Artificial Neural Networks and Linear Discriminant Analysis: A Valuable Combination in the Selection of New Antibacterial Compounds

Miguel Murcia-Soler,[†] Facundo Pérez-Giménez,*[,†] Francisco J. García-March,[†] Ma Teresa Salabert-Salvador,[†] Wladimiro Díaz-Villanueva,[‡] María José Castro-Bleda,[§] and Angel Villanueva-Pareja[†]

Department of Physical Chemistry, Faculty of Pharmacy, and Department of Computer Science, Faculty of Physics, Universitat de València, Av. Vicent Andrés Estellés, s/n. 46100 Burjassot, Valencia, Spain, and Department of Information Systems and Computation, Universitat Politècnica de València, Camí de Vera, s/n. 46071 Valencia, Spain

Received November 11, 2003

A set of topological descriptors has been used to discriminate between antibacterial and nonantibacterial drugs. Topological descriptors are simple integers calculated from the molecular structure represented in SMILES format. The methods used for antibacterial activity discrimination were linear discriminant analysis (LDA) and artificial neural networks of a multilayer perceptron (MLP) type. The following plot frequency distribution diagrams were used: a function of the number of drugs within a value interval of the discriminant function and the output value of the neural network versus these values. Pharmacological distribution diagrams (PDD) were used as a visualizing technique for the identification of antibacterial agents. The results confirmed the discriminative capacity of the topological descriptors proposed. The combined use of LDA and MLP in the guided search and the selection of new structures with theoretical antibacterial activity proved highly effective, as shown by the in vitro activity and toxicity assays conducted.

1. INTRODUCTION

In the past two decades QSAR (quantitative structure−activity relationship) models have been extensively used as important tools in molecular design, especially in developing new drugs. These methods allow for the estimatation of the properties of new chemical compounds without having to perform their syntheses or assays.

QSAR methods explore the relationship between a specific property and the chemical structure of a series of molecules,[1] thus making it possible to observe the variation of this property in the analyzed series and to use this information to predict the values of the property in other compounds in the same therapeutic category.

The common basis of QSAR studies is the analysis of a data set consisting of a list of compounds with their molecular structures and biological activities; this is to say, the biological properties of molecules depend on their structures. QSAR is essentially the search for a model which correlates activity to independent variables.[2]

The first step in QSAR modeling is the representation of molecular structures by numbers termed molecular descriptors. They are computerized by using quantum-chemical, information-theoretic, graph-theoretic, or geometric approaches.[2,3] Chemical graph theory is largely applied to the characterization of chemical structures as well as to structure−property and structure−activity correlations by means of the so-called topological indices (TIs). TIs are numerical quantities based on various invariants or characteristics of molecular graphs, their formulation being based upon the characterization of chemical structure by graph theory.[4−6] Although several TIs have been proposed, only a few of them have been widely used in QSAR models.

In this paper, an interesting approach to QSAR analysis is observed, namely the use of a set of TIs as simple integers applied to individual atoms and bonds in molecules. The important common feature to all those descriptors is the independence of their numerical values in renumbering atoms in a chemical structure.[7−9] These descriptors encode information about atom type, bonds, degree vertex, distances between pairs of atoms, etc., and so they constitute an alternative to the use of molecular descriptors in QSAR studies, not only for the calculation process but also for a simpler interpretation[10−12] in the prediction of biological properties for a homogeneous collection of chemicals, so that such models are generally applicable.

Finding structure−activity relationships is essentially a pattern recognition process, and historically, QSAR models have been developed using linear methods such as linear discriminant analysis (LDA); however, several nonlinear QSAR methods have been proposed in recent years. Artificial neural networks (ANNs) is one group of methods that are increasingly being used in drug design for QSAR studies.[13−17] This method is capable of recognizing highly nonlinear structure−activity relationships; in contrast, LDA approaches can only capture linear relationships between molecular characteristics and the structural or functional features to be predicted.[18,19]

* Corresponding author phone: (34) 963 54 48 94; fax: (34) 963 54 48 92; e-mail: Facundo.Perez@uv.es.

[†] Department of Physical Chemistry, Faculty of Pharmacy, Universitat de València.

[‡] Department of Computer Science, Faculty of Physics, Universitat de València.

[§] Universitat Politècnica de València.

10.1021/ci030340e CCC: $27.50
Published on Web 04/13/2004

Table 1. Description of the 62 Topological Indices

description	number and notation of indices
group of number and atom type indices	1 to 7(A^C, A^N, A^O, A^S, A^F, A^{Cl}, A^{Br})
group of number and bond type indices	8 to 10 (B^1, B^2, B^3)
group of degree vertex indices	11 to 14 (V^1 to V^4)
group of conjugated double bonds indices	15 and 16 (B^{D1}, B^{D2})
group of sumatory of distance indices from each atom type	17 to 24: distance from N (D^{N1} to D^{N8}) 25 to 32: distance from O (D^{O1} to D^{O8}) 33 to 36: distance from S (D^{S1} to D^{S4}) 37 to 41: distance from F (D^{F2} to D^{F6}) 42 to 47: distance from Cl (D^{Cl2} to D^{Cl7}) 48 to 50: distance from Br (D^{Br2} to D^{Br4})
group of sumatory of general distance indices	51 to 62 (D^1 to D^{12})

Our research is aimed at discriminating between antibacterial and nonantibacterial compounds by topological methods and demonstrating the discriminative ability of the abovementioned group of simple topological descriptors in order to select new antibacterial agents from new structures. To this end, the methods used for antibacterial activity discrimination are linear discriminant analysis and artificial neural networks.

In the study, a total of 217 molecules with well-known antibacterial activity and 216 compounds without this activity were used. The method developed is a three-step process for the selection of new antibacterial compounds. First, calculation of topological indices for each molecule in the study. Second, indices are processed by the artificial neural network; once its training is completed, the most suitable network architecture is determined to ensure an adequate discrimination process. A linear discriminant analysis is performed at the same time, which provides a discriminant function for the classification of active and inactive compounds. Third, using the discriminant function obtained, a guided search for structures with theoretically antibacterial activity is conducted among the compounds in the chemical database, which are then classified by the discriminant function and the previously trained neural network.

Finally, pharmacological and toxicological tests are carried out to determine and verify antibacterial activity and toxicity in the compounds selected.

2. METHODS

The selection of the molecules that make up the discrimination groups—with and without antibacterial activity—was achieved based on data from the 12th edition of Merck Index, on CD-Rom.[20]

A total of 217 molecules with antibacterial activity were selected for the active group from different antibiotics groups such as aminoglycosides, β-lactams, macrolides, tetracyclines, nitrofurans, quinolones, sulfonamides, sulfones, etc. Molecules for the nonactive group (216 nonantibacterial compounds) were selected from different therapeutic categories, such as analgesic narcotics, analgesic nonnarcotics, antihistaminics, antihyperlipoproteinemics, antihypertensives, diuretics, hypnotics, antidiabetics, and antidepressants.

Both the active and nonactive molecule groups were divided into training and test sets, accounting for 70% and 30% of the molecules, respectively, so that the different therapeutic categories were represented in both sets.

A set of topological and structural indices[10] was used for the discrimination of antibacterial activity. The chosen set of molecular descriptors should adequately capture the phenomena underlying the properties of each compound. It is also important for descriptors to be obtained without much computational effort, since they have to be computed for every molecule whose properties need to be discriminated, as is the case with molecules with antibacterial activity.

Topological indices were drawn from the representation of the molecule in SMILES, and its hydrogen-suppressed graph was made including information on atom and bond type.

From the hydrogen-suppressed graph with atom and bond type information, breadth first search (BFS) and depth first search (DFS) traversals were used in order to calculate the topological indices,[21] this saving time and calculation memory in comparison with matrices commonly used for topological index representation and calculation. The indices are related to the atoms' number and type, bonds' number and type, conjugated double bonds, distance among selected atoms' types, and other general distances. Vertices (atoms) in this structure were arbitrarily assigned numbers.

Table 1 lists the 62 topological indices and their descriptions.[22]

Bond a_{ij} in the diagram has value 1 when there is an edge between vertices i and j, otherwise it is zero.

$$(A)_{ij} = \begin{cases} a_{if} & if \ \ i \neq j \\ 0 & if \ \ i = j \end{cases} \tag{1}$$

Degree vertex or topological valence δ_i for the atom is equal to the number of bonds (edges) that come up to each atom.

$$\delta_i = \sum_{j=1}^{n} a_{ij} \tag{2}$$

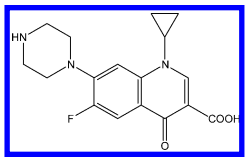
Distance is the length of the shortest path, d_{ij}, between the vertices in the graph, where D_{ij} is the number of steps in the shortest path (i.e. minimum number of edges) in a graph between vertices i and j.[23]

$$(D)_{ij} = \begin{cases} d_{if} & if \ \ i \neq j \\ 0 & if \ \ i = j \end{cases} \tag{3}$$

We selected 62 indices, 14 of them including simple information on the molecule: total number of atoms of a certain element (carbon, nitrogen, oxygen, sulfur, fluorine, chlorine, and bromine), total number of bonds of a certain type (single, double or triple), and number of atoms with a specific vertex degree (vertex degree equal to 1, 2, 3, or 4).

The remaining 48 indices contain different topological data, such as the number of double bonds at distance 1 or 2

Table 2. Chemical Structure, Smiles Notation, and Indices Calculated Values for Ciprofloxacin



no.	index	value	no.	index	value	no.	index	value	no.	index	value
1	A^{C}	17	**17**	D^{N1}	8	**33**	D^{S1}	0	**48**	D^{Br2}	0
2	A^{N}	3	**18**	D^{N2}	11	**34**	D^{S2}	0	**49**	D^{Br3}	0
3	A^{O}	3	**19**	D^{N3}	9	**35**	D^{S3}	0	**50**	D^{Br4}	0
4	A^{S}	0	**20**	D^{N4}	8	**36**	D^{S4}	0	**51**	D^{1}	27
5	A^{F}	1	**21**	D^{N5}	8	**37**	D^{F2}	2	**52**	D^{2}	37
6	A^{Cl}	0	**22**	D^{N6}	9	**38**	D^{F3}	3	**53**	D^{3}	43
7	A^{Br}	0	**23**	D^{N7}	4	**39**	D^{F4}	4	**54**	D^{4}	42
8	B^{1}	21	**24**	D^{N8}	5	**40**	D^{F5}	5	**55**	D^{5}	39
9	B^{2}	6	**25**	D^{O1}	3	**41**	D^{F6}	4	**56**	D^{6}	33
10	B^{3}	0	**26**	D^{O2}	6	**42**	D^{Cl2}	0	**57**	D^{7}	23
11	V^{1}	4	**27**	D^{O3}	8	**43**	D^{Cl3}	0	**58**	D^{8}	15
12	V^{2}	10	**28**	D^{O4}	11	**44**	D^{Cl4}	0	**59**	D^{9}	10
13	V^{3}	10	**29**	D^{O5}	9	**45**	D^{Cl5}	0	**60**	D^{10}	5
14	V^{4}	0	**30**	D^{O6}	11	**46**	D^{Cl6}	0	**61**	D^{11}	2
15	B^{D1}	6	**31**	D^{O7}	6	**47**	D^{Cl7}	0	**62**	D^{12}	0
16	B^{D2}	4	**32**	D^{O8}	4						

and minimum distance between pairs of atoms, which are counted as the number of bonds between atoms. Indices were classified into six groups associated with the most frequent elements in molecules with pharmacological activity (nitrogen, oxygen, sulfur, fluorine, chlorine, bromine) and a general group in which distances between atom pairs were considered without identifying the atom type.[12]

In calculating atom number and type indices, a straightforward function searches the SMILES chain, the counter moving up every time an element is found that coincides with the searched atom.[22]

To obtain bond type and number atoms, a depth graph search is carried out by means of a DFS algorithm, this increasing the counter every time a bond type is found. The same operation is conducted when trying to draw vertex degree indices, but in this case the number of bonds converging on the atom and increasing the corresponding counter is checked up every time an atom is reached.

A DFS algorithm is also carried out in drawing conjugated double bonds, by means of which, bonds of a specific type are detected; a BFS algorithm is performed next with a view to determining those bonds at a given distance. The BFS algorithm has the advantage of moving between nodes through the shortest possible way.

The different distance indices (general and atom type) are calculated in the same way, with a depth search (DFS) followed by a breadth search (BFS) starting from each atom type.

For example purposes, Table 2 includes the indices calculated for the ciprofloxacine molecule, a largely used antibacterial from the quinolone group.

Once the whole set of indices to be used was defined, the process of establishing the structure−activity relationships permitting antibacterial activity discrimination started, using linear discriminant analysis (LDA) and artificial neural networks (ANNs) to that end. The same training and test sets of molecules were used for both methods.

Linear discriminant analysis (LDA) was carried out by means of the BMDP statistical package,[24] using topological indices as independent variables.[2] LDA was used to select the (pharmacological and/or structural) parameters that identify the antibacterial or nonantibacterial nature of molecules. The criteria for the selection of the best LDA equation included a comparison of the tabulated F and Wilk's U statistical values, the determination of the percentage of molecules correctly classified, and the prediction of the classification of molecules not included in the training process (cross-validation).

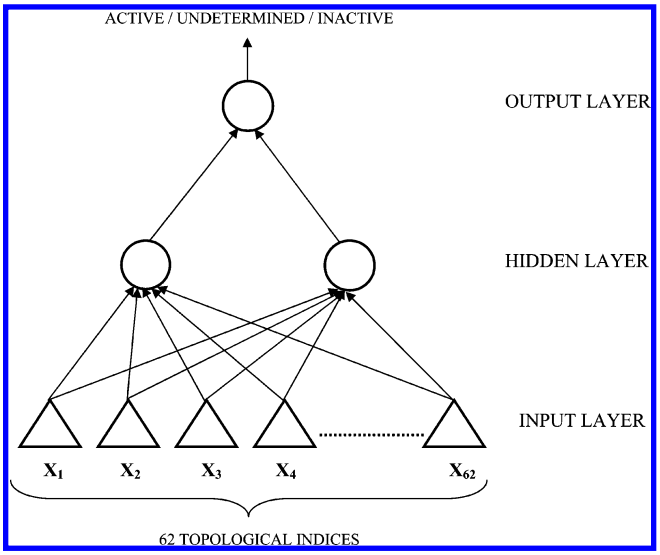


Figure 1. Schematic representation of a multilayer perceptron with one hidden layer.

The same process was replicated using an artificial neural network as the discrimination method. Multilayer perceptrons (MLPs) were used for classification purposes, this being a feed-forward neural network with a multilayer structure.[25,26] Each layer is made up of a number of units, and each unit in a single layer is connected to all units in the next layer. All connections between two units in adjacent layers are assigned a weight, namely a positive or negative real number that multiplies the signal from the preceding unit. Each unit sums its various weighted inputs until some preset level (which depends on the activation function employed) is reached, and at this point it fires and sends its signal to the units in the next layer.

The number of input units was set by the number of topological descriptors of the molecules (62 topological indices). Input data was discretized by dividing by the maximum value of all the indices. The hidden layer had two units. There was only one output unit corresponding to the property being classified: a +1 value was assigned to the active molecules and −1 to inactive ones. Therefore, we used the hyperbolic tangent function—defined in the interval [−1,1]—as the activation function. Figure 1 shows a diagram of an MLP with one hidden layer.

MLP training was conducted by using the neural net software package "SNNS: Stuttgart Neural Network Simulator", developed by the University of Stuttgart.[27] The validation criterion was used to stop the learning process and to select the best configuration. The training process followed as the classification error rate of the validation data decreased (down to a maximum number of training epochs, in this case, 10 000 epochs).

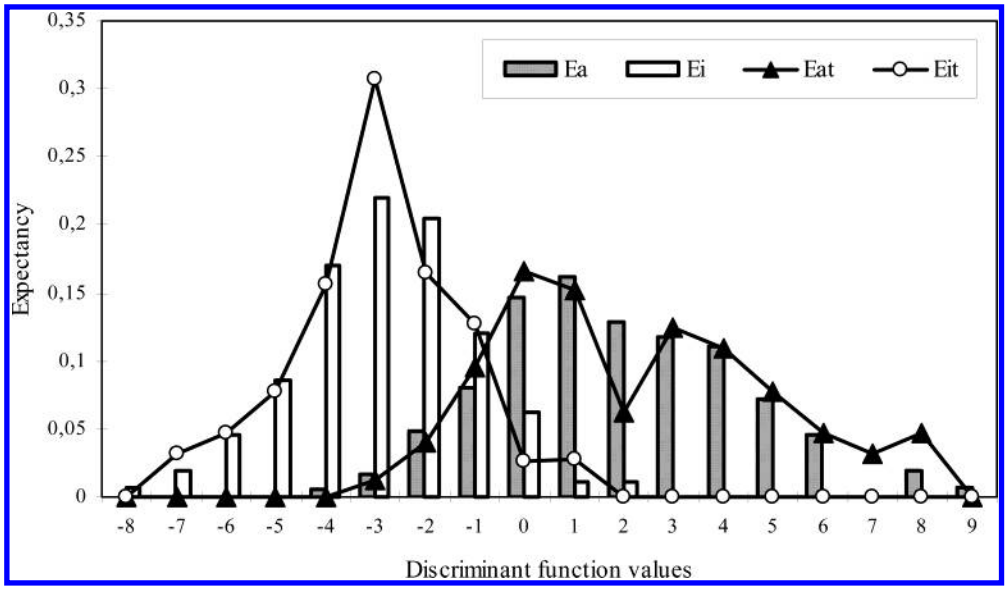


Figure 2. Pharmacological distribution diagram (PDD) for the LDA of antibacterial activity (E_a and E_{at}, activity expectancy of reference and test groups, respectively; E_i and E_{it}, inactivity expectancy of reference and test groups, respectively).

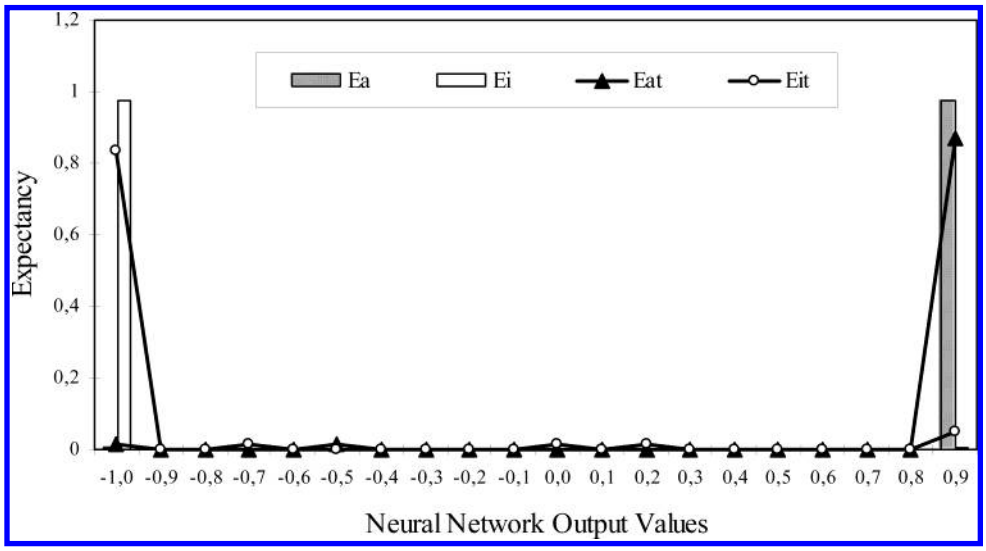


Figure 3. Pharmacological distribution diagram (PDD) for the MLP of antibacterial activity (E_a and E_{at}, activity expectancy of reference and test groups, respectively; E_i and E_{it}, inactivity expectancy of reference and test groups, respectively).

Table 3. Best Obtained Equation by Linear Discriminant Analysis (LDA) Applied to Antibacterial Activity[a]

parameter	coefficient	*F* to remove	*U* statistical
V^2	−0.3680	42.2422	0.6605
D^{N8}	0.2602	99.4499	0.7529
D^{O5}	0.1057	38.4833	0.5234
D^{S2}	0.5739	35.6434	0.5906
D^{F5}	0.4680	19.5497	0.4617
D^{Cl4}	−0.6088	19.1343	0.4919
constant	0.0742		

[a] Discriminant function: $\Delta P = -0.3680\ V^2 + 0.2602\ D^{N8} + 0.1057\ D^{O5} + 0.5739\ D^{S2} + 0.4680\ D^{F5} - 0.6088\ D^{Cl4} + 0.0742$.

Both the discriminant function and the neural network are capable of describing pharmacological activity patterns as well as nonactivity patterns. In other words, both methods pinpoint not only active drugs according to their distribution but also inactive compounds. When applied to the discrimination of concrete pharmacological actions, we call them *pharmacological distribution diagram* (PDD).[28]

Then, using the obtained discriminant function, a guided search was conducted for those structures with antibacterial activity that met the topological requirements set by the variables with a positive contribution to that function, from compounds in a large chemical database (Available Chemicals Directory).[29]

The topological-structural indices of the selected compounds were calculated and then the compounds were classified as active or nonactive by the discriminant function and the artificial neural network at consecutive stages. The compounds theoretically classified as active by the neural network and with better discriminant function values underwent microbiological assays in order to determine their minimal inhibitory concentration (MIC) versus different gram-positive microorganism strains (*Enterococcus faecalis* ATCC 29212 and *Staphylococcus aureus* ATCC 6538P) and gram-negative ones (*Escherichia coli*ATCC 10536 and *Pseudomonas aeruginosa* ATCC 27853), supplied as lyophiles by the Spanish Type Culture Collection[30] of the Department of Microbiology from Valencia University.

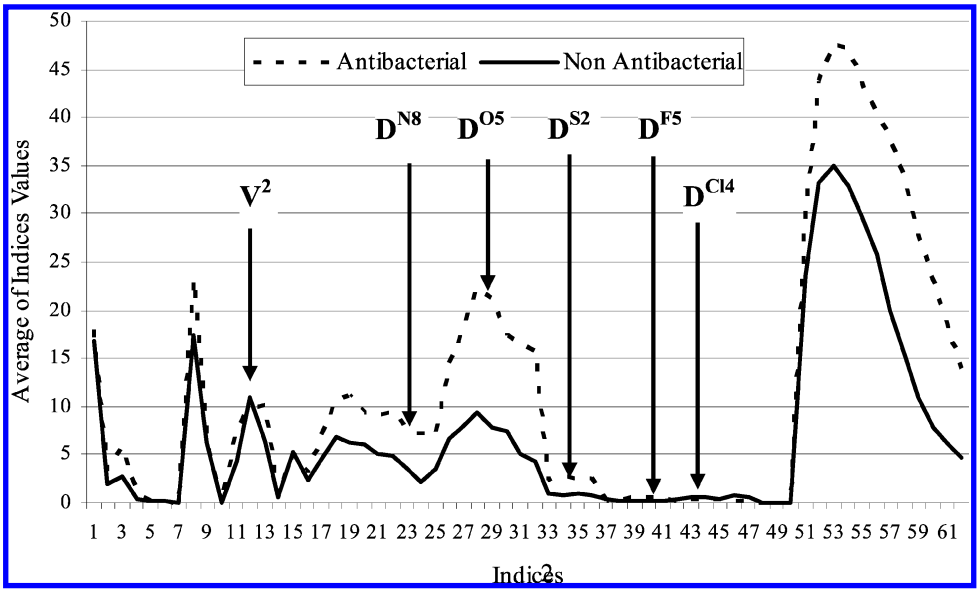


Figure 4. Molecular spectrum of antibacterial and nonantibacterial compounds.

To that end, microtiter plates were used where serial dilutions were made for each individual product. The products were added to the microorganism inoculate in a liquid culture medium (tripticasa-soy broth); each plate also included the dilutions of two reference drugs, namely cephalosporin C and nalidixic acid.

Once the microbiological assay was completed, those molecules showing antibacterial activity then underwent a toxicity test in order to determine their lethal dose 50 using blood-unrelated Swiss mice (25−30 g weight) to whom the products were administered intraperitoneally.

3. RESULTS AND DISCUSSION

The indices corresponding to the 433 compounds, 217 with antibacterial activity and 216 without this activity, employed in the discrimination process, were calculated using the methodology presented above.

This group of molecules was divided into two parts: a discrimination set (70% of the compounds) and a test set (30% of the compounds) for both the active and inactive molecules.

The best discrimination function was obtained with variables V^2, D^{N8}, D^{O5}, D^{S2}, D^{F5}, and D^{Cl4}. Table 1 shows the discriminant function (ΔP), obtained as the difference between the variables defining the groups of active and inactive molecules, together with the values for F-Snedecor and Wilk's U statistical parameters used with each variable.

In the active group, molecules with discriminant function values higher than 0.5 ($\Delta P > 0.5$) were classified as active, while $\Delta P < 0$ corresponded to inactive molecules; compounds with ΔP values between 0 and 0.5 were classified as undetermined activity. The discriminant conditions imposed are intended to minimize the percentage of error, that is, to give the lowest possible number of false positives, although "a priori" this may force us to discard a greater number of active compounds. The same criterion was applied to the compound group with no antibacterial activity; $\Delta P < -0.5$ molecules were classed as inactive, those with a ΔP value between 0 and −0.5 as undetermined, and those with a ΔP value over 0 were mistakenly classified as active.

The same discrimination process was carried out using artificial neural networks. After training the MLP models, the following classification criterion was applied: if the molecule was inactive and the output achieved with the MLP was within the interval [−1, −0.5], it was considered correct; if the output was within the interval [−0.5, 0], the result was classed as undetermined; finally, if the output was within the interval [0, 1], it was an error. When testing an active molecule, the classification criterion was similar: it was considered to be correctly classified when the MLP output value was between 1 and 0.5; if the output was found within the interval [0.5, 0], it was classed as undetermined; and if the output was between 0 and −1, it was considered an error.

Table 2a,b shows the results obtained for the training set (153 active molecules and 152 inactive molecules respectively) using LDA (ΔP value and its probability) and MLP (output value and its probability).

Overall accuracy for the training set was 77.78% on the active group for LDA and 98.69% for MLP. The LDA percentage increased to 81.51% when the undetermined molecules were eliminated, while the MLP percentage remained constant because the number of undetermined molecules was zero. On the inactive group, overall accuracy was 86.84% for LDA and 98.68% for MLP. The LDA percentage increased to 89.90% when the undetermined molecules were ruled out, but the MLP percentage remained constant.

A cross-validation test was applied to the ΔP function and to the trained MLP (the test set was composed of 64 active molecules and 64 inactive molecules). Table 3a,b shows the results obtained using LDA (ΔP value and its probability) and the neural network model (output value and its probability) for both groups of molecules.

Overall accuracy was 75.00% on the active group for LDA and 95.31% for MLP. The LDA percentage increased to 81.36%, when the undetermined molecules were eliminated, while the MLP percentage remained constant because the number of undetermined molecules was zero. On the inactive group, overall accuracy was 90.63% for LDA and 87.50% for MLP. The LDA percentage increased to 93.55% when

Table 4. (a) Results Obtained for 153 Different Compounds with Antibacterial Activity Used as Active Group Training in the Linear Discriminant Analysis and the Neural Network Package and (b) Results Obtained for 152 Different Compounds without Antibacterial Activity Used as Inactive Group Training in the Linear Discriminant Analysis and the Neural Network Package

	discriminant function			neural network				discriminant function			neural network		
compound	ΔP value	classif	prob %	output value	classif	prob %	compound	ΔP value	classif	prob %	output value	classif	prob %
						(a)							
acediasulfone	0.576	+	53.05	1.000	+	100	minocycline	4.481	+	73.68	1.000	+	100
amdinocillin pivoxil	2.301	+	62.16	1.000	+	100	miokamycin	1.833	+	59.69	0.999	+	100
amikacin	6.298	+	83.29	1.000	+	100	nadifloxacin	1.882	+	59.95	1.000	+	99.97
ampicillin	3.254	+	67.20	1.000	+	100	nafcillin sodium	2.419	+	62.79	1.000	+	100
apalcillin	4.598	+	74.30	1.000	+	100	neomycin	5.456	+	78.84	1.000	+	100
apramycin	4.991	+	76.38	1.000	+	100	nifuradene	−1.300	-	43.13	1.000	+	100
aspoxicillin	6.543	+	84.59	1.000	+	100	nifurfoline	−1.358	-	42.82	1.000	+	100
azidamfenicol	0.999	+	55.28	1.000	+	100	nifurpirinol	−1.931	-	39.80	1.000	+	100
azlocillin	4.604	+	74.33	1.000	+	100	nifurprazine	−0.994	-	44.75	1.000	+	100
bacampicillin	3.197	+	66.90	1.000	+	100	nitrofurantoin	−0.827	-	45.63	1.000	+	100
benzylsulfamide	−2.354	-	37.56	0.976	+	98.80	norfloxacin	1.355	+	57.16	1.000	+	100
biapenem	4.089	+	71.61	1.000	+	100	ofloxacin	2.979	+	65.75	1.000	+	100
brodimoprim	0.415	U	52.19	1.000	+	100	oxacillin	2.575	+	63.61	1.000	+	100
butirosin	5.615	+	79.68	1.000	+	100	oxytetracycline	5.979	+	81.60	1.000	+	100
carbenicillin	3.108	+	66.43	1.000	+	100	paromomycin	5.785	+	80.58	1.000	+	100
carbomycin	2.411	+	62.74	0.999	+	99.97	pazufloxacin	1.267	+	56.70	0.973	+	100
carindacillin	1.841	+	59.73	1.000	+	100	pefloxacin	1.983	+	60.48	1.000	+	98.63
cefaclor	−0.745	-	46.06	1.000	+	100	penicillin G	2.106	+	61.13	1.000	+	100
cefatrizine	4.751	+	75.11	1.000	+	100	penicillin N	3.576	+	68.90	1.000	+	100
cefazedone	3.349	+	67.70	1.000	+	100	penicillin V	1.949	+	60.30	1.000	+	100
cefcapene pivoxil	6.125	+	82.37	1.000	+	100	phenethicillin potassium	2.317	+	62.25	1.000	+	99.99
cefdinir	4.981	+	76.33	1.000	+	100	phthalylsulfacetamide	0.742	+	53.92	1.000	+	100
cefetamet	5.135	+	77.14	1.000	+	100	pipacycline	4.529	+	73.94	1.000	+	100
cefixime	5.082	+	76.86	1.000	+	100	piromidic acid	−1.872	-	40.11	0.969	+	100
cefodizime	9.459	+	100.00	1.000	+	100	pivcefalexin	2.366	+	62.51	1.000	+	98.46
cefonicid	6.380	+	83.72	1.000	+	100	propicillin	1.949	+	60.30	1.000	+	100
ceforanide	5.690	+	80.07	1.000	+	100	quinacillin	4.984	+	76.34	1.000	+	100
cefpiramide	6.574	+	84.75	1.000	+	100	rifamide	4.596	+	74.29	1.000	+	99.98
cefpodoxime proxetil	5.605	+	79.62	1.000	+	100	rifamycin SV	3.304	+	67.46	0.999	+	99.96
cefroxadine	1.745	+	59.22	1.000	+	100	rifapentine	4.472	+	73.64	0.998	+	99.90
cefteram	5.696	+	80.11	1.000	+	100	rosaranicin	0.927	+	54.90	0.963	+	98.17
ceftibuten	4.767	+	75.20	1.000	+	100	rosoxacin	−1.979	-	39.54	0.993	+	99.63
ceftizoxime	4.662	+	74.64	1.000	+	100	salazosulfadimidine	−0.264	-	48.60	1.000	+	100
cefuroxime	3.106	+	66.42	1.000	+	100	sancycline	2.601	+	63.75	1.000	+	100
cefuzonam	8.924	+	97.17	1.000	+	100	sisomicin	3.614	+	69.11	1.000	+	100
cephalexin	1.690	+	58.93	1.000	+	100	solasulfone	5.403	+	78.56	0.974	+	98.71
cephalothin	3.261	+	67.24	1.000	+	100	spectinomycin	0.915	+	54.84	1.000	+	100
cephapirin sodium	2.785	+	64.72	0.984	+	99.21	spiramycin	0.658	+	53.48	0.989	+	99.47
cephradine	1.690	+	58.93	1.000	+	100	streptomycin	8.080	+	92.71	1.000	+	100
chloramphenicol	−1.042	-	44.49	1.000	+	100	succinylsulfathiazole	1.883	+	59.95	1.000	+	100
chlortetracycline	2.032	+	60.74	1.000	+	100	succisulfone	0.631	+	53.34	1.000	+	100
ciprofloxacin	0.987	+	55.21	1.000	+	100	sulfachlorpyridazine	−0.450	-	47.62	1.000	+	100
clarithromycin	3.726	+	69.70	1.000	+	99.98	sulfachrysoidine	0.376	U	51.99	1.000	+	100
clinafloxacin	−2.065	-	39.08	0.995	+	99.76	sulfadiazine	−0.209	-	48.90	1.000	+	100
clindamycin	−0.814	-	45.70	0.976	+	98.78	sulfadicramide	0.743	+	53.93	1.000	+	100
clomocycline	1.719	+	59.09	0.991	+	99.53	sulfadoxine	0.686	+	53.62	1.000	+	100
cloxacillin	1.223	+	56.46	1.000	+	100	sulfaethidole	2.455	+	62.98	1.000	+	100
cyclacillin	2.888	+	65.27	1.000	+	100	sulfaguanidine	1.208	+	56.39	1.000	+	100
dapsone	0.890	+	54.70	1.000	+	100	sulfalene	0.320	U	51.69	1.000	+	100
demeclocycline	2.324	+	62.29	1.000	+	100	sulfaloxic acid	0.584	+	53.09	1.000	+	100
dibekacin	3.934	+	70.79	1.000	+	100	sulfamethazine	0.527	+	52.79	1.000	+	100
dichloramine T	−1.326	-	42.99	0.983	+	99.14	sulfamethizole	2.823	+	64.92	1.000	+	100
difloxacin	1.607	+	58.49	1.000	+	100	sulfamethoxazole	0.584	+	53.09	1.000	+	100
dihydrostreptomycin	8.080	+	92.71	1.000	+	100	sulfamethoxypyridazine	−0.103	-	49.45	1.000	+	100
erythromycin	3.834	+	70.27	1.000	+	99.98	sulfamidochrysoidine	−0.838	-	45.57	1.000	+	100
fenbenicillin	1.315	+	56.95	1.000	+	100	sulfamoxole	0.846	+	54.47	1.000	+	100
fleroxacin	5.832	+	80.83	1.000	+	100	sulfanilamide	−0.038	-	49.80	1.000	+	100
florfenicol	1.295	+	56.84	1.000	+	100	sulfanilic acid	0.067	U	50.36	1.000	+	100
floxacillin	2.491	+	63.17	1.000	+	100	N4-sulfanilylsulfanilamide	1.779	+	59.40	1.000	+	100
flumequine	1.163	+	56.15	1.000	+	100	sulfanitran	0.734	+	53.88	1.000	+	100
fortimicin B	2.159	+	61.41	1.000	+	100	sulfaperine	0.159	U	50.84	1.000	+	100
furaltadone	−2.091	-	38.95	1.000	+	100	sulfaphenazole	−1.101	-	44.18	0.959	+	97.95
gentamicin	3.980	+	71.04	1.000	+	100	sulfapyrazine	−0.209	-	48.90	1.000	+	100
hexedine	−3.238	-	32.89	−1.000	-	0.00	sulfapyridine	−0.469	-	47.52	1.000	+	100
imipenem	2.312	+	62.22	1.000	+	100	sulfasymazine	0.311	U	51.65	1.000	+	100
isepamicin	6.501	+	84.36	1.000	+	100	sulfathiazole	1.621	+	58.57	1.000	+	100
lenampicillin	3.043	+	66.08	1.000	+	100	sulfathiourea	2.096	+	61.08	1.000	+	100
leucomycins	1.142	+	56.04	0.968	+	98.39	sulfisoxazole	1.058	+	55.59	1.000	+	100
lomefloxacin	4.536	+	73.98	1.000	+	100	talampicillin	2.673	+	64.13	1.000	+	100
lymecycline	4.482	+	73.69	0.999	+	99.96	taurolidine	0.522	+	52.76	1.000	+	100
mafenide	−0.406	-	47.85	0.942	+	97.08	temafloxacin	3.119	+	66.48	1.000	+	100
meclocycline	3.432	+	68.14	1.000	+	100	tetracycline	4.237	+	72.40	1.000	+	100
4′-(methylsulfamoyl)sulfanilanilide	1.984	+	60.49	1.000	+	100	thiazolsulfone	3.553	+	68.78	1.000	+	100
							tigemonam	4.679	+	74.73	1.000	+	100
mezlocillin	6.697	+	85.40	1.000	+	100	tosufloxacin	3.487	+	68.43	1.000	+	100
midecamycins	1.252	+	56.62	0.999	+	99.96	trimethoprim	0.153	U	50.81	1.000	+	100
miloxacin	−0.388	-	47.95	0.995	+	99.74	xibornol	−1.237	-	43.46	−1.000	-	0.00

	discriminant function (%)	neural network (%)	discriminant function (%)	neural network (%)
undetermined (U)	4.58	0.00	adjusted accuracy (excluded undetermined)	81.51 98.69
false inactivity	17.65	1.31		
overall accuracy	77.78	98.69		

Table 4 (Continued)

compound	discriminant function			neural network			compound	discriminant function			neural network		
	Δ*P* value	classif	prob %	output value	classif	prob %		Δ*P* value	classif	prob %	output value	classif	prob %
							(b)						
acecarbromal	−0.869	-	45.41	−1.000	-	0.00	guanoxan	−1.880	-	40.06	−1.000	-	0.00
acetaminophen	−1.449	-	42.34	−1.000	-	0.00	haloxazolam	−1.050	-	44.45	−1.000	-	0.00
acetohexamide	−1.040	-	44.50	−1.000	-	0.01	hexethal	−1.926	-	39.82	−1.000	-	0.00
adinazolam	−4.366	-	26.92	−1.000	-	0.00	hydrochlorothiazide	−1.094	-	44.21	−1.000	-	0.00
ahistan	−0.781	-	45.87	−1.000	-	0.00	hydroxyzine	−6.060	-	17.97	−1.000	-	0.00
alfentanil	−2.417	-	37.23	−1.000	-	0.00	ibufenac	−1.711	-	40.96	−1.000	-	0.00
althiazide	−0.579	-	46.94	−1.000	-	0.00	indeloxazine hydrochloride	−3.612	-	30.91	−1.000	-	0.00
aminochlorthenoxazin	−3.036	-	33.95	−1.000	-	0.00	indoramin	−4.931	-	23.93	−1.000	-	0.00
amlodipine	−2.436	-	37.12	−1.000	-	0.00	isomethadone	−3.813	-	29.85	−1.000	-	0.00
amobarbital	−1.343	-	42.90	−1.000	-	0.00	isoxicam	2.060	+	60.89	1.000	+	100
ampiroxicam	1.788	+	59.45	1.000	+	100	ketoprofen	−2.498	-	36.80	−1.000	-	0.00
amtolmetin guacil	−2.974	-	34.28	−1.000	-	0.00	levcromakalim	−0.300	U	48.41	−1.000	-	0.00
antipyrine	−1.922	-	39.84	−0.971	-	1.44	levophacetoperane	−3.339	-	32.35	−1.000	-	0.00
atorvastatin	−3.408	-	31.98	−1.000	-	0.00	lisinopril	−2.299	-	37.85	−1.000	-	0.00
azacosterol	−2.984	-	34.23	−1.000	-	0.00	lofentanil	−3.604	-	30.95	−1.000	-	0.00
azosemide	−1.751	-	40.74	−1.000	-	0.00	loxoprofen	−2.235	-	38.18	−1.000	-	0.00
benactyzine	−1.506	-	42.04	−0.995	-	0.27	mebhydroline	−4.189	-	27.86	−1.000	-	0.00
benazepril	−3.287	-	32.63	−1.000	-	0.00	mecloxamine	−3.851	-	29.65	−1.000	-	0.00
benorylate	−2.079	-	39.01	−1.000	-	0.00	metazocine	−1.711	-	40.96	−1.000	-	0.00
benzylmorphine	−4.023	-	28.74	−1.000	-	0.00	methyclothiazide	−1.507	-	42.03	−0.968	-	1.62
bevantolol	−4.031	-	28.69	−1.000	-	0.00	methyl-4-pyridyl ketone thiosemicarbazone	−0.573	-	46.97	−0.943	-	2.85
binifibrate	−3.999	-	28.86	−0.959	-	2.03							
brallobarbital	−0.920	-	45.14	−1.000	-	0.00	methyl-4-thiazoleethanol-5	0.273	+	51.44	−1.000	-	0.00
bromfenac	−1.658	-	41.24	−0.999	-	0.05	metopon	−1.813	-	40.42	−1.000	-	0.00
bucetin	−2.447	-	37.07	−1.000	-	0.00	metralindole	−2.400	-	37.31	−1.000	-	0.00
bunitrolol	−1.823	-	40.36	−1.000	-	0.00	mofebutazone	−2.603	-	36.24	−1.000	-	0.00
butacetin	−1.817	-	40.40	−1.000	-	0.00	moveltipril	0.009	+	50.05	−1.000	-	0.00
butethal	−1.711	-	40.96	−1.000	-	0.00	moxastine	−3.978	-	28.97	−1.000	-	0.00
buthiazide	−0.623	-	46.71	−1.000	-	0.00	muzolimine	−2.860	-	34.88	−1.000	-	0.00
butriptyline	−3.821	-	29.80	−1.000	-	0.00	nalbuphine	−2.862	-	34.87	−1.000	-	0.00
captopril	−0.190	U	49.00	−0.976	-	1.18	nefazodone	−5.009	-	23.52	−1.000	-	0.00
carbinoxamine	−4.219	-	27.70	−1.000	-	0.00	nicardipine	−0.518	-	47.26	−1.000	-	0.00
carteolol	−1.614	-	41.47	−1.000	-	0.00	nicoclonate	−2.954	-	34.39	−1.000	-	0.00
carvedilol	−5.139	-	22.84	−1.000	-	0.00	nitrazepam	−2.973	-	34.28	−1.000	-	0.00
chlorhexadol	−2.277	-	37.97	−1.000	-	0.00	nitrendipine	0.285	+	51.51	−1.000	-	0.00
chlorpheniramine	−4.322	-	27.15	−1.000	-	0.00	norpipanone	−6.021	-	18.18	−1.000	-	0.00
chlorthenoxazin	−3.876	-	29.51	−1.000	-	0.00	opipramol	−5.405	-	21.43	−1.000	-	0.00
clocinizine	−7.221	-	11.83	−1.000	-	0.00	oxaflozane	0.361	+	51.91	−1.000	-	0.00
clofenamide	−0.324	U	48.29	−1.000	-	0.00	oxametacine	−2.060	-	39.11	−1.000	-	0.00
clomestrone	−4.641	-	25.47	−1.000	-	0.00	parsalmide	−2.451	-	37.04	−1.000	-	0.00
crotethamide	−0.764	-	45.96	−1.000	-	0.00	phenobarbital	−2.235	-	38.18	−1.000	-	0.00
cyclothiazide	−1.515	-	41.99	−1.000	-	0.00	phenoperidine	−5.598	-	20.41	−1.000	-	0.00
cyproheptadine	−4.557	-	25.91	−1.000	-	0.00	phenopyrazone	−3.128	-	33.47	−1.000	-	0.00
delapril	−2.766	-	35.38	−1.000	-	0.00	phentolamine	−3.978	-	28.97	−1.000	-	0.00
desipramine	−4.557	-	25.91	−1.000	-	0.00	phenylbutazone	−3.652	-	30.69	−1.000	-	0.00
desomorphine	−2.709	-	35.68	−1.000	-	0.00	phenylbutyramide	−1.922	-	39.84	−1.000	-	0.00
dihydrocodeinone enol acetate	−1.345	-	42.89	−1.000	-	0.00	pizotyline	−2.728	-	35.58	−1.000	-	0.00
							polythiazide	2.135	+	61.28	−1.000	-	0.00
dioxadrol	−5.289	-	22.04	−1.000	-	0.00	prazosin	−2.150	-	38.63	−1.000	-	0.01
diphenylpyraline	−4.088	-	28.39	−1.000	-	0.00	propacetamol	−1.766	-	40.67	−1.000	-	0.00
dipipanone	−5.547	-	20.68	−1.000	-	0.00	propoxyphene	−3.760	-	30.12	−1.000	-	0.00
disulfamide	−0.564	-	47.02	−1.000	-	0.00	protizinic acid	0.060	+	50.32	−1.000	-	0.00
dothiepin	−2.467	-	36.96	−1.000	-	0.00	pyrisuccideanol	−0.922	-	45.12	−0.999	-	0.05
doxylamine	−3.978	-	28.97	−1.000	-	0.00	pyrrobutamine	−6.162	-	17.43	−1.000	-	0.00
enalapril	−2.449	-	37.05	−1.000	-	0.00	rescimetol	−2.445	-	37.07	−1.000	-	0.00
enfenamic acid	−3.919	-	29.29	−1.000	-	0.00	reserpine	−1.077	-	44.31	−1.000	-	0.00
eptazocine	−2.447	-	37.07	−1.000	-	0.00	ritanserin	−2.745	-	35.49	−1.000	-	0.00
etaqualone	−3.183	-	33.18	−1.000	-	0.00	ronifibrate	−4.526	-	26.08	−1.000	-	0.00
etodolac	−2.549	-	36.53	−1.000	-	0.00	salacetamide	−1.343	-	42.90	−0.999	-	0.06
etonitazene	−2.835	-	35.01	−0.996	-	0.20	salicylsulfuric acid	0.182	+	50.96	−0.972	-	1.41
etymemazine	−0.897	-	45.26	−1.000	-	0.00	secobarbital	−1.499	-	42.07	−1.000	-	0.00
febarbamate	−1.191	-	43.70	−1.000	-	0.00	sulfinalol	−2.088	-	38.96	−1.000	-	0.00
felodipine	−3.854	-	29.63	−1.000	-	0.00	suprofen	−0.619	-	46.73	−0.999	-	0.05
femoxetine	−4.551	-	25.94	−1.000	-	0.00	talniflumate	0.192	+	51.01	−1.000	-	0.02
fenbufen	−3.234	-	32.91	−1.000	-	0.00	tandospirone	−3.829	-	29.76	−1.000	-	0.00
feprazone	−3.284	-	32.64	−1.000	-	0.00	thenyldiamine	−3.194	-	33.12	−1.000	-	0.00
flufenamic acid	−0.007	U	49.96	−1.000	-	0.00	timolol	−0.315	U	48.33	−0.999	-	0.07
flunitrazepam	−0.834	-	45.59	−0.976	-	1.19	tolbutamide	−0.516	-	47.27	−1.000	-	0.00
fluoresone	0.635	+	53.36	−0.967	-	1.64	tolfenamic acid	−3.664	-	30.63	−1.000	-	0.00
flurbiprofen	−0.575	-	46.96	−1.000	-	0.00	tolmetin	−1.656	-	41.25	−0.999	-	0.05
furazabol	−2.144	-	38.67	−1.000	-	0.00	toloxatone	−1.762	-	40.69	−1.000	-	0.00
glibornuride	0.381	+	52.01	−0.998	-	0.12	tolpropamine	−3.713	-	30.37	−1.000	-	0.00
gliquidone	−1.150	-	43.92	−1.000	-	0.00	triamterene	−1.937	-	39.76	−0.989	-	0.54
glybuzole	1.674	+	58.85	−0.949	-	2.55	triazolam	−5.977	-	18.41	−1.000	-	0.00
glymidine	−2.144	-	38.67	−1.000	-	0.00	trimazosin	0.057	+	50.30	−1.000	-	0.01
glypinamide	−1.863	-	40.15	−1.000	-	0.00	tritoqualine	0.500	+	52.64	−0.958	-	2.11
guanadrel	−2.248	-	38.12	−1.000	-	0.00	vinbarbital	−1.343	-	42.90	−1.000	-	0.00

	discriminant function (%)	neural network (%)	discriminant function (%)	neural network (%)
undetermined (U)	3.29	0.00	adjusted accuracy (excluded undetermined)	89.90 98.68
false activity	9.87	1.32		
overall accuracy	86.84	98.68		

Table 5. (a) Results Obtained for 64 Different Compounds with Antibacterial Activity Used as Active Group Test in the Linear Discriminant Analysis and the Neural Network Package and (b) Results Obtained for 64 Different Compounds without Antibacterial Activity Used as Inactive Group Test in the Linear Discriminant Analysis and the Neural Network Package

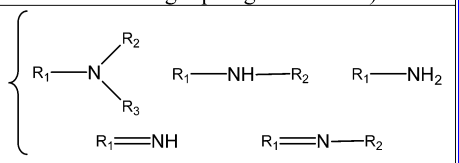
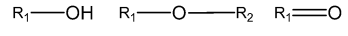
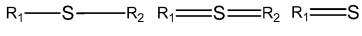
	discriminant function			neural network				discriminant function			neural network		
compound	ΔP value	classif	prob %	output value	classif	prob %	compound	ΔP value	classif	prob %	output value	classif	prob %
						(a)							
amdinocillin	1.624	+	58.58	1.000	+	100	nalidixic acid	−0.920	-	45.14	1.000	+	100
amoxicillin	3.728	+	69.71	0.998	+	99.92	nifuratel	−1.143	-	43.96	1.000	+	100
apicycline	6.474	+	84.22	1.000	+	100	nifurtoinol	−0.721	-	46.19	1.000	+	100
arbekacin	5.082	+	76.86	1.000	+	100	oxolinicacid	−0.916	-	45.16	−1.000	-	0
azidocillin	4.185	+	72.12	1.000	+	100	penamecillin	1.685	+	58.91	−1.000	-	0
azithromycin	3.661	+	69.35	1.000	+	100	penicillin O	3.624	+	69.15	0.996	+	99.82
benzylpenicillinic acid	2.106	+	61.13	0.999	+	99.97	phthatylsulfathiazole	1.464	+	57.74	1.000	+	100
carumonam	6.128	+	82.39	1.000	+	100	pipemidic acid	−1.459	-	42.29	0.992	+	99.61
cefamandole	4.962	+	76.23	1.000	+	100	pivampicillin	3.931	+	70.78	1.000	+	100
cefditoren	6.529	+	84.51	1.000	+	100	ribostamycin	4.727	+	74.98	1.000	+	100
cefmenoxime	8.611	+	95.52	1.000	+	100	rifampin	5.944	+	81.42	1.000	+	99.98
cefoperazone	8.558	+	95.23	1.000	+	100	rokitamycin	1.039	+	55.49	0.997	+	99.84
cefotiam	7.820	+	91.34	1.000	+	100	roxithromycin	2.771	+	64.64	0.999	+	99.96
cefprozil	1.899	+	60.04	1.000	+	100	sparfloxacin	5.840	+	80.87	1.000	+	100
ceftezole	7.040	+	87.21	1.000	+	100	sulbenicillin	4.573	+	74.17	1.000	+	100
ceftriaxone	8.776	+	96.39	1.000	+	100	sulfabenzamide	−0.412	-	47.82	1.000	+	100
cephacetrile sodium	2.432	+	62.85	1.000	+	100	sulfacytine	0.525	+	52.77	1.000	+	100
cephalosporinc	2.641	+	63.96	1.000	+	100	sulfadimethoxine	0.474	U	52.51	1.000	+	100
clometocillin	1.438	+	57.60	1.000	+	100	sulfaguanole	0.942	+	54.98	0.999	+	99.97
diathymosulfone	−0.802	-	45.76	0.989	+	99.43	sulfamerazine	0.159	U	50.84	1.000	+	100
dicloxacillin	−1.348	-	42.88	0.944	+	97.18	sulfamethomidine	0.525	+	52.77	1.000	+	100
doxycycline	5.028	+	76.58	1.000	+	100	4-sulfanilamidosalicylic acid	0.324	U	51.71	1.000	+	100
enoxacin	1.355	+	57.16	1.000	+	100							
N2-formylsulfisomidine	0.523	+	52.76	1.000	+	100	p-sulfanilylbenzylamine	0.522	+	52.76	1.000	+	100
fortimicin A	4.257	+	72.50	1.000	+	100	N-sulfanilyl-3,4-xylamide	0.430	U	52.27	1.000	+	100
grepafloxacin	1.722	+	59.10	1.000	+	100							
hetacillin	3.311	+	67.50	−0.409	-	29.55	sulfaproxyline	0.214	U	51.13	1.000	+	100
lincomycin	1.541	+	58.14	1.000	+	100	sulfasomizole	1.989	+	60.51	1.000	+	100
meropenem	4.671	+	74.69	1.000	+	100	sulfisomidine	0.787	+	54.16	1.000	+	100
methacycline	5.028	+	76.58	1.000	+	100	tetroxoprim	−0.528	-	47.21	0.943	+	97.16
methenamine	−2.134	-	38.72	1.000	+	100	thiamphenicol	−0.154	-	49.18	1.000	+	100
methicillinsodium	3.263	+	67.24	1.000	+	100	ticarcillin	4.624	+	74.44	1.000	+	100
micronomicin	3.247	+	67.16	1.000	+	100	trovafloxacin	3.855	+	70.37	1.000	+	100

	discriminant function (%)	neural network (%)		discriminant function (%)	neural network (%)
undetermined (U)	7.81	0.00	adjusted accuracy (excluded undetermined)	81.36	95.31
false inactivity	17.19	4.69			
overall accuracy	75.00	95.31			

compound	ΔP value	classif	prob %	output value	classif	prob %	compound	ΔP value	classif	prob %	output value	classif	prob %
						(b)							
aminopyrine	−1.554	-	41.78	0.042	+	52.09	mefenamic acid	−2.447	-	37.07	−0.997	-	0.17
aprobarbital	−0.975	-	44.85	−1.000	-	0.00	mepindolol	−1.563	-	41.74	0.252	+	62.61
aranidipine	0.139	+	50.73	−1.000	-	0.00	meptazinol	−3.288	-	32.62	−1.000	-	0.00
arbutin	−0.916	-	45.16	1.000	+	100	methotrimeprazine	−0.580	-	46.93	−1.000	-	0.00
bamipine	−4.665	-	25.34	−1.000	-	0.00	methyldiphenhydramine-p	−3.610	-	30.92	−1.000	-	0.00
bunazosin	−2.408	-	37.27	−1.000	-	0.00							
bupropion	−2.455	-	37.02	−1.000	-	0.00	methyprylon	−1.292	-	43.17	−1.000	-	0.00
chloralose-alfa	−3.571	-	31.13	−0.999	-	0.04	moclobemide	−4.373	-	26.89	−1.000	-	0.00
chlorpropamide	−0.757	-	46.00	−0.997	-	0.14	nilvadipine	1.537	+	58.13	−1.000	-	0.00
chlortalidone	−2.319	-	37.74	−1.000	-	0.00	noxiptilin	−2.882	-	34.77	−1.000	-	0.00
cicletanine	−2.901	-	34.67	−1.000	-	0.01	*p*-bromoacetanilide	−1.554	-	41.78	−1.000	-	0.00
clemastine	−4.485	-	26.29	−1.000	-	0.00	phenazopyridinehydrochloride	−2.565	-	36.44	1.000	+	100
clinofibrate	−5.124	-	22.91	−1.000	-	0.00							
clobenzepam	−3.880	-	29.49	−1.000	-	0.00	phenyltoloxamine	−3.978	-	28.97	−1.000	-	0.00
clometacin	−2.001	-	39.42	−1.000	-	0.00	pildralazine	−1.142	-	43.97	1.000	+	100
dezocine	−2.815	-	35.12	−1.000	-	0.00	pirazolac	−3.480	-	31.60	−1.000	-	0.00
dimetacrine	−2.933	-	34.50	−1.000	-	0.00	propallylonal	−0.396	U	47.91	−1.000	-	0.00
droxicam	1.747	+	59.24	1.000	+	100	propizepine	−2.297	-	37.86	−0.998	-	0.09
ebastine	−6.192	-	17.27	−1.000	-	0.00	quinethazone	−2.662	-	35.93	−1.000	-	0.00
enalaprilat	−1.973	-	39.57	−1.000	-	0.00	rescinnamine	−2.496	-	36.81	−1.000	-	0.00
ethchlorvynol	−3.224	-	32.96	−1.000	-	0.00	salicylamide0-acetic acid	−1.499	-	42.07	−1.000	-	0.00
ethylmethylthiambutene	−0.162	U	49.14	−1.000	-	0.00							
etozolin	0.270	+	51.42	−1.000	-	0.00	simfibrate	−4.972	-	23.72	−1.000	-	0.00
gemfibrozil	−1.973	-	39.57	−1.000	-	0.00	sufentanil	−2.990	-	34.20	−1.000	-	0.00
glafenine	−4.303	-	27.26	0.959	+	97.97	syrosingopine	−0.545	-	47.12	−1.000	-	0.01
glimepiride	−0.894	-	45.27	−1.000	-	0.01	thonzylamine	−3.823	-	29.79	−1.000	-	0.00
guanazodine	−2.717	-	35.64	−1.000	-	0.00	tilisolol	−0.880	-	45.35	1.000	+	100
heptabarbital	−2.603	-	36.24	−1.000	-	0.00	tofenacin	−3.978	-	28.97	−1.000	-	0.00
isothipendyl	−1.310	-	43.08	−1.000	-	0.00	tolazamide	−1.254	-	43.37	−0.999	-	0.05
isoxepac	−2.654	-	35.97	−1.000	-	0.00	tramadol	−2.605	-	36.23	−1.000	-	0.00
lofepramine	−5.906	-	18.78	−1.000	-	0.00	tripamide	−3.159	-	33.30	−0.683	-	15.86
lofexidine	−6.522	-	15.52	−1.000	-	0.00	vinylbital	−1.343	-	42.90	−1.000	-	0.00
lormetazepam	−5.174	-	22.65	−1.000	-	0.00	ximoprofen	−2.343	-	37.61	−1.000	-	0.00
l-tryptophan	−2.079	-	39.01	−1.000	-	0.00							

	discriminant function (%)	neural network (%)		discriminant function (%)	neural network (%)
undetermined (U)	3.13	0.00	adjusted accuracy (excluded undetermined)	93.55	87.50
false activity	6.25	12.50			
overall accuracy	90.63	87.50			

Table 6. Topological Indices with a Positive Contribution to the Discriminant Function and Functional Chemical Groups Containing Them

Indices with a positive contribution to the Discriminant Function	Chemical Groups (functional groups containing topological indices)
D^{N8}	
D^{O5}	
D^{S2}	
D^{F5}	

the undetermined molecules were ruled out, while the MLP percentage remained constant.

It should be pointed out that the accuracy percentage was fulfilled for both methods, although it was better for the neural network. The topological indices proposed are effective in the discrimination of complex pharmacological properties, such as antibacterial activity.

The discrimination of activity carried out shows that the values obtained for LDA and MLP on the antibacterial and nonantibacterial groups make it possible to separate both populations. Figures 2 and 3 show the histogram of frequencies (pharmacological distribution diagram)[28] obtained by depicting all LDA values and the output values of MLP. It can be seen that the top E_i (inactivity expectancy) and E_a (activity expectancy) values are distributed on both sides of $\Delta P = 0$ for LDA, and output value = 0 for MLP.

Positive LDA values were obtained for active compounds, with a maximum ΔP value around 1 and negative LDA values for inactive compounds with a maximum ΔP value around -3, in both the training and test groups.

Positive MLP values were obtained for active compounds with maximum output value 1 and negative MLP values for inactive compounds with maximum output value -1, in both the training and test groups. However, the selection of active molecules with ΔP and output value > 0.5 ensures the correct discrimination of this kind of molecules.

The good results achieved by both discrimination methods allow us to implement a combined use in selecting molecules with highly effective antibacterial activity: first performing a guided search for molecules meeting the topological requirements in the discriminant function obtained (LDA) and then selecting them through the previously trained neural network.

Figure 4 shows the line corresponding to the average index values of active and inactive molecules, called the molecular spectrum.

The molecular spectrum shows a different profile on the antibacterial and nonantibacterial groups, with clearly differentiated zones for certain indices, this being the case with the discriminant function variables (V^2, D^{N8}, D^{O5}, D^{S2}, D^{F5}, and D^{Cl4}). Indices pointing to the presence of heteroatoms (D^{N8}, D^{O5}, D^{S2}, and D^{F5}) positively contribute to the ΔP value; as shown by the molecular spectrum, the line corresponding to the molecules with antibacterial activity is above that of the compounds without this activity. Likewise, in the molecular spectrum, indices V^2 and D^{Cl4}, with a negative contribution to the ΔP value, have higher values for the compounds without antibacterial activity.

Table 4 shows the molecular fragments derived from the four indices which positively contribute to the discriminant function; they have been used as parameters in the search for new structures with theoretical antibacterial activity in a large chemical database[29] of molecules without known therapeutic activity. The molecules were classified by applying the discriminant function first and the neural network afterward, only those molecules that made it through both filters being considered active (ΔP and output value $>$ 0.5).

As a result, the search provided the 10 molecules included in Table 5; they were all applied the corresponding antibacterial activity tests in order to determine the minimal inhibitory concentration (MIC) versus strains of *Enterococcus faecalis* ATCC 29212, *Staphylococcus aureus* ATCC 6538P, *Escherichia coli* ATCC 10536, and *Pseudomonas aeruginosa* ATCC 27853, using microtiter plates where serial dilutions were made with a liquid culture medium (tripticase-soy broth).

The results obtained for the assayed products (see Table 5), which resulted in four of them being active. C.A.S. product 16707-41-8 is simultaneously active with *Escherichia coli* ATCC 10536, *Staphylococcus aureus* ATCC 6538P, and *Enterococcus faecalis* ATCC 29212, with MIC values that are clearly better than those of the reference drugs. C.A.S. product 13676-54-5 is active with *Staphylococcus aureus* ATCC 6538P and *Enterococcus faecalis* ATCC 29212; C.A.S. compound 2411-89-4 is active with *Escherichia coli* ATCC 10536 and *Pseudomonas aeruginosa* ATCC 27853, whereas the C.A.S. product 23911-25-3 seems to be active with *Enterococcus faecalis* ATCC 29212 and *Escherichia coli* ATCC 10536.

The lethal dose (LD_{50}) of all four products with antibacterial activity (see Table 6) evidenced their low toxicity, with LD_{50} values far higher than those obtained for the MICs of such products, this being the case with C.A.S. compound 16707-41-8. LD_{50} values for this compound are 10−20 times greater than the MIC values obtained with *Enterococcus faecalis* and *Staphylococcus aureus*. The same can be said about C.A.S. compound 13676-54-5, with LD_{50} values between 3 and 6 times higher than its MIC values for *Enterococcus faecalis* and *Staphylococcus aureus*.

Table 7. M.I.C. Values versus Different Microorganism Strains for the Selected Molecules by the Discriminant Function and Neural Network Trained and Reference Drugs

molecule	M.I.C. (μg/mL)			
	S. aureus	*E. faecalis*	*E. coli*	*P. aeruginosa*
quinoline yellow (C.A.S. 8004-92-0)	>1000	>1000	>1000	>1000
direct yellow 27 (C.A.S. 10190-68-8)	>1000	>1000	>1000	>1000
N-[4-(2-benzoxazolyl) phenyl]maleimide (C.A.S. 16707-41-8)	15.7	31.3	500	>1000
methylthymol blue, sodium salt (C.A.S. 1945-77-3)	>1000	>1000	>1000	>1000
1,1′-(methylenedi-4,1-phenylene) bismaleimide (C.A.S. 13676-54-5)	125	250	>500	>500
glycine, N,N′-[3H-2,1-benzoxathiol-3-ylidenebis[(6-hydroxy-5-methyl-3, 1-phenylene)methylene]]bis[N-(carboxymethyl)-, S,S-dioxide, tetrasodium salt (C.A.S. 3618-43-7)	>1000	>1000	>1000	>1000
2′,3′-di-O-benzoyluridine (C.A.S. 50408-20-3)	>1000	>1000	>1000	>1000
o-cresolphthalein complexone (C.A.S. 2411-89-4)	>1000	>1000	1000	1000
ethylenediaminetetraacetic dianhydride (C.A.S. 23911-25-3)	>1000	1000	500	>1000
5-notroorotic acid potasium, salt monohydrate (C.A.S. 60779-49-9)	>1000	>1000	>1000	>1000
cephalosporin C	31.3	>1000	31.3	>500
nalidixic acid	62.5	1000	15.7	500

Table 8. Lethal Dose 50 (LD_{50}) for the Molecules with Antibacterial Activity

molecule	LD_{50} (mg/kg)
o-cresolphthalein complexone (C.A.S. 2411-89-4)	>1500
ethylenediaminetetraacetic dianhydride (C.A.S. 23911-25-3)	<1000
N-[4-(2-benzoxazolyl)phenyl]maleimide (C.A.S. 16707-41-8)	313 ± 51
1,1′-(methylenedi-4,1-phenylene) bismaleimide (C.A.S. 13676-54-5)	713 ± 368

4. CONCLUSIONS

The computer package applied was developed for the calculation of 62 topological-structural indices in a series of molecules with and without antibacterial activity. Classification results verify the discriminative ability of the topological descriptors proposed and suggest that the molecular descriptor set adequately captures the phenomena underlying compound properties. These indices can consequently constitute a viable alternative to the use of molecular descriptors in QSAR studies.

The molecular spectrum obtained for molecule groups with and without antibacterial activity evidences the existence of descriptor sets with higher values for antibacterial activity structures, these descriptors also being present in the discriminative function with a positive coefficient.

The guided structure search and the subsequent use of the discriminant function and a trained neural network as successive filters allowed us to select those molecules with an active response to different microorganism strains, with highly satisfactory MIC and LD_{50} values. This validates the proposed method, which could be used as a very effective drug design tool.

ACKNOWLEDGMENT

The authors wish to thank José María Valls Sánchez, Computer Science Graduate, for the software developed for the calculation indices, and *Generalitat Valenciana* (Regional Government), GV01-48 (Science and Technology Office) for their financial support.

REFERENCES AND NOTES

(1) Darvas, F.; Erdos, I.; Teglas, G. *QSAR in Drug Design and Toxicology*; Elsevier: Amsterdam, 1987.

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

(3) Basak, S. C.; Bertelsen, S.; Grunwald, G. Application of Graph Theoretical Parameters in Quantifying Molecular Similarity and Structure−Activity Studies. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 270−276.

(4) Balaban, A. T.; Motoc, I.; Bonchev, D.; Mekenyan, O. Topological Indices for Structure−Activity Correlations. *Top. Curr. Chem.* **1983**, *114*, 21−55.

(5) Trinajstic, N. *Chemical Graph Theory*; CRC Press: Boca Raton, FL, 1983.

(6) Balaban, A. T. Applications of Graph Theory in Chemistry. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 334−343.

(7) Stankevich, M. I.; Stankevich, I. V.; Zefirov, N. S. Topological Indexes in Organic Chemistry *Usp. Khimii* **1988**, *57*, 337−366.

(8) Rouvray, D. H. Should We Have Designs on Topological Indexes? *Stud. Phys. Theor. Chem.* **1983**, *28*, 159−177.

(9) Balaban, A. Chemical Graphs. XXXIV. Five New Topological Indices for the Branching of Tree-like Graphs. *Theor. Chim. Acta* **1979**, *53*, 355−375.

(10) Jaen-Oltra, J.; Salabert-Salvador, M. T.; García-March, F. J.; Pérez-Giménez, F.; Tomás-Vert, F. Artificial neural network applied to prediction of fluorquinolone antibacterial activity by topological methods. *J. Med. Chem.* **2000**, *43*, 1143−1148.

(11) Tomás-Vert, F.; Pérez-Giménez, F.; Salabert-Salvador, M. T.; García-March, F. J.; Jaén-Oltra, J. Artificial neural network applied to the discrimination of antibacterial activity by topological methods. *J. Mol. Struct. (THEOCHEM)* **2000**, *504*, 249−259.

(12) Murcia-Soler, M.; Pérez-Giménez, F.; García-March, F. J.; Salabert-Salvador, M. T.; Díaz-Villanueva, W.; Medina-Casamayor, P. Dis-

crimination and selection of new potential antibacterial compounds using simple topological descriptors. *J. Mol. Graph. Model.* **2003**, *21*, 375−390.
(13) Kovesdi, I.; Dominguez-Rodriguez, M. F.; Orfi, L.; Naray-Szabo, G.; Varro, A.; Papp, J. G.; Matyus, P. Application of neural networks in structure−activity relationships. *Med. Res. Rev.* **1999**, *19*, 249−269.
(14) Jalali-Heravi, M.; Parastar, F. Use of artificial neural networks in QSAR study of anti-HIV activity for a large group of HEPT derivatives. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 147−154.
(15) Manallack, D. T.; Livingstone, D. J. Neural networks in drug discovery: have they lived up to their promise? *Eur. J. Med. Chem.* **1999**, *34*, 195−208.
(16) Tetko, I. V.; Kovalishyn, V. V.; Livingstone, D. J. Volume Learning Algorithm Artificial Neural Networks for 3D QSAR Studies. *J. Med. Chem.* **2001**, *44*, 2411−2420.
(17) Micheli, A.; Sperduti, A.; Starita, A. Analysis of the Internal Representations Developed by Neural Networks for Structures Applied to Quantitative Structure−Activity Relationship Studies of Benzodiazepines. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 202−218.
(18) Devillers, J. *Neural Networks in QSAR and Drug Design*; Academic Press: London, 1996.
(19) Viswanadhan, V. N.; Mueller, G. A.; Basak, S. C.; Weinstein, J. N. Comparison of a Neural Net-Based QSAR Algorithm (PCANN) with Hologram and Multiple Linear Regression-Based Calcium Channel Antagonists. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 505−511.
(20) *The Merck Index*, 12th ed. on CD-Rom; Budavari, S., Ed.; 1996.
(21) Valls Sánchez, J. M. Drug Design by molecular topology: graphical interface development, Graduation Project, Computer Science Engineer Graduate, University of Valencia, 2001.
(22) Murcia-Soler, M.; Pérez-Giménez, F.; García-March, F. J.; Salabert-Salvador, M. T.; Díaz-Villanueva, W.; Castro-Bleda, M. J. Drugs and Nondrugs: An Effective Discrimination with Topological Methods and Artificial Neural Networks. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1688−1702.
(23) Randić, M.; Kleiner, A. F.; DeAlba, L. M.; Distance/Distance Matrices. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 277−286.
(24) BMDP Statistical Program, University of California, Los Angeles, 1990.
(25) Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. Learning internal representations by error propagation. In *PDP: Computational models of cognition and perception*; Rumelhart, D. E., McClelland, J. L., Eds.; I. MIT Press: 1986.
(26) Bishop, C. M. *Neural Networks for Pattern Recognition*; Oxford University Press: 1995.
(27) Zell, A.; et al. *SNNS Stuttgart Neural Network Simulator V 4.2, Institute for Parallel and Distributed High Performance Systems (IPVR)*; University of Stuttgart, 1995.
(28) Gálvez, J.; García-Domenech, R.; de Gregorio Alapont, C.; de Julián-Ortiz, J. V.; Popa, L. Pharmacological distribution diagrams: A tool for *de novo* drug design. *J. Mol. Graphics* **1996**, *14*, 272−276.
(29) Available Chemicals Directory is available from MDL Information Systems Inc., San Leandro, CA.
(30) Dawson, C.; Belloch, C.; G×bb-López, M. D.; Uruburu, F. *Spanish Type Culture Collection, Catalogue of strains*, 3rd ed.; Department of Microbiology, University of Valencia: 1990.

CI030340E