Designing Antibacterial Compounds through a Topological Substructural Approach

Enrique Molina,[†] Humberto González Díaz,[‡,§] Maykel Pérez González,*[,‡,||] Elismary Rodríguez,[†] and Eugenio Uriarte[§]

Department of Chemistry and Pharmacy, Faculty of Engineering Chemistry and Pharmacy, University of Camagüey, 74650 Camagüey, Cuba, Chemical Bioactives Center, Central University of Las Villas, Santa Clara, 54830 Villa Clara, Cuba, Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela, 15706 Santiago de Compostela, Spain, Service Unit, Experimental Sugar Cane Station "Villa Clara-Cienfuegos", Ranchuelo, Villa Clara 53100, Cuba

Received September 9, 2003

A novel application of TOPological Substructural MOlecular DEsign (TOPS-MODE) was carried out in antibacterial drugs using computer-aided molecular design. Two series of compounds, one containing antibacterial and the other containing non-antibacterial compounds, were processed by a *k*-means cluster analysis in order to design training and predicting series. All clusters had a *p*-level < 0.005. Afterward, a linear classification function has been derived toward discrimination between antibacterial and non-antibacterial compounds. The model correctly classifies 94% of active and 86% of inactive compounds in the training series. More specifically, the model showed a global good classification of 91%, i.e., 263 cases out of 289. In predicting series, the model has shown overall predictabilities of 91 and 83% for active and inactive compounds, respectively. Thereby, the model has a global percentage of good classification of 89%. The TOPS-MODE approach, also, similarly compares with respect to one of the most useful models for antimicrobials selection reported to date.

INTRODUCTION

The principal use of antibacterial drugs is the prevention and control of infectious process.[1] However, novel antibacterial compounds with more selectivity and less toxicity are needed for the future. The necessity for novel specific antibacterial drugs is evidenced by the microorganisms' development of resistance to drugs.[2−8] For this reason many researchers worldwide have been interested in the search and evaluation of novel antimicrobial compounds.[9−16] Specifically, our center has been working fundamentally in synthesis, biological testing, and development of antibacterial drugs from nitrovinylfuran compounds.[17−20] These chemicals have been known for a long time to have antimicrobial and other activities.[2−27] Hence, the broad diversity of possible chemical structures to test imposed on us the necessity for development of an alternative technique to classical trial and error screenings.

There are essentially two known methods for the "discovery" of novel drugs: the "rational design" and the "massive screening". In this connection, computer-aided drug design methods have been rapidly developed in the last years.[28] In particular, graph-theoretical methods may be very useful in quantitative structure−activity relationship (QSAR) problems, to perform a rational analysis of different pharmacological activities.[29] In the context of novel in silico, graph-theoretical and topological methods for modeling physicochemical and biological properties of a chemical, there has been introduced the TOPological Substructural MOlecular DEsign (TOPS-MODE) approach.[30−44] The TOPS-MODE has been applied to the description of physicochemical properties of organic compounds. So many applications for the design of biologically active compounds have been also described. Thereby, the aims of this work are first to find rationality in the search of novel antibacterial drugs using the TOPS-MODE approach and, second, to continue the validation of the method for describing the biological activity of a heterogeneous series of compounds.

LINEAR DISCRIMINANT ANALYSIS AND TOPS-MODE APPROACH

Here, we use the TOPS-MODE approach to obtain the molecular descriptors with which we developed the QSAR function. The mathematical details of the method have been largely reported elsewhere;[30−40] thus we will outline only the fundamental remarks. This method was introduced as TOSS-MODE,[30−40] but its own author renamed it since 2001.[41−43]

Briefly, this method codifies the molecular structure by means of the edge adjacency matrix **E** (likewise called the bond adjacency matrix **B**). The **E** or **B** matrix is a square table of order *m* (the number of chemical bonds in the molecule).[44] The elements of such a matrix (e_{ij}) are equal to 1 if bonds *i* and *j* are adjacent (it means that an atom exists that participates either in bond *i* or in bond *j*) or 0 otherwise. To codify information related to heteroatoms, the TOPS-MODE approach uses $\mathbf{B}(\mathbf{w}_{ij})$ weighted matrices instead of **B**. The weights ($\mathbf{w}_{ij}$) are chemically meaningful numbers such as bond distances, bond dipole, bond polarizabilities, or even mathematical expressions involving atomic weights.[41−44]

* Corresponding author phone: 53-42-281473; Fax: 53-42-28113; e-mail: mpgonzalez76@yahoo.es.
[†] University of Camagüey.
[‡] Central University of Las Villas.
[§] University of Santiago de Compostela.
[||] Experimental Sugar Cane Station "Villa Clara-Cienfuegos".

10.1021/ci0342019 CCC: $27.50
Published on Web 02/24/2004

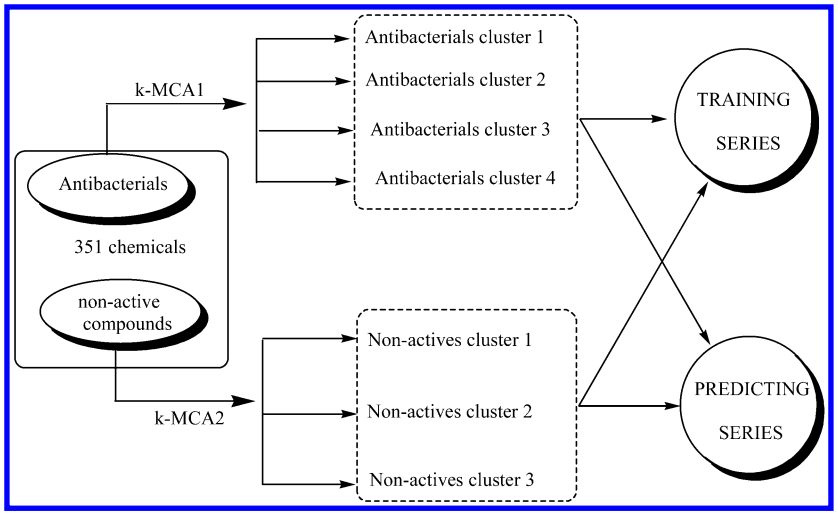

Figure 1. Training and predicting series design throughout *k*-MCA.

These weights are introduced in the main diagonal of matrix **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 **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 matrix **B(w$_{ij}$)**^{k}.

Linear discriminant analysis (LDA) has been the election statistical technique in most of the QSAR studies carried out using the TOPS-MODE approach.[35,38−40,44] In the present work, a similar expression for the QSAR is derived:

$$\text{Antib-Actv} = b + b_0\mu_0 + b_1\mu_1 + b_2\mu_2 + ... + b_k\mu_k \quad (1)$$

where Antib-Actv (acronym of antibacterial activity) is an indicator variable. This variable reaches the values Antib-Actv = 1 for antibacterial compounds or Antib-Actv = −1 for the nonactive ones. Deciding whether a compound may be classified as an antibacterial or not is based on the information extracted from the literature.[45] The **B(w$_{ij}$)** matrix was weighted in the main diagonal with the standard bond dipole moments.[30−44,46] The calculation of μ_k was carried out by means of the software package Modes Lab 1.0 b.[47]

In eq 1, b_k are the coefficients of the classification function, determined by least-squares as implemented on the LDA modulus of STATISTICA 6.0.[48] Forward stepwise was fixed as the strategy for variable selection.[49] To develop the QSAR for antibacterial/non-antibacterial 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 must important steps in computer-aided search of novel antibacterial drugs is to design a representative and randomized training and predicting series. With this aim, we select a large data set of 351 compounds having great structural variability: 213 of them are active (antibacterial) and the other ones are inactive compounds.[45] The inactive compounds were selected from drugs of 11 different pharmacological groups: antiinflammatory, antihistaminic, tuberculostatic, neuroleptic, diuretic, antihypertensive, analgesic, anticonvulsant, anticoagulant, sedative/hypnotic, and antiasthmatic. Later, two *k*-means cluster analyses (*k*-MCAs) were performed for active and inactive series of compounds.[50]

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]	F[c]	p-level[d]
Statistics for Active Compound Clusters				
μ_3	7.43	0.62	832.7	0.00
μ_4	7.32	0.60	853.5	0.00
μ_5	7.98	0.60	930.1	0.00
Statistics for Inactive Compound Clusters				
μ_6	1.22	0.21	398.5	0.00
μ_7	1.21	0.20	401.5	0.00
μ_8	1.17	0.20	399.3	0.00

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

K-MEANS CLUSTER ANALYSIS

The *k*-MCA may be used in training and predicting series design.[50,51] The idea consists of carrying out a partition of either an active or nonactive series of compounds in several statistically representative classes of chemicals. Hence, one may select from all these classes the members of training a predicting series. This procedure ensures that any chemical class (as determined by the clusters derived from *k*-MCA) will be represented in both compounds series (training and predicting). It permits design 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 the other for the inactive compounds) were carry out to select a representative sample for the prediction and training sets.

A first *k*-MCA (*k*-MCA1) splits antibacterial compounds in four clusters with 27, 88, 81, and 18 members and standard deviations of 0.22, 0.12, 0.11, and 0.32, respectively. On the other hand, the series of nonactive compounds was partitioned into three clusters (*k*-MCA2) with 15, 33, and

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

name	ΔP	name	ΔP	name	ΔP
enheptin-P[b]	−0.31	miloxacin	0.99	azodermin	0.61
benzazone-VII	0.39	nifuroxazide	0.95	norfloxacin	0.98
nitrofurfurilo	0.19	capillen[b]	−0.43	ceprochin	0.12
furaguanidine	0.15	nifurpirinol	0.72	benzonaftol	0.62
citric acid[b]	−0.14	cinoxacin	0.98	rosoxacin	0.98
desderman	0.80	nalidixicum ac.	0.78	balsalazide	1.00
dibromsalicylamide	0.73	geosot[b]	−0.46	pefloxacin	0.99
furazonal	0.45	arbutin	0.75	isoconazole nitrate	1.00
nifuraldezone	0.91	nifurpipone	0.89	magnolol	0.36
nifurthiazole	0.90	impregon	1.00	honokiol	0.32
nitrofurantoin	0.94	bromchlorofenum	0.99	ofloxacin	1.00
furazolidone	0.79	anabial	0.99	Su-7692[b]	−0.10
nifuradene	0.68	clobromsalan	0.95	xibornol[b]	−0.38
methylacetopyronone	0.38	triclocarban	0.98	xenysalate	0.47
nidroxyzone	0.76	nifurstyrenate	0.93	epiroprim	0.62
broxyquinoline	0.61	oxolinic acid	0.98	amonal	0.56
clioquinol	0.57	nitrofen	0.71	T 615	0.74
chlorquinol	0.66	SC 28538	0.88	ethylhydrocupreine	0.87
criminal	0.96	brodimoprim	0.91	santacyl	0.10
nitroxoline	0.61	nifurfoline	0.99	aminoacrichine	0.93
actofuran	0.73	diaveridine	0.66	pyrrolomycin A	0.48
furacril	0.52	furaltadone	0.98	mesalazine[b]	−0.30
furalazine	0.51	amonal B	0.44	AM 833	1.00
nifurtoinol	0.95	T 638	0.89	cryptotanshinone	0.77
micotiazone	0.69	flumequine	0.98	tetrenolin	0.46
furmethoxadone	0.71	tioxacin	0.61	pyrimethamine	0.14
nifurimide	0.64	droxacin	0.96	demethylthiolutin	0.42
broquinaldol	0.73	nifuralide	0.84	patulin	0.23
chlorquinaldol	0.77	terizidone	0.93	thiolutin	0.26
tilbroquinol	0.25	cefazolin	0.86	hydroxifumigatin	0.82
nifurprazine	0.38	etoxazene[b]	−0.40	pentizidone	0.16
akritoin	0.95	piromidic acid	0.85	showdomycin	0.86
tiliquinol[b]	−0.46	pipemidic acid	0.92	mizoribine	0.77
miran	0.55	metioprim	0.92	negamycin	0.32
dioxidin[b]	−0.40	trimethoprim	0.91	cordycepin	0.56
nifurdazil	0.81	butazopyridine[b]	−0.35	bromamphenicol	0.94
linatine	0.58	cicliomenol	U	fluoramphenicol	0.94
cloxiquine acetate	0.80	amonal A	0.45	neamine	0.72
furylfuramide	0.91	NF 167	0.90	ceftezole	0.98
nifurpyrimidine	0.74	mebroxine	0.96	ceftizoxime	0.97
nitrofurfurylideneisoniazid	0.93	phenylaspriodine	0.96	cetofenicol	0.94
antibrucellin	0.88	acetylsalol	0.81	chloramphenicol gly.	0.99
cloponone	0.98	enoxacin	0.98	actinobolin	0.83
nifurvidine	0.80	heptelidic acid	0.35	indolmycin	0.14
acefuralazine	0.94	aditoprime	0.78	anisomycin	0.85
mefuralazine	0.88	crytolepine	0.23	lomondomycin	1.00
furamizole	0.90	ipsalazide	1.00	chloramphenicol	1.00
nifurzide	0.91	nifurquinazol	0.96	FCE 20054	1.00
everninocin	1.00	cefradine	0.94	congocidine	0.99
sporaricin B	0.56	tomaymycin	0.98	clindamycin	0.96
diploicin	1.00	furoxacillin	0.28	atranorin	1.00
cefetrizole	0.94	cefapirin	1.00	cefathiamidine	0.95
cefuroxime	1.00	sparsophenicol	1.00	mirincamycin	0.96
cefalexin	0.84	astromicin	0.93	cefivitril	1.00
cefoxitin	0.95	quinacillin	0.40	cefedrolor	0.99
cefotaxime	1.00	cefaloram	1.00	trospectomycin	0.20

[a] *U* represents an unclassified compound; unclassified compounds are those that have a difference in their posterior probabilities of classification lower than 0.05. [b] Misclassified compounds.

90 members and standard deviations of 0.25, 0.12, and 0.07, respectively. 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 between SS and within SS (standard deviation between and within clusters), the respective Fisher ratio, and their *p*-level of significance considered to be lower than 0.05.[50−52] All spectral moments (from μ_0 to μ_{15}) were used in both analyses; all variables show *p*-levels < 0.05 for the Fisher test; the results are depicted in Table 1.

There is a main conclusion that may be withdrawn from *k*-MCA: the structural diversity of several to-date known antibacterial compounds (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 or pharmacological signification of these clusters seem to be speculative. Meanly, if it is considered that *k*-MCAs based partitions of data which

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

name	ΔP	name	ΔP	name	ΔP
bromobutanol	−1.00	butoctamide	−0.92	metrafazoline[a]	0.40
bason	−1.00	metomidate	−0.25	methylis nicotinas	−0.37
bromisoval	−0.64	glutethimi	−0.90	vasactin	−0.91
carbromode	−1.00	hoechst 264	−0.88	manozodil	−0.97
baldrianol	−0.81	alonimid[a]	0.26	tifemoxone	−0.21
calciidiethylacetase	−0.97	etomidate	−0.50	pinacidil	−0.89
amylurea	−0.95	fepitrizola	0.45	isoxsuprine	−0.36
mepentamate	−1.00	yodo-hachemina	−0.10	labetalol[a]	0.33
ectylurea	−0.51	nicatol	−0.46	timonacic	−0.46
ibrotamide	−1.00	ac. clofibricum	−0.91	mercaptopurine	−0.50
proptylurea	−0.70	clofibrate	−0.92	dacarbazine	−0.61
somnamid	−0.85	mandenol	−0.72	ethoxene	−0.99
chloral-PABA	−0.68	imanixil	−0.56	cycloleucine	−0.98
bromophenazone[a]	0.21	enprofylline	−0.06	caracemide[a]	0.24
trichloroisobutyl salicylate	−0.98	isamoxole	−0.93	trimelamol[a]	0.13
propionylphenetidin	−0.64	dimabefylline[a]	0.65	meradin	−0.54
bromoaminoacetate	−0.81	isalon DIWAG	−0.97	Ba 21 381	−0.94
isopral	−1.00	ileton Pliva	−0.77	acluracil	−0.40
quisqualamine	−0.59	tiamenidine	−0.86	citenazone	−0.88
γ-hidroxibutirato	−0.83	oxdralazine	−0.85	abbott-29 590	−0.62
carbocloral	−0.95	clonidinea	0.18	strinoline	−0.06
alcabrol	−0.91	guanoclor sulfate[a]	0.43	emorfazone[a]	0.36
valerium Paul Thibault	−0.94	gemedin	−0.92	SM 1704	−0.14
amylene hydrate	−1.00	nicopholine[a]	0.14	furofenac[a]	0.16
clomethiazole	−0.76	metirosine	−0.94	flutiazin	−0.91
cetohexazine	−0.63	bupicomide	−0.71	metiazinicum ac.	−0.49
elaldehydum	−0.95	hyperium	−0.91	fenharmane[a]	0.06
aponal	−0.99	guanethidine	−0.92	bamipine	−0.52
ethchlorvynol	−1.00	mydrial	−0.91	dithiaden	−0.29
penthrichloral	−0.93	oxedrine	−0.84	antazoline	−0.52
barbital(969)	−0.91	abiadin	−0.31	modaline sulfate	−0.80
acecarbromal	−0.78	proxamine	−0.41	rofelodine	−0.35
capuride	−0.85	formetanate	−0.19	nomifensine	−0.60
subdamine	−0.94	indanazoline	−0.60	clorprothixene	−0.21
fepiron	−0.48	tetryzoline	−0.78	methopromazine	−0.77
femerazol	−0.78	tramazoline	−0.49	tilozepine[a]	0.27
GYKI 21 622	−0.83	coumazoline	−0.39	dehydroclothepine	−0.22
provalamidum	−0.99	isopropylmethamine	−0.55	phensuximide	−0.56
centazolone[a]	0.47	oxymetazoline	−0.92	mephenytion	−0.97
ethadione	−0.99	ethosuximide	−0.96	SOG-18	−0.86
brosuximide[a]	0.24				

[a] Misclassified compounds.

consider not only four but also five or six clusters are statistically significant too (not reported results). 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 as few parameters (b_k) as possible.

To derive a discriminant function that permits the classification of chemicals as active (antibacterial) or inactive (non-antibacterial), we will 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:

$$\begin{aligned}\text{Actv} = &-1.014\mu_0 + 1.762\mu_1 - 0.032\mu_5 + 0.152\mu_2 - \\ &0.149\mu_3 + 0.123\mu_4 + 6.46 \times 10^{-5}\mu_8 - 5.04N = \\ &289U = 0.454D^2 = 4.939F = 48.24 \quad (2)\end{aligned}$$

In this model, the coefficient U is the Wilk statistics, D^2 is the squared Mahalanobis distance, and F is the Fisher ratio. The Wilks U-statistics for the overall discrimination can take values in the range from 0 (perfect discrimination) to 1 (no discrimination). For the discrimination of active/inactive compounds studied here, the model classifies correctly 93% (157/168) of active and 86% (107/124) of inactive compounds in the training series; for a global good classification of 91%. The percentages of false actives and false inactive compounds in the training series were 7 (12/168) and 12% (17/124), respectively, and no statistical outliers were detected. The previous statement is based on two facts, all misclassified chemicals (accordingly to posterior probabilities and Mahalanobis's distance) did not given rise to model improvement after being left out from the model. Additionally, *k*-MCA demonstrates that no group of chemicals (possible outliers) exists that appreciable differentiates from the remnant ones. False actives are those inactive compounds that the model classifies as actives, and the false inactives are those actives classified as inactive by the model. In Tables 2 and 3, the compounds classifications using the above-given model are depicted.

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

name	ΔP	name	ΔP	name	ΔP
		Active Compounds			
nihydrazone	0.68	amifloxacin	0.99	cefmetazole	1.00
diiodohydroxyquinoline[a]	−0.47	S 825 455	0.96	polygodial	0.56
peonol[a]	−0.26	benzylis	0.27	kalafungin	0.99
nifurethazone	0.66	fumigatin	0.73	cefalotin	0.99
nifurmazole	0.96	MSD-819	0.31	cefadroxil	0.94
nifurizone	0.92	minimycin	0.92	althiomycin	0.97
DJ 6783	0.94	FCE 22101	0.37	cefroxadine	0.97
N 176	0.99	chloramphenicol	0.90	cephalosporin C	1.00
ormetoprim	0.80	imipenen*	−0.50	cefuracetime	1.00
ethacridine lactate[a]	−0.50	cefacetrile	1.00	sporaricin A	0.88
tetroxoprim	0.98	antimycin	0.92	usnic acid	0.91
aurocitrin	0.61	PS-5	0.63	deoxyfrenocilin	1.00
talmetoprim	1.00	citromycetin	1.00	lincomycin	0.90
metioxate hydrocloride	0.91	kasugamycin	0.99	cefrotil	0.99
L 280 401	1.00	cefaclor	0.94	cefixime	1.00
		Inactive Compounds			
carbromal	−0.97	Ag 307	−0.96	praxadine	−0.95
thiourethane	−0.95	guancidine	−0.98	fenamole	−0.93
hedonal	−0.95	azepexole	−0.86	tetridamine	−0.89
brallobarbital	−0.45	phenamazoline	−0.85	B. W. 775C	−0.98
allylthiourea	−0.96	pyridylcarbinol	−0.86	KB 1 043[a]	0.36
quinetalate[a]	0.20	NSC-143 019	−0.88	clotiapine[a]	0.18

[a] Misclassified compounds.

In those tables, $\Delta P = [P(\text{actv}) - P(\text{nonactv})]$, where P(actv) is the a posteriori probability with which the model classifies a compound as active. Conversely, P(nonactv) is the posteriori probability with which the model classifies a compound as nonactive. This value (ΔP) takes positive values when P(actv) > P(nonactv) and negative otherwise. Therefore, when ΔP is positive (negative) the compound was classified as antibacterial (non-antibacterial). When ΔP is in the range $-0.05 < \Delta P < +0.05$, the compound was considered unclassified.[53]

One of the most important criteria for the acceptance or not of a discriminant model, such as model 2, is based on the statistics for the external prediction series.[51,53,54] Model 2 classifies correctly 91 and 83% of active and inactive compounds in the prediction series, respectively, which represents an overall predictability of 89%. 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 the active or inactive group.

COMPARISON WITH OTHER APPROACHES

Certainly, there must have been previously published several QSARs on antimicrobial activity, e.g., ref 55. However, those models, which give rise to a good discrimination of this activity in large and heterogeneous series of organic compounds, are uncommon. Particularly, García-Domenech and de Julián-Ortiz reported a good model in this sense.[56] These authors make use of the connectivity indices (topological indices)[57−59] to find both LDA and ANN (artificial neural network) based split rules that make possible classification of chemicals according to the presence or not of antimicrobial activity. The ANN model is by definition a not-linear model henceforth; the comparison with the model reported here is meaningless. Conversely, it is perfectly possible to go ahead with the comparison of both LDA models (see Table 5). However, strict comparisons between the methodologies are not possible, due to the same data sets not being present and therefore limiting our discussion at the data selection in both cases.

Table 5. Comparison between TOPS-MODE and Connectivity Approaches for Antimicrobial Activity

statistics	connectivity	TOPS-MODE
training series over all predictability (%)	94	91
N for training series (actives)	64 (34)	289 (174)
predicting series over all predictability (%)	92	89
N for predicting series (actives)	47 (26)	63 (45)
N total	111	352
Wilks'λ (U-statistics)	0.2762	0.454
F	20.9	48.24
D^2	not reported	4.939
p-level	>0.05	>0.05
cluster analysis design of series	no	yes
structural composition of training series	3 groups[a]	broader range
structural composition of predicting series	7 groups[b]	broader range

[a] Quinolones, sulfonamides, and cephalosporins. [b] Diaminopyridine (one compound), cephamicins (two compounds), oxacephems (one compound), and sulfones (one compound) in addition to those groups used in training series.

The connectivity function has shown an overall predictability of 94%, which seems to be bigger than the TOPS-MODE function predictability (91%). Nevertheless, it is remarkable that the TOPS-MODE model was derived from training series (289/64) 4 $^1/_2$ times bigger than the series used by García-Domenech and de Julián-Ortiz. Specifically, TOPS-MODE training series have more than 5 times the number of active compounds with respect to Valencia's group model (174/34). However, both models significantly recognized the existence of active and inactive chemicals groups. In this sense, comparison of the absolute magnitudes of F, U, and/or D^2 values is meaningless. Unfortunately, García-Domenech and de Julián-Ortiz did not report D^2. Consequently, it is not possible to compare this parameter for both models. Anyhow, in this particular case it is not necessary because in a two group classification problem U have a Fisher distribution as well as D^2. Thereby, we can

use here *U*-statistics to test the hypothesis of the groups' separation. In both cases *p*-level was <0.05, indicating that there is a probability of error lower than 5% after accepting the hypothesis of separation of groups.[50,53,56]

Validation of the models is the other major bottleneck in QSAR. Both models were successfully validated by means of external prediction series. Again, the connectivity indices function has a bigger global predictability but uses a very much reduced number of compounds than the TOPS-MODE approach (see Table 5).

Another remarkable problem, especially in the case of heterogeneous series of chemicals classification, is the spectrum of structural patterns considered. In the present work, compound diversity is demonstrated not only by visual detection of the number of chemical families of compounds but also by use of *k*-MCA.[50] Without doubts, the TOPS-MODE model reported here considered a broader diversity of chemical families (compare Tables 1−5) in special if one takes into account that Domenech and de Julián-Ortiz add a few compounds of only four families to the predicting series.

CONCLUDING REMARKS

Despite some criticism, there is an increasing necessity for topological-indices-based QSAR models in order to rationalize the drug discovery process.[54,60] 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.[60,61] On the other hand, most recently published papers in this area make use of reduced or homologous series of compounds to fit the QSAR for antimicrobial drug screening. This fact determines that these are not general models, which could be used to predict the biological activity of heterogeneous series of compounds. In the present paper the TOPS-MODE approach has been largely probed to generate good predictive linear models in order to account for antimicrobial activity of a broader range of molecular structural patterns. Henceforth, we can assert that the TOPS-MODE approach may be used as an efficient alternative to massive screening of antimicrobial drugs.

ACKNOWLEDGMENT

The authors would like to express their gratitude to Ph.D. Jose Luis Garcia and the Cuban Ministry of Higher Education for financial support and kindness. H.D.G. and E.M. thank the University of "Santiago de Compostela" in Spain for kind hospitality and the Xunta the Galicia (Grant PR405A2001/65-0) for partial financial support, which has made possible the realization of up-to-date revision of antimicrobial and QSAR matters. The authors sincerely acknowledge the comments of Dr. R. Garcia-Domenech, and, specifically, M.P.G. thanks the owner of the software Modeslab 1.0 for donation of this valuable tool for the realization of this work. Finally, yet importantly, the authors sincerely acknowledge the comments of the anonymous reviewers that helped to improve the quality of the final manuscript.

REFERENCES AND NOTES

(1) Goodman, A.; Rall, T. W.; Nies, A. S.; Taylor, P. *Las Bases Farmacológicas de la Terapéutica*; Editorial Media Panamericana; Madrid, España, 1998; pp 7−11.

(2) Reese, E. R.; Betts, F. R.; Gumustop, B. *Handbook of Antibiotics*; Lippincott Williams and Wilkins: London, 2000.

(3) Aeschlimann, J. R.; Dresser, L. D.; Kaatz, G. W.; Rybak, M. J. Effects of NorA Inhibitors on in Vitro Antibacterial Activities and Postantibiotic Effects of Levofloxacin, Ciprofloxacin, and Norfloxacin in Genetically Related Strains of *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **1999**, *43*, 335−340.

(4) Spangler, S. K.; Lin, G.; Jacobs, M. R.; Appelbaum, P. C. Postantibiotic Effect of Sanfetrinem Compared with Those of Six Other Agents against 12 Penicillin-Susceptible and -Resistant Pneumococci. *Antimicrob. Agents Chemother.* **1997**, *41*, 2173−2176.

(5) Keil, S.; Wiedemann, B. Antimicrobial Effects of Continuous versus Intermittent Administration of Carbapenem Antibiotics in an in Vitro Dynamic Model. *Antimicrob. Agents Chemother.* **1997**, *41*, 1215−1219.

(6) Periti, P. Tobramycin-Clinical Pharmacology and Chemotherapy. *J. Chemother.* **1996**, *8*, 3−30.

(7) Xiong, Y. Q.; Caillon, J.; Drugeon, H.; Potel, G.; Baron, D. Influence of pH on Adaptive Resistance of *Pseudomonas aeruginosa* to Aminoglycosides and Their Postantibiotic Effects. *Antimicrob. Agents Chemother.* **1996**, *40*, 35−39.

(8) Finch, R. G. Antibacterial Activity of Quinupristin/Dalfopristin. Rationale for Clinical Use. *Drugs* **1996**, *51*, 31−37.

(9) Jung, F.; Delvare, C.; Boucherot, D.; Hamon, A. Synthesis and Structure−Activity Relationship of New Cephalosporins with Amino Heterocycles at C-7. Dependence of the Antibacterial Spectrum and β-Lactamase Stability on the pK_a of the C-7 Heterocycle. *J. Med. Chem.* **1991**, *34*, 1110−1116.

(10) Domagala, J. M.; Bridges, A. J.; Culbertson, T. P.; Gambino, L.; Hagen, S. E.; Karrick, G. Synthesis and Biological Activity of 5-Hydroxyquinolones, and the Overwhelming Influence of the Remote N_1-Substituent in Determining the Structure−Activity Relationship. *J. Med. Chem.* **1991**, *34* 1142−1154.

(11) Hagen, S. E.; Domagala, J. M.; Heifetz, C. L.; Johnson, J. Synthesis and Biological Activity of 5-Alkyl-1,7,8-trisubstituted-6-fluoroquinoline-3-carboxylic Acids. *J. Med. Chem.* **1991**, *34*, 1155−1161.

(12) Desos, P.; Lepagnol, J. M.; Morain, P.; Lestage, P.; Cordi, A. A. Structure−Activity Relationships in a Series of 2(1*H*)-Quinolones Bearing Different Acidic Function in the 3-Position: 6,7-Dichloro-2(1*H*)-oxoquinoline-3-phosphonic Acid, a New Potent and Selective AMPA/Kainate Antagonist with Neuroprotective Properties. *J. Med. Chem.* **1996**, *39*, 197−206.

(13) Maccia, M.; et al. Synthesis and Antimicrobial Activity of 7β-(*S*)- and 7β-[(*R*)-3-(methyleneaminoxy)-2-methylpropionamido] Substituted Cephalosporanic Acid Derivatives. *Farmaco* **1996**, *51*, 283−286.

(14) Kus, C.; et al. Synthesis and Antimicrobial Activity of Some New Piperidinyl Benzimidazoles. *Farmaco* **1996**, *51*, 413−417.

(15) Rho, T. C.; Bae, E. A.; Kim, D. H.; Oh, W. K.; Kim, B. Y.; Ahn, J. S.; Lee, H. S. Anti-*Helicobacter pylori* Activity of Quinolone Alkaloids from *Evodiae fructus*. *Biol. Pharm. Bull.* **1999**, *22*, 1141−1143.

(16) Gootz, T. D.; Zaniewski, R. P.; Haskell, S. L.; Kaczmarek, F. S.; Maurice, A. E. Activities of Trovafloxacin Compared with Those of Other Fluoroquinolones against Purified Topoisomerases and gyrA and grlA Mutants of *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **1999**, *43*, 1845−1855.

(17) Castañedo, N.; Goisueta, R.; Pérez, J.; González, J.; Silveira, E.; Cuesta, M.; Martínez, A.; Lugo, E.; Estrada, E.; Carta, A.; Navia, O.; Delgado, M. Cuban Pat. 22446, 1994. Eur. Pat. Appl. No. 955000567, 1994. Can. Pat. Appl. No. 2,147,594, 1994. Jpn. Pat. Appl. No. 222002, 1994. U.S. Pat. Appl. No. 60008011, 1994.

(18) Toledo, F.; Sanchez, L.; Lopez, E.; Martinez, J.; Brito, R. Deteccion e idetificacion del producto intermedio en la sintesis del bioactivo 1-(fur-2-il)-2-nitroeteno. *Afinidad.* **1998**, *478*, 443−447.

(19) Blondeau, J. M.; Castañedo, N.; González, O.; Medina, R.; Silveira, E. *In Vitro* Evaluation of G-1: A Novel Antimicrobial Compound. *Antimicrob. Agents Chemother.* **1999**, *11*, 1663−1669.

(20) Weuffen, W.; Kramer, A.; Groschel, D.; Bulka, E. Nitrofurane. In *Handbuch der Antisektik, Band II Antiseptika*; Weuffen, W., Berencsi, G., Groschel, D., Kemter, B., Kramer, A., Krasilnikow, A. P., Eds.; VEB Verlag: Berlin, 1979; pp 12−189.

(21) McCoy, W. F.; Thornburgh, S. Industrial Antimicrobial Uses for 2-(2-Bromo-2-nitroethenyl)furan, U.S. Pat. 5,090,990, 1992.

(22) Miyaji, T. Nitrofurans Compounds as Food Additives. In *Mechanism of Toxicity and Metabolism*, Vol. 6; Karki, N. T., Ed.; Pergamon: Oxford, U.K., 1976; pp 127−136.

(23) *Payazone Nitrovin, Growth Promoting Agent for Pigs*; Cyanamid International: Rostock, Germany, 1973; p 19.

(24) Chou, H. M.; Huangfu, M.; Chou, H. Clinical Evaluation of F 30066 in the Treatment of Acute Schistosomiasis. *Chin. Med. J.* **1963**, *82*, 250−257.

(25) Hulbert, P. B.; Bueding, E.; Robinson, C. H. Structure and Antischistosomal Activity in the Nitrofuryl Series. *J. Med. Chem.* **1973**, *16*, 72−78.

(26) Kelloval, G.; Sturdik, E.; Stibrany, L.; Drobnica, L.; Augustin, J. Antimicrobial Effect of Esters and Amides of 3-(5-Nitro-2-furyl)acrylic Acid. *Folia Microbiol.* **1984**, *29*, 23−34.
(27) Dore, J. Ch.; Viel, C. Antitumoral Chemotherapy: X. Cytotoxic and Antitumoral Activity of β-Nitrostyrenes and Nitrovinyl Derivatives. *Farmaco* **1975**, *30*, 81−109.
(28) Hansch, C.; Leo, A. *QSAR: Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, 1995.
(29) Kubinyi, H. *QSAR: Hansch Analysis and Related Approaches*; VCH: New York, 1993.
(30) 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.
(31) 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.
(32) 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.
(33) Estrada, E. Modeling the Diamagnetic Susceptibilities of Organic Compounds by a Substructural Graph Theoretical Approach. *J. Chem. Soc., Faraday Trans.* **1998**, *94*, 1407−1411.
(34) 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.
(35) 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.
(36) Estrada, E.; Gutiérrez, Y. Modeling Chromatographic Parameters by a Novel Graph Theoretical Sub-structural Approach. *J. Chromatogr., A* **1999**, *858*, 187−199.
(37) 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.
(38) Estrada, E.; Peña, A. In Silico Studies for the Rational Discovery of Anticonvulsant Compounds. *Bioorg. Med. Chem.* **2000**, *8*, 2755−2770.
(39) 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.
(40) González, M. P.; González, H. D.; Molina, R. R.; Cabrera, M. A.; Ramos, R. A. TOPS-MODE Based QSAR Derived from Heterogeneous Series of Compounds. Applications to the Design of New Herbicides. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1192−1199.
(41) 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.
(42) Estrada, E.; Molina, E.; Uriarte, E. Quantitative Structure−Toxicity Relationships Using TOPS-MODE. 2. Neurotoxicity of a Noncongeneric Series of Solvents. *SAR QSAR Environ. Res.* **2001**, *12*, 445−459.
(43) Estrada, E.; Uriarte, E. Recent Advances on the Role of Topological Indices in Drug Design Discovery Research. *Curr. Med. Chem.* **2001**, *8*, 1573−1588.
(44) Estrada, E. Edge Adjacency Relationships and a Novel Topological Index Related to Molecular Volume. *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 31−33.
(45) Negwer, N. *Organic-Chemical Drugs and Their Synonyms*; Akademie-Verlag: Berlin, 1987.
(46) Ferguson, L. N. *Textbook of Organic Chemistry*; Van Nostrand: Princeton, NJ, 1978.
(47) Gutierrez, Y.; Estrada. E. *Modes Lab*, version 1.0 b, Santiago de Compostela, España, 2002.
(48) Statsoft, Inc. (2001). STASTISTICA (data analysis software system), version 6.0. www.statsoft.com.
(49) Van Waterbeemd, H. Discriminant Analysis for Activity Prediction. In *Chemometric Methods in Molecular Design*; Van Waterbeemd, H., Ed.; Method and Principles in Medicinal Chemistry, Vol. 2; Manhnhold, R., Krogsgaard-Larsen, Timmerman, H., Eds.; VCH: Weinhiem, Germany, 1995; pp 265−282.
(50) Kowalski, R. B.; Wold, S. Pattern Recognition in Chemistry. In *Handbook of Statistics*; Krishnaiah, P. R., Kanal, L. N., Eds.; North Holland: Amsterdam, 1982; pp 673−697.
(51) Mc Farland, J. W.; Gans, D. J. Cluster Significance Analysis. In *Chemometric Methods in Molecular Design*; Van Waterbeemd, H., Ed.; Method and Principles in Medicinal Chemistry, Vol. 2; Manhnhold, R., Krogsgaard-Larsen, Timmerman, H., Eds.; VCH: Weinhiem, Germany, 1995; pp 295−307.
(52) Johnson, R. A.; Wichern D. W. *Applied Multivariate Statistical Analysis*; Prentice-Hall: Upper Saddle River, NJ, 1988.
(53) 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.
(54) Franke, R. *Theoretical Drug Design Methods*; Elsevier: Amsterdam, 1984.
(55) 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.
(56) 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.
(57) Gough, J.; Hall, L. H. Modeling the Toxicity of Amide Herbicides Using the Electrotopological State. *Environ. Toxicol. Chem.* **1999**, *18*, 1069−1075.
(58) Kier, L. B.; Hall, L. H. In *The Electrotopological State: Structure Modeling for QSAR and Database Analysis, in Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach: Reading, U.K., 1999.
(59) 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.
(60) 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.
(61) 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.

CI0342019