TOPS-MODE Based QSARs Derived from Heterogeneous Series of Compounds. Applications to the Design of New Herbicides

Maykel Pérez González,*[,†,‡] Humberto Gonzalez Díaz,[†,§] Reinaldo Molina Ruiz,[†] Miguel A. Cabrera,[†] and Ronal Ramos de Armas[†,||]

Chemical Bioactives Center, Central University of Las Villas, Santa Clara, C. P. 54830, Villa Clara, Cuba, Experimental Sugar Cane Station "Villa Clara-Cienfuegos", C. P. 53100, Service Unit, Cuba, Faculty of Pharmacy, University of Santiago de Compostela, 15706, Spain, and Department of Chemistry, Faculty of Chemistry and Pharmacy, Central University of Las Villas, Santa Clara, 54830 Villa Clara, Cuba

Received March 2, 2003

A new application of *TOP*ological *S*ub-structural *MO*lecular *DE*sign (TOPS-MODE) was carried out in herbicides using computer-aided molecular design. Two series of compounds, one containing herbicide and the other containing nonherbicide compounds, were processed by a k-Means Cluster Analysis in order to design the training and prediction sets. A linear classification function to discriminate the herbicides from the nonherbicide compounds was developed. The model correctly and clearly classified 88% of active and 94% of inactive compounds in the training set. More specifically, the model showed a good global classification of 91%, i.e., (168 cases out of 185). While in the prediction set, they showed an overall predictability of 91% and 92% for active and inactive compounds, being the global percentage of good classification of 92%. To assess the range of model applicability, a virtual screening of structurally heterogeneous series of herbicidal compounds was carried out. Two hundred eighty-four out of 332 were correctly classified (86%). Furthermore this paper describes a fragment analysis in order to determine the contribution of several fragments toward herbicidal property; also the present of halogens in the selected fragments were analyzed. It seems that the present TOPS-MODE based QSAR is the first alternate general "in silico" technique to experimentation in herbicides discovery.

INTRODUCTION

Not only is the pharmaceutical industry under increasing pressure to develop more effective and safe chemicals[1] but also there are many other fields of applied sciences in which the computational design of novel compounds becomes a forefront problem. In this sense, the design of herbicides has been one of the major problems in the environmental incidence. Thus, not only herbicidal action is desired to design compounds but also specific soil sorption properties may be studied by QSAR tools.[2] The problem becomes more important considering the possibility of existence of herbicide-resistant mutants.[3] However, almost all Quantitative Structure Activity Relationships (QSARs) deals with a homologous series of organic compounds.

For instance, Murai et al. studied SL-950 (Nicosulfuron) derivatives as post emergence application herbicides for corn which has a novel type of pyridylsulfonylurea structure. The analogues of SL-950 were synthesized, and their QSAR analysis was carried out to understand the drug-receptor interaction.[4] Next Clark et al. investigated pyrazole nitrophenyl ethers by comparative molecular field analysis.[5] Highlights also, the QSAR developed between the physicochemical parameters of the 5-substituent of imidazolinone herbicide analogues, imazapyr and root absorption, translocation, inhibition of acetohydroxyacid synthase (AHAS), and herbicidal activity of the analogues.[6] In general, QSAR on herbicides action topics lack the range of applicability mainly by the above-mentioned use of very reduced and homologous series of chemicals.[7]

On the other hand, Graph-Theoretical methods have shown to be very useful in QSAR problems in order to perform a rational analysis of different pharmacological, toxicological, and other activities.[9,10] In the context of the Graph-Theoretical and Topological methods for modeling physicochemical and biological properties of chemical there has been introduced the *TOP*ological *S*ub-structural *MO*lecular *DE*sign (TOPS-MODE) approach. The TOPS-MODE has been applied to the description of physicochemical properties of organic compounds. Several applications for the design of biologically active compounds have been described.[11−25] Thereby, the aim of this work is to find rationality in the search of novel herbicides using TOPS-MODE approach. Second, to continue the validation of the methods for describing biological activity of heterogeneous series of compounds.

LINEAR DISCRIMINANT ANALYSIS AND TOPS-MODE APPROACH

Here, we use the TOPS-MODE approach to obtain molecular descriptors through which we developed the QSAR function. The mathematical details of the method have been largely reported,[11−23] thus we will outline only the fundamental remarks.

* Corresponding author e-mail: maykel@cbq.uclv.edu.cu (M.P.G.).
† Chemical Bioactives Center, Central University of Las Villas.
‡ Experimental Sugar Cane Station "Villa Clara-Cienfuegos".
§ University of Santiago de Compostela.
|| Faculty of Chemistry and Pharmacy, Central University of Las Villas.

10.1021/ci034039+ CCC: $25.00
Published on Web 06/12/2003

Briefly, this method codifies the molecular structure by means of the edge adjacency matrix **E** (likewise called bond adjacency matrix **B**). The **E** or **B** matrix is a square table of order *m* (the number of chemical bonds in the molecule).[24] The elements of such a matrix (e_{ij}) are equal to 1 if the bonds i and j are adjacent (it means that an atom exists, which participates either in the bond i or in the bond j) or 0 otherwise. To codify information related to heteroatoms, the TOPS-MODE approach use $\mathbf{B(w_{ij})}$ weighted matrices instead of **B**. The weights ($\mathbf{w_{ij}}$) are chemically meaningful numbers such as bond distances, bond dipole, bond polarizabilities, or even mathematical expressions involving atomic weights such as hydrophobicity.[11−25] These weights are introduced in the main-diagonal of the matrix $\mathbf{B(w_{ij})}$. Afterward, the spectral moments of this matrix may be used as molecular fingerprints in QSAR studies in order to codify molecular structure. By definition, the expression "spectral moments" must be understood as the sum of the elements in the natural powers of $\mathbf{B(w_{ij})}$. It means that the spectral moment of order k (μ_k) is the sum of the main diagonal elements (e_{ii}) of the matrix $\mathbf{B(w_{ij})^k}$. In the present work the $\mathbf{B(w_{ij})}$ matrix was weighted in the main diagonal with the parameter $w_{ij} = (h_i/\delta_i) + (h_j/\delta_j)$, which characterizes the bond hydrophobicity. In this expression h_i is the standard hydrophobicity of the atom I bonded with j and δ_i is the vertex or atom degree.[21,22,24] Such a parameter μ_1 is equal to the sum of atom hydrophobicities in the molecule. The calculation of the μ_k was carried out by means of the software package Modes Lab 1.0 b.[25]

Linear discriminant analysis (LDA) has been the election statistical technique in most of the QSAR studies carried out using TOPS-MODE.[18,19,23,24] In the present work, a similar expression for the QSAR is derived

$$A. = b + b_0\mu_0 + b_1\mu_1 + b_2\mu_2 + \ldots\ldots + b_k\mu_k \quad (1)$$

where A. (Acronym of Herbicide Activity) is an indicator variable. This variable reaches the values A. = 1 for herbicides compounds or A. = −1 for the nonactive ones. Deciding whether a compound may be classified as an herbicide is based on the information extracted from the literature.[27−31]

In eq 1 the b_k are the coefficients of the classification function determined by least-squares as implemented on the LDA modulus of STATISTICA 6.0.[32] Forward stepwise was fixed as the strategy for variable selection.[33] To develop the QSAR for herbicides/nonherbicides compound discrimination, we use the first 15 μ_k as molecular descriptors. Examining Wilk's U-statistic, Mahalanobis distance, the percentage of good classification and the proportion between the cases and variables in the equation determined the quality of the model. Additionally, calculating the percentages of good classification in the external prediction series carried out the validation of the model. Compounds in the external prediction series were never used to develop the classification function.

One of the most important step in computer-aided search of novel herbicides is to design a representative, randomized training and predicting series. With this aim we select a large data set of 246 compounds having great structural variability: 135 of them are active (herbicides) and the others are inactive.[26−31] Later, two k-Means Cluster Analysis (k-MCA) were performed for active and inactive series of compounds.[34]

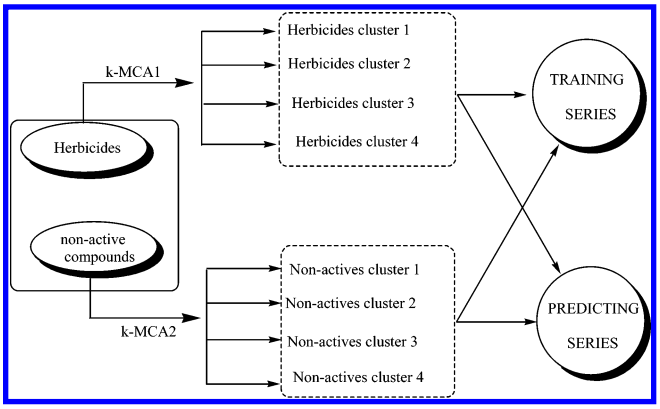

Figure 1. Training and predicting series design throughout k-MCA.

COMPUTATION OF FRAGMENT CONTRIBUTIONS

Each of the μ_k spectral moments given in eq 1 contains structural information on the molecules that can be directly obtained by the following computational approach.[16] In this approach Estrada et al. calculated the spectral moment for all the fragments contained in a given substructure and by difference of these moments obtained the contribution of the substructure.

The general algorithm followed in this computational approach is as follows. First, we select the substructures whose contribution to the moments we would like to determine. Then we generate all the fragments (subgraphs) which were contained in the corresponding substructure and calculate the spectral moments for both the substructure and all their fragments. The contribution of the substructure of the spectral moments is finally obtained as the difference between the spectral moments of the substructure and all their fragments.

Having the contributions of the different structural fragments in which we are interested, we only need to substitute these contributions into the quantitative model developed to describe the property studied, e.g. model (1) in which we obtain the quantitative contribution of the different fragments to *P*.

K-MEANS CLUSTER ANALYSIS

The k-MCA may be used in training and predicting series design.[34,35] The idea consists of carrying out a partition of either active or nonactive series of compound in several statistically representative classes of chemicals. Thence, one may select from the member of all these classes of training and predicting series. This procedure ensures that any chemical classes (as determined by the clusters derived from k-MCA) will be represented in both compounds series (training and predicting). It permits the designing of both training and predicting series, which are representative of the entire "experimental universe". Figure 1 graphically illustrates the above-described procedure where two independent cluster analyses (one for the active compounds and other for the inactive compounds) were carried out to select a representative sample for the prediction and training sets.

A first k-MCA (k-MCA1) splits herbicides in four clusters with 25, 66, 32, 12 members and Standard Deviations of 0.22, 0.33, 0.13, and 0.11, respectively. On the other hand, the series of nonactive compounds was partitioned into four clusters (k-MCA2) with 12, 39, 8, 52 members and Standard Deviations of 0.12, 0.25, 0.12, and 0.27, respectively.

Table 1. Main Results of the k-Means Cluster Analysis for Active and Inactive Compounds

Variance Analysis				
spectral moments	between SS[a]	within SS[b]	Fisher ratio (F)	p-level[c]
Statistics for Active Compound Clusters (k-MCA 1)				
μ_2	40.71	26.96	65.93	**0.00**
μ_5	52.61	31.26	73.48	**0.00**
μ_{11}	13.85	10.55	57.29	**0.00**
μ_{12}	22.34	28.42	34.31	**0.00**
μ_{15}	89.78	39.68	98.80	**0.00**
Statistics for Inactive Compound Clusters (k-MCA 2)				
μ_2	212.37	75.69	100.07	**0.00**
μ_5	71.56	51.37	41.59	**0.00**
μ_{11}	87.58	20.72	83.05	**0.00**
μ_{12}	86.43	63.82	48.30	**0.00**
μ_{15}	30.09	35.41	30.31	**0.00**

[a] Variability between groups. [b] Variability within groups. [c] Level of significance.

Selection 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 of the standard deviation between and within clusters, the respective Fisher ratio and their p-level of significance considered to be lower than 0.05.[35,36] All spectral moments (from μ_0 to μ_{15}) were used in both analysis; all variables show p-levels < 0.05 for the Fisher test, and 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 herbicides (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. Mainly, if it is considered that k-MCAs based partitions of data which consider not only 4 but also 5 or 6 clusters are statistically significant too (results not reported). However, the use of the k-MCA analysis here points to a structurally representative distribution of chemicals into training and predicting series.

DEVELOPMENT OF THE DISCRIMINANT FUNCTION

Once we perform a random and representative selection of training series it could be used to fit the discriminant function. The model selection was subjected to the principle of parsimony. Then we chose a function with high statistical significance but having few parameters (b_k) as possible.

To derive a discriminant function that permits the classification of chemicals as active (herbicides) or inactive (nonherbicides) we use the linear discriminant analysis in which spectral moments are used as independent variables. The classification model obtained is given below together with the statistical parameters of the LDA:

$$A. = -0.395 \cdot \mu_2 + 0.076 \cdot \mu_5 - 7.20 \cdot 10^{-4} \cdot \mu_{11} + 3.20 \cdot 10^{-4} \cdot \mu_{12} - 2.01 \cdot 10^{-5} \cdot \mu_{14} + 1.34 \cdot 10^{-6} \mu_{15} + 1.86 \quad (2)$$

$$N = 185,\ U = 0.46,\ F = 33.81,\ D^2 = 4.67$$

Table 2. Classification and Name of the Active Compounds in the Training Series

name	P(H)[b]	name	P(H)[b]
glufosinate-ammonium	0.839	flufenacet	0.920
2,4,5-T	0.731	flufenpyr	0.991
3,4-DA	0.973	flumetsulam	0.958
3,4-DP	0.919	flumiclorac	0.963
4-CPB	0.976	flumipropyn[a]	0.196
aclonifen	0.778	fosamine	0.734
alachlor	0.803	glufosinate	0.528
alloxydim[a]	0.446	halosafen	0.752
alorac[a]	0.309	haloxydine	0.991
amiprofos-methyl	0.975	imazamox	0.981
anilofos	0.991	imazapyr[a]	0.353
beflubutamid	0.985	lactofen ethyl[a]	0.374
benoxacor	0.892	linuron	0.979
bensulide	0.832	MCPA-thioethyl	0.908
benzadox	0.957	mecoprop	0.899
benzfendizone[a]	0.179	medinoterb	0.972
benzobicyclon	0.882	mefenpyr	0.984
benzofluor	0.800	mephenate	0.953
bethoxazin	0.989	mesosulfuron	0.874
bispyribac	0.755	metolachlor	0.864
buthiuron	0.933	metoxuron	0.596
butroxydim	0.759	metsulfuron	0.814
CDEA	0.443	monalide	0.717
chlomethoxyfen	0.609	naphthalic anhydride	0.950
chloranocryl	0.985	napropamide	0.604
chlorazine	0.977	neburon	0.641
chlorfenac	0.998	nipyraclofen	0.923
chlorflurazole	0.974	phenmedipham	0.993
chloridazon	0.991	phenobenzuron[a]	0.494
chlorsulfuron	0.626	picolinafen	0.926
chlorthiamid	0.924	pretilachlor	0.958
cliodinate	0.939	procyazine	0.251
cresol	0.988	profluazol	0.999
cyanazine	0.862	prometon	0.862
cyclosulfamuron	0.995	propachlor	0.996
cyhalofop	0.899	quinclorac	0.765
cyperquat	0.879	quinoclamine	0.973
dichlormid	0.915	quizalofop	0.645
dichlorprop-P	0.539	rimsulfuron	0.970
diclosulam	0.981	secbumeton[a]	0.345
dietholate	0.979	siduron	0.969
dinitramine	0.757	tebuthiuron[a]	0.401
dinoprop	0.981	terbucarb	0.980
diquat	0.974	terbumeton	0.990
dithiopyr	0.692	thiazopyr	0.997
DSMA	0.993	thidiazuron	0.967
ethoxysulfuron	0.981	thiobencarb[a]	0.181
etnipromid	0.716	trifopsime	0.911
EXD	0.958	tritosulfuron	0.967
fenchlorazole	0.669	xylachlor	0.994
fenoxaprop-P	0.998	xeletol	0.936

[a] Misclassified compounds. [b] P(H): The posterior probability that a case belongs to a herbicidal group. It is basically proportional to the Mahalanobis distance from that group centroid.

In this model the coefficient U is the Wilk's statistics, D^2 is the squared Mahalanobis distance, and F is the Fisher ratio. The Wilk's U-statistics for the overall discrimination can take values on the range from 0 (perfect discrimination) to 1 (no discrimination). For the discrimination of active/inactive compounds studied here, the model classified correctly 88.11% of active and 94.04% of inactive compounds in the training series, for a global good classification of 90.81%. The percentages of false actives and false inactive compounds in the training series were 5.94% and 11.89%, respectively, and statistical outliers were not detected. The previous statement was based on two facts; all misclassified

Table 3. Classification and Names of the Inactive Compounds in the Training Series

name	P(H)[b]	name	P(H)[b]
acebutolol[a]	0.936	cetamolol	0.005
acecarbromal	0.181	cilazaprilo[a]	0.509
acediasulfone	0.294	cinarizina	0.057
aceglutamide	0.092	cinoxacin	0.188
acenocoumarol	0.078	clindamycin[a]	0.498
acetohexamide[a]	0.569	corticosterone	0.365
acetyldigitoxin	0.321	cycloserine	0.008
acexamic acid	0.002	dilazep	0.186
aciclovir	0.163	dinitrato de etilenglicol	0.062
aclarubicin	0.092	diphenhydramine	0.056
actarit	0.101	efedrina	0.324
adiphenine	0.399	espiraprilo[a]	0.518
adrenalone	0.283	ethiofos	0.138
alacepril	0.466	etilefrina	0.475
alfentanil	0.180	etodolac	0.361
alibendol	0.054	fenilpropanolamina	0.214
allantoin	0.236	fenitoina	0.447
alprostadilo	0.063	fenobarbital	0.028
ambazone	0.013	fisosotigmina	0.021
amfenac	0.263	glipizide	0.202
amineptine	0.422	guanetidina	0.111
antipyrin	0.088	heparin	0.316
bametan	0.359	hexanitrato de manitol[a]	0.956
benidipina	0.302	iloprost	0.000
betanidina[a]	0.526	indenolol	0.031
bevantolol	0.186	indoramina	0.376
bis(chloroethyl)-nitrosourea	0.308	levarterenol	0.108
		lotamoxef	0.470
bleomycin A2	0.191	metilergonovina	0.041
capreomycin	1.38E−06	metoprolol	0.142
carazolol	1.15E−05	metoxamina	0.409
cefaclor	0.205	mezlocillin[a]	0.483
cefamandole	0.082	naftopidil	0.284
cefpodoxime proxetil	0.041	netilmicin	0.185
cefuroxime	0.033	sulfimpirazone	0.076
isosorbide nitrate	0.018	tenoposide	0.022
oxifedrina	0.067	tetranitrato de penta-eritritol	0.105
penicillin	0.263		
propatil nitrato	0.262	trimetazidina	0.003
proziquantel	0.024	vencuronium	0.239
ribavirin	0.228	vinburnina	0.023
rifampicina	0.147	zatebradine	0.156
sirivudine	0.062	zidovudine	0.438

[a] Misclassified compounds. [b] P(H): The posterior probability that a case belongs to a herbicidal group. It is basically proportional to the Mahalanobis distance from that group centroid.

chemicals (accordingly to posterior probabilities and Mahalanobis's distance) do not rise to model improvement after leaving-out from it. Additionally, k-MCA demonstrates that any group of chemicals did not exist (possible outliers) that differentiate appreciably from the remnant ones. False actives are those inactive compounds that model classifies as actives, and the false inactive are those actives classified as inactive by the model. In Tables 2 and 3, the compounds classification using the above-given model is depicted.

One of the most important criteria for the acceptance or not for a discriminant model, such as model (2), is based on the statistics for the external prediction series.[37−43] Model (2) classified correctly 91.17% and 92.01% of active and inactive compounds in the prediction series, respectively, which represents an overall predictability of 91.52%. In Table 4 we give the classification of compounds in the prediction series together with their Difference between the Posteriori Probability Percentage of Classification in Active or Inactive Group.

Table 4. Classification and Names of the Compounds in External Prediction Series

name	P(H)[b]	name	P(H)[b]
Active Compounds			
2,4-D	0.9342	fenoprop	0.9928
2,4-DEB	0.9203	fenthiaprop	0.9325
acetochlor	0.5502	furylosyfen	0.9638
ametryn	0.9811	hexachloroacetone	0.9945
basagran	0.7431	hexazinone	0.7911
benfluralin	0.9866	mama	0.9811
bifenox	0.9806	metam	0.8452
butenachlor[a]	0.2661	monolinuron	0.8126
chlorbufam	0.9143	morfamquat	0.7918
cinmethylin	0.7546	pethoxamid[a]	0.4073
cisanilide[a]	0.4468	profoxydim	0.6360
clofop	0.9421	propaquizafop	0.9981
dichlobenil	0.9651	propham	0.8707
diethamquat	0.7878	propoxycarbazone	0.9987
dinoseb	0.9605	terbutryn	0.9974
DMPA	0.9979	thenylchlor	0.6233
eglinazine	0.8877	tripropindan	0.8992
Inactive Compounds			
6-aminohexanoic acid	0.2636	dimetrofina	0.3557
acarbose	0.0007	ergonovina	0.2340
acetophenazine	0.3428	esmolol	0.4303
acetylsalicylic acid[a]	0.5735	etosuximide	0.4049
acyclovir	0.0925	glicerol 1,2-dinitrato	0.0537
alprenolol	0.3636	hexobendina	0.1070
aminocaproic acid	0.2636	mepindolol	0.4420
barnidipina[a]	0.5655	naftidrofuril	0.2695
buflomedilo	0.2782	oxprenolol	0.2306
cardilate	0.0026	pancuronium	0.0250
cefatrizine	0.0274	propafenone	0.0391
cefoperazone	0.0462	sufentanil	0.0373
cilnidipina	0.1994	vincamina	0.0397
digitoxin	0.0015		

[a] Misclassified compounds. [b] P(H): The posterior probability that a case belongs to a herbicidal group. It is basically proportional to the Mahalanobis distance from that group centroid.

THE RANGE OF APPLICABILITY OF THE MODEL

Applicability of the models is the other major bottleneck in QSAR. As outlined in Introduction almost all herbicides QSARs have been derived from a homologous series of organic compounds. It means that you cannot predict the potentialities as herbicide of any organic chemical despite its molecular structure. Considering that the present model was derived from a heterogeneous series of organic compounds it could be used as a general purpose model for herbicides computational discovery. To demonstrate the power of the present model to recognize as herbicides a broad range of molecular structures we carry out a virtual screening experiment. A total of 284 herbicides out of 332 were correctly classified (85.54%). These chemicals were never used neither in training nor in predicting series. The present result strongly demonstrates that around 86 chemicals out of 100 submitted to biological assay will be in fact herbicides.

STRUCTURAL INTERPRETATION

As we previously explain, the TOPS-MODE approach is able to compute the contribution of any structural fragment (real or hypothetical) to the biological property or activity studied.[10,11,16] In the present case, we can find the positive and negative contributions of such fragments to the development of the herbicides activity. These fragments will be named here as active and inactive, respectively. The presence

Table 5. Names of Herbicidal Compounds Selected from the Literature Used in the Virtual Screening and the Range of Probabilities of Classifications According the Discriminant Model Obtained

		$1 > P > 0.99$		
diphenamid	profluralin	chlornitrofen	ioxynil	ethalfluralin
hexaflurate	fluazolate	OCH	fluoronitrofen	ethoxyfen
tridiphane	ipazine	prometryn	CPMF	cyanatryn
triaziflam	triflusulfuron	isomethiozin	fluothiuron	nitrofluorfen
cyprazine	pentachlorophenol	2,4-DEP	fomesafen	bromoxynil
oxadiazon	amibuzin	erbon	sulfentrazone	flupropacil
terbuthylazine	metosulam	dipropetryn	bromofenoxim	nitrofen
dimethametryn	butralin			
		$0.99 > P > 0.98$		
atrazine	dinoterb	acifluorfen	oxadiargyl	MAA
methalpropalin	chloroxynil	metribuzin	bifenox methyl	Msma
tri-allate	fluoromidine	haloxyfop	2,3,6-TBA	metflurazon
prosulfuron	cyprazole	pyriclor	dichlorprop	atraton
disul	pyributicarb	trifluralin	cacodylic acid	ethiozin
cypromid	desmetryn	fluchloralin	CMA	fluorodifen
		$0.98 > P > 0.97$		
chlorazifop	benzipram	aziprotryne	isopropalin	sulfallate
butylate	fenclorim	iodosulfuron	oxyfluorfen	isoxaflutole
diclofop	paraquat	pyridate	trimeturon	triclopyr
butamifos	tricamba	ametridione	dinofenate	quizalofop-P
pyrazolynate	proglinazine	chlorthal	fluridone	o-dichlorobenzene
fenteracol	thiazafluron	lactofen	mecoprop-P	
		$0.97 > P > 0.95$		
pyridafol	bromobutide	furyloxyfen	2,4,5-TB	prosulfalin
proxan	tebutam	flumetsulam	trifloxysulfuron	dalapon
iodobonil	iprymidam	carfentrazone	chlorfenprop	pyrithiobac
isoxachlortole	di-allate	tritac	clopyralid	dimefuron
oxaziclomefone	pendimethalin	chlorpropham	chloramben	TCA
DNOC	mesoprazine	picloram	methiuron	fluazifop
clodinafop	prodiamine	benzofenap	pyraflufen	fluazifop-P
quinmerac	dinosam	anisaron	diuron	
		$0.95 > P > 0.90$		
4-CPP	diflufenican	flucarbazone	BCPC	isoxapyrifop
metamifop	swep	clomeprop	delachlor	sebuthylazine
simazine	dicamba	methyl isothiocyanate	CPPC	dichlormate
penoxsulam	karbutilate	difenopenten	flupyrsulfuron	fluoroglycofen
trifop	norflurazon	butafenacil	propyzamide	flurochloridone
isopolinate	nitralin	bromobonil	benzoylprop	chlorotoluron
florasulam	cloprop	pentanochlor	methoprotryne	chlorbromuron
propanil	buthidazole			
		$0.90 > P > 0.86$		
cloransulam	tribenuron	cafenstrole	monuron	chloroxuron
flupoxam	flamprop	parafluron	chloropon	tetrafluron
isouron	flamprop-M	benzfendizone methyl	metamitron	primisulfuron
2,4-DB	MCPA	flurazole	methazole	furilazole
simetryn	bromacil	carboxazole	3,4-DB	amidosulfuron
buturon	fluxofenim	fenoxaprop	clomazone	fluroxypyr
simeton	dazomet	fluometuron	pentoxazone	oryzalin
metobenzuron	dicyclonon	MCPB		
		$0.86 > P > 0.73$		
benfuresate	EBEP	daimuron	mefluidide	imazamethabenz
pyrazoxyfen	esprocarb	4-CPA	methyl bromide	quinonamid
pyriftalid	flazasulfuron	thidiazimin	pyriminobac	chloreturon
isoproturon	metobromuron	etinofen	CEPC	flumezin
flupropanate	credazine	terbuchlor	monisouron	diflufenzopyr
tioclorim	tralkoxydim	glyphosate	methyldymron	propazine
orbencarb	ethofumesate	etobenzanid	imazapic	isocarbamid
dimexano	flurtamone	bensulfuron	bentazone	
		$0.73 > P > 0.52$		
phenisopham	imazaquin	dimethachlor	fenasulam	halosulfuron
tiocarbazil	benazolin	imazethapyr	bilanafos	dimethenamid
imazosulfuron	chloroacetic acid	acrolein	prynachlor	dimethenamid-P
azafenidin	SMA	chlorflurenol	difenoxuron	allyl alcohol
thifensulfuron	brompyrazon	ethiolate	sulcotrione	sulfometuron
chlorimuron	flufenican	ethametsulfuron	fentrazamide	barban
triasulfuron	methiobencarb	naproanilide	cloquintocet	dimepiperate
		$0.52 > P > 0.23$		
azimsulfuron	noruron	sulfosulfuron	isoxaben	asulam
pebulate	EPTC	dimidazon	mefenacet	indanofan
sethoxydim	clethodim	cinidon-ethyl	prosulfocarb	trietazine
fenuron	pyraclonil	sulglycapin	pyrazosulfuron	butachlor
oxasulfuron	phenmedipham-ethyl	cinosulfuron	mesotrione	chlorprocarb
fluthiacet	metazachlor	perfluidone	cloproxydim	asulam ethyl
epronaz	difenzoquat	methometon	molinate	endothal
piperophos	vernolate	nicosulfuron	cycloate	amicarbazone
naptalam	amitrole	carbasulam	propisochlor	pydanon
diethatyl	foramsulfuron	oxapyrazon		

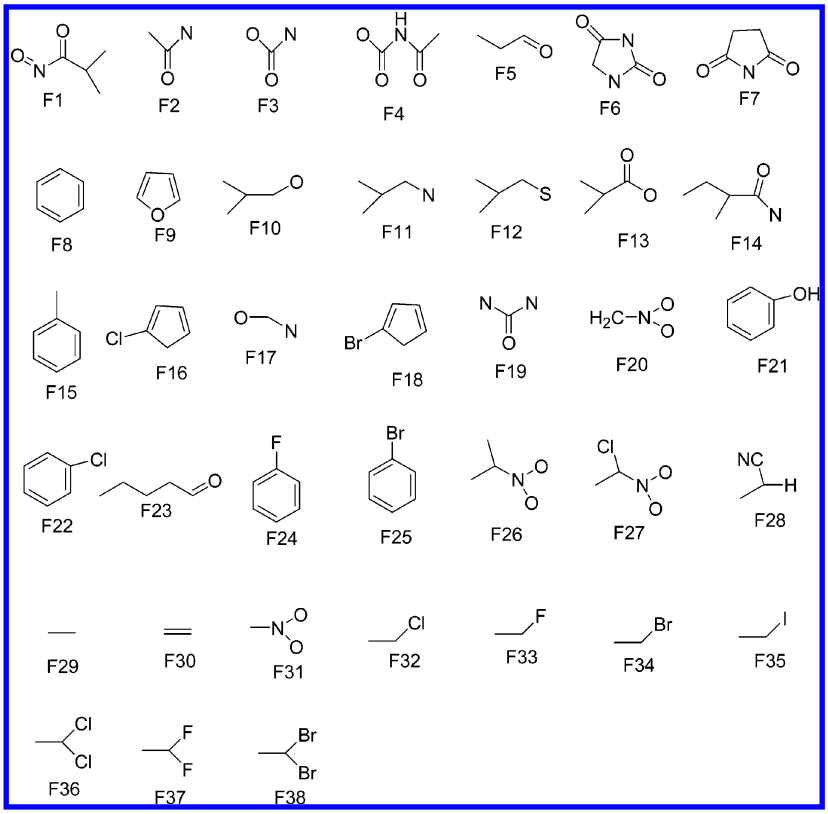

Figure 2. Structures of selected fragments for which their contributions to the herbicidal activity are evaluated.

Table 6. Contribution of Some Selected Fragments to the Herbicidal Activity

fragment	contribution	fragment	contribution	fragment	contribution
F1	−2.31	F14	−1.85	F27	−2.68
F2	1.02	F15	−1.73	F28	0.49
F3	0.66	F16	−1.33	F29	1.52
F4	−0.64	F17	1.12	F30	1.62
F5	0.81	F18	−1.77	F31	1.87
F6	−2.48	F19	0.66	F32	1.68
F7	−2.64	F20	1.87	F33	1.58
F8	0.17	F21	−0.89	F34	2.01
F9	0.36	F22	−3.28	F35	2.29
F10	1.75	F23	0.27	F36	2.45
F11	1.75	F24	−1.59	F37	2.81
F12	1.68	F25	−4.56	F38	2.42
F13	−0.66	F26	−2.76		

of active fragments does not presuppose the development of the herbicide activity per se, because it is well known that the activity is the consequence of the sum of contributions of all fragments in the molecule.

In Figure 2, we show the structure of a series of fragments selected from our database. The contributions of the herbicide activity of these fragments were computed by using the model (2). These quantitative contributions are given in Table 6.

As can be seen in Figure 2 and Table 6 there are a few active fragments that do not commonly appear in herbicides compounds, such as the F6 and F7; these cyclic fragments are present in anticonvulsant compounds.[18] If we analyze the series of fragments from F33 to F35 we can find an interesting behavior. These fragments not only are very common in nonherbicides compounds but also are still abundant in herbicides. As a herbicide must be simple in its chemical structure, the contribution of such fragments (such as F33 to F35) must be taken into account carefully. As can be seen from the former series, the higher the electronegative halogen substituent the increase the contribution of the fragment in herbicidal activity. Besides model 2 predicts also a positive contribution of hydrophobicity in herbicidal activity, property which also increases on this series.

However there are three fragments that need special attention. When a further halogen type atom is added to the fragment, an opposite effect is observed. This fact can be explained taken into account the influence that four (with its high electronegativity) exert their own mode of action on an herbicidal compound. Anyhow the difference between fragment 36 and 37 regarding their contribution to herbicidal activity must be taken into account. We cannot reach the same conclusion from the comparison between fragments 36 and 38. It seems that hydrophobicity has a higher influence on fragment 38 as the molecular volume (of the halogen) increases and its electronegativity decreases.

Finally, analyzing the contributions of fragment 30, it seems that (at first sight) the number of unsaturated bonds has a positive effect on the herbicidal activity. The same is true for the fragment 29 as well. These cannot be taken for granted because this interaction of these fragments with the rest of the molecule is also an important aspect in order to explain properly this phenomenon. Based on all these results we can conclude the higher and more specific fragments should be studied in order to obtain better herbicidal compounds by the TOPS-MODE approach.

CONCLUDING REMARKS

Despite some criticism, there is an increase in necessity of topological-indices-based QSAR models in order to

rationalize the drug discovery process. In this sense, the TOPS-MODE approach has been extended not only to the discovery of novel leads but also to the study of the physicochemical and absorption properties of drugs.[49,50] On the other hand, QSAR make use of reduced or homologous series of compounds. Consequently, it decays the model capacity to predict the activity of different structural features. In the present paper the TOPS-MODE approach has been largely, probe to generate good predictive linear models in order to account for herbicide activity. Thence, we can assert that the TOPS-MODE approach may be used as an efficient alternative to massive screening of drugs, pesticides, and herbicides.

ACKNOWLEDGMENT

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. González D.H. thanks the University of "Santiago de Compostela" in Spain for kind hospitality and the Xunta the Galicia (PR405A2001/65-0) for partial financial support. Additionally Maykel Pérez thanks the collaboration of the Dr. Aleksandar Sabljic for his significant contribution to the development of this work and 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) Sadowski J.; Kubinyi H. A scoring scheme for discriminating between drugs and nondrugs. *J. Med. Chem.* **1998**, *41*, 3325−3329.
(2) Reddy, K. N.; Locke, M. A. Prediction of soil sorption (Koc) of herbicides using semiempirical molecular properties. *Weed Sci.* **1994**, *42*, 3, 453−461.
(3) Huppatz. J. L. Quantifying the inhibitor-target site interactions of photosystem II herbicides. *Weed Sci.* **1996**, *44*, 3, 743−748.
(4) Murai, S., Nakamura, Y.; Akagi, T.; Sakashita, N.; Haga, T. Synthesis and quantitative structure−activity relationships of pyridylsulfonylurea herbicides. *ACS Symp. Ser.* **1992**, *504*, 43−55.
(5) Clark, R. D. Synthesis and QSAR of herbicide 3-pyrazolyl alpha,-alpha, alpha-trifluorotolyl ethers. *J. Agric. Food Chem.* **1996**, *44*, 11, 3643−3652.
(6) Little, D. L.; Ladner, D. W.; Shaner, D. L.; Ilnicki, R. D. Modeling root absorption and translocation of 5-substituted analogues of the imidazolinone herbicide, imazapyr. *Pesticide Sci.* **1994**, *41*, 3, 171−185.
(7) Bargar, T. M.; Secor, J.; Markley, L. D.; Shaw, B. A.; Erickson, J. A. A comparative molecular field analysis study of obtusifoliol 14 alpha-methyl demethylase inhibitors. *Pesticide Sci.* **1999**, *55*, 11, 1059−1069.
(8) Hansch. C.; Leo, A. *QSAR: Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, 1995.
(9) Kubinyi, H. *QSAR: Hansch Analysis and Related Approaches*; VCH: New York, 1993.
(10) Estrada, E. Spectral moments of the edge adjacency matrix in molecular graphs. 1. Definition and applications to the prediction of physical properties of alkanes. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 844−849.
(11) Estrada, E. Spectral moments of the edge adjacency matrix in molecular graphs. 2. Molecules containing heteroatoms and QSAR applications. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 320−328.
(12) Estrada, E. Spectral moments of the edge adjacency matrix in molecular graphs. 3. Molecules containing cycles. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 23−27.
(13) Estrada, E. Modeling the diamagnetic susceptibilities of organic compounds by a substructural graph theoretical approach. *J. Chem. Soc., Faraday Trans*. **1998**, *94*, 1407−1411.
(14) Estrada, E. Generalized Spectral Moments of the interacted line graph sequence. A novel approach to QSPR studies. *J. Chem. Inf. Comput. Sci*. **1998**, *39*, 90−95.
(15) Estrada, E.; Peña, A.; García-Domenech, R. Designing sedative/hypnotic compounds from a novel structural graph-theoretical approach. *J. Comput.-Aided Mol. Des*. **1998**, *12*, 583−595.
(16) Estrada, E.; Gutiérrez, Y. Modeling chromatographic parameters by a novel graph theoretical sub-structural approach. *J. Chromatogr. A* **1999**, *858*, 187−199.
(17) Estrada, E.; Gutiérrez, Y.; González, H. Modeling diamagnetic and magnetooptic properties of organic compounds with the TOSS-MODE approach. *J. Chem. Inf. Comput. Sci*. **2000**, *40*, 1386-1399.
(18) Estrada, E.; Peña, A. In Silico studies for the rational discovery of anticonvulsant compounds. *Bioorg. Med. Chem*. **2000**, *8*, 2755−2770.
(19) Estrada, E.; Uriarte, E.; Montero, A.; Tejeira, M.; Santana, L.; De Clercq, E. A novel approach to the rational selection and design of anticancer compounds. *J. Med. Chem*. **2000**, *43*, 1975−1985.
(20) Estrada, E.; Molina, E. QSPR/QSAR by graph theoretical descriptors beyond the frontiers. In *QSAR/QSPR studies by molecular descriptors*; Diudea, M., Ed.; Nova Science: New York, 2000; pp 83−107.
(21) Estrada, E.; Uriarte E. Quantitative structure-toxicity relationships using TOPS-MODE. 1. Nitrobenzene toxicity to tetrahymena pyriformis. *SAR QSAR Environ. Res.* **2001**, *12*, 309−324.
(22) Estrada, E. et al. *SAR QSAR Environ. Res*. **2003**, *14*, 145−163.
(23) Estrada, E.; Uriarte, E. Recent Advances on the role of Topological Indices in Drug Design Discovery Research. *Curr. Med. Chem*. **2001**, *8*, 1573−1588.
(24) Estrada, E. Edge Adjacency relationship and a novel topological index related to molecular volume. *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 31−33.
(25) Gutierrez, Y.; Estrada, E. Modes Lab, 2002, version 1.0 b.
(26) Negwer, N. *Organic-Chemical Drugs and their Synonyms*; Akademie-Verlag: Berlín, 1987.
(27) Asthon F. M.; Crafts A. S. *Mode of action of herbicide*; Wiley-Interscience Publications: New York, 1981; p 525.
(28) Thomson, W. T. *Agricultural chemicals: Herbicides book II*, 14th ed.; Thomson Publications: revised 1998.
(29) Sabljic, A.; Gusten H.; Verhaar H.; Hermens J. Qsar modeling of soils sorption. Improvements and systematics of log K_{oc} vs log K_{ow} correlations. *Chemosphere* **1995**, *31*, 4489−4514.
(30) Gramatica P.; Corradi M.; Consonni Modelling and prediction of soil sorption coefficients of nonionic organic pesticides by molecular descriptors. *Chemosphere* **2000**, *41*, 763−777.
(31) Baker J.; Mihelcic J.; Sabljic A. Reliable QSAR for estimating K_{oc} for persistent organic pollutants: correlations with molecular connectivity indices. *Chemosphere* **2001**, *45*, 213−221.
(32) Statsoft, Inc. STATISTICA 6.0, 2002, version 6.0.
(33) Van Waterbeemd, H. Discriminant analysis for activity prediction. In *Method and Principles in Medicinal Chemistry*; Manhnhold, R., Krogsgaard-Larsen, Timmerman, H., Eds.; VCH: Weinhiem, 1995; Vol. 2 (Chemometric methods in molecular design, Van Waterbeemd, H., Ed.), pp 265−282.
(34) 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.
(35) McFarland, J. W.; Gans, D. J. Cluster Significance Analysis. In *Method and Principles in Medicinal Chemistry*; Manhnhold, R., Krogsgaard-Larsen, Timmerman, H., Eds.; VCH: Weinhiem, 1995; Vol. 2 (Chemometric methods in molecular design Van Waterbeemd, H., Ed.), pp 295−307.
(36) Johnson, R. A.; Wichern D. W. *Applied Multivariate Statistical Analysis*; Prentice-Hall: NJ, 1988.
(37) González, H.; Olazabal, E.; Castañedo, N.; Hernádez, I.; Morales, A.; Serrano, H. S.; González, J.; Ramos, R. Markovian Chemicals "*in Silico*" Design (MARCH-INSIDE), a promising approach for Computer Aided Molecular Design II: Experimental and Theoretical Assessment of a Novel Method for Virtual Screening of Fasciolicides. *J. Mol. Mod*. **2002**, *1*, 83−95.
(38) Franke, R. *Theoretical Drug Design Methods*; Elsevier: Amsterdam, 1984.
(39) García-Domenech, R.; Gálvez, J.; Moliner, R.; García-March, F. Prediction and Interpretation of some Pharmacological Properties of cephalosporins Using Molecular connectivity. *Drug. Invest.* **1991**, *3*, 344−350.
(40) Gracía-Domenech, R.; de Julián-Ortiz, J. V. Antimicrobial Activity Characterization in a Heterogeneous Group of Compounds, *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 445−449.
(41) Kier, L. B.; Hall, L. H. The Nature of Structure−Activity Relationships and Their Relation to Molecular Connectivity. *Eur. J. Med. Chem-Chim. Ther*. **1977**, *12*, 307−312.
(42) Kier, L. B.; Hall, L. H. The Relation of Molecular Connectivity to Molecular Volume and Biological Activity. *Eur. J. Med. Chem-Chim. Ther*. **1981**, *16*, 399−407.
(43) Gálves, J.; Gracía-Domenech, R.; de Julián-Ortiz, J. V.; Soler, R. Topological Approach to drug design. *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 272−284.
(44) Randić, M. Resolution of ambiguities in quantitative structure−property studies by use of orthogonal descriptors. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 311−320.

(45) Randić, M. Orthogonal molecular descriptors. *New. J. Chem*. **1991**, *15*, 517−525.
(46) Randić, M. Correlation of enthalpy of octanes with orthogonal connectivity indices *J. Mol. Struct. (THEOCHEM)* **1991**, *233*, 45−59.
(47) Estrada, E.; Perdomo, I.; Torres-Labandeira, J. Combination of 2D-, 3D-Connectivity and Quantum Chemical Descriptors in QSPR. Complexation of α- and *β*-Cyclodextrin with Benzene Derivatives. *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 1561−1568.
(48) Randić, M. Fitting of Nonlinear Regressions by Orthogonalized Power Series. *J Comput. Chem*. **1993**, *4*, 363−370.
(49) Estrada, E.; González, H. D. What Are the Limits of Applicability for Graph Theoretic Descriptors in QSPR/QSAR? Modeling Dipole Moments of Aromatic Compounds with TOPS-MODE Descriptors. *J. Chem. Inf. Comput. Sci*. **2003**, *43*, 75−84.
(50) Cabrera, M. A.; González, H. D.; Teruel, C.; Pla-Delfina, J. M.; del Val Bermejo, M. A ToSS MoDe Approach to Determine Physicochemical and Absorption Properties in 6-Fluoroquinolones Derivatives: Experimental Validation. *Eur. J. Pharm. Biopharm*. **2002**, *53*, 317−325.

CI034039+