Topological Approach to Drug Design

J. Gálvez,*,† R. García-Domenech, J. V. de Julián-Ortiz, and R. Soler

Unit Research into Molecular Connectivity and Drug Design, Department of Physical Chemistry, Faculty of Pharmacy, University of Valencia, Spain

Received September 14, 1994®

In this paper we demonstrate that by an adequate combination of different topological indices it is possible to select and design new active compounds in different therapeutical scopes, with a very high efficiency level. Particularly successful in the search of new "lead drugs", the results show the surprising ability of the topological methods to describe molecular structures.

1. INTRODUCTION

In a recent paper[1] we showed that by an adequate choice of topological indices, mainly connectivity and some of their linear combinations, and using both linear discriminant analysis (LDA) as well as correlating selected biological properties, it is possible to predict minor (nonnarcotic) analgesic activity for a wide set of compounds with a very significant degree of accuracy.

In this paper those results are generalized in such a way that by using connectivity indices[2] and combinations of them as well as the "charge" [3] and "geometrical" indices[4] it is easy to discriminate the pharmacological activity for very different groups of drugs, including minor analgesics, antiviral, bronchodilator, antifungal, antihyperlipoproteinemic, hypoglycemic, or betablocker agents, with such a level of accuracy that, in most cases, only a low number of variables are required for the efficient discrimination of a wide set of compounds. Moreover, it is feasible to design new active molecules for a selected pharmacological field, by using topological indices in an inverse way as compared with the traditional method, i.e., instead of using them to predict the property/ies value/s for a given compound (or group of them), they are used to obtain the structure of the compound/s which present a preestablished theoretical value for the studied property.

The format of this paper will be as follows: An outline and definition of our methodology is described in section 2. The results obtained together with the discussion are included in section 3. Finally, the main conclusions are presented in section 4.

2. METHODOLOGY

Nowadays, the more common methods used for the design of new compounds showing preestablished properties or new drugs, imply the use of physicochemical descriptors in those called QSAR methods (Hansch, Free−Wilson, ...)[5] with the eventual addition of topological descriptors as a complement as well as the quantum mechanical calculations[6] or those based on molecular mechanics.[7]

In the case of drug design, although some of these methods are able to build a chemical that precisely fits its molecular target in the organism, they are usually limited to predefined structures ("pharmacophores"), which are later refined by successive steps following a process named "pharmacomodulation". However, when the objective is to find new "lead drugs", i.e., drugs not derived from a known active structure, the preceding methods are not very versatile.

Table 1. Contribution by Different Molecular Fragments to the value of S

Group	Contribution	Group	Contribution
	2		28
	6		14
	12		36
	20		
	10		18
	18		49.5
	24		

Other alternatives are possible to attack the problem. Pattern recognition methods, which are classified as "nonparametric" or "heuristic", use the molecular topology for the search of particular structural patterns which would be able to store a given biological or pharmacological action. They are strictly based on geometrical parameters and include no statistical considerations. Their ability is seriously limited by the set of variables used, and they lead to poor results. Two of the more important algorithms have been the named LLM[8] and KNN.[9]

In a recent paper[1] we showed that by using connectivity indices and some of their linear combinations, as well as a "E" shape factor, it is possible to predict not only the analgesic activity but also the analgesic potency, with a high level of accuracy (over 80%). Although those results demonstrate the ability of the Kier and Hall's descriptors on the prediction of the analgesic activity, the need to use a

† Current address: c/Ciscar, 40 46005-Valencia, Spain.

® Abstract published in *Advance ACS Abstracts,* February 1, 1995.

0095-2338/95/1635-0272$09.00/0

Table 2. Values of the Topological Descriptors (Excepting χ_s) for a Representative Set of Drugs

INDICES	ALCLOFENAC	BENORYLATE	ACYCLOVIR	ZIDOVUDINE	ACEFYLLINE	ISOPRENALINE	CLOFIBRATE	LOVASTATIN	BUFORMIN	TOLBUTAMIDE	BUFRALOL	PROPRANOLOL
G1	4.25	7.50	3.75	5.50	5.50	4.75	5.50	8.25	3.50	5.50	6.50	3.75
$G1^V$	9.45	16.05	12.43	14.02	14.90	10.23	12.95	17.10	2.55	6.80	13.45	11.18
G2	5.89	10.89	10.94	9.00	12.39	5.33	6.11	11.22	4.11	5.67	8.42	4.33
$G2^V$	6.49	11.38	11.34	13.28	13.87	4.72	7.08	17.12	4.94	5.71	9.02	5.32
G3	1.50	2.50	2.85	3.32	4.14	2.31	2.50	4.81	0.81	2.63	3.02	1.44
$G3^V$	1.48	3.16	2.92	5.04	5.51	1.10	3.40	7.73	1.55	2.71	3.32	1.78
G4	0.92	1.60	1.76	1.85	1.77	1.64	1.36	2.75	0.49	1.69	2.16	1.54
$G4^V$	1.11	2.57	1.59	2.45	2.25	1.01	1.75	4.42	1.00	1.89	1.84	1.10
G5	0.42	1.03	0.75	0.91	0.93	0.68	0.60	1.46	0.24	0.93	1.16	0.78
$G5^V$	0.62	1.28	0.82	1.27	0.88	0.66	0.92	2.20	0.56	1.28	1.15	0.57
J1	0.33	0.34	0.25	0.31	0.34	0.34	0.37	0.29	0.35	0.32	0.36	0.21
$J1^V$	0.73	0.73	0.83	0.78	0.93	0.73	0.86	0.61	0.25	0.40	0.75	0.62
J2	0.45	0.49	0.73	0.50	0.77	0.38	0.41	0.40	0.41	0.33	0.47	0.24
$J2^V$	0.50	0.52	0.76	0.74	0.87	0.34	0.47	0.61	0.49	0.34	0.50	0.30
J3	0.12	0.11	0.19	0.18	0.26	0.17	0.17	0.17	0.08	0.15	0.17	0.08
$J3^V$	0.11	0.14	0.19	0.28	0.34	0.08	0.23	0.28	0.16	0.16	0.18	0.10
J4	0.07	0.07	0.12	0.10	0.11	0.12	0.09	0.10	0.05	0.10	0.12	0.09
$J4^V$	0.09	0.12	0.11	0.14	0.14	0.07	0.12	0.16	0.10	0.11	0.10	0.06
J5	0.03	0.05	0.05	0.05	0.06	0.05	0.04	0.05	0.02	0.05	0.06	0.04
$J5^V$	0.05	0.06	0.05	0.07	0.06	0.05	0.06	0.08	0.06	0.08	0.06	0.03
$0\chi - 0\chi^V$	0.60	1.70	2.17	2.22	1.97	1.87	0.69	1.75	1.25	1.71	1.06	1.06
$0\chi / 0\chi^V$	1.07	1.13	1.26	1.22	1.21	1.21	1.07	1.10	1.19	1.16	1.09	1.09
$1\chi - 1\chi^V$	0.51	1.25	1.88	1.79	1.62	1.14	0.71	1.26	0.88	0.95	0.87	0.95
$1\chi / 1\chi^V$	1.11	1.18	1.41	1.31	1.34	1.23	1.13	1.11	1.24	1.15	1.13	1.14
$2\chi - 2\chi^V$	0.66	1.84	2.00	1.96	1.90	1.39	1.14	1.81	1.15	1.62	1.25	1.25
$2\chi / 2\chi^V$	1.21	1.39	1.65	1.46	1.52	1.36	1.27	1.19	1.52	1.36	1.21	1.25
$3\chi p - 3\chi p^V$	0.42	1.09	1.64	2.02	1.68	1.17	1.05	1.53	0.59	0.91	1.19	1.08
$3\chi c - 3\chi c^V$	0.16	0.53	0.29	0.39	0.46	0.39	0.40	0.53	0.29	0.80	0.51	0.32
$3\chi p / 3\chi p^V$	1.24	1.40	1.84	1.67	1.67	1.56	1.47	1.22	1.54	1.33	1.38	1.37
$3\chi c / 3\chi c^V$	1.47	2.03	1.93	1.66	1.80	1.65	1.41	1.38	3.17	2.39	1.31	1.51
$4\chi p - 4\chi p^V$	0.29	0.96	1.44	1.80	1.34	0.61	0.70	1.68	0.57	0.87	0.85	0.74
$4\chi c - 4\chi c^V$	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.15	0.10	0.00
$4\chi pc - 4\chi pc^V$	0.15	0.59	0.57	1.13	1.08	0.78	0.97	0.98	0.23	0.93	0.83	0.50
$4\chi p / 4\chi p^V$	1.26	1.52	2.11	1.89	1.80	1.50	1.55	1.33	2.00	1.56	1.41	1.38
$4\chi c / 4\chi c^V$	0.00	0.00	0.00	0.00	0.00	0.00	1.73	0.00	0.00	5.78	1.41	0.00
$4\chi pc / 4\chi pc^V$	1.37	1.83	2.00	1.97	1.93	2.01	1.93	1.36	3.03	2.10	1.60	1.59
W	371	1373	475	730	502	419	489	2293	201	708	767	817
N	14	23	16	19	17	15	16	29	11	18	19	19
V3	4	8	4	6	5	5	4	11	0	4	5	9
V4	4	7	4	4	5	3	4	3	2	4	5	3
Pr1	2	5	2	4	2	1	2	6	0	2	2	5
Pr2	2	4	2	4	3	3	2	9	0	2	3	4
Pr3	2	3	2	3	3	3	2	11	0	2	1	6
L	10	13	9	10	8	9	10	13	8	12	10	11
S	74	144	127	130	128	75	76	211	10	78	130	140
E	0.74	0.85	1.57	1.30	2.00	0.93	0.76	1.25	0.16	0.54	1.30	1.16

high number of variables to obtain good discriminate levels, led us to introduce new topological descriptors. The molecular charge distribution plays an important role in many biological and pharmacological activities. It can be evaluated through physicochemical parameters such as dipole moment and electronic polarizability. In a previous work,[3] "topological charge indices" (C.I.) were defined, and their ability to evaluate the charge transfers between pairs of atoms and the global charge transfer was demonstrated: A set of heterogeneous hydrocarbon compounds was chosen, and the C.I. were correlated against the dipole moment with a good correlation index.

Table 3. Discriminant Functions and Overall Classification from the Linear Discriminant Analysis on a Group of Compounds (43 with Analgesic Activity and 108 Inactives)

	data-base group classification				test group ("cross-validation" analysis) classification			
discriminant function	no. of compds	actives	inactives	% success	no. of compds	actives	inactives	% success
$A_1 = 0.152 - 5.895\ {}^4\chi_c^v$	actives: 43	40	3	93.0	actives: 50	47	3	94.0
	inactives: 108	71	37	34.3	inactives: 102	75	27	26.5
$A_2 = -2.267 - 2.527G_4 + 28.4\ J_4^v - 0.928\ (3\chi_p - 3\chi_p^v) + 1.027V4$	actives: 43	35	8	81.4	actives: 50	34	16	68.0
	inactives: 108	20	88	81.5	inactives: 102	23	79	77.5
A_1 and A_2	actives: 43	34	9	79.1	actives: 50	34	16	68.0
	inactives: 108	9	99	91.7	inactives: 102	16	86	84.3

Table 4. Results Obtained by the Discriminant Analysis on Analgesic Drugs[a]

base group				test group			
active compds	classif	inactive compds	classif	active compds	classif	inactive compds	classif
2,4-dimethylacetophenone	+	bromosalicylchloranilide	−	zomepirac	+	flucytosine	+
2-amino-4-picoline	+	buclosamide	−	acetanilide	+	fungichromin	−
diclofenac	+	butoconazole	−	flufenamic acid	−	griseofulvin	+
ASA	+	candicidin D	−	alclofenac	+	halethazole	+
acemetacin	+	chlordantoin	−	aminochlorthenoxacin	+	hexetidine	−
alminoprofen	+	chlormidazole	+	4-hydroxyisophthalic acid	+	miconazole	−
Ameclofen	+	dermostatin	−	ammonium salicylate	−	naftifine	−
mefenamic acid	+	dimazole dihydochloride	−	antipyrine	+	natamycin	−
aminopyrine	+	econazole	−	bucolome	+	nifuratel	−
aminopromazine	+	enilconazole	−	bufexamac	−	nystatin	−
benorylate	+	exalamide	−	bumadizon	+	siccanin	−
antrafenine	−	filipin	−	butibufen	+	sulbentine	−
apazone	+	fluconazole	−	carprofen	+	sulconazole	−
acetylsalicylsalicyclic acid	+	bromovirin	−	cinchophen	+	tenonitrozole	−
aniline	−	bucyclovir	−	cumetacin	+	terbinafine	−
benzydamine	−	cytarabine	−	clopirac	+	terconazole	−
benzpiperylon	+	descyclovir	−	1-(*p*-chlorophenyl)propanol	−	9-(3-hydroxypropoxy)	
bucloxi acid	+	desoxyacylovir	−	chlortenoxicam	+	guanine	−
etodolac	−	idoxuridine	−	droxicam	−	9-(1,3-dihydroxy-2-	−
phenacetin	+	methyl galate	−	epirizol	+	propoxymethyl)guanine	
fenbufen	+	ribavirin	−	etersalate	+	9-(4-hydroxybutyl)guanine	+
fencolfenac	+	trifluridine	−	ethenzamide	−	9-(4-hydroxy-3-(hydroxy-	−
phenylbutazone	+	vidarabine	−	phenol	−	methyl)butyl)guanine	
fenoprofen	+	zidovudine	−	fentiazac	+	9-arabinofuranosylpurine	−
difenpiramide	+	albuterol	−	feprazone	+	theophylline	+
diflunisal	+	bambuterol	−	flurbiprofen	+	medibazine	−
isoxicam	−	bitolterol	−	glafenine	+	methoxyphenamine	−
niflumic acid	−	carbuterol	−	glucametacin	−	tretoquinol	−
oxaceprol	−	clenbuterol	−	ibufenac	+	epinephrine	−
oxametacine	+	clorprenaline	−	ibuprofen	+	broxaterol	−
ramifenazone	+	dioxethedrine	−	ibuproxam	+	butylnoradernaline	−
salicylic acid	+	ephedrine	−	indomethacin	+	colterol	−
tiaprofenic acid	+	eprozinol	−	indoprofen	+	diphenylorciprenaline	−
tolmetin	+	etafedrine	−	isonixin	+	dimethylnorepinephrine	−
propyfenazone	+	fenoterol	−	kebuzone	+	dioxyephedrine	−
sulindac	+	formoterol	−	ketoprofen	+	isobutylmethylxantine	+
tenoxicam	−	hexoprenaline	−	ketorolac	+	imazodam	+
methyl salicylate	+	acebutolol	−	metofoline	−	bopindolol	−
salsalate	+	acefylline	+	morazone	−	bucumolol	−
salverine	−	enprofylline	+	naproxen	+	bufuralol	−
benorylate	+	theobromine	+	oxaprocin	+	bunitrolol	−
benoxaprofen	+	alprenolol	−	oxypizone	+	bupranolol	−
clidanac	+	amosulalol	−	paracetamol	+	butidrine hydrochloride	−
		arotinilol	−	piroxicam	−	butofilolol	−
		atenolol	−	pirprofen	−	dilevalol	−
		befunolol	−	proglumetacin	−	carazolol	−
		betaxolol	−	viminol	−	carteolol	−
		bevantolol	−			carvedilol	−
		bisoprolol	−			celiprolol	−

[a] For inactive compounds only a representative group is shown. A compound is classified as active (+) if $A_1 > 0$ and $A_2 > 0$, and as inactive (−) in other case.

The "topological charge indices" (C.I.), *Gk* and *Jk* are defined as

$$Gk = \sum_{i=1,j=i+1}^{i=N-1,j=N} |CTij|\delta(k,Dij)$$

and

$$Jk = \frac{Gk}{(N-1)}$$

where N = number of vertices (atoms different to hydrogen).

Table 5. Discriminant Functions and Overall Classification from the Linear Discriminant Analysis on a Group of Compounds (23 with Antiviral Activity and 238 Inactives)

	data-base group classification				test group ("cross-validation" analysis) classification			
discriminant function	no. of compds	actives	inactives	% success	no. of compds	actives	inactives	% success
$V_1 = 2.785 - 1.171^0\chi^v + 2.108^3\chi_p$	actives: 23	21	2	91.3	actives: 20	18	2	90.0
	inactives: 238	68	170	71.4	inactives: 200	60	140	70.0
$V_2 = -0.486 - 0.737J_1 + 0.196J_1^v + 0.793J_3^v + 0.28^1\chi/^1\chi^v$	actives: 23	15	8	65.2	actives: 20	16	4	80.0
	inactives: 238	15	223	93.7	inactives: 200	18	182	91.0
V_1 and V_2	actives: 23	15	8	65.2	actives: 20	16	4	80.0
	inactives: 238	12	226	95.0	inactives: 200	12	188	94.0

Table 6. Results Obtained by the Discriminant Analysis on Antiviral Drugs[a]

base group				test group			
active compds	classif	inactive compds	classif	active compds	classif	inactive compds	classif
bromovirin	−	butoconazole	−	acycloguanosine	+	dimazole dihydrochloride	−
bucyclovir	+	chlordantoin	−	acyclovir	+	enilconazole	−
descyclovir	+	dermostatin	−	amantadine	−	filipin	−
desoxyacyclovir	+	econazole	−	gancyclovir	+	flucytosine	−
9-(4-hydroxybutyl)guanine	+	exalamide	−	9-(3-hydroxypropoxy)G	+	griseofulvin	−
3HM-HBG	+	fluconazole	+	9-(1,3-dihydroxy-2-propoxymethyl)G	−	carbuterol	−
9-arabinofuranosyl purine	+	epinephrine	−	hidoxuridine	−	clorprenaline	−
methyl galate	−	albuterol	−	ribavirin	+	diphenylorciprenaline	−
trifluridine	−	bitolterol	−	rimantadine	−	dioxethedrine	−
vidarabine	+	butylnorepinephrine	−	arildone	−	ephedrine	−
3′-Az-2′,3′-ddU	+	clenbuterol	−	2′-d-5-hydroxyU	+	theobromine	+
2′-d-5-CF_3U	−	acebutolol	−	2′-d-5-methoxyU	+	befunolol	−
2′-d-5-BrU	−	alprenolol	−	2′-d-5-ethoxyU	+	vevantolol	−
2′-d-5-IU	−	atenolol	−	2′-d-5-(2-propynyloxy)U	+	bopindolol	−
3′-Az-2′,3′-dd-5-(2-propynyloxy)U	+	betaxolol	−	2′-dC	+	bufuralol	−
2′-d-5-cyanomethyloxyU	+	bisoprolol	−	2′-d-5-FC	+	bupranolol	−
2′-d-5-thiocyanoU	−	ASA	−	2′-d-5-MeC	+	ASA	−
3′-Az-2′,3′-dd-5-MethioU	−	alclofenac	−	zidovudine	+	benorylate	−
3′-(3-oxo-1-propenyl)T	+	a,omp[urome	−	cytarabine	+	butibufen	−
3′-Az-3′-dd-6-azaT	+	aniline	−	2′-d-5-FU	+	carprofen	−
2′,3′-ddC	−	apazone	−			diflunisal	−
MC 540	−						
inosine	+						

[a] For inactive compounds, only a representative group is shown. A compound is classified as active (+) if $V_1 > 0$ and $V_2 > 0$, and as inactive (−) in other case.

CTij = mij − mji. "*m*" stands for the elements of the **M** matrix

$$\mathbf{M} = A \times D^*$$

being A = adjacency (N·N) matrix, D^* = inverse square distance matrix, in which their diagonal entries are assigned as 0, and δ = Kronecker's delta. Thus, *Gk* represents the sum of all the *CTij* terms, with *Dij* = *K*, being *Dij* the entries of the topological distance matrix. In the valence C.I. terms, the presence of heteroatoms is taken into account by introducing their electronegativity values (according to Pauling's scale taking chlorine as standard value = 2) in the corresponding entry of the main diagonal of the adjacency matrix.

Since geometrical factors, such as the molecular shape, may condition the pharmacological activity, a simple set of descriptors named "geometrical indices" (G.I.)[4,10] were also introduced.

First, the shape index "*E*" is defined, for alicyclic compounds as

$$E = \frac{\sum_i ni(di + 1)}{L}$$

Where "*ni*" is the number of vertices placed at a "*di*" topological distance from the "main path", being the latter the set of edges joining the two more separated vertices in the graph by the shortest way. "*L*" represents the graph length, i.e., the number of edges constituting the main path. It is clear that the lower the "*E*" value the more elongated the graph. If the graph is assimilated to an ellipse, "*E*" would represent the excentricity.

In the case of cyclic compounds, "E" is calculated by

$$E = \frac{S}{L^2}$$

Where "*S*" is the "surface parameter", calculated as the sum of the "*S*" values for all the rings and alicyclic fragments in the molecule, which, in turn, are obtained as the products

$$S = E \times L^2$$

Thus, for example, the fragment (C—C—C) would contribute as

$$S = 1.5 \times 2^2 = 6$$

The hexagonal ring contributes to the global value as $S_h = E_h{\cdot}L_h^2$, where the subindex "h" means "hexagonal".

Since it can be assumed that a regular hexagonal ring has circular symmetry, we can consider its E_h value as that of the graph

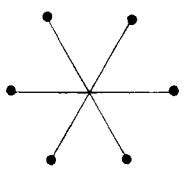
Table 7. Results Obtained from the Discriminant Functions V_1 and V_2 on a Set of Nucleosides and Glycosides (27 with Antiviral Activity (+) and 21 without it(−)). Active Compounds: $V_1 > 0$ and $V_2 > 1.65$[a,b]

compounds with antiviral activity				compounds without antiviral activity			
compd	V_1	V_2	classif	compd	V_1	V_2	classif
bromovirine	0.25	1.67	+	AICAR	3.09	−0.78	−
cytarabine	3.48	2.19	+	AIR	3.36	−1.39	−
idoxuridine	0.36	1.63	−	bucladesine	0.82	−0.88	−
9-arabinopurine	3.82	2.45	+	citicoline	0.43	−2.38	−
ribavirine	3.61	2.28	+	cyclic AMP	4.68	0.00	−
trifluridine	2.90	2.16	+	cyclic GMP	4.71	−0.28	−
vidarabine	3.97	2.03	+	2′-cytidylic acid	2.37	−0.92	−
AzddU	1.65	3.17	+	3′-cytidylic acid	2.34	−1.05	−
d-5-CF_3U	1.90	2.56	+	fludarabine	2.96	1.67	+
d-5-FU	2.22	2.12	+	3′-guanylic acid	2.86	−1.27	−
d-5-BrU	0.34	2.23	+	5′-guanylic acid	3.25	−1.48	−
d-5-IU	−1.36	2.28	−	inosinic acid	3.34	−0.22	−
d-5-OHU	2.14	2.15	+	NMN	1.69	−1.73	−
d-5-OMeU	1.49	2.89	+	nucleocidin	4.10	−1.90	−
d-5-OEtU	0.96	2.96	+	pentostatin	3.50	5.49	+
d-5-(2-propynyloxy)U	0.91	2.33	+	puromycin	−0.07	−0.24	−
Azdd-5-(2-propynyloxy)U	0.83	2.33	+	showdomycin	3.55	0.42	−
d-5-cyanomethyloxyU	1.06	2.81	+	sodium ascorbate	4.01	0.32	−
d-5-SCNU	0.67	1.85	+	L-streptose	4.36	2.15	+
Azdd-5-SMeU	0.79	2.83	+	5′-uridylic acid	2.99	−1.65	−
dC	1.40	3.50	+	xanthosine	4.32	1.72	+
dFC	2.04	2.45	+				
dMeC	1.32	3.41	+				
3-(3-oxopropenyl)T	0.48	3.72	+				
Azd-azaT	1.64	3.26	+				
ddC	1.81	2.90	+				
zidovudine	1.49	3.13	+				

[a] Success on actives: 25/27 = 92.6%; success on inactives: 17/21 = 81%. [b] Abbreviations: d, 2′-deoxy; dd, 2′,3′-dideoxy; Az, 3′-azido; Me, methyl; Et, ethyl; U, uridine; C, cytidine; T, thymidine.

Table 8. Discriminant Functions and Overall Classification from the Linear Discriminant Analysis on a Group of Compounds (190 with Bronchodilator Activity and 170 Inactives)

	data-base group classification				test group ("cross-validation" analysis) classification			
discriminant function	no. of compds	actives	inactives	% success	no. of compds	actives	inactives	% success
$D_1 = -0.305 - 0.1535\ {}^0\chi + 1.515\ {}^4\chi_p{}^v - 1.251\ {}^4\chi_{pc}{}^v$	actives: 190	128	62	67.4	actives: 175	109	66	62.3
	inactives: 170	42	128	75.3	inactives: 165	39	126	76.4
$D_2 = -0.553 - 0.769G_1 + 0.124G_1{}^v + 0.198\ G_2{}^v - 7.44J_3{}^v + 0.575G_4 - 28.677G_4{}^v - 42.913G_5 + 63.573G_5{}^v$	actives: 190	146	44	76.8	actives: 175	129	46	73.7
	inactives: 170	47	123	72.4	inactives: 165	64	101	61.2
D_1 and D_2	actives: 190	105	85	55.3	actives: 175	87	88	49.7
	inactives: 170	17	153	90.0	inactives: 165	14	151	91.5



$$E_h = \frac{(3 \times 1) + (4 \times 2)}{2} = 5.5$$

and therefore

$$S_h = 5{,}5 \times 3^2 = 49{,}5$$

Table 1 illustrates the values for different cyclic and alicyclic fragments.

The rest of the geometrical indices are V3 = number of vertices with valence 3 (double bonds are counted as 2), Tnr = number of "nonramified" terminal vertices (i.e., number of terminal vertices showing valence 1 linked to vertices with valence 2), V4 = number of vertices with valence 4 or higher (double bonds are counted as 2), Pr1 = number of pairs of adjacent (separated by one edge) ramifications, Pr2 = number of pairs of ramifications separated by two edges, and Pr3 = number of pairs of ramifications separated by three edges. In addition, the vertices number (N) as well as the Wiener path number (w) have also been included.

In order to obtain a more descriptive C.I., the resonance effect has to be taken into account. The original procedure for the C.I. calculation results in distinct values for those resonant species showing different electronegativity substituents in the "ortho" or "meta" positions in the aromatic ring. This is topologically correct since both forms are not equivalent. However, due to the existence of an hybrid, according to the classical resonance theory, the aromatic bonds must be assigned a mean value between 1 and 2. The value 1.5 is thus introduced in the entries of the adjacency matrix for each bond. In a future work, we will deal with this interesting subject in detail.

Therefore, each compound is characterized by a set of 62 indices, including connectivity ones up to the fourth order as well as combinations of them, which implies a pretty good chemical and topological description. Table 2 illustrates the values of the whole set of indices (excepting χ_s) for a representative group of drugs.

Table 9. Results Obtained by the Discriminant Analysis on Bronchodilator Drugs[a]

base group				test group			
active compds	classif	inactive compds	classif	active compds	classif	inactiove compds	classif
5,8-difluorotrimetoquinol	+	candicidin D	−	3-butylxanthine	−	chlordantoin	−
5-iodotrimetoquinol	+	dermostatin	−	3-isobutylxanthine	−	fungichromin	−
epinephrine	−	econazole	+	3-methylxanthine	−	naftifine	+
albuterol	−	enilconazole	+	3-hydroxytulobuterol	−	sulbentine	−
amitriptyline	+	fuconazole	−	3-O-methylrimiterol	−	desoxyacylovir	−
azanator	+	griseofulvin	−	AA-497	+	9-(2-hydroxyethoxy)guanine	−
tiotropium	+	miconazole	−	ABC-99	+	2′-deoxy-5-methoxyuridine	−
bambuterol	−	descyclovir	−	AC-119	+	atenolol	−
bitolterol	−	idoxuridine	−	acephylline	−	bisoprolol	−
broxaterol	−	methyl galate	−	bamiphylline	−	bupranolol	−
butylnorepinephrine	−	vidaravine	−	BB-1502	+	epanolol	−
carbuterol	−	inosine	−	caffein	−	levobunolol	−
clenbuterol	−	zidovudine	−	clemastine	+	nadolol	−
chlorpheniramine	+	arotinilol	−	dametralast	−	talinolol	−
clorprenaline	−	betaxolol	−	ethylnorepinephrine	−	sotalol	−
colterol	−	bopindolol	−	ethyltetrahydroharmine	+	xibenolol	−
darodipine	+	bunitrolol	−	phenothiazine	+	acemetacin	−
deptropine	−	butofilolol	−	fenprinast	+	meclofenamic acid	−
diazepine	+	carteolol	−	phentolamine	+	benoxaprofen	−
diphenylorciprenaline	−	ASA	−	flutopium	+	bufexamac	−
diltiazem	+	aminopyrine	−	glycopyrrolate	−	cinmetacin	+
eprozinol	+	aniline	−	hydroxyzyne	+	diclofenac	−
hexoprenaline	+	apazone	−	isobutylmethylxanthine	−	etersalate	−
imazodam	+	butibufen	−	mabuterol	−	fenclofenac	−
medibazine	+	tenoxicam	+	nileprost	−	feprazone	−

[a] Only a representative number of compounds of each group is showed. A compound is classified as active (+) if $D_1 > 0$ and $D_2 > 0$, and as inactive (−) in other case.

Table 10. Discriminant Functions and Overall Classification from the Linear Discriminant Analysis on a Group of Compounds (30 with Antifungal Activity and 98 Inactives)

	data-base group classification				test group ("cross-validation" analysis) classification			
discriminant function	no. of compds	actives	inactives	% success	no. of compds	actives	inactives	% success
$F_1 = -2.092 + 0.131G_1^v$	actives: 30	16	14	53.3	actives: 29	10	19	34.5
	inactives: 98	24	74	75.5	inactives: 100	18	82	82.0
$F_2 = -2.754 + 1.547\,{}^4\chi_p^v +$	actives: 30	20	10	66.6	actives: 29	17	12	58.6
$15.513J_2 - 10.718J_2^v - 0.872V4$	inactives: 98	13	85	86.7	inactives: 100	19	81	81.0
F_1 and F_2	actives: 30	13	17	43.3	actives: 29	9	20	31.0
	inactives: 98	6	92	93.9	inactives: 100	4	96	96.0

Once the indices have been calculated, the discriminant functions are obtained by applying the "linear discriminant analysis" (LDA)[11] to wide groups of compounds including both, with and without activity, taking care to include as much structural heterogeneity as possible in order to generalize the validity of the results. The variables used are chosen in a stepwise manner (stepwise discriminant analysis), so that at each step the variable adding the most to the separation of the groups, is entered into the discriminant function and those variables are removed as necessary to obtain a smaller Wilk's λ parameter than was previously obtained at an earlier step with the same number of variables in the function. (The main conditions were tolerance = 0.010; F to enter = 4.000; F to remove = 3.996.) Finally, a cross-validation analysis is carried out to test the accuracy level of the function.

3. RESULTS AND DISCUSSION

Tables 3 and 4 shows the results obtained for the analgesic discrimination. As may be seen, the equation using χ connectivity indices led to a degree of success of about 94% in the test group; however, about 73% of false actives were included. In addition, the activity probability values (not shown here) were about 50% for both active and inactive compounds. Hence, this discriminant function may not be considered as a good one. By contrast, the general discriminant function, in which the indices G4, $J4^v$, $({}^3\chi_p - {}^3\chi_p^v)$, and V4 appear is much more effective in discriminating the nonanalgesic compounds. This implies a greater accuracy as we may assure the exclusion of false active molecules. Moreover, if both discriminant functions are consecutively used, the efficiency of the general function remains unalterated regarding the active compounds discrimination (68%) but notably increases in that of inactives (up to 84.3%). This result is particularly interesting since it suggests the existence of a complementary character between connectivity and charge (and the other descriptors used) indices.

The case of antiviral (Tables 5 and 6) is still more curious. In fact, a good level of accuracy is obtained with just two χ indices (90% and 70% for actives and inactives in the test group, respectively), while these results are nearly inverted when the C.I. function is used in the test group (80% and 91%, respectively). The combined use of both functions, as occurred with analgesics, led to a sustantive improvement in the inactive discrimination, without lowering that of actives. Once more, a complementary character between the two types of descriptors is pointed out.

Although there is no doubt that discriminant analysis may be considered as an efficient tool, their results must be very

Table 11. Results Obtained by the Discriminant Analysis on Antifungal Drugs[a]

base group				test group			
active compds	classif	inactive compds	classif	active compds	classif	inactive compds	classif
bromosalicylchloranilide	−	methyl galate	−	benzoic acid	−	9-(4-hydroxybutyl)guanine	−
buclosamide	−	arildone	−	amorolfine	−	trifluridine	−
butoconazole	−	2′-deoxy-5-hydroxyuridine	−	amphotericin B	+	2′-deoxy-5-fluorouridine	−
candicidin	+	5-iodotrimetoquinol	+	bifonazole	−	3-butylxanthine	−
chlordantoin	−	6-thiocaffein	−	biphenamine	−	AA-497	−
chlormidazole	−	AH-3021	−	chlorphenesin	−	butylnorepinephrine	−
dermostatin	+	bambuterol	−	ciclopirox	−	dimethylnorepinephrine	−
diamthazole	−	bitolterol	−	cloconazole	−	ethylnorepinephrine	−
econazole	+	carbuterol	−	clotrimazole	−	i[ratrp[oi,	+
enilconazole	−	clenbuterol	−	cloxyquin	−	methylepinephrine	−
exalamide	−	clorprenaline	−	isoconazole	+	norephedrine	−
filipin	+	darodipine	−	itraconazole	+	RS-11635	−
fluconazole	+	diphenylorciprenaline	−	ketoconazole	+	betaxolol	−
flucytosine	−	etafedrine	−	loflucarban	−	bupranolol	−
fungichromin	+	phentolamine	−	mepartricin	+	celiprolol	−
griseofulvin	−	formoterol	−	omoconazole	+	nadoxolol	−
halethazole	−	hydroxyzyne	−	isoconazole nitrate	+	nifelalol	−
hexetidine	−	indoramin	−	pecilocin	−	pronethalol	−
miconazole	+	L-648051	−	perimycin	+	xibenolol	−
naflifine	−	methylpropylxanthine	−	pyrithione	−	alclofenac	−
natamycin	+	nisbuterol	−	pyrrolnitrin	−	benoxaprofen	−
nifuratel	−	arotinilol	+	salicylanilide	−	antrafenine	−
nystatin	+	bucumolol	−	tioconazole	+	bucolome	−
oligomycin D	+	carazolol	−	tolciclate	−	clidanac	−
siccanin	+	esmolol	−	tolindate	−	droxicam	−
sulbentine	−	acetanmilide	−	tolnaflate	−	fenclofenac	−
sulconazole	+	aminopyrine	−	tubercidin	−	glafenine	−
tenonitrozole	−	ammonium salicylate	−	ujothion	−	isonixin	−
terbinafine	−	benorylate	−	viridin	−	naproxen	−
terconazole	+	bufexamac	−			piroxicam	−

[a] For inactive compounds, only a representative group is shown. A compound is classified as active (+) if $F_1 > 0$ and $F_2 > 0$, and as inactive (−) in other case.

Table 12. Discriminant Functions and Overall Classification from the Linear Discriminant Analysis on a Group of Compounds (24 with Hypolipidemic Activity and 158 Inactives)

	data-base group classification				test group ("cross-validation" analysis) classification			
discriminant function	no. of compds	actives	inactives	% success	no. of compds	actives	inactives	% success
$L_1 = -1.243 + 1.146\,{}^3\chi_c^v$	actives: 24	16	8	66.7	actives: 23	12	11	52.2
	inactives: 158	29	129	81.6	inactives: 166	25	141	84.9
$L_2 = -2.456 + 0.793G_1 - 0.321G_2 +$	actives: 24	18	6	75.0	actives: 23	11	12	47.8
$1.095G_3^v - 2.053G_5 - 21.37J_4$	inactives: 158	27	131	82.9	inactives: 166	24	142	85.5
L_1 and L_2	actives: 24	15	9	62.5	actives: 29	11	12	47.8
	inactives: 158	14	144	91.1	inactives: 100	10	156	94.0

carefully handled in order to avoid false accuracy derived, for example, from the structural similarity. As a conclusive proof, two groups of nucleosides, one of them showing antiviral activity and the other not (this last including also other glycosides), were discriminated as test groups. The obtained results are shown in Table 7. In this case the compounds are classified as actives if $V_1 > 0$ and $V_2 > 1.65$. As it may be seen, the discriminant equation obtained by χ terms is unable to distinguish between both groups. On the other hand, the C.I. function is able to correctly classify over 80% among the inactives and more than 90% among the actives (overall accuracy 87.5%). These results are conclusive, clearly demonstrating the excellent discriminant ability of the C.I. as compared to χ, especially in the case of very similar structures.

The discriminant functions obtained for other pharmacological groups, such as bronchodilator, antifungal, antihyperlipoproteinemic, hypoglycemic and betablocker drugs, together with the cross-validation analysis for representative sets of compounds, are shown in Tables 8−17 Tables 9, 11, 13, 15, and 17 corresponding to Tables 8, 10, 12, 14, and 16, respectively, show, for each pharmacological scope, the compounds classification for a representative set of drugs. Tables 18−20 summarize, in the cases of analgesic, antiviral, and bronchodilator drugs, the compounds classification with regard to all the investigated pharmacological activities, in order to easily verify the discriminant efficiency of the function.

The accuracy level is excellent in all cases, although in the group of hypoglycemics the combined use of both functions significantly decreases the discriminant efficiency in the active compounds, which may be solved by using the function H_2 as a single discriminate function. Furthermore, all the equations were significant with respect to stability (within a range of 10%) and randomness, which was to be expected considering both the high number and the heterogeneity of the correlated compounds. Nevertheless, it is important to point out that χ and C.I. indices are not orthogonal descriptors,[12] which implies that different combinations of them may describe the same molecular property, showing consequently a similar degree of discriminant efficiency. In fact, our current results indicate that the

Table 13. Results Obtained by the Discriminant Analysis on Hypolipidemic Drugs[a]

base group				test group			
active compds	classif	inactive compds	classif	active compds	classif	inactive compds	classif
beclobrate	+	enilconazole	−	bezafibrate	+	chlordantoin	+
binifibrate	+	naflifine	−	lovastatin	+	griseofulvin	−
ciprofibrate	−	siccanin	+	clinofibrate	+	miconazole	−
pravastatin	+	terconazole	−	simvastatin	+	sulconazole	−
clofibrate	−	bromovirin	−	clofibric acid	−	descyclovir	−
etofibrate	+	bucyclovir	−	nicotinamide	−	ribavirin	−
fenofibrate	+	inosine	−	gemfibrozil	+	trifluridine	−
nicofibrate	+	epinephrine	−	acipimox	−	acephylline	−
pirifibrate	+	caffeine	−	niceritrol	+	norephedrine	−
ronifibrate	+	ephedrine	−	nicoclonate	−	theobromine	−
simfibrate	+	L-648051	−	theofibrate	+	amosulalol	−
nicomol	+	bisoprolol	−	oxiniacine	−	atenolol	−
acifran	−	xibenolol	+	azacosterol	+	bupranolol	−
benflurex	−	meclofenamic acid	−	benzalbutyramide	−	antrafenine	−
clomestrone	+	aniline	−	eicosapentaenoic acid	−	piroxicam	−
eritadenine	−	bufexamac	−	furazabol	+	phenol	−
meglutol	+	diclofenac	−	melinamide	−	ASA	−
mytatrienediol	+	chlorpropham	−	omithin	−	benfluralin	−
oryzanol	+	emiglite	−	pantethin	+	phenformin	−
pentaerythritol	−	glipizide	−	phenylbutyramide	−	glisoxepid	−
pirozadile	−	tolbutamide	−	probucol	+	tolazamide	−
sitosterol	+	buclosamide	−	sultosilic acid	−	clenbuterol	−
tiadenol	−	cytarabine	−	triparanol	−	salsalate	−
xenbucin	−	albuterol	−				

[a] For inactive compounds only a representative group is shown. A compound is classified as active (+) if $L_1 > 0$ and $L_2 > 0$, and as inactive (−) in other case.

Table 14. Discriminant Functions and Overall Classification from the Linear Discriminant Analysis on a Group of Compounds (16 with Hypoglycemic Activity and 149 Inactives)

	data-base group classification				test group ("cross-validation" analysis) classification			
discriminant function	no. of compds	actives	inactives	% success	no. of compds	actives	inactives	% success
$H_1 = -1.897 - 1.193\ {}^3\chi_c^v + 0.734\ {}^4\chi_p$	actives: 16	7	9	43.8	actives: 16	4	12	25
	inactives: 149	51	98	65.8	inactives: 152	40	1123	73.7
$H_2 = -2.41 - 4.366J_1^v + 10.25J_2 -$	actives: 16	14	2	87.5	actives: 16	12	4	75.0
$26.61J_3 + 57.48J_5^v$	inactives: 149	18	131	87.9	inactives: 152	24	128	84.2
H_1 and H_2	actives: 16	1	15	6.3	actives: 16	1	15	6.3
	inactives: 149	5	144	96.6	inactives: 152	7	145	95.4

Table 15. Results Obtained by the Discriminant Analysis on Hypoglycemic Drugs[a]

base group				test group			
active compds	classif	inactive compds	classif	active compds	classif	inactive compds	classif
ASA	−	terconazole	−	buformin	−	enilconazole	−
acarbose	−	miconazole	−	carbutamide	−	dermostatin	−
acetohexamide	−	sulconazole	−	chlorguanide	−	nifuratel	−
benfluorex	−	inosine	−	chlorpropamide	−	methyl galate	−
dichloroacetic acid	−	cytarabine	−	phenbutamide	+	ribavirin	−
emiglite	−	desoxyacyclovir	−	glibenclyclamide	−	zidovudine	−
fenfluramine	−	ABC-99	−	glibormuride	−	clenbuterol	−
phenformin	−	caffein	−	glybuzole	−	glycopyrrolate	−
gliclazide	−	norephedrine	−	gliquidone	−	ipratropium	−
glymidine	−	levobunolol	−	metformin	−	celiprolol	−
glipentide	−	betaxolol	−	midaglizide	−	butofilolol	−
glipizide	−	nifenalol	−	miglitol	−	bufuralol	−
glisoxepid	+	bufexamal	−	retomoxyl	−	diclofenac	−
imipramine	−	ibuprofen	−	tolazamide	−	bucolome	+
linoglir	−	tiaprofen	−	tolbutamide	−	clidanac	−
methahexamide	−	eicosapentaenoic acid	−	tolcycloamide	−	nicofibrate	−

[a] For inactive compounds only a representative group is shown. A compound is classified as active (+) if $H_1 > 0$ and $H_2 > 0$, and as inactive (−) in other case.

efficiency of the above mentioned functions, may be even improved either including additional ones or by using new topological descriptors.

The discriminant efficiency may also be notably improved by correlating specific properties closely related to the pharmacological mechanism of action. Since these properties usually change within a given range for active compounds, they may be used as discriminant functions. Significant examples of these properties are the inhibition concentrations for cyclooxigenase, DNA-girase, and DNA-polimerase, for analgesics, quinolone-type antibacterials, and herpes antiviral agents, respectively, as well as beta receptor's affinity

Table 16. Discriminant Functions and Overall Classification from the Linear Discriminant Analysis on a Group of Compounds (25 with betablocker Activity and 118 Inactives)

	data-base group classification				test group ("cross-validation" analysis) classification			
discriminant function	no. of compds	actives	inactives	% success	no. of compds	actives	inactives	% success
$B_1 = 7.989 - 3.015\ {}^0\chi^v + 9.104\ {}^2\chi^v - 5.514\ {}^4\chi_p - 25.83\ {}^4\chi_c - 5.54\ ({}^0\chi - {}^0\chi^v) + 16.90\ ({}^2\chi - {}^2\chi^v) - 23.03\ {}^2\chi/{}^2\chi^v + 11.8\ {}^3\chi_p/{}^3\chi_p{}^v - 6.39\ ({}^4\chi_{pc} - {}^4\chi_{pc}{}^v)$	actives: 25	23	2	92.0	actives: 25	18	7	72.0
	inactives: 118	8	110	93.2	inactives: 118	18	100	84.8
$B_2 = -5.3 + 0.657G_1{}^v + 0.263G_2{}^v + 1.294G_3{}^v - 5.693g_4{}^v - 78.17J_3 + 117.50J_4$	actives: 25	22	3	88.0	actives: 25	18	7	72.0
	inactives: 118	12	106	88.8	inactives: 118	16	102	86.4
B_1 and B_2	actives: 25	20	5	80.0	actives: 25	15	10	60.0
	inactives: 118	4	114	96.6	inactives: 118	6	112	94.9

Table 17. Results Obtained by the Discriminant Analysis on Betablocker Drugs[a]

base group				test group			
active compds	classif	inactive compds	classif	active compds	classif	inactive compds	classif
befinolol	+	benorylate	−	acebutolol	+	aminochlorthenoxazin	−
betaxolol	+	benoxaprofen	−	alprenolol	+	2-amino-4-picoline	−
bevantolol	−	alclofenac	−	amosulalol	−	aminopropylon	−
bisoprolol	+	alminoprofen	−	arotinolol	−	aminopyrine	−
bopindolol	+	acetaminophen	−	atenolol	+	amonium salicylate	−
bucumolol	+	acetaminosalol	−	bufetolol	+	antipyrine	−
carazolol	+	acetanilide	−	bufuralol	−	antrafenine	+
carteolol	+	salicylic acid	−	bunitrolol	+	apazone	−
carvedilol	−	azapropazone	−	bupranolol	+	aspirin	−
celiprolol	+	benzoic acid	−	butidrine HCl	−	diplosal acetate	−
cetamolol	+	acabel	−	butofilolol	+	caffeine	−
cloranolol	+	bitolterol	−	dilevalol	−	acefylline	−
levobunolol	+	eprozinol	−	epanolol	+	carbuterol	−
mepindolol	+	etafedrine	−	esmolol	+	albuterol	−
metipranolol	−	etamiphyllin	−	indenolol	+	clenbuterol	−
metoprolol	+	ethylnorepinephrine	−	labetalol	−	clorprenaline	−
moprolol	+	etofylline	−	nadoxolol	−	dioxethedrine	−
nadolol	+	fenoterol	−	nifenalol	−	doxofylline	−
pindolol	+	mabuterol	−	nipradilol	+	dyphylline	−
practolol	+	medibazine	+	oxprenolol	+	enprophylline	−
pronethalol	−	metaproterenol	−	penbutolol	−	ephedrine	−
propranolol	+	methoxyphenamine	−	talinolol	+	epinephrine	−
sotalol	+	*N*-methylephedrine	−	tertatolol	+	flutropium bromide	−
sulfinalol	−	*N*-methylepinephrine	−	timolol	−	formoterol	−
xibenolol	+	oxitropium bromide	−	toliprolol	+	hexoprenaline	−
		phenylpropanolamine HCl	−			indoramin	−
		pinacidil	−			ipatropium bromide	−
		pirbuterol	−			amitriptyline	−
		procaterol	−			isohezarine	−
		protokylol	+			isoproterenol	−
		bamifylline	−			proxyphylline	−
		tolubuterol	−			reproterol	−
		bromovincamine	−			rimiterol	−
		bumadizon	−			soterenol	+
		cytarabine	−			atropine	−
		rolipram	−			terbutaline	−
		diflunisal	−			1-theobromineacetic	−
		etodolac	−			theophylline	−

[a] For the inactive compounds only a representative group is illustrated. A compound is classified as active (+) if B_1 and $B_2 > 0$, and as inactive (−) if B_1 or $B_2 < 0$.

constants in the case of betablocker drugs. For instance, the cyclooxigenase inhibition concentration-50 (μM) by analgesics, in bovine seminal vesicle,[13] fits reasonably well to the equation

$$\log CI_{50} = 0.3181 \cdot G1^v + 6.3413 \cdot J1 - 0.6792 \cdot V4 + 1.4248 \cdot E - 2.2474$$

$$N = 20,\ r = 0.908,\ SE = 0.49,\ F = 17.69,\ p < 0.0001$$

This function values for most analgesics stand between 0 and 3. Since many nonanalgesics possess values within this range and few analgesics show values outside it, this function may be considered as a necessary but not sufficient condition.

Another significant example is that of the mean values (as contrasted with different bacterial strains) for the minimum inhibition concentration MIC (μM) of a heterogeneous set of antibacterial drugs including quinolones, sulfamides, cephalosporins, etc.), which fits to the equation

$$\log MIC = -0.3409 \cdot G1^v + 2.3756 \cdot ({}^0\chi - {}^0\chi^v) - 37.6889 \cdot ({}^0\chi/{}^0\chi^v) + 7.2370 \cdot ({}^2\chi/{}^2\chi^v) - 0.9171 \cdot ({}^3\chi_c - {}^3\chi_c{}^v) + 36.1217$$

$$N = 35,\ r = 0.961,\ SE = 0.36,\ F = 69.72,\ p < 0.0001$$

In this case the optimal range is between −1 and 2. As

Table 18. Comparison of the Results Given for Every Discrimination Applied to Analgesics

	activity ( )						
compd	analgesic	antiviral	bronchodilator	antifungal	hypolipidemic	hypoglycemic	betablocker
2,4-dimethylacetophenone	+	−	−	−	−	−	−
2-amino-4-picoline	+	−	−	−	−	−	−
2-(1-propenyl)phenol	−	−	−	−	−	−	−
ASA	+	−	−	−	−	−	−
acemetacin	+	−	−	−	−	+	−
acetanilide	+	−	−	−	−	−	−
flufenamic acid	−	−	−	−	−	−	−
alclofenac	+	−	−	−	−	−	−
alminoprofen	+	−	−	−	−	−	−
ameclofen	+	−	−	−	−	−	−
mefenamic acid	+	−	−	−	−	−	−
aminopyrine	+	−	−	−	−	−	−
aminopromazine	+	−	−	−	−	−	−
benorylate	+	−	−	−	−	−	−
benoxaprofen	+	−	−	−	−	−	−
cinchophen	+	−	−	−	−	−	−
cinmetacin	+	−	+	−	−	+	−
clopirac	+	−	−	−	−	−	−
1-(*p*-chlorophenyl)propanol	−	−	−	−	−	−	−
chlortenoxicam	+	−	−	−	−	−	−
diclofenac	+	−	−	−	−	−	−
difenpiramide	+	−	−	−	−	−	−
diflunisal	+	−	−	−	−	−	−
droxicam	−	−	−	−	−	−	−
epirizol	+	−	−	−	−	−	−
etersalate	+	−	−	−	−	−	+
ethenzamide	+	−	−	−	−	−	−
glucametacin	−	−	−	−	+	−	−
ibufenac	+	−	−	−	−	−	−
ibuprofen	+	−	−	−	−	−	−
ibuproxam	+	−	−	−	−	−	−
indomethacin	+	−	−	−	−	−	−
indoprofen	+	−	−	−	−	−	−
isonixin	+	−	−	−	−	−	−
isoxicam	−	−	−	−	−	−	−
kebuzone	+	−	−	−	−	+	−
ketoprofen	+	−	−	−	−	−	−
ketorolac	−	−	−	−	−	−	−
proglumetacin	−	−	+	−	+	−	+
ramifenazone	+	−	−	−	−	−	−
salicylic acid	+	−	−	−	−	−	−
methyl salycilate	+	−	−	−	−	−	−
salsalate	+	−	−	−	−	−	−
salverine	−	−	+	−	−	−	−
sulfadiacine	−	−	−	−	−	−	−
sulindac	+	−	−	−	−	−	−
tenoxicam	−	−	+	−	−	−	−
tiaprofenic acid	+	−	−	−	−	−	−
tolmetin	−	+	−	−	−	−	−

occurs with analgesics, although many nonantibacterial compounds show values within this range, few antibacterials show them outside it.

All these results clearly demonstrate the excellent effectiveness of the adequately chosen topological descriptors in the prediction and discrimination of the pharmacological activity on diverse therapeutical scopes. Moreover, our own results on other fields[14] suggest the possibility of application to other biological actions and, even more general, to any kind of structure-based physical or chemical properties.[15,16]

Although these results are very interesting by themselves, it is possible to take a stepwise approach and wonder if the topological descriptors could be applied to new compounds design or particularly to drug design. In fact, our research team has developed an algorithm with this objective;[17] it is based on the use of topological functions in an ***inverse way as compared to the conventional method***, i.e., instead of using them to predict the value/s for a given property/ies shown by a selected compound/s, the functions are used to obtain the compound/s showing a predetermined value for that property/ies. This is possible because, ***in contrast to physicochemical descriptors, topological descriptors derived from the adjacency matrix are not just simple structure-related parameters, but they are an algebraic description of the structure itself.***

The algorithm functions by assembling the molecular fragments in order to build the complete compound, calculating its topological indices and the resulting values for the limitant selected property/ies. The designed compound will be either accepted or rejected, depending on whether these values are within the preestablished range or not. The system, that allows to improve known active compounds as well as design other completely new, works at various levels in such a way that it first selects, within a wide range, the possible basic structures which may be able to show the desired pharmacological action. Later, by successive approximations, the selected structure is refined becoming another one with improved values of the selected properties.

Table 19. Comparison of the Results Given for Every Discrimination Applied to Antivirals

	activity						
compd	analgesic	antiviral	bronchodilator	antifungal	hypolipidemic	hypoglycemic	betablocker
bromovirin	−	−	+	−	−	−	−
buciclovir	−	+	−	−	−	−	−
citarabine	−	+	−	−	−	−	−
desciclovir	−	+	−	−	−	−	−
desoxiaciclovir	−	+	−	−	−	−	−
9-(3-hydroxypropoxy)G	−	+	−	−	−	−	−
DHPG	−	−	−	−	−	−	−
3HM-HBG	−	+	−	−	−	−	−
9-(4-hydroxybutyl)guanine	−	+	−	−	−	−	−
idoxuridine	−	−	−	−	−	−	−
9-arabinosylpurine	−	+	−	−	−	−	−
methyl galate	−	−	−	−	−	−	−
ribavirin	−	+	−	−	−	−	−
trifluridine	−	−	−	−	−	−	−
vidarabine	−	+	−	−	−	−	−
rimantadine	−	−	−	−	+	−	−
foscarnet	−	−	−	−	−	−	−
arildone	−	−	+	−	−	−	−
3′-Az-2′,3′-ddU	−	+	−	−	−	−	−
2′-d-5-CF3U	−	−	−	−	−	−	−
2′-d-5-FU	−	+	−	−	−	−	−
2′-d-5-BrU	−	−	−	−	−	−	−
2′-d-5-IU	−	−	−	−	−	−	−
2′-d-5-hydroxyU	−	+	−	−	−	−	−
2′-d-5-methoxyU Vir26	−	+	−	−	−	−	−
2′-d-5-ethoxyU	−	+	−	−	−	−	−
2′-d-5-(2-propynyloxy)U	−	+	−	−	−	−	−
3′-Az-dd-5-(2-propynyloxy)U	−	+	−	−	−	−	−
2′-d-5-cyanomethyloxyU	−	+	−	−	−	−	−
2′-d-5-thiocyanoU	−	−	−	−	−	−	−
3′-Az-2′,3′-dd-5-methioU	−	−	−	−	−	−	−
2′-dC	−	+	−	−	−	−	−
2′-d-5-FC	−	+	−	−	−	−	−
2′-d-5-MeC	−	+	−	−	−	−	−
3′-(3-oxo-1-propenyl)T	−	+	−	−	−	−	−
3′-Az-3′-dd-6-azaT	−	+	−	−	−	−	−
2′,3′-ddC	−	−	+	−	−	−	−
MC-540	−	−	−	−	−	+	−
inosine	−	+	−	−	−	−	−
zidovudine	−	+	−	−	−	−	−

The results categorically confirm the capacity of the topological descriptors used on drug design. Thus, a group of 20 compounds were selected as theoretical new analgesics, and they were synthesized or commercially obtained. After the experimental assays, a set of 12 showed significant analgesic activity (which means *at least* analgesic potency similar to ASA). Two of the new analgesic compounds (2-(1-propenylphenol) and 2,4-dimethylacetophenone) have been patented.[18] Other active compounds were 3-methylpyridazine, *p*-chlorobenzohydrazide, sulfadiazine (a well-known commercial antibacterial), and 1-(*p*-chlorophenyl)propanol.

In the antiviral design the results were still better. After testing over 12 000 commercial compounds, a selected group of 17 was chosen for experimental tests. The "in vitro" assays, as contrasted to herpes simplex-I virus on cellular cultures, result in the existence of 12 significantly active compounds (which means inhibition levels similar to phoscarnet), such as nitrofurantoine, 1-chloro-2,4-dinitrobenzene, 5-methylcytidine, 1,2,3-triazol-4,5-dicarboxylic acid, cordicepine, nebularine, and inosine.[19] The importance of these results may be better evaluated considering that according to Spink's calculations,[20] the probability of successful random selection is about 1/400 000 for anticonvulsants, and this statistic would be even more unfavorable in the case of antiviral agents. It must be emphasized that among the selected compounds there were several which may be considered as new "lead drugs", such as 2-(1-propenylphenol) and 2,4-dimethylacetophenone as analgesics or nitrofurantoine and 1,2,3-triazol-4,5-dicarboxylic acid as antivirals.

New active compounds have also been obtained as antibacterial, for example, 1-chloro-2,4-dinitrobenzene, 3-chloro-5-nitroindazole, and 3-methyl-1-phenyl-2-pyrazolin-5-one; the first showing also a potent antifungal activity as well as in the field of hypoglycemic agents (for example 4-(4-methyl-5-oxo-2-pyrazolin-1-yl)benzoic acid and 1-(mesitylene-2-sulfonyl)-1*H*-1,2,4-triazol)).[21]

In the group of hyperlipoproteinemics, although no experimental assays have been carried out so far, the cross-validation results (see Tables 12 and 13) allow us to expect a good level of success. Furthermore, several of the more widely used drugs show $^4\chi_c^v$ values higher than 0.1, which is related to the presence of quaternary ramifications containing certain heteroatoms, such as oxygen. The optimal value for this index seems to be 0.102, which allows the prediction of activity for such different compounds as the aryloxyalkanoic acid derivatives (for example clofibrate) or vitamin E. Something similar happens in the bronchodilator group in which a especially interesting case is that of ethyltetrahydroharmine (see Table 20), for which the predicted antiviral and antifungal activities have been confirmed in literature.

Table 20. Comparison of the Results Given for Every discimination Applied to Bronchodilators

	activity						
compd	analgesic	antiviral	bronchodilator	antifungal	hypolipidemic	hypoglycemic	betablocker
acephylline	+	+	−	−	−	−	−
acetyltetrahydroharmine	+	−	+	−	−	−	−
albuterol	−	−	−	−	−	−	−
amitriptyline	−	−	+	−	−	−	−
azanator	+	−	+	−	−	+	−
butylnorepinephrine	−	−	−	−	−	−	−
caffein	+	+	−	−	−	−	−
carbuterol	−	−	−	−	−	−	−
clemastine	−	−	+	−	−	−	−
clembuterol	−	−	−	−	−	−	−
clorprenaline	−	−	−	−	−	−	−
deptropine	−	−	+	−	−	−	−
diphenylorciprenaline	−	−	−	−	−	−	−
diltiazem	−	−	+	−	−	−	−
dimethylnorepinephrine	−	−	−	−	−	−	−
dioxethedrine	−	−	−	−	−	−	−
dioxyephedrine	−	−	−	−	−	−	−
epinephrine	−	−	−	−	−	−	−
eprozinol	−	−	+	+	−	−	−
etafedrine	−	−	−	−	−	−	−
ethylnorepinephrine	−	−	−	−	−	−	−
ethyltetrahydroharmine	−	+	+	+	−	−	−
fenoterol	−	−	−	−	−	−	−
phenothiazine	−	−	+	+	−	−	−
fenprinast	−	−	+	−	+	−	−
phentolamine	+	−	+	−	−	−	−
flutropium	−	−	+	+	+	−	−
formoterol	−	−	−	−	−	−	−
glycopyrrolate	−	−	−	−	+	−	−
hexoprenaline	−	−	+	−	−	−	−
hydroxyzyne	−	−	+	+	−	−	+
isobutylmethylxanthine	+	−	−	−	−	−	−
imazodam	+	−	+	−	−	−	−
indoramin	−	−	−	−	−	−	−
ipratropium	−	−	−	+	+	−	−
isoetharine	−	−	−	−	−	−	−
isoproterenol	−	−	−	−	−	−	−
mabuterol	−	−	−	−	−	−	−
methoxyphenamine	−	−	−	−	−	−	−
methylpropylxanthine	+	+	−	−	−	−	−
nileprost	−	−	−	+	−	−	−
nisbuterol	−	−	−	−	+	−	+
norephedrine	−	−	−	−	−	−	−
propyltetrahyhdroharmine	−	−	+	+	−	−	−
reproterol	−	−	−	−	−	−	−
rimiterol	−	−	+	−	−	−	−
theobromine	+	+	−	−	−	−	−
theophylline	+	+	−	−	−	−	−
tetrahydroharmine	−	+	+	−	−	−	−

4. CONCLUSIONS

This paper clearly demonstrates that by an adequate choice of topological descriptors it is possible to not only predict different pharmacological activities but also to design new active compounds (including "lead drugs") in several therapeutical scopes, with a surprising level of efficiency, especially considering the simplicity of the calculations.

These results suggest that any pharmacological activity must be related to common basic mechanisms: the fit to its molecular "target", followed by a charge-transfer process. The first step would be taken into account by connectivity and G.I. indices, and the second by the topological charge indices. This hypothesis would be in accordance with the classical receptor theory[22] as well as with LFERs approach.[23] The fact that the compounds chosen were taken from varied pharmacological groups with different mechanisms of action as well as the complementary character shown by connectivity and topological charge indices, reinforces this hypothesis.

In any case, considering the well-known difficulty for the description of the molecular pharmacological mechanisms, our results demonstrate, without any doubt, that molecular topology, in spite of its limitations, ought to be considered not just as an excellent tool for molecular and drug design but as a ***real alternative approach to the study of chemical bonds, whose theoretical physicochemical basis is still to be developed***. Thus, molecular topology offers the possibility to describe chemical bonds from very simple algebraical notions.

ACKNOWLEDGMENT

This study has been supported by CICYT, SAF92-0684 (Spanish Ministery of Science and Education). The authors are indebted to Jose Luis Temprado and Inmaculada Rios for their technical assistance and Dr. Salabert and Dr. Muñoz for their help in experimental tests.

REFERENCES AND NOTES

(1) Gálvez, J.; Garcia, R.; de Julian-Ortiz, V. J.; Soler, R. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1198−1203.

(2) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976; pp 46−61.

(3) Gálvez, J.; Garcia, R.; Salabert, M. T.; Soler, R. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 520−525.

(4) Garcia, R.; Gálvez, J.; Moliner, R.; Garcia, F. *Drug Invest.* **1991**, *3*(5), 344−350.

(5) Franke, R. *Theoretical Drug Design Methods*; Elsevier: Amsterdam, Oxford, New York, Tokyo, 1984; Pharmacochemistry Library, Vol. 7, pp 25−28 and 256−265.

(6) Richards, W. G. *Quantum Pharmacology*; Butterworth: London, 1977.

(7) Burket, U.; Allinger, N. L. *Molecular Mechanics*; ACS Monograph 177; American Chemical Society: Washington, DC, 1982.

(8) Nilsson, N. J. *Learnig Machines*; McGraw-Hill, New York, 1965.

(9) Cover, T. M.; Hart, P. E. *IEEE Trans. Inf. Theory* **1967**, IT-13, 21.

(10) Moliner, R.; Garcia, F.; Gálvez, J.; Garcia, R.; Serrano, C. *An. R. Acad. Farm.* **1991**, *57*, 287−298.

(11) Collous, T. *Discriminant Analysis and Applications*; Academic Press: New York, 1973.

(12) Randic, M. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 311−320.

(13) Garcia, F. *Diseño por topologia molecular, síntesis y ensayos farmacológicos de nuevos fármacos analgésicos*; Doctoral Thesis, Faculty of Pharmacy, University of Valencia, Spain, 1991; p 67.

(14) Garcia, R.; Gálvez, J.; Anton, G. M.; Mari, J. I. *Rev. Toxicol.* **1992**, *9*, 15−17.

(15) Soler, R.; Garcia, F.; Antón, G. M. et al. *J. Chromatogr.* **1992**, *607*, 91−95.

(16) Soler, R.; Garcia, F.; Gálvez, J.; Garcia, R. *An. Quim.* **1992**, *88*, 382−389.

(17) Gálvez, J.; Garcia, R.; Bernal, J. M.; Garcia, F. *An. R. Acad. Farm.* **1991**, *57*, 533−546.

(18) Gálvez, J.; Garcia, R.; Moliner, R.; Garcia, F. Spanish patents Ns.P9101034 and P9101134, 1991.

(19) Muñoz, C.; Julian-Ortiz, J. V.; Gimeno, C.; Catalán, V.; Gálvez, J. *Rev. Esp. Quimioter* **1994**, *7*, 279−280.

(20) Spinks, A. *Chem. Industry* **1973**, *12*, 885.

(21) Antón, G. M.; Garcia, R.; Pérez, F. et al. *Arzneim.-Forsch./Drug Res.* **1994**, *44*(II), 821−826.

(22) Franke, R. *Theoretical Drug Design Methods*; Elsevier: Amsterdam, Oxford, New York, Tokyo, 1984, Pharmacochemistry Library, Vol. 7, pp 316−322 and 97.

(23) Hetnarski, B.; O'Brien, R. D. *J. Med. Chem.* **1975**, *18*, 29.

CI940106B