Getting Discriminant Functions of Antibacterial Activity from Physicochemical and Topological Parameters

Rama K. Mishra[†]

Department of Chemistry, Sambalpur University, Jyoti-Vihar-768 019 India

R. Garcia-Domenech and J. Galvez*

Drug Design & Molecular Connectivity Research Unit, Department of Physical Chemistry, Faculty of Pharmacy, University of Valencia, Avenida V.A. Estellés s.n. 46100-Burjasot, Valencia, Spain

Received September 8, 2000

Linear discriminant analysis has been demonstrated to be a very useful tool in the selection and design of new drugs. Up to now we have used it through the search of a topological pattern of activity. In this work our goal is to calculate a complete set of physicochemical parameters using semiempirical (quantum chemical) calculations as well as topological indices (TIs) and try to find out any discriminant function for antibacterial activity through the combined use of both types of descriptors. The physicochemical parameters, such as heat of formation, HOMO, LUMO, dipole moment, polarizability, hyperpolarizability, PM3 generated IR vibrational frequencies, etc., were calculated using PM3 Hamiltonian implemented within the MOPAC97 package. Among the TIs, connectivity as well as topological charge indices stands as the most representatives. The obtained results suggest that one of the maxima and minima vibrational frequencies play an important role in the antibacterial activity. These frequencies are associated with the torsional molecular vibration (N3) and the stretching vibration (N5) of X−H groups (X = C, N, O). Furthermore, the differences between the maxima and minima values showed an even better discriminant ability than the values themselves. The additional use of the topological indices provided a clear improvement in the discriminant function and also provided a straightforward way to predict the values of such frequencies, so that the results can be applied to a large set of compounds searching for new candidates as antibacterials

INTRODUCTION

The search for drugs showing adequate biological characteristics making them able to later be used in pharmacological development with therapeutical purposes is a goal in which many research groups are involved nowadays. Despite the significant advances in this field, there are still many diseases, which do not yet have an effective therapy, for instance the parasitosis (malaria, Chagas, etc.), fungii diseases (infections caused by *Candida albicans*), viral (AIDS), etc. Particularly, the infectious diseases have become the main target since the resistances to antibiotics were detected, as for instance resistance from *Mycobacterium avium* and *tuberculosis*. This is why we have selected this group for developing the present study.

Although it is possible to take advantage of the enormous chemical information that exists about compounds synthesized and/or isolated (its number estimated to be more than few millions) or from those contained in virtual libraries derived from the use of combinatorial chemistry (in limitless number), it seems necessary to get such information by using methodologies able to select, in a short time, those potential candidates showing the best characteristics for further development. Nowadays, different methods are available for such a goal, all of them based upon the relationships between the chemical structure and the properties (physical, chemical, and biological) of molecules. Several kinds of formalisms have been used in these studies, such as molecular mechanics,[1] quantum chemical descriptors,[2] similarity/dissimilarity approaches,[3] topological descriptors,[4−7] 3D-QSAR,[8,9] and so forth.

The *Molecular Connectivity and Drug Design Research Unit* has demonstrated the efficiency of the topological approaches in the selection and design of new drugs.[10−13] In this mathematical formalism a molecule is assimilated to a graph, where each vertex represents one atom and each axis one bond. Starting from the interconnections among the different vertexes, an adjacency topological matrix can be built up whose elements *ij*, take the values either 1 or 0, whether the vertex *i* is connected to the vertex *j* or not, respectively. The manipulation of this matrix gives origin to a set of topological indices or descriptors that characterize each graph and can be used to carry out QSPR[14−16] and QSAR[17−19] studies. However, one point seems to be critical in this matter: the selection of the adequate set of topological indices. This is not an easy task since numerous TIs are described in the literature. A good choice of using indices related to well-defined physicochemical magnitudes.[20,21]

The objective of this work is to test the usefulness of the joint use of specific physicochemical parameters obtained

* Corresponding author phone: +34-96-386-4891; fax: +34-96-386-4892; e-mail: jgalvez@uv.es.

[†] Present address: Laboratory of Molecular Modeling and Design, Department of Medicinal, Chemistry & Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, IL 60612.

10.1021/ci000303c CCC: $20.00
Published on Web 02/20/2001

from quantum chemical calculations as well as TIs. Various physicochemical parameters such as heat of formation, polarizability, hyperpolarizability, vibrational frequencies, etc. have been used jointly with connectivity,[22,23] topological charge indices,[24] and geometrical indices[25] in order to get solid models able to predict the antibacterial activity.

MATERIAL AND METHODS

Compounds Selection. To carry out the linear discriminant analysis a set of 59 compounds (24 active and 35 inactive) classified by the Merck Index[26] within different therapeutic categories, including both antibacterial and nonantibacterial drugs, were selected. The set was very carefully chosen in order to include as much structural heterogeneity as possible. In the case of antibacterials we had tried to take into account drugs acting through different mechanisms of action, including antibiotics, chemotherapics, bacteriostatics, and bactericides.

Physicochemical Parameters. To generate the physicochemical parameters, we had made use of the MOPAC97 package.[27] All the calculations were performed using standard PM3 Hamiltonian implemented in MOPAC97. Molecular geometries were completely optimized by employing BFGS algorithm[28] within a given point group. In fact, we had started our calculations with one initial geometry, considering all the molecules to have closed-shell singlet and setting the SCF convergence criteria to be 10^{-24}. Further, making use of the keyword FORCE the optimization of all the active and inactive sets of molecules had been performed to ensure the presence of any imaginary vibrational frequency. The standard heat of formation (ΔH_f), total energy (T = sum of the electronic and the core−core repulsion energy), HOMO (Ho), LUMO (Lu), dipole moment (D), molecular length (L_1), surface area (S), volume (V), the first three minimal vibrational frequencies (N1, N2, and N3), and the last three maximal vibrational frequencies (N4, N5, and N6) in the PM3 generated IR spectra were obtained from the above semiempiraical calculations. Again the optimized geometries of all the molecules were used in the time dependent coupled perturbed Hartree−Fock (TDCPHF) method[29] in order to generate the average of the linear polarizability (α) and the average of the first- and the second-order hyperpolarizabilities (β and γ) in the static field (zero frequency).

Topological Indices. As pointed out above, connectivity indices[22,23] as well as topological charge indices[24] and geometrical indices[25] were included in this study. All these descriptors have in common that their determination has been carried out from the adjacency topological matrix obtained from the graph of depleted hydrogens.

Linear Discriminant Analysis. The objective of the linear discriminant analysis, LDA, which is considered as a *heuristic* algorithm able to distinguish between two or more categories or objects, is to find a linear function to discriminate between the active and inactive compounds as for their different descriptor's values. Two sets of compounds: the first one with a proven pharmacological activity (in our case, antibacterial) and the second one composed of inactive compounds were considered for the analysis. The discriminant ability was tested by the percentage of correct classifications in each group. In addition, a "cross-validation" test, in which each case is eliminated from the data set and then the regression analysis was carried out again with the $N-1$ remaining cases, was being predicted later than the value for the case eliminated. The procedure was repeated as many times as there are cases in the data. From the residuals obtained, the standard error of estimate $SE_{(CV)}$ had been determined for the cross-validation.

Table 1. Symbols and Definitions for Quantum and Topological Indices

index symbol	definition	refs
ΔH_f	heat of formation at standard state (kcal/mol)	27
T	total energy (eV)	27
HO	highest occupied molecular orbital (eV)	27
LU	lowest unoccupied molecular orbital (eV)	27
D	dipole moment (debyes)	27
L1	molecular length (A°)	27
S	molecular surface area (A°^{2})	27
V	molecular volume (A°^{3})	27
N1−N3	the three minimum IR frequencies (cm^{-1})	27
N4−N6	the three maximum IR frequencies (cm^{-1})	27
a	the average linear polarizability (AU)	29
b	the average first hyperpolarizability (AU)	29
g	the average second hyperpolarizability (AU)	29
$^m\chi_p$	path connectivity index of order $m = 0-4$	22
$^m\chi^v_p$	path valence connectivity index or order $m = 0-4$	22
$^m\chi_c$	cluster connectivity index of order $m = 0-4$	22
$^m\chi^v_c$	cluster valence connectivity index or order $m = 0-4$	22
$^m\chi_{pc}$	path-cluster connectivity index of order $m = 0-4$	22
$^m\chi^v_{pc}$	path-cluster valence connectivity index or order $m = 0-4$	22
G_m	charge index or order $m = 0-5$	24
J_m	bound charge index of order $m = 0-5$	24
G^v_m	valence charge index or order $m = 0-5$	24
J^v_m	valence bound charge index of order $m = 0-5$	24
PR_i	number of pairs of ramifications separated by i edges	24
W	Wiener index	24
L	graph length	24

Furthermore, a test set was included in such an order to check the validity of the selected discriminant functions.

LDA was performed by using the BMDP 7M package.[30] The selection of the descriptors was based on the F-Snedecor parameter, and the classification criterion was the shortest Mahalanobis distance (distance of each case to the mean of all cases used in the regression equation). 7M chooses the variables used in computing the linear classification functions in a stepwise manner: at each step the variable that adds the most to the separation of the groups is entered into (or the variable that adds the least is removed from) the discriminant function. The quality of the discriminant function is evaluated by the parameter Wilk's lambda or U-statistic, which is a multivariate analysis of a variance statistic that tests the equality of group means for the variable(s) in the discriminant function.

Multilinear Regression. Since, as outlined later, the only QDs which demonstrated to play an important role in the antibacterial activity were N5 and N3, which are the second highest (N5) and the third minimal (N3) frequencies, multilinear regression analysis was used to obtain the connectivity function relating these QDs with TIs. The Furnival-Wilson algorithm[31] was followed to find subsets of descriptors, and it was selected initially for the equations according to the criterion of minimum value of the Mallow's Cp parameter.[32]

The utility of expressing the N5 and N3 as a function of the TIs is to the extent that their predictive capability is for larger sets of compounds for which the calculation of TIs is much faster and easier than the calculation of QDs.

Table 2. Values of the Optical Properties with the Quantum Descriptors

compound	ΔH_f	T	HO	LU	D	L1	S	V	N1
				Active Group					
acedapsone	−95.04	−3826.8	9.39	0.61	6.16	13.77	107.5	520.5	15.3
amoxicillin	−152.55	−4329.7	9.38	0.4	5.1	12.75	90.4	491.8	14.2
azidamfenicol	−38.72	−3778.2	10.41	1.41	4.31	11.11	106.4	450.2	12.6
brodimoprim	−17.72	−3360.7	8.76	0.09	3.56	7.9	63.4	375.5	13.4
carbenicillin	−197	−4564.3	9.6	0.48	3.8	11.6	107.6	599.6	13
cefazolin	49.09	−4900.3	9.43	1.5	6.32	17.41	129.0	508.3	10.3
cefotaxime	−141.73	−5393.8	9.29	1	4.82	14.73	142.4	739.3	9.8
ceftriaxone	−61.8	−6316.8	9.23	1.49	7.37	19.05	177.5	1352	4.3
cefuroxime	−174.66	−5322.9	9.42	0.82	5.08	14.83	149.2	1024	9.7
chloramphenicol	−123.56	−3912	10.16	1.51	4.68	11.75	66.5	329.2	13.6
cloxacillin	−84.22	−4983.8	9.77	0.81	3.23	12.3	141.5	729.9	5.5
demeclocyclin	−287.08	−5802.8	9.2	1.21	0.91	8.71	86.2	444.1	20.5
dicloxacillin	−88.79	−5285.1	9.8	0.85	2.78	12.37	100.2	747.5	7.6
doxycycline	−280.44	−5650.7	9.33	1	3.18	12.56	117.9	711.2	22.2
methicin	−181.57	−4594.3	9.54	0.39	6.48	12.91	108.4	636.6	6.3
minocycline	−243.83	−5684.3	9.2	0.95	3.35	13.14	155.6	1008	18.4
oxytetracycline	−318.12	−5944.2	9.23	1.13	4.09	12.68	118.6	718.5	22.6
rifampin	−375.44	−10160	8.31	1.09	3.72	19.76	243.0	1942	12.1
rifaximin	−339.81	−9627.6	7.86	1.03	2.58	17.39	253.7	2151	18
sisomicin	−278.82	−5689.6	9.14	−0.72	1.15	15.84	139.7	663.6	8.4
streptomycin	−448.32	−7749.6	9.26	0.43	4.1	13.32	177.4	1447	6.6
sulfoxon	−160.53	−4548.8	9.17	0.81	4.59	13.26	149.8	800.1	10.5
tetracycline	−285.7	−5650.9	9.41	1.05	3.17	12.49	123.4	655.3	22.1
tetroxoprim	−99.82	−4057.6	8.74	0.07	3.04	15.38	132.3	769.8	12.1

compound	N2	N3	N4	N5	N6	α	β	γ
				Active Group				
acedapsone	19.3	24.9	3177.4	3364.6	3364.6	181.8	131.1	51139
amoxicillin	19	30.1	3527.2	3851.1	3886	169.9	177.6	24469
azidamfenicol	21.1	24.8	3357	3857	3900.5	140	174.6	23410.8
brodimoprim	15.3	33.9	3431.5	3525.9	3534.7	164.9	239.9	27491.5
carbenicillin	18.9	26.1	3347.6	3847.2	3848.2	172.5	81	23312.9
cefazolin	14.4	19.3	3163.4	3366.5	3847.2	230.7	1248.6	85935.1
cefotaxime	10.5	18.1	3428.2	3534.6	3848.6	221.9	245.3	50649.1
ceftriaxone	7.4	11.1	3424.2	3533.1	3888	287.8	1208.6	101904.1
cefuroxime	13	18.1	3399	3514.7	3849.4	204.4	293.2	51283.5
chloramphenicol	17.7	23.9	3354.3	3880.6	3900	131.7	222.5	23052
cloxacillin	11.3	16.9	3170.3	3338.1	3875	203.6	179.7	34176.8
demeclocyclin	26.3	36.4	3733.9	3872.4	3880.2	223.2	481.5	50350.5
dicloxacillin	9.7	15.1	3170.5	3329.6	3875.2	212.3	245.8	38594.6
doxycycline	35.8	41	3729.9	3857.9	3880.8	220	187.8	42813.9
methicin	13	20.5	3170.6	3372.7	3874.1	0	0	0
minocycline	25.2	32.9	3688.4	3738.7	3880.4	223.8	230.8	48951
oxytetracycline	36.8	39	3813.6	3845.7	3880.8	224	216.9	42872.7
rifampin	16.9	22.4	3867.5	3873.9	3893.3	434.1	415.2	109886.2
rifaximin	23.2	25.8	3857.5	3881.5	3893.2	432	861.9	143172
sisomicin	11	20.5	3867.6	3876.1	3895	192.9	122.6	21390.4
streptomycin	11.1	14.3	3881.3	3884.6	3894.9	238.2	240.3	28426
sulfoxon	17.6	18.7	3369.9	3855.1	3856.1	222	383.2	80950.5
tetracycline	33.5	37.5	3736.1	3852.2	3880.6	220.6	203.7	42443.2
tetroxoprim	17.1	19.5	3431	3526.1	3534.6	186.7	257.7	31989.6

compound	ΔH_f	T	HO	LU	D	L1	S	V	N1
				Inactive Group					
acetohexamide	−61.71	−3768	9.85	1.32	3.83	14.68	112.0	545.5	10
aminopyrine	20.61	−2581.6	9	0.21	3.06	10.37	64.1	344.1	22.4
antrafenine	−287.37	−7834.3	8.84	1.24	2.84	15.81	160.0	1348.8	9.6
apazone	−25.05	−3439.9	8.72	0.72	0.52	12.17	137.2	629.5	23.1
azacosterol	−88.85	−4261.5	8.75	−1.02	2.59	16.33	97.3	539.2	23.9
benfluorex	−200.66	−4630.9	9.3	0.55	3.72	11.42	100.3	613.6	9.5
benorylate	−146.57	−3875.4	9.21	0.72	7.32	13.62	89.2	336.3	17
benoxaprofen	−62.71	−3439.7	9.29	1.24	3.92	13.79	76.7	348.1	10.1
benzpiperylon	42.68	−3771	8.89	0.61	3.13	13.45	145.1	744.5	23.1
benzydamine	43.07	−3384.4	8.66	0.39	2.3	11.29	98.1	575.9	23.7
bucloxic acid	−135.16	−3387.7	9.81	0.88	5.4	12.87	72.6	309.2	16.7
clemizole	55.76	−3362.4	8.79	0.38	3.93	11.11	122.5	526.9	24.6
clofibric acid	−114.24	−2552.2	9.28	0.18	2.67	9.87	31.3	112.0	12.9
chlorothen	56.65	−2924.8	8.65	0.68	1.38	10.64	105.3	467.7	13.2
clorpropamide	−99.09	−3093.6	10.0	1.14	6.82	12.47	64.8	333.3	15.7
doxylamine	18.7	−2970.3	9.13	0.16	2.54	11	111.3	525.4	28
eicosapentaenoic acid	−66.76	−3419.5	9.66	−0.71	4.26	11.11	110.2	774.8	9.3
pheniramine	44.95	−2528.0	9.02	0.06	2.39	10.5	104.2	494.8	30.8

Table 2 (Continued)

compound	ΔH_f	T	HO	LU	D	L1	S	V	N1
				Inactive Group					
glibornuride	−143.49	−4219.5	9.86	0.78	5.31	12.81	100.4	514.2	11.9
glybuzole	14.07	−3097.1	9.41	1.22	7.58	13.02	77.3	356.5	2.5
gliclazide	−81.31	−3654.8	9.4	0.77	4.7	13.5	102.7	574.3	14.9
glymide	−75.97	−3617.2	8.91	0.83	4.92	13.89	90.4	393.3	8.9
glipizide	−110.43	−5076.8	9.92	0.93	6.99	20.96	139.8	823.4	10.4
indomethacin	−111.21	−4151	8.85	0.9	3.93	14.07	108.2	468.5	15.1
miglitol	−240.35	−2335.2	11.2	−0.52	4.36	7.15	29.0	101.9	27.7
nicomol	−221.7	−7811.4	10.2	1.14	3.65	15.93	241.0	1294.3	13.4
nicotine	19.99	−1725.2	9.13	0.07	2.78	8.29	42.9	153.4	47.7
methapyrilene	56.31	−2951.1	7.76	0.09	3.13	7.26	57.5	296.7	20.9
probucol	−108.81	−5405.7	8.5	0.47	3.97	15.22	149.9	1170.9	8.4
salicylic acid	−108.74	−1803	9.79	0.49	5.69	6.23	32.8	74.5	54.1
β-sitosterol	−122.48	−4504.4	9.52	−1	1.61	16.8	116.3	546.4	24.6
tolbutamide	−108.22	−3091.5	9.69	0.5	5.93	12.76	74.9	318.3	15.1
tripelennamine	51.19	−2705.6	8.41	−0.29	2.76	9.34	84.2	429.2	25.9
triprolidine	52.79	−2914.2	9.02	0.18	2.51	11.47	120.1	601.7	20.7
dibekacin	−334.19	−5865.3	9.3	−1.57	3.63	14.02	141.0	809.6	6.3

compound	N2	N3	N4	N5	N6	α	β	γ
				Inactive Group				
acetohexamide	17.6	21.3	3175.4	3355.1	3356.7	170.7	97.3	27390
aminopyrine	31.5	41.1	3151.2	3136.6	3134.7	131.1	185.5	21335.4
antrafenine	15	15.1	3068.5	3073.2	3325	287.5	580.9	57454.2
apazone	28.7	52.6	3135.5	3170.1	3181.1	170.6	187.5	48011.9
azacosterol	34.3	41.2	3153.8	3160.7	3879.7	192.4	100.7	16621.7
benfluorex	17.9	21.3	3082.2	3177.3	3360	162.5	142.4	18241.8
benorylate	22.6	28.5	3167.6	3170.3	3353.8	170.3	347.3	40426.6
benoxaprofen	18.8	31.8	3088.1	3178.4	3877.5	174.9	1071.5	128712.
benzpiperylon	28.6	35.1	3115.8	3133	3388.2	199.8	128.9	33328.4
benzydamine	25.9	34.9	3079.1	3128.7	3132.7	174.6	236.6	23888.7
bucloxic acid	19.9	30.8	3045.7	3051.8	3883.5	139.3	380.2	23360.9
clemizole	32.3	34.8	3067.5	3072.1	3080.9	185.1	474.6	37275.7
clofibric acid	41.5	50.1	3174.8	3177.1	3849.4	98.9	308.5	18449.3
chlorothen	25.3	26.7	3116.8	3131.2	3132.9	156.8	530.4	43534.1
clorpropamide	24.1	35.3	3180.7	3357.8	3360	132.5	516.5	36884.2
doxylamine	30.6	44.4	3133.5	3150.8	3174.6	151.3	65.1	14641.1
eicosapentaenoic acid	22	26.1	3086.1	3183.1	3885.4	161.5	54.3	14578.3
pheniramine	31.9	38.4	3083.5	3117.9	3132.9	171.4	123.0	40736.0
glibornuride	20.1	41.5	3337.8	3359.9	3826.7	175.7	19.4	21992.4
glybuzole	15.6	25.6	3177.8	3178.1	3381.1	160.1	503.1	29377.8
gliclazide	23.7	33.9	3169.9	3326.8	3346	159.7	105	25776.8
glymide	16.1	19.4	3074.2	3143.9	3357.3	165.6	507.2	30351.5
glipizide	13	18.5	3342	3344	3352	229.1	173.1	39324.2
indomethacin	23	32.2	3135	3145	3848.9	201.6	478.3	66229.9
miglitol	38.5	68.1	3886.7	3866.8	3875.3	57.6	50.4	3964.6
nicomol	19.3	21.1	3076.7	3094.6	3780.1	323.2	77.6	44062.3
nicotine	77.8	106	3064.7	3078	3120.3	86.5	30.9	6712.7
methapyrilene	28.1	39.1	3066.1	3076.3	3076.9	171.4	123	40736
probucol	14.4	37.7	3182.7	3822.2	3870.2	281.6	256	63212.7
salicylic acid	111.8	188.5	3074.6	3883.3	3889.9	65.7	145.4	5374.1
β-sitosterol	34.7	42.1	3164.1	3166.6	3868.7	207.2	16	16677.1
tolbutamide	17.6	26.5	3175.5	3315.4	3332.1	136.7	23.5	24199.2
tripelennamine	34.8	45	3086.9	3118.2	3128.7	149.7	558.8	21494.3
triprolidine	22	24.1	3077.4	3083.6	3170.6	165.6	232.4	29013.4
dibekacin	13.3	18.8	3865.4	3873.9	3882.4	186.0	57.6	20632.7

RESULTS AND DISCUSSION

Tables 1 and 2 illustrate the names of all the drugs used in this study as well as their values of QDs and TIs. From these values, a preliminary search of the descriptors showing the highest discriminant efficiency was carried out. The results obtained are illustrated in Table 3, in which the values of the F function for each one of the variables is outlined.

From these results it is clear that once the two QDs, N5 and N3, have been removed no more significant QDs are found, and it can be concluded that both the above QD parameters play an important role in the antibacterial activity. Moreover, the results obtained including all the descriptors (TDs and QDs) (not outlined here) demonstrate that both QDs are the most significant ones.

The discriminant functions selected were as follows:

$$DF1 = (0.08*N5) - 29.1; \quad N = 59;$$
$$F = 64.138 \text{ U-statistic (Wilks' lambda)} = 0.4705$$

(Active set: 19 molecules are in the training set and five molecules are in the test set; inactive set: 23 are in the training set and 12 are in the test set.)

$$DF2 = (0.014*N5) - (0.103*N3) - 48.2; \quad N = 59;$$
$$F = 54.197 \text{ U-statistic (Wilks' lambda)} = 0.3406$$

Table 3. Values of the *F* Parameter for (A) Step 0: Including All the QDs, (B) Step 1: after Extracting N5, and (C) Step 2: after Extracting N5 and N3

		F parameter		
no.	descriptor	step 0	step 1	step 2
1	ΔH_f	19.05	0.05	0.82
2	*T*	22.05	6.84	0.35
3	HO	0.09	1.16	0.85
4	LU	2.33	1.87	0.49
5	D	0.12	0.01	0.05
6	L1	2.65	3.21	0.30
7	S	8.35	5.79	1.06
8	V	9.29	4.91	0.00
9	N1	7.43	10.94	0.58
10	N2	6.83	15.49	1.56
11	N3	6.46	21.35	
12	N4	47.09	1.85	0.00
13	N5	64.14		
14	N6	30.00	2.63	1.65
15	α	5.99	2.13	0.10
16	*β*	0.84	2.73	0.91
17	*γ*	4.86	2.79	0.23

(Active set: 16 are in the training set and eight are in the test set; inactive set: 27 are in the training set and eight are in the test set.)

$$DF3 = 0.01*(N5-N3) - 33.90; \quad N = 59;$$
$$F = 48.427 \text{ U-statistic (Wilks' lambda)} = 0.45235$$

(Active set: 19 are in the training set and five are in the test set; inactive set: 23 are in the training set and 12 are in the test set.)

Table 4 shows the classification matrix for the complete set of compounds as well as the overall accuracy, for the three DFs. It may be mentioned that BMDP 7M takes care of the number of compounds to remain in each training or test set, and the division is made in such a way so as to have the most significant results. As can be realized, a mean overall accuracy of over 85% is achieved in both, training and test sets, in all the cases. Further, Table 5 describes the classification of the compounds derived from the DF3 function.

Moreover, as illustrated in Table 6, this discriminant ability may be extended through the topological description of (N5−N3), which was the best descriptor for the one-variable set. Indeed, the regression equation relating to this difference with the TIs was a four-variable one:

$$N5-N3 = (72.75*G3) - (48.14*G3v) + (2527.18*S0v) - (734.515*S4c) + 433.55 \quad (1)$$

$$N = 37; \quad r = 0.9845; \quad Cp = 5.0; \quad F = 253.45; \quad SE = 53.99; \quad p = 0.0000$$

Figure 1 depicts the plot of residual versus deleted residual with four outliers.

The results obtained by applying the (N5−N3) topological function (eq 1), together with the other TIs to discriminate a wide set of antibacterial (242) and nonantibacterial (221) compounds, lead to the following equation:

$$DF4 = (-2.36*Pr2) + (0.008*(N5-N3)) + (0.07643*S) + 18.2$$

$$N = 463; \quad F = 116.57; \quad \text{U-statistic (Wilks' lambda)} = 0.5657$$

Moreover, the application of this discriminant function to a wide test set including 65 actives and 68 inactives reproduced rather well the overall accuracy of the training set (over 84%).

All these results demonstrated two important features: The first one is the key role that the frequencies N5 and N3 as well as their difference (N5−N3) seems to play in the general antibacterial activity, and the second one is the possibility to achieve a topological description of such frequencies, so that this predictive ability can be easily extended to wide sets of compounds looking for potential new antibacterials. The reason for this is that the used TIs are much more easily calculated than the QDs.

The influence of these frequencies in the antibacterial activity is a new feature, since no reference was found in the literature about it. Although, of course, more research is necessary to establish the latest reasons explaining this, it can be proposed that the N−H, C−H, and O−H functional groups play a decisive role in the drug-receptor interaction, so that any factor able to modify the vibrational frequencies of these groups should be taken into account. A possible structural feature influencing these frequencies is the exist-

Table 4. Discriminant Function and Classification Matrix Obtained by Linear Discriminant Analysis Study

		classification matrix		
discriminant function	group	% correct	actives	inactives
DF1 = (0.008*N5) − 29.16				
N = 59				
F = 64.138	active	84.2	16	3
U-statistic (Wilks' lambda) = 0.4705	inactive	82.6	4	19
	test active	80.0	4	1
	test inactive	83.3	2	10
DF2 = (0.014*N5) − (0.103*N3) − 48.2				
N = 59				
F = 54.197	active	81.2	13	3
U-statistic (Wilks' lambda) = 0.3406	inactive	96.3	1	26
	test active	87.5	7	1
	test inactive	87.5	1	7
DF3 = (0.01*(N5−N3)) − 33.90				
N = 59				
F = 48.427	active	84.2	16	3
U-statistic (Wilks' lambda) =0. 45235	inactive	87.0	3	20
	test active	80.0	4	1
	test inactive	100	0	12

Table 5. Results Obtained in the LDA Study and Classification of the Compounds from Discriminant Function DF3

active group			inactive group		
compound	prob.	class	compound	prob.	class.
			Training Group		
amoxicillin	0.974	A	antrafenine	0.980	I
azidamfenicol	0.976	A	apazone	0.964	I
acedapsone	0.246	I	azacosterol	0.964	I
brodimopprim	0.593	A	benfluorex	0.949	I
carbenicillin	0.974	A	benorylate	0.955	I
cefazolin	0.260	I	benoxaprofen	0.953	I
cefotaxime	0.649	A	bucloxic acid	0.986	I
ceftriaxone	0.661	A	clofibric acid	0.961	I
chloramphenicol	0.981	A	chlorothen	0.969	I
demeclocycline	0.977	A	clorpropamide	0.784	I
			doxylamine	0.968	I
dicloxacillin	0.203	I	glibornuride	0.975	I
methicin	0.269	I	glibornuride	0.791	I
oxytetracycline	0.970	A	glybuzole	0.951	I
rifampin	0.980	A	gliclazide	0.829	I
rifaximin	0.981	A	glymide	0.962	I
sisomicin	0.981	A	indomethacin	0.966	I
streptomycin	0.984	A	miglitol	0.033	A
tetracycline	0.972	A	nicomol	0.977	I
tetroxoprim	0.627	A	probucol	0.037	A
			salicylic acid	0.086	A
			tripelennamine	0.977	I
			triprolidine	0.980	I
			pheniramine	0.970	I
			debekacin	0.019	A
			Test Group		
cefuroxime	0.603	A	benfluorex	0.949	I
cloxacillin	0.214	I	acetohexamide	0.765	I
doxycycline	0.972	A	aminopyrine	0.971	I
minocycline	0.922	A	benzpiperylon	0.971	I
sulfoxon	0.977	A	benzydamine	0.972	I
			clemizole	0.98	I
			eicosapentaenoic acid	0.949	I
			glipizide	0.779	I
			nicotine	0.991	I
			methapyrilene	0.984	I
			β-sitosterol	0.962	I
			tolbutamide	0.835	I

ence of hydrogen bonding between the drug groups and the receptor ones. Obviously the existence of hydrogen bonding between groups from the drug and from the receptor should modify decisively the vibrational frequency values and hence the antibacterial activity. The surprising character of this influence must be emphasized since a simple and well-known physicochemical property, as it is the vibrational frequency of a given functional group, is playing such an important role in the antibacterial activity. It is all the more surprising because the antibacterial group includes drugs acting through different mechanisms of action.

CONCLUSIONS

Molecular topologies along with the physicochemical parameters derived from PM3 calculations have been shown to demonstrate one useful methodology for the search of new compounds with antibacterial activity, since they can characterize aspects closely related to their mechanism of action. Indeed, we have realized that the vibrational frequency of the functional group and the torsional vibrational frequency are playing the crucial role to differentiate the active and inactive compounds.

Using the multilinear regression and the LDA, a pattern of topological similarity of antibacterial activity has been obtained. This pattern has been applied successfully for the search of drugs that together with other contrasted pharmacological activities can also show antibacterial activity.

An added advantage is that this screening can be carried out using large databases and with low time-consuming.

Table 6. Results Obtained by the Prediction of (N5−N3) through Topological Descriptors[a]

compound	val(obs)	val(cal)	residuals.	residual(cv)
acedapsone	3339.7	3403.9	−64.2	−70.0
acetohexamide	3333.8	3351.6	−17.8	−20.3
aminopyrine	3095.5	3162.3	−66.8	−70.4
apazone	3117.5	3137.1	−19.6	−22.4
azacosterol	3119.5	3141.2	−21.7	−23.3
azidamfenicol	3832.2	3858.9	−26.7	−36.7
benfluorex	3156.0	3142.6	13.4	14.3
benoxaprofen	3146.6	3201.2	−54.6	−57.3
benzpiperylon	3097.9	3067.5	30.4	32.9
benzydamine	3093.8	3117.9	−24.1	−25.6
cefazolin	3347.2	3276.1	71.1	76.2
cefotaxime	3516.5	3499.0	17.5	18.5
ceftriaxone	3522.0	3502.3	19.7	21.8
clemizole	3037.3	3081.3	−44.0	−49.1
clofibric acid	3127.0	3197.4	−70.4	−75.2
chlorothen	3104.5	2947.4	157.1	175.0
chlorpropamide	3322.5	3306.2	16.3	18.9
demeclocycline	3836.0	3811.0	25.0	30.6
dibekacin	3855.1	3755.9	99.2	136.0
doxycycline	3816.9	3825.4	−8.5	−10.1
glibornuride	3318.4	3252.6	65.8	74.9
glybuzole	3152.5	3158.7	−6.2	−7.2
gliclazide	3292.9	3294.8	−1.9	−2.2
glipizide	3325.5	3269.4	56.1	64.8
indomethacin	3112.8	3185.3	−72.5	−76.3
minocycline	3705.8	3706.6	−0.8	−0.9
nicomol	3073.5	3139.4	−65.9	−136.9
oxytetracycline	3806.7	3894.1	−87.4	−112.6
methapyrilene	3037.2	3038.1	−0.9	−1.1
salicylic acid	3694.8	3705.2	−10.4	−14.5
β-sitosterol	3124.5	3122.7	1.8	2.0
streptomycin	3870.3	3851.9	18.4	24.1
tetracycline	3814.7	3755.3	59.4	72.1
tetroxoprim	3506.6	3508.5	−1.9	−2.1
tolbutamide	3288.9	3311.8	−22.9	−26.4
tripelennamine	3073.2	3060.0	13.2	14.3
triprolidine	3059.5	3034.7	24.8	27.4

[a] Cross-validation study.

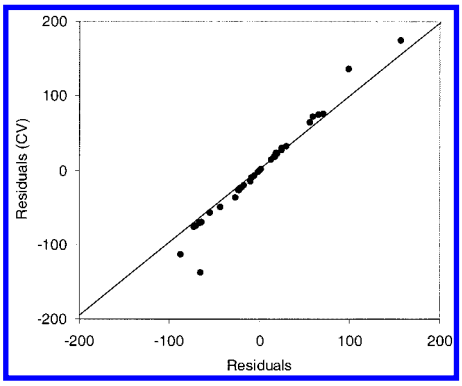

Figure 1. Plot of residuals (CV) versus residuals for the prediction of N5−N3.

ACKNOWLEDGMENT

This study has been supported by GV99-91-1-12 (Generalitat Valenciana).

REFERENCES AND NOTES

(1) Siebel, G. L.; Kollman, P. A. *Comprehensive Medicinal Chemistry (Vol. IV): Quantitative drug design*; Hansch, C., Sammes, P. G., Taylor, J. B., Ramdsen, C. A., Eds.; Pergamon: 1990; pp 125−138.

(2) Weinstein, H.; Osman, R.; Green, J. P. *Computer-assisted drug design*; Olson, E. C., Cristoffersen, R. E., Eds.; ACS Symposium Series; American Chemical Society: Washington, DC, 1979; pp 161−170.

(3) Johnson, M. A.; Maggiora, G. M.; *Concepts and applications of molecular similarity*; John Wiley-Intersciences: New York, 1990.

(4) Hall, L. H.; Kier, L. B. *Reviews in computational chemistry*; Kipkowitz, K. B., Boyd, D. B., Eds.; VCH: 1991; Vol. I, pp 367−422.

(5) Basak, S. C.; Grunwald, G. D.; Niemi, G. I. *Chemical topology to three-dimensional geometry*; Balaban, A. T., Eds.; Plenum Press: New York, 1997; pp 73−116.

(6) Galvez, J.; Garcia, R.; Julian-Ortiz, J. V. de; Soler, R. Topological approach to drug design. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 272−284.

(7) Estrada, E.; Pena, P.; Garcia-Domenech, R. Designin sedative hypnotic compounds from a novel substructural graph-theoretical approach. *J. Comput. Aid. Mol. Des.* **1998**, *12*, 583−595.

(8) Cramer III, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.

(9) Ren, B. A new topological index for QSPR of alkanes. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 139−143.

(10) Galvez, J.; Garcia-Domenech, R.; De Gregorio, C.; De Julian-Ortiz, J. V.; Popa, L. Pharmacological distribution diagrams: a tool for de novo drug design. *J. Mol. Graphics* **1996**, *14*, 272−276.

(11) Julian-Ortiz, J. V. de; Galvez, J.; Muñoz-Collado, C.; Garcia-Domenech, R.; Jimeno-Cardona, C. Virtual combinatorial syntheses and computational screening of new potential anti-herpes compounds. *J. Med. Chem.* **1999**, *42*, 3308−3314.

(12) Gozalbes, R.; Galvez, J.; Moreno, A.; Garcia-Domenech, R. Discovery of new antimalarial compounds by use of molecular connectivity techniques. *J. Pharm. Pharmacol.* **1999**, *51*, 111−117.

(13) Rios-Santamarina, I.; Garcia-Domenech, R.; Galvez, J. New bronchodilators selected by molecular topology. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 477−482.

(14) Garcia-Domenech, R.; Villanueva, A.; Gálvez, J.; Gozalbes, R. Application de la topologie moléculaire a la prediction de la viscosite liquide des composes organiques. *J. Chim. Phys.* **1999**, *96*, 1172−1185.

(15) Ovidiu, I.; Ivanciuc, T.; Balaban, A. Quantitative structure−property relationship study of normal boling points for halogen-/oxygen-/sulfur-containing organic compounds using the CODESSA program. *Tetrahedron* **1998**, *54*, 9129−9142.

(16) Hosoya, H.; Gotoh, M.; Murakami, M.; Ikeda, S. Topological index and thermodynamic properties. 5. How can we explain the topological dependency of thermodynamic properties of alkanes with the topology of graphs? *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 192−196.

(17) Gao, H.; Williams, C.; Labute, P.; Bajorath, J. Binary quantitative structure−activity relationship (QSAR) analixyx of estrogen receptor ligands. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 164−168.

(18) De Gregorio, C.; Kier, L. B.; Hall, L. H. QSAR modeling with the electrotopological state indices: corticosteroids. *J. Comput. Aid. Mol. Des.* **1998**, *12*, 557−561.

(19) Gozalbes, R.; Galvez, J.; Garcia-Domenech, R.; Derouin, F. Molecular search of new acrive drugs against toxoplasma gondii. *SAR QSAR Environ. Res.* **1999**, *10*, 47−60.

(20) Galvez, J. On a topological interpretation of electronic and vibrational molecular energies. *J. Mol. Struct. (Theochem*) **1998**, *429*, 255−264

(21) Gálvez, J.; Garcia-Domenech, R.; de Gregorio-Alapont, C. Indices of differences of path lengths: Novel topological descriptors derived from electronic interferences in graphs. *J. Comput. Aid. Mol. Des.* (in press).

(22) Kier, L. B.; Hall, L. H. General definition of valence delta-values for molecular connectivity. *J. Pharm. Sci.* **1983**, *72*, 1170−1173.

(23) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure−Activity Analysis*; Research Studies Press: Letchworth, England, 1986; pp 225−246.

(24) Galvez, J.; Garcia-Domenech, R.; Salabert, M. T.; Soler, R. Charge indices. New topological descriptors. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 520−525.

(25) García, R.; Gálvez, J.; Moliner, R.; García, F Prediction and interpretation of some pharmacological Properties of caphalosporins using molecular connectivity. *Drug Invest.* **1991**, *3*, 344.

(26) *The Merck Index*, 12×bb ed.; Budaravi, S., Ed.; 1996.

(27) Stewart, J. J. P. *MOPAC97 Manual*; Fujitsu Limited: 1997.

(28) Thiel, W. J. Fast semiempirical geometry optimization. *J. Mol. Struct.* (*Theochem*) **1988**, 163, 415.

(29) Karna, S. P.; Dupuis, M. Frequency dependent nonlinear optical properties of molecules: formulation and implementation of HONDO program. *J. Comput. Chem*. **1991**, *12*, 487.

(30) Dixon, W. J. *BMDP Statistical software*; Univeristy of California: Berkeley, CA, 1990.

(31) Furnival, G. M. All possible regression with less computation. *Technometrics* **1971**, *13*, 403−408.

(32) Hocking, R. R. On yates order in fractional factorial designs. *Technometrics* **1972**, *14*(4), 967−970.

CI000303C