-furyl)thiazole	88.47	+	**63**	Isoniazid	68.97	+
16	2-Hydrazino-4-(5-nitro-2-furyl)thiazole	88.22	+	**64**	Lead acetate, basic	29.66	–
17	2-Hydrazino-4-(*p*-aminophenyl)thiazole	85.31	+	**65**	MeA-α-C acetate	47.75	NC
18	2-Hydrazino-4-(*p*-nitrophenyl)thiazole	89.03	+	**66**	Melphalan	43.91	–
19	2-Naphthylamine	84.30	+	**67**	Metronidazole	87.12	+
20	3-(5-Nitro-2-furyl)-imidazo(1,2-α)pyridine	89.99	+	**68**	*N*-[4-(5-Nitro-2-furyl)-2-thiazolyl]formamide	79.14	+
21	3-Aminotriazole	51.06	NC	**69**	*N*-[5-(5-Nitro-2-furyl)-1,3,4-thiadiazol-2-yl]acetamide	65.93	+
22	3-Nitro-3-hexene	94.74	+	**70**	Nitrobenzene	90.45	+
23	4-Chloro-4′-aminodiphenylether	68.31	+	**71**	Nitroso-2-oxopropylethanolamine	79.76	+
24	5-Azacytidine	21.39	–	**72**	Nitrosodibutylamine	95.57	+
25	Acetaldehyde	81.05	+	**73**	*N*-Nitroso-2,3-dihydroxypropyl-2-hydroxypropylamine	85.48	+
26	Acetamide	60.75	+	**74**	*N*-Nitrosoallyl-2-oxopropylamine	82.08	+
27	Acetaminophen	49.74	NC	**75**	*N*-Nitrosomethyl-2,3-dihydroxypropylamine	89.86	+
28	AF-2	81.69	+	**76**	*N*-Nitroso-*N*-methylurea	56.84	+
29	Aldrin	98.12	+	**77**	*N*-Nitrosopiperidine	93.67	+
30	α-1,2,3,4,5,6-Hexachlorocyclohexane	37.09	–	**78**	*N*-Nitrosopyrrolidine	92.56	+
31	Aramite	55.23	+	**79**	*o*-Phenylenediamine·2HCl	81.71	+
32	Azobenzene	88.99	+	**80**	*o*-Toluidine·HCl	85.44	+
33	Benzidine	86.24	+	**81**	*p*,*p*′-DDE	59.70	+
34	Benzo(*a*)pyrene	75.92	+	**82**	Phenobarbital, sodium	11.68	–
35	β-Propiolactone	64.15	+	**83**	PhIP·HCl	81.27	+
36	Bis-(chloromethyl)ether	85.72	+	**84**	Piperonyl butoxide	73.72	+
37	Bis-2-hydroxyethyldithiocarbamic acid, potassium	79.15	+	**85**	Propylthiouracil	41.01	–
38	Caffeic acid	65.64	+	**86**	Sesamol	72.00	+
39	Captafol	47.80	NC	**87**	Sterigmatocystin	24.10	–
40	Captan	53.42	+	**88**	Streptozotocin	35.03	–
41	Catechol	82.44	+	**89**	Styrene oxide	83.43	+
42	Chlorambucil	58.86	+	**90**	Thiouracil	39.88	–
43	Chloroform	49.18	NC	**91**	Trichloroethylene	66.09	+
44	Chloromethyl methyl ether	86.86	+	**92**	Trp-P-2 acetate	47.42	–
45	Chrysazin	43.38	–	**93**	Uracil	25.49	–
46	Cyclophosphamide	60.76	+	**94**	Urethane	51.38	NC
47	D&C red no. 5	34.41	–	**95**	Vinyl acetate	60.81	+
48	DDT	35.62	–	**96**	Vinyl chloride	85.46	+
Noncarcinogenic compounds							
97	1,4-Dichlorobenzene	71.23	+	**125**	FD&C blue no. 1	2.88	–
98	1-[(5-Nitrofurfurylidene)amino]hydantoin	20.99	–	**126**	FD&C blue no. 2	13.13	–
99	1-Chloro-2,4-dinitrobenzene	81.51	+	**127**	FD&C red no. 3	63.47	+
100	2,3,4,5,6-Pentachlorophenol	21.50	–	**128**	FD&C yellow no. 5	6.18	–
101	2,4,5-Trichlorophenoxyacetic acid	21.70	–	**129**	FD&C yellow no. 6	44.58	–
102	3-Nitro-4-hydroxyphenylarsonic acid	52.12	NC	**130**	Fenvalerate	13.20	–
103	5,5-Diphenylhydantoin	23.32	–	**131**	Fluoxetine·HCl	50.76	NC
104	6-Dimethylamino-4,4-diphenyl-3-heptanone·HCl	24.26	–	**132**	Hexamethylenetetramine	16.33	–
105	Adipamide	35.72	–	**133**	Isopropyl-*N*-(3-chlorophenyl)carbamate	30.19	–
106	Aspirin	26.10	–	**134**	Methotrexate	2.32	–
107	Benzoate, sodium	69.65	+	**135**	Misoprostol	1.27	–
108	Benzoguanamine	86.87	+	**136**	Monochloroacetic acid	60.76	+

(*continued on next page*)

Table 2 (*continued*)

No.	Compounds	Prob	Class	No.	Compounds	Prob	Class
109	Black PN	1.69	–	**137**	Nefiracetam	13.89	–
110	C.I. pigment yellow 12	1.81	–	**138**	*N*-Nitrosocimetidine	47.24	–
111	C.I. pigment yellow 16	0.34	–	**139**	Oxprenolol·HCl	72.77	+
112	C.I. pigment yellow 83	0.24	–	**140**	Phenyl isothiocyanate	80.22	+
113	Cadmium acetate	33.23	–	**141**	Prazepam	7.34	–
114	Caffeine	9.33	–	**142**	Praziquantel	4.39	–
115	Chlorodifluoromethane	60.57	+	**143**	Propranolol·HCl	70.04	+
116	Chromium (iii) acetate	13.89	–	**144**	Quercetin dihydrate	28.39	–
117	Compound 50-892	30.37	–	**145**	Rutin trihydrate	0.15	–
118	Diazepam	28.23	–	**146**	Sorbic acid	81.52	+
119	Dichlorodifluoromethane	40.48	–	**147**	Sotalol·HCl	42.14	–
120	Dimethylformamide	60.94	+	**148**	Tetramethylthiuram disulfide	43.06	–
121	Erythorbate, sodium	51.22	NC	**149**	Tilidine fumarate	6.38	–
122	Estazolam	34.63	–	**150**	Trichlorofluoromethane	37.58	–
123	Ethynodiol diacetate	0.73	–	**151**	Trifluralin, technical grade	29.63	–
124	Etodolac	27.48	–	**152**	Urea	49.69	NC

+ Positive values are for compounds with carcinogenic activity; – Negative values are for compounds with noncarcinogenic activity; NC: non-classified drugs.

Table 3. Results of the classification of the compounds in the external prediction set, according to the third carcinogenic criterion

No.	Compounds	Prob	Class	No.	Compounds	Prob	Class
Carcinogenic compounds							
1	[4-Chloro-6-(2,3-xylidino)-2-pyrimidinylthio]acetic acid	63.59	+	**13**	Methyl *tert*-butyl ether	68.93	+
2	2-Acetylaminofluorene	66.43	+	**14**	*N*-[4-(5-Nitro-2-furyl)-2-thiazolyl]acetamide	62.15	+
3	3,3′,4,4′-Tetraaminobiphenyl·4HCl	72.80	+	**15**	*N*-Hydroxy-2-acetylaminofluorene	56.06	+
4	Auramine-O	45.60	–	**16**	Nitroso-2,3-dihydroxypropyl-2-oxopropylamine	75.96	+
5	Butylated hydroxyanisole	60.59	+	**17**	*N*-Nitrosodimethylamine	88.55	+
6	Carbon tetrachloride	35.07	–	**18**	*N*-Nitrosomorpholine	91.78	+
7	Ciprofibrate	3.86	–	**19**	Phenacetin	52.53	NC
8	Dichloroacetylene	86.54	+	**20**	Safrole	85.39	+
9	Doxylamine succinate	12.75	–	**21**	Tamoxifen citrate	4.25	–
10	Formic acid 2-[4-(5-nitro-2-furyl)-2-thiazolyl]hydrazide	80.71	+	**22**	Thioacetamide	76.41	+
11	IQ	71.72	+	**23**	Trp-P-1 acetate	42.60	–
12	MeIQx	68.65	+				
Noncarcinogenic Compounds							
24	1,1,1-Trichloroethane	20.66	–	**31**	FD&C green no. 3	1.94	–
25	1,2-Dichloroethane	88.77	+	**32**	Flecainide acetate	10.89	–
26	Barium acetate	33.23	–	**33**	Maleic hydrazide	38.53	–
27	Benzalazine	25.25	–	**34**	Oxamyl	25.95	–
28	Cyclohexylamine·HCl	89.98	+	**35**	Saccharin	29.62	–
29	Deltamethrin	4.81	–	**36**	Toluene diisocyanate	33.15	–
30	Diphenyl-*p*-phenylenediamine	82.67	+	**37**	Triprolidine·HCl monohydrate	89.82	+

+ Positive values are for compounds with carcinogenic activity; – Negative values are for compounds with noncarcinogenic activity; NC: non-classified drugs.

tems, evidencing that a considerable proportion of rodent noncarcinogens (>50% of the data) were predicted as positives by more than 50% of the approaches (false positives). These results are agreed with reported in the literature,[10] evidencing that the presence of so many 'difficult' chemicals confirms the fact that the sample of chemicals originally selected for experimentation (and then for the prediction exercise) was willingly—and obviously—biased toward suspected chemicals.

Table 4. Results of the cross-validation *leave-group-out* procedure (removing 20% of the complete data set)

	λ	% Class carcinogenic compounds	% Class noncarcinogenic compounds	% Global class
1	0.649	77.78	69.03	74.56
2	0.609	81.94	71.42	78.07
3	0.663	75.00	78.57	76.32
4	0.668	80.56	69.05	76.32
Average	0.647	78.82	72.01	76.31

4. Discussion

The variables in the model 3 (Eq. 5) encoded specific structure information. As can be seen the variables in this model are related to the hydrophobicity and the

Table 5. Results of the rodent carcinogenicity prediction, by the TOPS-MODE, for the first NTP comparative exercise

Compounds	Carcinogenic activity	Prob	Class (Model 3)	Class (Model 3)[a]
dl-Amphetamine sulfate	–	66.32	+	+
Promethazine hydrochloride	–	55.83	+	+
Resorcinol	–	82.43	+	+
Polysorbate 80	np			
γ-Butyrolactone	–	68.93	+	+
Manganese sulfate monohydrate	–	89.93	+	
Monochloroacetic acid	–	60.76	+	+
p-Nitrophenol	–	86.56	+	
Tricresyl phosphate	–	46.76	–	–
Ethylene glycol	–	90.77	+	
Theophylline	–	6.24	–	–
4,4′-Thiobis(6-*t*-butyl-*m*-cresol)	–	60.74	+	+
Chloramine	–	84.34	+	+
4,4′-Diamino-2,2′-stilbene disulfonic acid	–	22.23	–	–
Methyl bromide	–	88.57	+	
Sodium azide	–	85.73	+	
C.I. pigment red 23	–	37.13	–	–
4′-Hydroxyacetanilide	–	45.74	–	–
Titanocene dichloride	–	0.00	–	
HC yellow 4	–	75.95	+	+
p-Nitroaniline	–	86.28	+	
Naphthalene	+	89.11	+	
o-Benzyl-*p*-chlorophenol	+	86.47	+	
2,2-Bis(bromomethyl)-1,3-propanediol	+	77.13	+	+
tert-Butyl alcohol	+	67.71	+	+
3,4-Dihydrocoumarin	+	64.19	+	+
Mercuric chloride	np			
Methylphenidate hydrochloride	+	64.54	+	
Triamterene	+	77.97	+	
Diphenylhydantoin	+	26.25	–	
Pentachloroanisole	+	20.93	–	–
p-Nitrobenzoic acid	+	68.33	+	
Tris(2-chloroethyl)phosphate	+	76.72	+	
Direct blue 218	+	1.89	–	–
C.I. pigment red 3	+	76.69	+	+
2,4-Diaminophenol dihydrochloride	+	75.31	+	
Salicylazosulfopyridine	+	62.15	+	+
C.I. acid red 114	+	4.69	–	
C.I. direct blue 15	+	1.91	–	
Coumarin	+	62.21	+	+
2.3-Dibromo-1-propanol	+	84.91	+	+
3.3′-Dimethylbenzidine	+	81.18	+	
o-Nitroanisole	+	86.73	+	+
1,2,3-Trichloropropane	+	84.15	+	+

np: not performed.

[a] Classification of chemicals, by Model 3, according to third carcinogenic criterion.

bond dipole moment. In general way, the variables weighted with the bond dipole moment have a negative contribution to the carcinogenic property meanwhile the variables weighted with hydrophobicity have a positive contribution. This result agrees with that reported in the literature[7,8] where have been widely demonstrated that hydrophobicity is the key factor for carcinogenic activity. It is explained due to the hydrophobic compounds have a better capacity to cross the biological membrane and to reach different tissues and organs in the body, with a high probability to affect this zones. Carcinogenic activity also depends of the bond dipole moment, decreasing it with an increasing of this property. This results is explained by the fact when the total polarity of molecules is increased, the hydrophobicity is lower, affecting the capacity to permeate the biological membranes.

One of the most advantages that the present approach brings to QSTR and QSAR studies is concerned with the structural interpretability of the models. The interpretability of QSTR/QSAR models in terms of structural contributions to the studied property can provide important insights into possible mechanisms participating in the interaction of these molecules with the biological media. The detection of those fragment with positive contribution to the studied property play an important role, because they constituted a possible structural alert, which have a special interest in toxicological studies. Nevertheless, by the present approach, both positive and/or negative fragment contribution can be detected and their influence can be interpreted in terms of physicochemical or biological processes.[8] The structural fragments selected are illustrated in Figure 1 and their contributions to carcinogenicity activity appear in

Table 6. Results of the rodent carcinogenicity prediction, by the TOPS-MODE, for the second NTP comparative exercise

Compounds	Carcinogenic activity	Prob	Class (Model 3)	Class (Model 3)[a]
Scopolaminehydrobromide trihydrate	–	2.60	–	–
Codeine	–	12.87	–	–
Emodin	–	31.26	–	–
tert-Butylhydroquinone	–	64.05	+	+
iso-Butyraldehyde	–	84.63	+	+
1-Chloro-2-propanol	–	86.10	+	+
Xylene sulfonic acid, sodium salt	–	87.48	+	+
Cinnamaldehyde	–	91.48	+	+
Sodium nitrite	–	92.85	+	+
Citral	–	93.32	+	+
Oxymethalone	+	0.47	–	
Primaclone	+	25.80	–	
Phenolphthalein	+	44.97	–	–
D&C yellow no.11	+	56.73	+	
Anthraquinone	+	59.41	+	
1,2-Dihydro-2,2,4-trimethylquinoline	+	69.54	+	
Chloroprene	+	82.81	+	+
Isobutene	+	87.04	+	
Methyleugenol	+	87.96	+	+
Nitromethane	+	89.59	+	+
Cobalt sulfate heptahydrate	+	89.91	+	+
Pyridine	+	91.11	+	+
Diethanolamine	+	91.92	+	
Furfuryl alcohol	+	92.35	+	+
Tetrahydrofuran	+	92.37	+	+
Ethylene glycol monobutyl ether	+	93.14	+	
Ethylbenzene	+	93.20	+	+
Gallium arsenide	+	np		
Molybdenum trioxide	+	np		
Vanadium pentoxide	+	np		

np: not performed.
[a] Classification of chemicals, by model 3, according to third carcinogenic criterion.

Table 7. First NTP comparative exercise on the prediction of rodent carcinogenicity

Prediction method[a]	Prediction accuracy (%)
Ten	75
RAS	68
Wei	65
Ke	65
DER	59
TOP	57
Ben	57
Sty	57
Deh	56
TOPS-MODE	55 (58)[b]
Lij	55
COM	53
CAS	49

[a] For details see Ref. 56.
[b] The prediction accuracy when the third carcinogenic criterion was considered.

Table 9. The fragment contributions were calculated by the ModesLab software.[38]

The role of structural factor in the prediction of carcinogenic activity and in the explanation of the toxicological mechanism is discussed here, using some examples of chemical carcinogens (see Fig. 2)[60] that were well classified by Eq. 5.

The polynuclear aromatic hydrocarbons represent one of the most extensively studied classes of chemical carcinogens. Among them one of the most potent carcinogens is the benzo(α)pyrene. It has low water solubility and its elimination generally takes the form of conversion into more water-soluble substance, which is more readily excreted. For this reason the benzo(α)pyrene is converted to epoxide through a Phase I reaction. By the action of water, this substance form benzo(α)pyrene-7,8-dihydro-7,8-diol in human lung by cytochrome P4501A1-mediated process. This diol undergoes further epoxidation (in a regioselective and stereoselective way) to yield dihydroxy epoxides, benzo(α)pyrene-7,8-dihydrodiol-9,10-oxide (epoxide hydrolase and other cytochrome P450, catalyzed).[2] This dihydroxy epoxide is believed to be the actual carcinogen formed by metabolism of benzo(α)pyrene.

Our predictions explain the above mentioned well. It is known that this polynuclear aromatic hydrocarbon has a high aromaticity. For that reason, reactions related with the rupture of the internal double bond of this compound will not be favored from an energetic point of view, due to these bonds are involved in the aromaticity of the molecule. This fact is explained well by the model; because the internal bonds represented for the fragment F_1 have a negative contribution to the carcinogenic activity (see in Table 9 the negative value for this fragment).

Table 8. Second NTP comparative exercise on the prediction of rodent carcinogenicity

Prediction method[a]	Prediction accuracy (%)
Oncologic	65
SHE	65
R1	64
TOPS-MODE	63 (58)[b]
Huff et al.	62
R2	61
Benigni et al.	61
Tennant et al.	60
Ashby	57
Bootman	53
FALS	50
Benigni, old	48
RASH	45
Ke (estimated)	44
COMPACT	43
DERECK	43
Salmonella	33
Purdy	32
PROGOL	29
CAS	25

[a] For details see Ref. 57.
[b] The prediction accuracy when the third carcinogenic criterion was considered.

Table 9. The contributions of different structural fragments to the carcinogenic activity

Studied fragments	Fragment contributions
F_1	−0.38
F_2	0.29
F_3	0.38
F_4	0.19
F_5	0.10
F_6	0.47
F_7	0.74
F_8	−0.36
F_9	1.18
F_{10}	0.20
F_{11}	0.94
F_{12}	0.13
F_{13}	0.14
F_{14}	−0.44
F_{15}	0.22

The chemical reactions of this compound take place in the external bonds of the molecule, and according to our model this is a potential carcinogenic area. This external zone is represented for the fragments F_2, F_3, F_4, and F_5, which have a contribution positive to the property (see Table 9). It is important to remark that the TOPS-MODE approach recognize the more external aromatic ring as the molecular zone with the highest carcinogenic contribution (F_3), and this fact is in relation with the metabolism pathway of this compound. On the other hand we have seen that the carcinogenic activity is increased in the following order F_5, F_4, F_2, and F_3. It is a logic result due to the intermediate formed (a pyrene derivative) has a lower energy than any other formed in other position. This result is agreed with the

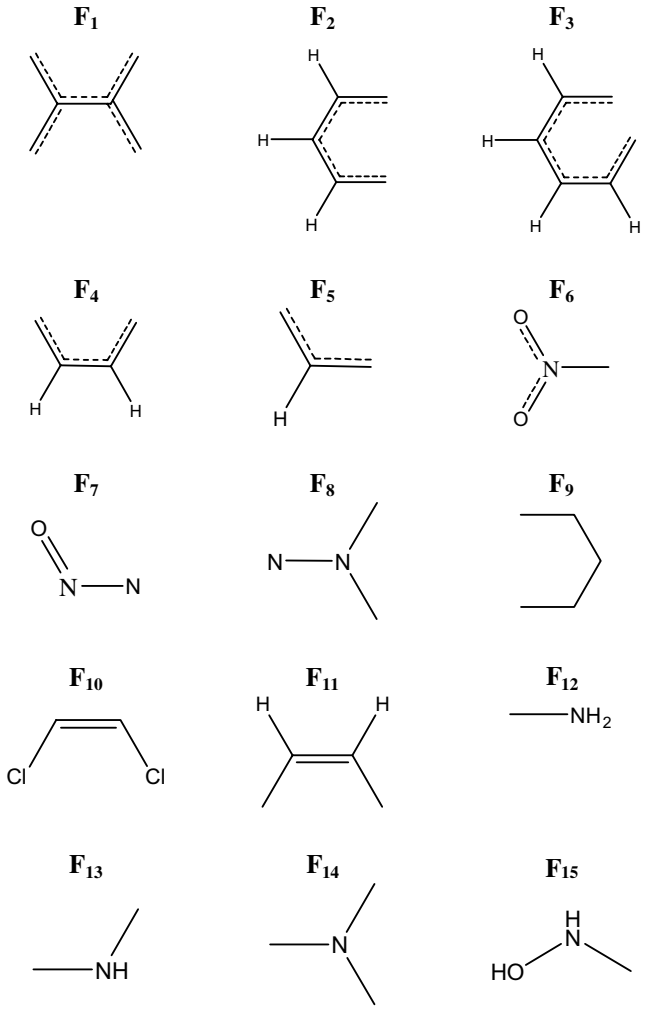

Figure 1. Structural of selected fragment for which their contribution to the carcinogenic was calculated.

formation of the benzo(α)pyrene-7,8-dihydrodiol-9,10-oxide, the metabolite with the carcinogenic activity.

As can be seen in the nitrobenzene (see Fig. 2), the fragments F_6 and F_2 appears in this compound. Both substructures are predicted to have positive contribution to carcinogenic activity, but the fragment F_6 has a higher contribution than the fragment F_2. This fact agrees with the main metabolic pathway proposal for this compound, the nitro group reduction. The nitro groups are unable to react directly with proteins or nucleoside base of DNA, but are metabolically activated to reactive species. The nitro group reduction is a well-known mechanism of biological activation for nitro aromatic compounds and it is related with the mutagenic and carcinogenic activity. Nitro groups of aryl compounds can be reduced to *N*-hydroxylarylamines, a reaction catalyzed by both microsomal and cytosolic enzymes. Further, this group is become in nitrenium cations, which react with nucleophilic substance.[51,60]

The TOPS-MODE approach recognizes the fragment F_2 with positive carcinogenic contribution too, but this contribution is not significant from the mechanistic

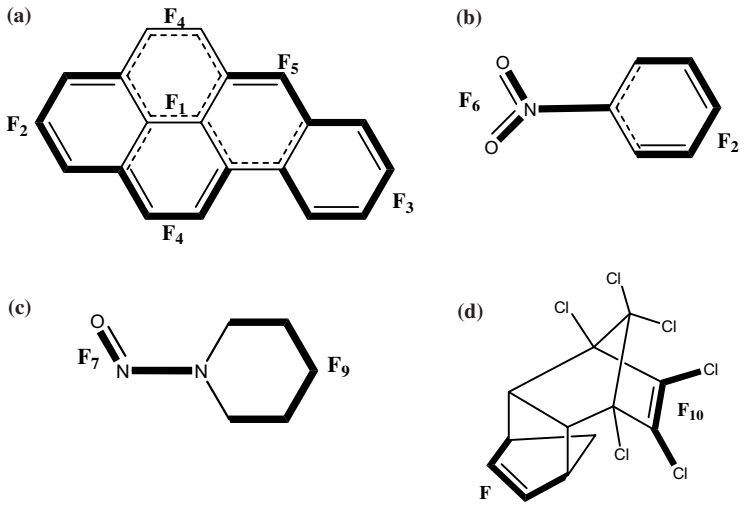

Figure 2. Bond contribution to carcinogenic activity of some carcinogenic compounds (identification of structural alerts in red color): (a) benzo(α)pyrene; (b) nitrobenzene; (c) *N*-nitrosopiperidine; (d) aldrin.

point of view. This substructure (F_2) represent a part of aromatic ring, it knows that this zone could suffer epoxidation reaction, as was detailed previously. In this case the oxidation reaction is disfavored due to electroacceptor role of the nitro group.

The nitrosamines are well-known potent carcinogens. The nitrosamines are chemically stable compound requiring metabolic activation for their carcinogenic effect.[60] The key activation pathway is through the cytochrome P-450 enzyme, catalyzing the hydroxylation of the α-carbon of the nitroso group to form α-hydroxylnitrosamines. *N*-Nitrosopiperidine is a potent esophageal carcinogen as well as induces liver tumors.[59] The metabolic activation of *N*-nitrosopiperidine leading to DNA adducts formation. The unstable α-hydroxyl *N*-nitrosopiperidine spontaneously decomposes to electrophilic intermediates as diazoalcane, diazonium salt, or carbonium ion, that react with nucleophilic sites of DNA. The fragments F_7 and F_9, with positive contribution to the carcinogenicity activity, appear in the structure of *N*-nitrosopiperidine compound (see Table 9 and Fig. 1) and both are involved in the biotransformation of this *N*-nitrosamine.[36] Also the highest positive value of the fragment F_9 could be explained by their contribution to the lipophilicity of this molecule, being a main condition for the transport and the easy interaction with enzymes.[31]

The aldrin is considered, by our model (Eq. 5), to be a carcinogenic compound with more than 95% classification probability. This compound has an isolated double bond in a cyclic system and its epoxidation has been reported to be the first and main step in its biotransformation.[56] This reaction is mediated by mixed-function oxidases, sometimes referred as aldrin-epoxidase, to form dieldrin from aldrin. The dieldrin can originate different metabolites in the biotransformation process. Some authors have also suggested that aldrin/dieldrin exposure induces hepatocarcinogenesis in mice through nongenotoxic mechanisms.[56]

According to biotransformation mentioned previously for the aldrin, the fragments F_{10} and F_{11} are the most related in the epoxidation reaction. Both substructures have positive contribution, being the fragment F_{11} the highest contribution. It is a logic result due to the epoxidation of olefinsis an oxidation reaction, through hydroperoxo-iron as an electrophilic oxidant,[1] which attacks the zones with high electronic density. The double bond of the nonsubstituted olefin (F_{11}) is a fragment with a high electronic density in relation to the alkene substituted by chloride atoms (F_{10}); for this reason the epoxidation reaction is in the fragment F_{11}. Of this way the theoretical result support the experimental facts.

The amines are a family of mutagenic and carcinogenic compounds. The most studied are the aromatic amines, due to the high level of use and industrial importance. These compounds are metabolized to reactive electrophiles in order to exert their carcinogenic potential. The aromatic amines are typically converted to *N*-hydroxyarylamines by a initial N-oxidation.[6] The TOPS-MODE approach recognize the fragments F_{12} and F_{13} with positive carcinogenic contribution, these substructures represent primary and secondary amines. This result agrees with the main activation pathway for this family, the reaction of N-oxidation. However, the F_{14} substructure (a tertiary amine) contributes negatively to the carcinogenic property. This fact could be explained taking into consideration that for this fragment is more difficult any kind of molecular interactions with cytochrome P450, decreasing the potentialities to have a carcinogenic activity.

The potentialities of the TOPS-MODE approach to identified substructures (through the analysis of fragment contributions), in the prediction of carcinogenic activity, have been widely demonstrated in the present paper, nevertheless this tool can be useful in the evaluation of carcinogenic/mutagenic behavior of novel compounds.

5. Conclusions

The prediction of carcinogenicity has been a goal of pharmaceutical companies due to the importance of this property during the drug development process. For this reason, several 'in silico' methods have been applied in order to predict this property in the early stage of drug development and some of them have become important tools to select new drug candidates. In this study, the TOPS-MODE has been a successful approach to predict the carcinogenic activity for a noncongeneric serie. The procedure has shown that a good discriminant model can be obtained using the hydrophobicity and the bond dipole moment as a weight in the diagonal entries of the bond matrix. The linear model developed in the current work is easily calculated and suitable for the rapid prediction of carcinogenicity, and the validation and cross-validation of the final model support this claim. This suggests that the present method should be regarded as one of the choices for lead optimization programs in the drug discovery process.

On the other hand, the TOPS-MODE approach permitted the identification and quantification of fragment contributions that were responsible for the carcinogenic activity. This possibility will play an important role in the generation of new chemical entities as well as in the optimization process and in the design of 'safety' drugs.

Acknowledgements

The authors acknowledge Software MODESLAB 1.0 for their free contribution to the development of this work. Also thanks to Dr. Romualdo Benigni for his useful comments on the manuscript.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2005.01.035.

References and notes

1. Alfin, D. N.; Dermot, F.; Minor, J. Epoxidation of Olefins by Cytochrome P450: Evidence from Site-Specific Mutagenesis for Hydroperoxo-iron as an Electrophilic Oxidant. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 3555–3560.
2. Bartsch, H. DNA Adducts in Human Carcinogenesis: Etiological Relevance and Structure–Activity Relationship. *Mutat. Res.* **1996**, *340*, 67–79.
3. Benigni, R.; Passerini, L. Carcinogenicity of the Aromatic Amines: from Structure–Activity Relationships to Mechanisms of Action and Risk Assessment. *Mutat. Res.* **2002**, *511*, 191–206.
4. Benigni, R. QSARS of Mutagens and Carcinogens: Two Case Studies Illustrating Problems in the Construction of Models for Non-congeneric Chemicals. *Mutat. Res.* **1996**, *371*, 29–46.
5. Benigni, R. The First US National Toxicology Program Exercise on the Prediction of Rodent Carcinogenicity: Definitive Results. *Mutat. Res.* **1997**, *387*, 35–45.
6. Benigni, R.; Alessandro, G.; Franke, R.; Gruska, A. Quantitative Structure–Activity Relationships of Mutagenic and Carcinogenic Aromatic Amines. *Chem. Rev.* **2000**, *100*, 3697–3714.
7. Benigni, R.; Andreoli, C.; Cotta-Ramusino, M.; Giorgi, F.; Gallo, G. The Electronic Properties of Carcinogens, and Their Role in SAR Studies of Non-congeneric Chemicals. *Toxicol. Modeling* **1995**, *1*, 157–167.
8. Benigni, R.; Giuliani, A. Putting the Predictive Toxicology Challenge into Perspective: Reflections on the Results. *Bioinformatic* **2003**, *19*, 1194–1200.
9. Benigni, R.; Richard, A. M. Quantitative Structure-Based Modeling Applied to Characterization and Prediction of Chemical Toxicity. *Methods* **1998**, *14*, 264–276.
10. Benigni, R.; Zito, R. The Second National Toxicology Program Comparative Exercise on the Prediction of Rodent Carcinogenicity: Definitive Results. *Mutat. Res.* **2004**, *566*, 49–63.
11. Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68*, 441–451.
12. Bristol, D. W.; Wachsman, J. T.; Greenwell, A. The NIEHS Predictive-Toxicology Evaluation Project: Chem-carcinogenicity Bioassay. *Environ. Health Perspect.* **1996**, *104*, 1001–1010.
13. Cabrera, M. A.; González, H.; Fernández, C.; Plá-Delfina, J. M.; Bermejo, M. A Novel Approach to Determining Physicochemical and Absorption Properties of 6-Fluoroquinolone Derivatives: Experimental Assessment. *Eur. J. Pharm. Biopharm.* **2002**, *53*, 317–325.
14. Cabrera, M. A.; Ruiz, A.; Fernández, C.; González, I.; Bermejo, M. A Topological-Substructural Molecular Design (TOPS-MODE) Approach to Determining Pharmacokinetics and Pharmacological Properties of 6-Fluoroquinolone Derivatives. *Eur. J. Pharm. Biopharm.* **2003**, *56*, 197–206.
15. Dearden, J. C.; Barratt, M. D.; Benigni, R.; Bristol, D. W.; Combes, R. D.; Cronin, M. T. D.; Judson, P. N.; Payne, M. P.; Richard, A. M.; Tichy, M.; Worth, A. P.; Yourick, J. J. The Development and Validation of Expert Systems for Predicting Toxicity. *Alternat. Lab. Anim.* **1997**, *25*, 223–252.
16. Enslein, K., Ed., 1999. Still Another Rodent Carcinogenesis Model. HDI Computational Toxicology NEWS No. 22.
17. Enslein, K.; Gombar, V. K.; Blake, B. W. Use of SAR in Computer-Assisted Prediction of Carcinogenicity and Mutagenicity of Chemicals by the TOPKAT Program. *Mutat. Res.* **1994**, *305*, 47–61.
18. Enslein, K. H.; Borgstedt, H.; Tomb, M. E.; Blake, B. W.; Hart, H. B. A Structure–Activity Prediction Model of Carcinogenicity Based on NCI/NTP Assays and Food Additives. *Toxicol. Indus. Health* **1987**, *3*, 267–287.
19. Ertl, P.; Rohde, B.; Selzer, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* **2000**, *43*, 3714–3717.
20. Estrada, E. Edge Adjacency Relationships and a Novel Topological Index Related to Molecular Volume. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 31–33.
21. Estrada, E. Spectral Moments of Edge Adjacency Matrix in Molecular Graph. 1. Definition and Application to the Prediction of Physical Properties of Alkanes. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 844–849.
22. Estrada, E. Spectral Moments of Edge Adjacency Matrix in Molecular Graph. 2. Molecules Containing

Heteroatoms and QSAR Applications. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 320–328.
23. 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.
24. Estrada, E.; Molina, E.; Uriarte, E. Quantitative Structure–Toxicity Relationship Using TOPS-MODE.2. Neurotoxicity of a Non-congeneric Series of Solvents. *SAR QSAR Environ. Res.* **2001**, *12*, 445–459.
25. Estrada, E.; Peña, A. In Silico Studies for the Rational Discovery of Anticonvulsant Compounds. *Bioorg. Med. Chem.* **2000**, *8*, 2755–2770.
26. Estrada, E.; Peña, A.; García-Domenech, R. Designing Sedative/Hynotic Compounds from a Novel Substructural Graph–Theoretical Approach. *J. Comput. Aided Mol. Des.* **1998**, *12*, 583–595.
27. Estrada, E.; Uriarte, E. Quantitative Structure–Toxicity Relationship Using TOPS-MODE.1. Nitrobenzene Toxicity to Tetrahymena Pyriformis. *SAR QSAR Environ. Res.* **2001**, *12*, 309–324.
28. Estrada, E.; Uriarte, E.; Gutierrez, Y.; González, H. Quantitative Structure–Toxicity Relationships Using TOPS-MODE. 3. Structural Factors Influencing the Permeability of Commercial Solvents Through Living Human Skin. *SAR QSAR Environ. Res.* **2003**, *14*, 145–163.
29. Estrada, E.; Uriarte, E.; Montero, A.; Teijeira, M.; Santana, L.; De Clercq, E. A Novel Approach for the Virtual Screening and Rational Design of Anticancer Compounds. *J. Med. Chem.* **2000**, *43*, 1975–1985.
30. Ferguson, L. N. *The Modern Structural Theory of Organic Chemistry*; Prentice-Hall: Englewoods Cliffs, 1963.
31. Franke, R.; Gruska, A.; Giuliani, A.; Benigni, R. Prediction of Rodent Carcinogenicity of Aromatic Amines: a Quantitative Structure–Activity Relationships Model. *Carcinogenesis* **2001**, *22*, 1561–1571.
32. Gasteiger, J.; Marsilli, M. A New Model for Calculating Atomic Charge in Molecules. *Tetrahedron Lett.* **1978**, *34*, 3181–3184.
33. Ghose, A. K.; Crippen, G. M. Atomic Physicochemical Parameters for Three-Dimensional-Structure-Directed Quantitative Structure–Activity Relationships. 2. Modeling Dispersive and Hydrophobic Interactions. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 21–35.
34. Gold, L. S.; Manley, N. B.; Slone, T. H.; Rohrbach, L. Supplement to the Carcinogenic Potency Database (CPDB): Results of Animal Bioassays Published in the General Literature in 1993 to 1994 and by the National Toxicology Program in 1995 to 1996. *Environ. Health Perspect. Suppl.* **1999**; *107* (Suppl. 4), 527–602.
35. Gold, L. S.; Zeiger, E. *Handbook of Carcinogenic Potency and Genotoxicity Databases*; CRC: Boca Roca, FL, 1997.
36. González, S.; Gaspar, J.; Calle, E.; Pereira, S.; Mariano, A.; Rueff, J.; Casado, J. Stereochemical Effects in the Metabolic Activation of Nitrosopiperidines: Correlations with Genotoxicity. *Mutat. Res.* **2004**, *558*, 45–51.
37. Gottmann, E.; Kramer, S.; Pfahringer, B.; Helma, Ch. Data Quality in Predictive Toxicology: Reproducibility of Rodent Carcinogenicity Experiments. *Environ. Health Perspect.* **2001**, *109*, 509–514.
38. Gutiérrez, Y.; Estrada, E. MODESLAB 1.0 (Molecular DEScriptors LABoratory) for Windows. Universidad de Santiago de Compostela, España; 2002.
39. Klopman, G. MULTI-CASE 1. A Hierarchical Computer Automated Structure Evaluation Program. *Quant. Struct.-Act. Relat.* **1992**, *11*, 176–184.
40. Klopman, G.; Rosenkranz, H. S. Approaches to SAR in Carcinogenesis and Mutagenesis Prediction of Carcinogenicity/Mutagenicity Using MULTI-CASE. *Mutat. Res.* **1994**, *305*, 33–46.
41. Miller, K. J. Additivity Methods in Molecular Polarizability. *J. Am. Chem. Soc.* **1990**, *112*, 8533–8542.
42. Morales, A.; Pérez, M.; Rieumont, J. TOPS-MODE Approach to Predict Mutagenicity in Dental Monomers. *Polymer* **2004**, *45*, 2045–2050.
43. Pérez, M.; Morales, A.; González, H. A TOPS-MODE Approach to Predict Permeability Coefficients. *Polymer* **2004**, *45*, 2073–2079.
44. Pérez, M.; Morales, A.; Molina, R.; García, J. A Topological Sub-structural Approach of the Mutagenic Activity in Dental Monomers. 1. Aromatic Epoxides. *Polymer* **2004**, *45*, 2773–2779.
45. Potapov, V. M. *Stereochemistry*; MIR: Moscow, 1978.
46. Richard, A. M. Application of SAR Methods to Non-congeneric Data Bases Associated with Carcinogenicity and Mutagenicity: Issues and Approaches. *Mutat. Res.* **1994**, *305*, 73–97.
47. Richard, A. M. Application of Artificial Intelligence and Computer-Based Methods to Predicting Chemical Toxicity. *Knowl. Eng. Rev.* **1999**, *14*, 1–12.
48. Richard, A. M.; Benigni, R. AI and SAR Approaches for Predicting Chemical Carcinogenicity: Survey and Status Report. *SAR QSAR Environ. Res.* **2002**, *13*, 1–19.
49. Rosekranz, H. S.; Klopman, G. Structural Basis of Carcinogenicity in Rodents of Genotoxicants and Non-genotoxicants. *Mutat. Res.* **1990**, *228*, 105–124.
50. Sanderson, D. M.; Earnshaw, C. G. Computer Prediction of Possible Toxic Action from Chemical Structure; the DEREK System. *Human Exp. Toxicol.* **1991**, *10*, 261–273.
51. Sawatari, K.; Nakanishi, Y.; Matsushima, T. Relationships between Chemical Structures and Mutagenicity: A Preliminary Survey for a Database Mutagenicity Test Results of New Work Place Chemicals. *Ind. Health* **2001**, *39*, 341–345.
52. Schultz, T. W.; Cronin, M. T. D.; Walker, J. D.; Aptula, A. O. Quantitative Structure–Activity Relationships (QSARs) in Toxicology: a Historical Perspective. *Theochem* **2003**, *622*, 1–22.
53. Smithing, M. P.; Darvas, F. HazardExpert: An Expert System for Predicting Chemical Toxicity. In *Food Safety Assessment*; Finlay, S. F., Robinson, S. F., Armstrong, D. J., Eds.; American Chemical Society: Washington, 1992; pp 191–200.
54. StatSoft, Inc., 2002. STATISTICA 6.0, version 6.0.
55. Toivonen, H.; Srinivasan, A.; King, R. D.; Kramer, S.; Helma, Ch. Statistical Evaluation of the Predictive Toxicology Challenge 2000–2001. *Bioinformatics* **2003**, *19*, 1183–1193.
56. United States Environmental Protection Agency, 2003. Health Effects Support Document for Aldrin/Dieldrin.
57. Wang, R.; Gao, Y.; Lai, L. Calculating Partition Coefficient by Atom-Additive Method. *Perspect. Drug Discov. Des.* **2000**, *19*, 47–66.
58. Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36.
59. Wong, H.; Murphy, Sh.; Wang, M.; Hecht, S. Comparative Metabolism of *N*-Nitrosopiperidine and *N*-Nitrosopyrrolidine by Rat Liver and Esophageal Microsomes and Cytochrome P450 2A3. *Carcinogenesis* **2003**, *24*, 291–300.
60. Woo, Y. T.; Lai, D.; Argus, M.; Arcus, J. Development of Structure Activity Relationship Rules for Predicting Carcinogenic Potential of Chemicals. *Toxicol. Lett.* **1995**, *79*, 219–228.