Drugs and Nondrugs: An Effective Discrimination with Topological Methods and Artificial Neural Networks

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

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 March 26, 2003

A set of topological and structural descriptors has been used to discriminate general pharmacological activity. To that end, we selected a group of molecules with proven pharmacological activity including different therapeutic categories, and another molecule group without any activity. As a method for pharmacological activity discrimination, an artificial neural network was used, dividing molecules into active and inactive, to train the network and externally validate it. The following plot frequency distribution diagrams were used: a function of the number of drugs within a value interval, 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 drug and nondrug molecules. The results confirmed the discriminative capacity of the topological descriptors proposed.

1. INTRODUCTION

Because of the high costs of drug development, the major aim of the pharmaceutical industry has shifted from the trial-and-error process of drug discovery to a rational, structure-based drug design. There are many potential drug candidates that present good results for in vitro assays but fail when subjected to in vivo assays. Consequently, being able to predict whether a chemical compound is "drug-like" or "nondrug-like" could be a useful tool for the selection and development of new potential drug candidates.

A successful and reliable drug design process could reduce the time and cost of developing useful pharmacological agents. Computational methods are used for the prediction of "drug-likeness", which is nothing but the identification and elimination of candidate molecules that are unlikely to survive the later stages of discovery and development. Drug-likeness could be predicted by a neural network-based approach.[1]

There is growing interest in the application of neural network systems to a wide range of chemical problems to predict whether a chemical compound is drug-like or nondrug-like. Ajay et al. used a Bayesian neural network to distinguish between drugs and nondrugs, evidencing the models' generalization ability, since 80% of the molecules in the MACCS-II Drug Data Report (MDDR) were classified as drug-like.[2] Sadowski and Kubinyi developed a scoring scheme for the rapid and automatic classification of molecules into drugs and nondrugs. The approach revealed certain features in molecules that either qualify or disqualify them as drugs.[3] More recently, Frimurer et al. used a feed-forward neural network technique to classify chemical compounds into potentially drug-like and nondrug-like compounds, with 88% success in both MDDR and the Available Chemicals Directory (ACD).[4]

A graph-theoretical approach to the problem of discriminating drugs from nondrugs was developed by Gálvez et al.,[5] trying to avoid the bias caused by different molecular sizes and complexity between the two groups, and discriminating particular activity among compounds recognized as drugs. The method was set up by using graph-theoretical descriptors such as subgraph Randic−Kier−Hall indices, topological charge indices, quotients of connectivity indices, Wiener path number, etc., and linear discriminant analysis was used to find classification functions by linear combination of descriptors. Classification successes of 83.1% and 80.8% were obtained for drugs and nondrugs, respectively, showing that it is possible to achieve a pattern of general pharmacological activity based on molecular topology.

In graph theory, molecular structures are represented as hydrogen-suppressed molecular graphs whose vertices and edges represent skeletal atoms and chemical bonds, respectively. Graph theory has been applied largely to the characterization of chemical structures, as well as to structure−property and structure−activity correlations, by means of the so-called topological indices. These topological indices are numerical steric descriptors that contain encoded information on the number of atoms and their structural environment derived from the hydrogen-suppressed molecular formula.[6] At present, the use of these descriptors covers most of the main research areas of drug development: virtual screening, drug design, combinatorial library design, QSAR, structure-pharmacokinetics, structure−toxicity relationships, etc.[7]

* Corresponding author phone: (34) 963 54 48 94; fax: (34) 963 54 48 92; e-mail: Facundo.Perez@uv.es.
[†] Department of Physical Chemistry, Universitat de València.
[‡] Department of Computer Science, Universitat de València.
[§] Universitat Politècnica de València.

10.1021/ci0302862 CCC: $25.00
Published on Web 07/08/2003

Our focus is to apply artificial neural networks as a method for discriminating drugs from nondrugs, using a set of topological descriptors as simple integers applied to individual atoms and bonds in molecules. The important common feature in all such descriptors is the independence of their numerical values in renumbering atoms in a chemical structure.[8−11] The descriptors encode information on atom type, bonds, degree vertex, distances between pairs of atoms, etc., and so constitute an alternative to the use of molecular descriptors in drugability studies, not only for the calculation process but also for a simpler interpretation.[12−14]

For both the drug group including the different therapeutic categories and the nondrug group, a sufficiently heterogeneous molecule set was used so as to achieve considerable structural diversity.

2. METHODS

Developed in our research unit and having proved highly effective in discriminating specific pharmacological features such as antibacterial activity,[12−14] a set of topological and structural indices[12] was used for the discrimination of a more general property, namely pharmacological activity without distinction of specific therapeutic activity. The chosen set of molecular descriptors should adequately capture the phenomena underlying the properties of the compound. It is also important for descriptors to be obtained without much computational effort, as they have to be computed for every molecule whose properties need to be discriminated, as is the case with molecules with pharmacological 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 to calculate the topological indices;[15] this saved time and computer memory in comparison with matrixes 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.

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

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

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

$$\delta_i = \sum_{j=1}^{n} a_{ij} \quad (2)$$

Distance is the length of the shortest path, d_{ij}, between the vertices in the graph,

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

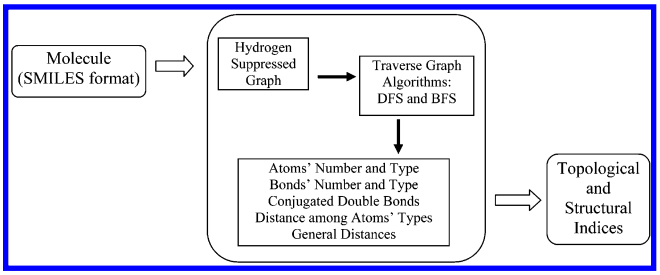

Figure 1. Calculation diagram of topological indices from molecular SMILES representation.

where D_{ij} is the number of steps in the shortest path (i.e., the minimum number of edges) in a graph between vertices i and j.[16]

Figure 1 shows a chart of the calculation of topological indices from the molecular SMILES representation. A total of 62 indices were selected. Of these, 14 included 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 (simple, 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 contained different topological data, such as the number of double bonds at distance 1 or 2 and the 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. Table 1 lists the 62 topological indices and their descriptions.

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

To obtain bond type and number atoms, a depth graph search is carried out by means of a DFS algorithm, which increases 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 DFS followed by a BFS starting from each atom type.

The indices do not only meet a condition common to all topological indices, i.e., being graph invariants, but also offer a second property: they are simple integers.[12−14] This is advantageous not only when index values are correlated with the physicochemical and pharmacological properties of a set of molecules, but also when it comes to inverting the direction of the calculation and obtaining structures that own the mentioned properties and fulfill the topological requirements by the discriminant function. The above-mentioned

Table 1. Description and Number of Indices in Output Files in Calculation Program

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})

Table 2. Distribution of Drugs and Nondrugs Compounds in Training Test and Validation Sets

therapeutic category	molecules of training group	molecules of test group	molecules of validation group	total
drugs	259	110	61	430
1. analgesic	43	19	8	70
2. antibacterial	61	26	12	99
3. antidepresant	22	8	5	35
4. antidiabetic	5	2	2	9
5. antifungal	15	7	4	26
6. antihypertensive	34	15	9	58
7. antihistaminic	18	7	5	30
8. antiinflammatory	13	5	4	22
9. diuretic	15	7	4	26
10. antihyperlipoproteinemic	11	5	3	19
11. sedative	22	9	5	36
nondrugs	158	67	25	250

methodology allowed us to design or select new structures with the pharmacological activity studied.

Once the whole set of indices to be used was defined, the process of establishing the structure−activity relationships started, using an artificial neural network for this purpose.

The set of molecules used in the classification method is made up of 430 molecules with proven pharmacological activity, within different therapeutic categories, and belonging to different groups of analgesic, antibacterial, antidepressant, antidiabetic, antifungal, antihypertensive, antihistaminic, antiinflammatory, diuretic, antihyperlipoproteinemic, and sedative compounds classified by the Merck Index,[17] versus 250 compounds for which no pharmacological activity was described (see Table 2).

Samples were split into three groups: a training set (approximately 60−65% of the samples), a test set (approximately 25−30% of the samples) and a validation group (approximately 10−15% of the samples). The training set was composed of 259 molecules with antibacterial activity and 158 inactive molecules. The test set comprised 110 active molecules and 67 inactive molecules. The validation set included 61 active molecules and 25 inactive molecules. This group actually constitutes the external validation of the network trained by the training and test sets, as none of these molecules had taken part in the network-training process.

The artificial neural network used as the discrimination method is a multilayer perceptrons (MLPs) feed-forward neural network with a multilayer structure.[18,19] 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 adds up its various weighted inputs until some pre-set 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.

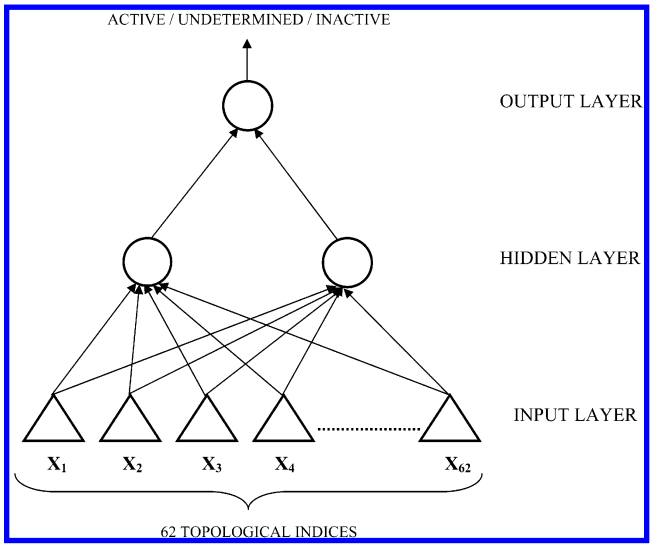

Figure 2. Schematic representation of a multilayer perceptron with one hidden layer.

The number of input units was set by the number of topological descriptors of the molecules (62 topological indices). Input data were discretized by dividing by the maximum value of all the indices.

There was only one output unit corresponding to the property being classified: a +1 value was assigned to the active molecules and a −1 was assigned to the inactive ones. Therefore, we used the hyperbolic tangent function — defined in the interval [−1,1] — as the activation function. Figure 2 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 Simula-

Table 3. Results Obtained for 259 Different Compounds with Pharmacological Activity and 158 Different Inactive Compounds Used as Training Group in the Neural Network Package

compound	therapeutical category	output value	classification	probability %
1. Alfentanil	analgesic	0.99988	+	99.994
2. Benzylmorphine	analgesic	0.99113	+	99.557
3. Desomorphine	analgesic	0.03745	U	51.873
4. Dihydrocodeine	analgesic	0.46379	U	73.190
5. Dimethylthiambutene	analgesic	0.99967	+	99.984
6. Ethoheptazine	analgesic	0.98906	+	99.453
7. Etonitazene	analgesic	0.99323	+	99.662
8. Hydromorphone	analgesic	0.80058	+	90.029
9. Levorphanol	analgesic	0.97340	+	98.670
10. Methadone Hydrochloride	analgesic	0.53674	+	76.837
11. Nalbuphine	analgesic	0.99961	+	99.981
12. Norpipanone	analgesic	0.99981	+	99.991
13. Phenadoxone	analgesic	0.99181	+	99.591
14. Propiram	analgesic	0.98526	+	99.263
15. Tilidine	analgesic	0.98528	+	99.264
16. Acetaminophen	analgesic	−0.76880	−	11.560
17. AcetylsalicylsalicylicAcid	analgesic	0.57202	+	78.601
18. Alclofenac	analgesic	0.99942	+	99.971
19. Antipyrine	analgesic	−0.88254	−	5.873
20. Aspirin	analgesic	0.66778	+	83.389
21. Benoxaprofen	analgesic	0.99739	+	99.870
22. 5-Bromosalicylic Acid Acetate	analgesic	0.60871	+	80.436
23. Butacetin	analgesic	0.99800	+	99.900
24. Ciramadol	analgesic	0.99927	+	99.964
25. Crotethamide	analgesic	0.59461	+	79.731
26. Diflunisal	analgesic	0.94502	+	97.251
27. Epirizole	analgesic	0.58294	+	79.147
28. Etodolac	analgesic	0.96990	+	98.495
29. Flufenamic Acid	analgesic	0.99986	+	99.993
30. Flurbiprofen	analgesic	0.95453	+	97.727
31. Indomethacin	analgesic	0.87789	+	93.895
32. Ketorolac	analgesic	0.42617	U	71.309
33. Lornoxicam	analgesic	0.99990	+	99.995
34. Mofezolac	analgesic	0.91792	+	95.896
35. Nefopam	analgesic	0.99707	+	99.854
36. Parsalmide	analgesic	0.97932	+	98.966
37. Phenopyrazone	analgesic	0.99649	+	99.825
38. Propacetamol	analgesic	0.79548	+	89.774
39. Salicylamide	analgesic	−0.94523	−	2.739
40. Suprofen	analgesic	0.75524	+	87.762
41. Tenoxicam	analgesic	0.99772	+	99.886
42. Tolfenamic Acid	analgesic	0.82224	+	91.112
43. Viminol	analgesic	0.99925	+	99.963
44. Acediasulfone	antibacterial	0.99808	+	99.904
45. Amdinocillin Pivoxil	antibacterial	0.84017	+	92.009
46. Apalcillin	antibacterial	0.99972	+	99.986
47. Aspoxicillin	antibacterial	0.97466	+	98.733
48. Bacampicillin	antibacterial	0.99920	+	99.960
49. Benzylsulfamide	antibacterial	0.58842	+	79.421
50. Butirosin	antibacterial	0.99844	+	99.922
51. Cefatrizine	antibacterial	0.99835	+	99.918
52. Cefazedone	antibacterial	0.99994	+	99.997
53. Cefetamet	antibacterial	0.69595	+	84.798
54. Cefodizime	antibacterial	0.99988	+	99.994
55. Cefpiramide	antibacterial	0.99881	+	99.941
56. Cefpodoxime Proxetil	antibacterial	0.99982	+	99.991
57. Ceftibuten	antibacterial	0.97204	+	98.602
58. Cefuroxime	antibacterial	0.99901	+	99.951
59. Cephapirin Sodium	antibacterial	0.97134	+	98.567
60. Chlortetracycline	antibacterial	0.99477	+	99.739
61. Clarithromycin	antibacterial	0.99512	+	99.756
62. Clomocycline	antibacterial	0.99983	+	99.992
63. Cloxacillin	antibacterial	0.99993	+	99.997
64. Demeclocycline	antibacterial	0.99910	+	99.955
65. Dichloramine T	antibacterial	0.99964	+	99.982
66. Erythromycin	antibacterial	0.99906	+	99.953
67. Floxacillin	antibacterial	0.99995	+	99.998
68. Fortimicin B	antibacterial	0.58222	+	79.111
69. Hexedine	antibacterial	0.36982	U	68.491
70. Imipenem	antibacterial	0.92537	+	96.269
71. Leucomycins	antibacterial	0.99995	+	99.998
72. Lymecycline	antibacterial	0.99994	+	99.997

Table 3 (Continued)

compound	therapeutical category	output value	classification	probability %
73. 4′-(Methylsulfamoyl)sulfanilanilide	antibacterial	0.99632	+	99.816
74. Minocycline	antibacterial	0.99950	+	99.975
75. Nadifloxacin	antibacterial	0.99988	+	99.994
76. Nifuradene	antibacterial	0.54763	+	77.382
77. Nifurfoline	antibacterial	0.99922	+	99.961
78. Nitrofurantoin	antibacterial	0.77189	+	88.595
79. Ofloxacin	antibacterial	0.99989	+	99.995
80. Paromomycin	antibacterial	0.99515	+	99.758
81. Pazufloxacin	antibacterial	0.99991	+	99.996
82. Phenethicillin Potassium	antibacterial	0.77210	+	88.605
83. PiromidicAcid	antibacterial	0.98676	+	99.338
84. Pivcefalexin	antibacterial	0.92610	+	96.305
85. Rifamide	antibacterial	0.99991	+	99.996
86. Rifapentine	antibacterial	0.99995	+	99.998
87. Salazosulfadimidine	antibacterial	0.99980	+	99.990
88. Sancycline	antibacterial	0.75649	+	87.825
89. Spectinomycin	antibacterial	0.99987	+	99.994
90. Sulfachlorpyridazine	antibacterial	0.99890	+	99.945
91. Sulfachrysoidine	antibacterial	0.96265	+	98.133
92. Sulfadoxine	antibacterial	0.96187	+	98.094
93. Sulfaguanidine	antibacterial	0.54603	+	77.302
94. Sulfamethazine	antibacterial	0.57281	+	78.641
95. Sulfamethizole	antibacterial	0.97519	+	98.760
96. Sulfamidochrysoidine	antibacterial	0.78189	+	89.095
97. Sulfanilamide	antibacterial	−0.90138	−	4.931
98. Sulfanitran	antibacterial	0.96978	+	98.489
99. Sulfapyridine	antibacterial	0.96182	+	98.091
100. Sulfathiazole	antibacterial	0.99801	+	99.901
101. Talampicillin	antibacterial	0.99811	+	99.906
102. Taurolidine	antibacterial	0.85412	+	92.706
103. Thiazolsulfone	antibacterial	0.99433	+	99.717
104. Tosufloxacin	antibacterial	0.99995	+	99.998
105. Adinazolam	antidepressant	0.99995	+	99.998
106. Amitriptyline	antidepressant	0.97571	+	98.786
107. Benmoxine	antidepressant	0.48932	U	74.466
108. Butriptyline	antidepressant	0.99146	+	99.573
109. Desipramine	antidepressant	0.99886	+	99.943
110. Dothiepin	antidepressant	0.99845	+	99.923
111. Duloxetine	antidepressant	0.99953	+	99.977
112. Femoxetine	antidepressant	0.25867	U	62.934
113. Imipramine *N*-Oxide	antidepressant	0.99930	+	99.965
114. Indeloxazine Hydrochloride	antidepressant	0.91443	+	95.722
115. Iproniazid	antidepressant	−0.76954	−	11.523
116. Levophacetoperane	antidepressant	0.99935	+	99.968
117. Milnacipran	antidepressant	0.99369	+	99.685
118. Mirtazepine	antidepressant	0.99686	+	99.843
119. Nomifensine	antidepressant	0.99446	+	99.723
120. Oxypertine	antidepressant	0.95360	+	97.680
121. Pizotyline	antidepressant	0.99968	+	99.984
122. Quinupramine	antidepressant	0.74996	+	87.498
123. Tandospirone	antidepressant	0.99877	+	99.939
124. Thozalinone	antidepressant	0.09797	U	54.899
125. Toloxatone	antidepressant	−0.10890	−	44.555
126. Venlafaxine	antidepressant	0.99979	+	99.990
127. Carbutamide	antidiabetic	0.95873	+	97.937
128. Gliclazide	antidiabetic	0.99400	+	99.700
129. Gliquidone	antidiabetic	0.99989	+	99.995
130. Glybuzole	antidiabetic	0.99970	+	99.985
131. Tolbutamide	antidiabetic	0.98091	+	99.046
132. Amphotericin B	antifungal	0.99995	+	99.998
133. Bromosalicylchloranilide	antifungal	0.96891	+	98.446
134. Ciclopirox	antifungal	0.54851	+	77.426
135. Clotrimazole	antifungal	0.88604	+	94.302
136. Exalanide	antifungal	0.78494	+	89.247
137. Fluconazole	antifungal	0.99994	+	99.997
138. Hexetidine	antifungal	0.99550	+	99.775
139. Ketoconazole	antifungal	0.99995	+	99.998
140. Natamycin	antifungal	0.99992	+	99.996
141. Oligomycins	antifungal	0.99995	+	99.998
142. Salicylanilide	antifungal	−0.31001	−	34.500
143. Terbinafine	antifungal	0.99942	+	99.971
144. Tioconazole	antifungal	0.99995	+	99.998
145. Tolnaftate	antifungal	0.88293	+	94.147

Table 3 (Continued)

compound	therapeutical category	output value	classification	probability %
146. Viridin	antifungal	0.99127	+	99.564
147. Alprenolol	antihypertensive	0.90372	+	95.186
148. Amosulalol	antihypertensive	0.99969	+	99.985
149. Benidipine	antihypertensive	0.99994	+	99.997
150. Bevantolol	antihypertensive	0.98932	+	99.466
151. Bunitrolol	antihypertensive	0.79855	+	89.928
152. Cadralazine	antihypertensive	0.79574	+	89.787
153. Carteolol	antihypertensive	0.74557	+	87.279
154. Cetamolol	antihypertensive	0.91693	+	95.847
155. Deserpidine	antihypertensive	0.99936	+	99.968
156. Enalapril	antihypertensive	0.83185	+	91.593
157. Felodipine	antihypertensive	0.99995	+	99.998
158. Guanabenz	antihypertensive	0.99963	+	99.982
159. Guanochlor	antihypertensive	0.99930	+	99.965
160. Hydracarbazine	antihypertensive	−0.88933	−	5.534
161. Irbesartan	antihypertensive	0.99995	+	99.998
162. Labetalol	antihypertensive	0.96285	+	98.143
163. Manidipine	antihypertensive	0.99995	+	99.998
164. Methyldopa	antihypertensive	−0.57395	−	21.303
165. Metipranolol	antihypertensive	0.91104	+	95.552
166. Moveltipril	antihypertensive	0.97837	+	98.919
167. Naftopidil	antihypertensive	0.99962	+	99.981
168. Nicardipine	antihypertensive	0.99994	+	99.997
169. Nipradilol	antihypertensive	0.99235	+	99.618
170. Pargyline	antihypertensive	−0.17818	−	41.091
171. Perindopril	antihypertensive	0.63686	+	81.843
172. Pinacidil	antihypertensive	0.78759	+	89.380
173. Piperoxan	antihypertensive	0.84752	+	92.376
174. Pronethalol	antihypertensive	−0.66656	−	16.672
175. Rescimetol	antihypertensive	0.99989	+	99.995
176. Rilmenidine	antihypertensive	0.22127	U	61.064
177. Sulfinalol	antihypertensive	0.85062	+	92.531
178. Temocapril	antihypertensive	0.99979	+	99.990
179. Tolonidine	antihypertensive	0.99544	+	99.772
180. Urapidil	antihypertensive	0.93827	+	96.914
181. Acrivastine	antihistaminic	0.99915	+	99.958
182. Azatadine	antihistaminic	0.99906	+	99.953
183. Bromodiphenhydramine	antihistaminic	0.99119	+	99.560
184. Carbinoxamine	antihistaminic	0.99976	+	99.988
185. Chlorothen	antihistaminic	0.99976	+	99.988
186. Cinnarizine	antihistaminic	0.99994	+	99.997
187. Clocinizine	antihistaminic	0.99995	+	99.998
188. Deptropine	antihistaminic	0.75904	+	87.952
189. Diphenylpyraline	antihistaminic	0.99925	+	99.963
190. Etymemazine	antihistaminic	0.99824	+	99.912
191. Histapyrrodine	antihistaminic	0.99395	+	99.698
192. Hydroxyzine	antihistaminic	0.99994	+	99.997
193. Loratadine	antihistaminic	0.99988	+	99.994
194. Methapyrilene	antihistaminic	0.99387	+	99.694
195. Orphenadrine	antihistaminic	0.99379	+	99.690
196. Talastine	antihistaminic	0.99869	+	99.935
197. Tripelennamine	antihistaminic	0.92547	+	96.274
198. Tritoqualine	antihistaminic	0.99991	+	99.996
199. Amfenac	antiinflammatory	−0.13436	−	43.282
200. Clidanac	antiinflammatory	0.99935	+	99.968
201. Diclofenac	antiinflammatory	0.99928	+	99.964
202. FenclozicAcid	antiinflammatory	0.99904	+	99.952
203. Fentiazac	antiinflammatory	0.99981	+	99.991
204. Ibuprofen	antiinflammatory	−0.69205	−	15.398
205. Isoxicam	antiinflammatory	0.99625	+	99.813
206. MetiazinicAcid	antiinflammatory	0.89160	+	94.580
207. NiflumicAcid	antiinflammatory	0.99990	+	99.995
208. Oxaprozin	antiinflammatory	0.94922	+	97.461
209. Piroxicam	antiinflammatory	0.98484	+	99.242
210. Suxibuzone	antiinflammatory	0.99992	+	99.996
211. Tolmetin	antiinflammatory	0.50794	+	75.397
212. Althiazide	diuretic	0.99993	+	99.997
213. Amiloride	diuretic	0.62577	+	81.289
214. Bendroflumethiazide	diuretic	0.99995	+	99.998
215. Bumetanide	diuretic	0.99990	+	99.995
216. Clofenamide	diuretic	0.99934	+	99.967
217. Cyclopenthiazide	diuretic	0.99994	+	99.997
218. Ethoxzolanide	diuretic	0.96291	+	98.146

Table 3 (Continued)

compound	therapeutical category	output value	classification	probability %
219. Furosemide	diuretic	0.99992	+	99.996
220. Indapamide	diuretic	0.99975	+	99.988
221. Methazolamide	diuretic	0.95300	+	97.650
222. Metolazone	diuretic	0.99969	+	99.985
223. Perhexiline	diuretic	0.99655	+	99.828
224. Polythiazide	diuretic	0.99994	+	99.997
225. Theobromine	diuretic	−0.86448	−	6.776
226. Xipamide	diuretic	0.99882	+	99.941
227. Beclobrate	antihyperlipoproteinemic	0.99955	+	99.978
228. Binifibrate	antihyperlipoproteinemic	0.99994	+	99.997
229. Clofibrate	antihyperlipoproteinemic	0.99827	+	99.914
230. Eritadenine	antihyperlipoproteinemic	0.24634	U	62.317
231. Fenofibrate	antihyperlipoproteinemic	0.91647	+	95.824
232. Lovastatin	antihyperlipoproteinemic	0.99653	+	99.827
233. Nicofibrate	antihyperlipoproteinemic	0.99971	+	99.986
234. Phenylbutyramide-a	antihyperlipoproteinemic	−0.09969	−	45.016
235. Pirozadil	antihyperlipoproteinemic	0.60201	+	80.101
236. Ronifibrate	antihyperlipoproteinemic	0.99993	+	99.997
237. Triparanol	antihyperlipoproteinemic	0.99977	+	99.989
238. Allobarbital	sedative	0.99275	+	99.638
239. Barbital	sedative	−0.75062	−	12.469
240. Brotizolam	sedative	0.99995	+	99.998
241. Butallylonal	sedative	0.99974	+	99.987
242. Carbubarb	sedative	0.98357	+	99.179
243. Clomethiazole	sedative	0.23057	U	61.529
244. Doxylamine	sedative	0.99955	+	99.978
245. Estazolam	sedative	0.99992	+	99.996
246. Ethinamate	sedative	0.97604	+	98.802
247. Fenadiazole	sedative	0.86837	+	93.419
248. Haloxazolam	sedative	0.99792	+	99.896
249. Homofenazine	sedative	0.99993	+	99.997
250. Mecloxamine	sedative	0.99964	+	99.982
251. Mephobarbital	sedative	0.99807	+	99.904
252. Niaprazine	sedative	0.99789	+	99.895
253. Novonal	sedative	−0.83206	−	8.397
254. Pentobarbital	sedative	0.73870	+	86.935
255. Phenylmethylbarbituric Acid	sedative	0.95075	+	97.538
256. Propiomazine	sedative	0.95125	+	97.563
257. Sulfonmethane	sedative	0.98045	+	99.023
258. Tetrabarbital	sedative	0.99643	+	99.822
259. Vinbarbital	sedative	0.86148	+	93.074
260. Abietic acid	nondrug	−0.91604	−	4.198
261. Acetyleneurea	nondrug	−0.92967	−	3.517
262. *N*-Acetylsulfanilic acid	nondrug	−0.94845	−	2.578
263. Adrenolutin	nondrug	−0.99851	−	0.075
264. Alazopeptin	nondrug	−0.43466	U	28.267
265. Allicin	nondrug	−0.36934	U	31.533
266. Amidochlor	nondrug	0.99993	+	99.997
267. Angelic acid	nondrug	−0.91038	−	4.481
268. Anhalonine	nondrug	−0.90583	−	4.709
269. Aristolochic acid	nondrug	−0.99998	−	0.001
270. Armepavine	nondrug	−0.97391	−	1.305
271. Baptigenin	nondrug	−0.99985	−	0.008
272. Benzenesulfonyl chloride	nondrug	0.93634	+	96.817
273. Benzophenone-6	nondrug	−0.99949	−	0.026
274. Benzoylecgonine	nondrug	−0.48467	U	25.767
275. Bibenzyl	nondrug	−0.53710	−	23.145
276. Boldine	nondrug	−0.99999	−	0.000
277. Bromolysergide	nondrug	−0.63813	−	18.094
278. Bromphenol blue	nondrug	−0.99994	−	0.003
279. Caldariomycin	nondrug	−0.93005	−	3.498
280. Calmagilate	nondrug	−0.99655	−	0.172
281. Canadine	nondrug	−0.99993	−	0.004
282. Casimiroin	nondrug	−0.99027	−	0.486
283. Cetraric acid	nondrug	−1.00000	−	0.000
284. Chimaphilin	nondrug	−1.00000	−	0.000
285. Chlorothymol	nondrug	−0.97248	−	1.376
286. Corybulbine	nondrug	−1.00000	−	0.000
287. Cresol red	nondrug	−1.00000	−	0.000
288. Cyheptadine	nondrug	−0.99422	−	0.289
289. Daidzein	nondrug	−0.99999	−	0.000
290. Desaspidin BB	nondrug	−0.99998	−	0.001
291. Dioxypyramidon	nondrug	0.31586	+	65.793

Table 3 (Continued)

compound	therapeutical category	output value	classification	probability %
292. Duroquinone	nondrug	−1.00000	−	0.000
293. Dypnone	nondrug	−0.95448	−	2.276
294. Ellipticine	nondrug	−1.00000	−	0.000
295. Elliptone	nondrug	−1.00000	−	0.000
296. Equol	nondrug	−0.98728	−	0.636
297. Erdin	nondrug	−0.96837	−	1.582
298. Erythropterin	nondrug	−0.28923	U	35.539
299. Estragole	nondrug	−0.83959	−	8.021
300. Euparin	nondrug	−1.00000	−	0.000
301. Eupatorin	nondrug	−1.00000	−	0.000
302. Evodiamine	nondrug	−0.99653	−	0.173
303. Fagarine	nondrug	−0.99998	−	0.001
304. Fervenulin	nondrug	−0.81359	−	9.321
305. Flindersine	nondrug	−0.20484	U	39.758
306. 9H−Fluorene	nondrug	−0.99761	−	0.120
307. Fraxetin	nondrug	−0.99994	−	0.003
308. Fulvoplumierin	nondrug	−0.99636	−	0.182
309. Furethidine	nondrug	0.99992	+	99.996
310. Furasic acid	nondrug	−0.71500	−	14.250
311. Galangin	nondrug	−1.00000	−	0.000
312. Galegine	nondrug	−0.90137	−	4.932
313. Gallein	nondrug	−0.99627	−	0.187
314. Gentianine	nondrug	−0.72891	−	13.555
315. Gigantine	nondrug	−0.99983	−	0.008
316. Gladiolic acid	nondrug	−1.00000	−	0.000
317. Gravitole	nondrug	−0.16464	U	41.768
318. Guanine	nondrug	−0.85631	−	7.185
319. Harmaline	nondrug	−0.91011	−	4.495
320. Harman	nondrug	−0.99345	−	0.328
321. Hemipyocianine	nondrug	−0.97357	−	1.322
322. Hepaxanthin	nondrug	−0.67757	−	16.122
323. Hexylene glicol	nondrug	−0.79539	−	10.231
324. Holomycin	nondrug	−0.18180	U	40.910
325. Homochelidonine	nondrug	−0.99701	−	0.150
326. Hordenine	nondrug	−0.67278	−	16.361
327. Hydantoin	nondrug	−0.90174	−	4.913
328. Hypoxanthine	nondrug	−0.88672	−	5.664
329. Ichthyopterin	nondrug	−0.54645	−	22.678
330. Indanthrene	nondrug	−1.00000	−	0.000
331. 3-Indolylacetone	nondrug	−0.21472	U	39.264
332. Irisolone	nondrug	−1.00000	−	0.000
333. Isatide	nondrug	−0.98984	−	0.508
334. Isocorypalmine	nondrug	−0.99995	−	0.002
335. Isomethadol	nondrug	−0.69428	−	15.286
336. Javanicin	nondrug	−1.00000	−	0.000
337. Juglone	nondrug	−1.00000	−	0.000
338. Kaempherol	nondrug	−1.00000	−	0.000
339. Kyanmethin	nondrug	−0.95288	−	2.356
340. Laudanine	nondrug	−0.99175	−	0.413
341. Leonurine	nondrug	−0.41525	U	29.238
342. Limettin	nondrug	−0.99935	−	0.032
343. Linatine	nondrug	−0.57093	−	21.454
344. Lophopharine	nondrug	−0.97508	−	1.246
345. Lumazine	nondrug	−0.80574	−	9.713
346. Lumichrome	nondrug	−0.65915	−	17.043
347. Lupulon	nondrug	−0.99245	−	0.377
348. Luteolin	nondrug	−1.00000	−	0.000
349. Macromerine	nondrug	−0.55646	−	22.177
350. Magneson	nondrug	−0.82106	−	8.947
351. Maleuric acid	nondrug	−0.92774	−	3.613
352. Mangostin	nondrug	−0.92578	−	3.711
353. Medicagol	nondrug	−1.00000	−	0.000
354. Mestilbol	nondrug	−0.99281	−	0.360
355. Methopterin	nondrug	0.99857	+	99.929
356. Methylenedigallic acid	nondrug	−1.00000	−	0.000
357. Myricetin	nondrug	−0.99799	−	0.100
358. Nandinine	nondrug	−0.99934	−	0.033
359. Neucoproine	nondrug	−0.99995	−	0.002
360. Nicotelline	nondrug	0.99771	+	99.886
361. Normetanephrine	nondrug	−0.86684	−	6.658
362. Normidulin	nondrug	−1.00000	−	0.000
363. Olivacine	nondrug	−1.00000	−	0.000
364. Oosporein	nondrug	−1.00000	−	0.000

Table 3 (Continued)

compound	therapeutical category	output value	classification	probability %
365. Opianic acid	nondrug	−1.00000	−	0.000
366. Opromazine	nondrug	0.99992	+	99.996
367. Oroxylin A	nondrug	−1.00000	−	0.000
368. Otobain	nondrug	−0.99031	−	0.484
369. Oxenin	nondrug	−0.65125	−	17.438
370. Pantethine	nondrug	−0.32288	U	33.856
371. Papaveraldine	nondrug	−1.00000	−	0.000
372. Patulin	nondrug	−0.94673	−	2.664
373. Pectolinarigenin	nondrug	−1.00000	−	0.000
374. Perylene	nondrug	−1.00000	−	0.000
375. Peyonine	nondrug	−0.76157	−	11.922
376. Phenatine	nondrug	0.90705	+	95.353
377. Phloretin	nondrug	−0.99996	−	0.002
378. Pteroic acid	nondrug	−0.21917	U	39.042
379. Quebrachamine	nondrug	0.99849	+	99.925
380. Retene	nondrug	−1.00000	−	0.000
381. Reticuline	nondrug	−0.95110	−	2.445
382. Rhodopin	nondrug	0.99985	+	99.993
383. Rottlerin	nondrug	−0.95124	−	2.438
384. Rufigallol	nondrug	−1.00000	−	0.000
385. Rutecarpine	nondrug	−0.99538	−	0.231
386. Sakuranetin	nondrug	−1.00000	−	0.000
387. Santalol−alpha	nondrug	−0.98621	−	0.689
388. Santoic acid	nondrug	−1.00000	−	0.000
389. Scoparone	nondrug	−0.99887	−	0.056
390. Scopoletin	nondrug	−0.99049	−	0.476
391. Sikkimotoxin	nondrug	−0.99143	−	0.429
392. Skatole	nondrug	−0.96017	−	1.992
393. Spherophysine	nondrug	−0.52222	−	23.889
394. Stylopine	nondrug	−0.99685	−	0.157
395. Tectorigenin	nondrug	−1.00000	−	0.000
396. Tetrahydropalmatine	nondrug	−1.00000	−	0.000
397. Thymolphthalein	nondrug	−1.00000	−	0.000
398. Tilorone	nondrug	−0.50562	−	24.719
399. Tropacocaine	nondrug	−0.83597	−	8.202
400. Tropic acid	nondrug	−0.79620	−	10.190
401. Tsuduranine	nondrug	−0.99988	−	0.006
402. Tuberin	nondrug	−0.51364	−	24.318
403. Uramil	nondrug	−0.91318	−	4.341
404. Urazole	nondrug	−0.91944	−	4.028
405. Vanilmandelic acid	nondrug	−0.76524	−	11.738
406. Vasicine	nondrug	−0.70199	−	14.901
407. Veratrole	nondrug	−0.95929	−	2.036
408. Versalide	nondrug	−1.00000	−	0.000
409. Violacein	nondrug	−0.79029	−	10.486
410. Visnagin	nondrug	−1.00000	−	0.000
411. Vitamin A	nondrug	−0.97939	−	1.031
412. Xanthopterin	nondrug	−0.80265	−	9.868
413. Xanthoxylin	nondrug	−1.00000	−	0.000
414. Xanthyletin	nondrug	−0.99900	−	0.050
415. Xylenol blue	nondrug	−1.00000	−	0.000
416. Zeatin	nondrug	−0.82244	−	8.878
417. Zingerone	nondrug	−0.85639	−	7.181

	active group (%)	inactive group (%)
undetermined (U)	3.86	6.96
false inactivity or activity	6.56	6.33
overall accuracy	89.58	86.71
adjusted accuracy (excluded undetermined)	93.17	93.20

tor", developed by the University of Stuttgart.[20] To successfully use neural networks, a number of considerations had to be taken into account, such as the network topology, the training algorithm, and the selection of the algorithm's parameters.[18,19] Tests were conducted using different network topologies, increasing the number of hidden units/MLPs with a hidden layer (with 2, 4, 8, 16, and 32). A modified error back-propagation algorithm — standard back-propagation — was used in all cases. Standard back-propagation is the most common learning algorithm. The training process continued as the classification error rate of the validation data decreased (down to a maximum number of training epochs, in this case 10 000 epochs).

The neural network is capable of describing pharmacological activity patterns as well as nonactivity patterns. In other words, this method not only pinpoints active drugs according to their distribution but also identifies inactive compounds. When applied to the discrimination of concrete

Table 4. Results Obtained for 110 Different Compounds with Pharmacological Activity and 67 Different Inactive Compounds Used as Group Test in the Neural Network Package

compound	therapeutical category	output value	classification	probability %
1. Anileridine	analgesic	0.99977	+	99.989
2. Dioxaphetyl Butyrate	analgesic	0.99976	+	99.988
3. Hydrocodone	analgesic	0.19619	U	59.810
4. Meptazinol	analgesic	0.98147	+	99.074
5. Norlevorphanol	analgesic	0.98906	+	99.453
6. Phenazocine	analgesic	−0.67966	−	16.017
7. Sufentanil	analgesic	0.99993	+	99.997
8. Acetaminosalol	analgesic	0.94984	+	97.492
9. Aminopyrine	analgesic	−0.63567	−	18.217
10. Benzpiperylon	analgesic	0.99622	+	99.811
11. Bumadizon	analgesic	0.99937	+	99.969
12. Clometacin	analgesic	0.77230	+	88.615
13. Emorfazone	analgesic	0.67811	+	83.906
14. Fenoprofen	analgesic	0.76722	+	88.361
15. Glafenine	analgesic	0.99982	+	99.991
16. Morpholine	analgesic	−0.89431	−	5.284
17. Phenazopyridine Hydrochloride	analgesic	−0.56739	−	21.631
18. Propyphenazone	analgesic	−0.94631	−	2.685
19. Tramadol	analgesic	0.99846	+	99.923
20. Amdinocillin	antibacterial	−0.55694	−	22.153
21. Apicycline	antibacterial	0.99988	+	99.994
22. Azidocillin	antibacterial	0.99955	+	99.978
23. Benzylpenicillinic Acid	antibacterial	−0.22713	−	38.644
24. Cefmenoxime	antibacterial	0.99511	+	99.756
25. Cefotiam	antibacterial	0.99973	+	99.987
26. Cephacetrile Sodium	antibacterial	0.73553	+	86.777
27. Clometocillin	antibacterial	0.99978	+	99.989
28. Dicloxacillin	antibacterial	0.99995	+	99.998
29. Enoxacin	antibacterial	0.99973	+	99.987
30. Fortimicin A	antibacterial	0.62918	+	81.459
31. Meropenem	antibacterial	0.99731	+	99.866
32. Micronomicin	antibacterial	0.99932	+	99.966
33. Nifuratel	antibacterial	0.99596	+	99.798
34. OxolinicAcid	antibacterial	0.96522	+	98.261
35. Penicillin O	antibacterial	−0.25661	−	37.170
36. PipemidicAcid	antibacterial	0.98777	+	99.389
37. Ribostamycin	antibacterial	0.91088	+	95.544
38. Rokitamycin	antibacterial	0.99995	+	99.998
39. Sulfadimethoxine	antibacterial	0.78606	+	89.303
40. Sulfamerazine	antibacterial	0.79804	+	89.902
41. 4-SulfanilamidosalicylicAcid	antibacterial	−0.33928	−	33.036
42. *N*-Sulfanilyl-3,4-xylamide	antibacterial	−0.28235	−	35.883
43. Sulfasomizole	antibacterial	0.97915	+	98.958
44. Tetroxoprim	antibacterial	0.90606	+	95.303
45. Trovafloxacin	antibacterial	0.99995	+	99.998
46. Amitriptylinoxide	antidepressant	0.99597	+	99.799
47. Demexiptiline	antidepressant	0.99943	+	99.972
48. 5-Hidroxitriptófano(Oxitriptan)	antidepressant	−0.53496	−	23.252
49. Lofepramine	antidepressant	0.99993	+	99.997
50. Noxiptilin	antidepressant	0.99935	+	99.968
51. Prolintane	antidepressant	0.98178	+	99.089
52. Trazodone	antidepressant	0.99973	+	99.987
53. L-Tryptophan	antidepressant	−0.27746	−	36.127
54. Glibenclamide	antidiabetic	0.99995	+	99.998
55. Tolcyclamide	antidiabetic	0.99100	+	99.550
56. Amorolfine	antifungal	0.06459	U	53.230
57. Chlormidazole	antifungal	0.98018	+	99.009
58. Fungichromin	antifungal	0.99984	+	99.992
59. Isoconazole	antifungal	0.99995	+	99.998
60. Nystatin	antifungal	0.99995	+	99.998
61. Sulconazole	antifungal	0.99995	+	99.998
62. Tolindate	antifungal	0.87270	+	93.635
63. Ajmaline	antihypertensive	−0.99976	−	0.012
64. Atenolol	antihypertensive	0.85528	+	92.764
65. Celiprolol	antihypertensive	0.98386	+	99.193
66. Cilnidipine	antihypertensive	0.99991	+	99.996
67. Debrisoquin	antihypertensive	−0.85543	−	7.229
68. Enalaprilat	antihypertensive	0.33293	U	66.647
69. Guanoxabenz	antihypertensive	0.99958	+	99.979
70. Losartan	antihypertensive	0.99995	+	99.998
71. Mepindolol	antihypertensive	0.55543	+	77.772
72. Nadolol	antihypertensive	0.29118	U	64.559

Table 4 (Continued)

compound	therapeutical category	output value	classification	probability %
73. Oxprenolol	antihypertensive	0.96568	+	98.284
74. Propranolol	antihypertensive	0.64455	+	82.228
75. Rescinnamine	antihypertensive	0.99993	+	99.997
76. Terazosin	antihypertensive	0.99981	+	99.991
77. Tilisolol	antihypertensive	0.82178	+	91.089
78. Antazoline	antihistaminic	0.99823	+	99.912
79. Chlorcyclizine	antihistaminic	0.99990	+	99.995
80. Clemastine	antihistaminic	0.99978	+	99.989
81. Epinastine	antihistaminic	0.96125	+	98.063
82. Mequitazine	antihistaminic	0.98373	+	99.187
83. Phenindamine	antihistaminic	−0.84524	−	7.738
84. Pyrilamine	antihistaminic	0.63216	+	81.608
85. Butibufen	antiinflammatory	0.44950	U	72.475
86. Ibuproxam	antiinflammatory	−0.46821	−	26.590
87. Mefenamic Acid	antiinflammatory	−0.99909	−	0.045
88. Pirazolac	antiinflammatory	0.99990	+	99.995
89. Ximoprofen	antiinflammatory	−0.24933	−	37.534
90. Ambuside	diuretic	0.99994	+	99.997
91. Benzthiazide	diuretic	0.99995	+	99.998
92. Chlortalidone	diuretic	0.99462	+	99.731
93. Epithiazide	diuretic	0.99994	+	99.997
94. Hydroflumethiazide	diuretic	0.99995	+	99.998
95. Mefruside	diuretic	0.99993	+	99.997
96. Trichlormethiazide	diuretic	0.99995	+	99.998
97. Benfluorex	antihyperlipoproteinemic	0.99992	+	99.996
98. Bezafibrate	antihyperlipoproteinemic	0.99989	+	99.995
99. Fluvastatin	antihyperlipoproteinemic	0.99866	+	99.933
100. Nicomol	antihyperlipoproteinemic	0.99988	+	99.994
101. Pravastatin Sodium	antihyperlipoproteinemic	0.99847	+	99.924
102. Aprobarbital	sedative	0.92778	+	96.389
103. Capuride	sedative	−0.66301	−	16.850
104. Doxefazepam	sedative	0.99991	+	99.996
105. Heptabarbital	sedative	0.99589	+	99.795
106. Isovaleryl Diethylamide	sedative	−0.83897	−	8.052
107. Mecloqualone	sedative	−0.30823	−	34.589
108. Propallylonal	sedative	0.99924	+	99.962
109. Talbutal	sedative	0.99510	+	99.755
110. Vinylbital	sedative	0.95825	+	97.913
111. Acacetin	nondrug	−1.00000	−	0.000
112. Actiphenol	nondrug	−0.64814	−	17.593
113. Alizarin	nondrug	−1.00000	−	0.000
114. Amsonic acid	nondrug	−0.99921	−	0.039
115. Anserine	nondrug	0.34638	+	67.319
116. Atranorin	nondrug	−1.00000	−	0.000
117. Benzimidazole	nondrug	−0.91739	−	4.131
118. Bergenin	nondrug	−0.98862	−	0.569
119. Bostrycoidin	nondrug	−1.00000	−	0.000
120. Cadalene	nondrug	−0.99999	−	0.000
121. Chavicine	nondrug	0.95402	+	97.701
122. Collinomycin	nondrug	−1.00000	−	0.000
123. Cusparine	nondrug	−0.93960	−	3.020
124. Datiscetin	nondrug	−1.00000	−	0.000
125. Durene	nondrug	−0.99653	−	0.173
126. Echinicrome A	nondrug	−1.00000	−	0.000
127. Enviroxime	nondrug	0.99993	+	99.997
128. Erucic acid	nondrug	0.99958	+	99.979
129. 5-Ethyl-2-picdine	nondrug	−0.93599	−	3.201
130. Fisetin	nondrug	0.75631	+	87.816
131. Formononetin	nondrug	−1.00000	−	0.000
132. Fumigatin	nondrug	−1.00000	−	0.000
133. Fustin	nondrug	−0.90974	−	4.513
134. Galipine	nondrug	−0.99012	−	0.494
135. Gentisin	nondrug	−1.00000	−	0.000
136. Gramine	nondrug	0.12066	+	56.033
137. Hadacidin	nondrug	−0.90695	−	4.653
138. Heliosupine	nondrug	0.99689	+	99.845
139. Honokiol	nondrug	0.19561	+	59.781
140. Hydrobenzoin	nondrug	−0.94408	−	2.796
141. Imperatorin	nondrug	0.92402	+	96.201
142. Irigenin	nondrug	−1.00000	−	0.000
143. Isatropic acid	nondrug	0.51668	+	75.834
144. Isophthalic acid	nondrug	−0.95974	−	2.013
145. Julocrotine	nondrug	−0.23998	U	38.001

Table 4 (Continued)

compound	therapeutical category	output value	classification	probability %
146. Lactarovidin	nondrug	−1.00000	−	0.000
147. Leucopterin	nondrug	−0.93396	−	3.302
148. Lumiflavine	nondrug	−0.83349	−	8.326
149. Maclurin	nondrug	−0.99999	−	0.000
150. Magnolol	nondrug	−0.23774	U	38.113
151. Meconic acid	nondrug	−0.98893	−	0.554
152. Methiotriazamine	nondrug	0.68853	+	84.427
153. Metitepine	nondrug	0.99953	+	99.977
154. Naringenin	nondrug	−1.00000	−	0.000
155. Norbobelanine	nondrug	0.97728	+	98.864
156. Nybomicin	nondrug	−1.00000	−	0.000
157. Osthole	nondrug	−0.98566	−	0.717
158. Pamoic acid	nondrug	−1.00000	−	0.000
159. Parabanic acid	nondrug	−0.92255	−	3.873
160. Pentacene	nondrug	−1.00000	−	0.000
161. Phaseolin	nondrug	−0.99997	−	0.002
162. Physodic acid	nondrug	0.99983	+	99.992
163. Quinizarin green SS	nondrug	−0.99768	−	0.116
164. Rhamnetin	nondrug	−1.00000	−	0.000
165. Rubiadin	nondrug	−1.00000	−	0.000
166. Saponarin	nondrug	0.99098	+	99.549
167. Scutellarein	nondrug	−1.00000	−	0.000
168. Sparsiflorine	nondrug	−0.99877	−	0.061
169. Sudan III	nondrug	0.99977	+	99.989
170. Thidiazuron	nondrug	0.91069	+	95.535
171. Trimellitic anhydride	nondrug	−0.82610	−	8.695
172. Tryptamine	nondrug	−0.43495	U	28.253
173. Vanillic acid	nondrug	−0.98497	−	0.752
174. Viridicatin	nondrug	−0.99687	−	0.157
175. Vitamin K5	nondrug	−0.99983	−	0.008
176. Xanthurenic acid	nondrug	−0.99986	−	0.007
177. Yangonin	nondrug	−0.99985	−	0.008

	active group (%)	inactive group (%)
undetermined (U)	4.55	4.48
false inactivity or activity	19.09	25.37
overall accuracy	76.36	70.15
adjusted accuracy (excluded undetermined)	80.00	73.44

pharmacological actions, we call it a pharmacological distribution diagram (PDD).[21]

A PDD is a frequency distribution diagram of a dependent variable in which the ordinate represents the expectancies of the variable for every interval. Expectancies are defined as the likelihood of a compound to be active or inactive for a value of the discriminant function or output value of the neural network. They are obtained by means of the expressions below, where 100 appears in the denominator to avoid dividing by zero.

Activity expectancy:

$$E_a = \text{Percentage of active molecules/}(\text{Percentage of inactive molecules} + 100)$$

Inactivity expectancy:

$$E_i = \text{Percentage of inactive molecules/}(\text{Percentage of active molecules} + 100)$$

The main advantage of these diagrams is the fact that they allow us to visually determine those property intervals that are more likely to find new active compounds and those that are less prone to find inactive ones.

3. RESULTS AND DISCUSSION

The 62 topological indices corresponding to the 430 drugs with different therapeutic activity and the 250 different nondrug structures were calculated using the methodology presented above. The 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.

The best performance on the validation data was achieved using a MLP with one hidden layer (with two units). Table 3 shows the results obtained, the output value, and its probability for the training set (259 active molecules and 158 inactive molecules).

Overall accuracy for the training set was 89.58% in the active group. The percentage increased to 93.17% when the undetermined molecules were eliminated. In the inactive group, overall accuracy was 86.71% and the percentage increased to 93.20% when the undetermined molecules were ruled out.

Table 5. Results Obtained for 61 Different Compounds with Pharmacological Activity and 25 Different Inactive Compounds Used as a Validation Group in the Neural Network Package

compound	therapeutical category	output value	classification	probability %
1. Codeine	analgesic	0.19621	U	59.811
2. Fentanyl	analgesic	0.99976	+	99.988
3. Piritramide	analgesic	0.99995	+	99.998
4. Aceclofenac	analgesic	0.99931	+	99.966
5. Bromfenac	analgesic	0.78948	+	89.474
6. Enfenamic Acid	analgesic	0.32517	U	66.259
7. Naproxen	analgesic	−0.99975	−	0.012
8. Phenyl Salicylate	analgesic	−0.81187	−	9.407
9. Acetyl Sulfamethoxypyrazine	antibacterial	0.99976	+	99.988
10. Cefadroxil	antibacterial	−0.41182	−	29.409
11. Cefotaxime	antibacterial	0.99508	+	99.754
12. Cinoxacin	antibacterial	0.92276	+	96.138
13. Epicillin	antibacterial	0.85309	+	92.655
14. Josamycin	antibacterial	0.99993	+	99.997
15. Netilmicin	antibacterial	0.99919	+	99.960
16. Panipenem	antibacterial	0.88213	+	94.107
17. Rolitetracycline	antibacterial	0.99988	+	99.994
18. Sulfacetamide	antibacterial	0.68555	+	84.278
19. Sulfametrole	antibacterial	0.94163	+	97.082
20. Sultamicillin	antibacterial	0.99811	+	99.906
21. Citalopram	antidepressant	0.99923	+	99.962
22. Fluvoxamine	antidepressant	0.99914	+	99.957
23. Metapramine	antidepressant	0.89077	+	94.539
24. Propizepine	antidepressant	0.99501	+	99.751
25. Sulpiride	antidepressant	0.59301	+	79.651
26. Glisoxepid	antidiabetic	0.99982	+	99.991
27. Phenbutamide	antidiabetic	0.99779	+	99.890
28. Buclosanide	antifungal	0.93406	+	96.703
29. Fenticonazole	antifungal	0.99995	+	99.998
30. Omoconazole	antifungal	0.99995	+	99.998
31. Sertaconazole	antifungal	0.99995	+	99.998
32. Bethanidine	antihypertensive	−0.81382	−	9.309
33. Carazolol	antihypertensive	0.99330	+	99.665
34. Cilazapril	antihypertensive	0.94455	+	97.228
35. Guanethidine	antihypertensive	−0.78126	−	10.937
36. Ketanserin	antihypertensive	0.99952	+	99.976
37. Lercanidipine	antihypertensive	0.99995	+	99.998
38. Nitrendipine	antihypertensive	0.99922	+	99.961
39. Pildralazine	antihypertensive	−0.85553	−	7.224
40. Timolol	antihypertensive	0.98102	+	99.051
41. Cetirizine	antihistaminic	0.99994	+	99.997
42. Diphenhydramine	antihistaminic	0.99628	+	99.814
43. Fenethazine	antihistaminic	0.97964	+	98.982
44. Promethazine	antihistaminic	0.80472	+	90.236
45. Zolamine	antihistaminic	0.91235	+	95.618
46. Fenbufen	antiinflammatory	−0.74448	−	12.776
47. Feprazone	antiinflammatory	0.99943	+	99.972
48. MeclofenamicAcid	antiinflammatory	0.99642	+	99.821
49. Oxyphenbutazone	antiinflammatory	0.99905	+	99.953
50. Acefylline	diuretic	−0.48233	−	25.884
51. Ethiazide	diuretic	0.99988	+	99.994
52. Meticrane	diuretic	0.99294	+	99.647
53. Torsernide	diuretic	0.99983	+	99.992
54. Acifran	antihyperlipoproteinemic	−0.88186	−	5.907
55. Clomestrone	antihyperlipoproteinemic	−0.99502	−	0.249
56. Niceritrol	antihyperlipoproteinemic	0.99991	+	99.996
57. Acecarbromal	sedative	−0.38947	−	30.527
58. Cyclobarbital	sedative	0.99940	+	99.970
59. Hexethal	sedative	0.03101	U	51.551
60. Nealbarbital	sedative	0.99969	+	99.985
61. Proxibarbal	sedative	0.99128	+	99.564
62. Acetylbutyrolactone-alpha	nondrug	−0.79248	−	10.376
63. Ampyrone	nondrug	−0.93857	−	3.072
64. Benzilic acid	nondrug	0.59871	+	79.936
65. Butyl citrate	nondrug	0.98606	+	99.303
66. Cinnobarine	nondrug	−1.00000	−	0.000
67. Domesticine	nondrug	−0.99437	−	0.282
68. Eriodictyol	nondrug	−0.99994	−	0.003
69. Filicinic acid	nondrug	−0.99182	−	0.409
70. Fuscin	nondrug	−1.00000	−	0.000
71. Glycosine	nondrug	−0.18510	U	40.745
72. Hexazole	nondrug	0.11345	+	55.673

Table 5 (Continued)

compound	therapeutical category	output value	classification	probability %
73. Imidazole	nondrug	−0.85420	−	7.290
74. Isonicotinic acid	nondrug	−0.92952	−	3.524
75. Lepidine	nondrug	−0.98372	−	0.814
76. Lycomarasmine	nondrug	−0.48706	U	25.647
77. Meteloidine	nondrug	−0.94075	−	2.963
78. Noformicin	nondrug	−0.62711	−	18.645
79. Osajin	nondrug	−1.00000	−	0.000
80. Pellotine	nondrug	−0.99992	−	0.004
81. Quinaldic acid	nondrug	−0.97869	−	1.066
82. Safranal	nondrug	−0.99299	−	0.350
83. Sparassol	nondrug	−0.99992	−	0.004
84. Triafur	nondrug	0.42785	+	71.393
85. Usnic acid	nondrug	−0.99991	−	0.005
86. Vitamin A2	nondrug	−0.99727	−	0.136

	active group (%)	inactive group (%)
undetermined (U)	4.92	8.00
false inactivity or activity	18.03	16.00
overall accuracy	77.05	76.00
adjusted accuracy (excluded undetermined)	81.03	82.61

A cross-validation test was applied in the training process of the MLP (the test set was composed of 110 active molecules and 67 inactive molecules). Table 4 shows the results obtained (output value and its probability).

Overall accuracy was 76.36% in the active group and 70.15% in the inactive group. These percentages increased to 80.00% and 73.44%, respectively, when the undetermined molecules were eliminated.

External validation was conducted by classifying, through the trained network, a set of molecules that had never participated in the training and test processes. This group, called the validation set, consisted of 61 pharmacologically active molecules belonging to different therapeutic categories (the same ones that took part in the learning process) and 25 nondrug molecules. Table 5 shows the results obtained (output value and its probability).

Overall accuracy for the validation set was 77.05% in the active group. The percentage increased to 81.03% when the undetermined molecules were eliminated. In the inactive group, overall accuracy was 76.00% and the percentage increased to 82.61% when the undetermined molecules were ruled out.

In view of the above, the discrimination of activity carried out shows that the values obtained for MLP on the drug and nondrug groups make it possible to separate both populations. Figures 3 and 4 show the histogram of frequencies (pharmacological distribution diagram) obtained by depicting the output values of MLP for training and test molecule groups in the first case and for the validation group in the second case. In both cases, top E_i (inactivity expectancy) and E_a (activity expectancy) values are distributed on both sides of the output value 0, with no overlapping occurring in the values obtained for active and inactive molecules because positive values were obtained for active compounds (maximum output value = 1) and negative MLP values were obtained for inactive compounds (minimum output value = −1) in the training, tes,t and validation groups.

Table 6 shows the accuracy percentages for the different therapeutic categories that make up the active molecule group used in the training, test, and validation processes: in all

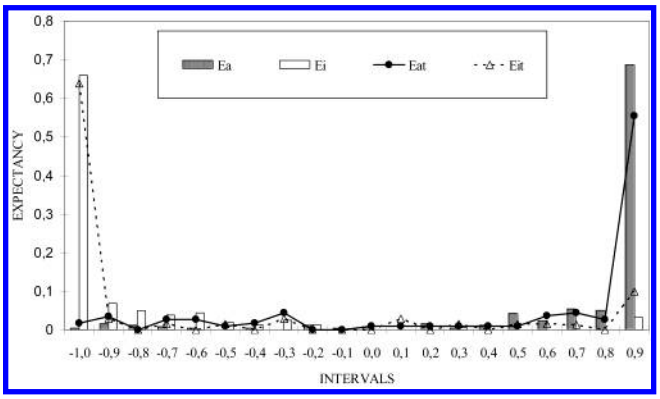

Figure 3. Pharmacological distribution diagram of MLP output values for pharmacological activity (E_a and E_at, activity expectancy of training and test groups, respectively; E_i and E_it, inactivity expectancy of training and test groups, respectively).

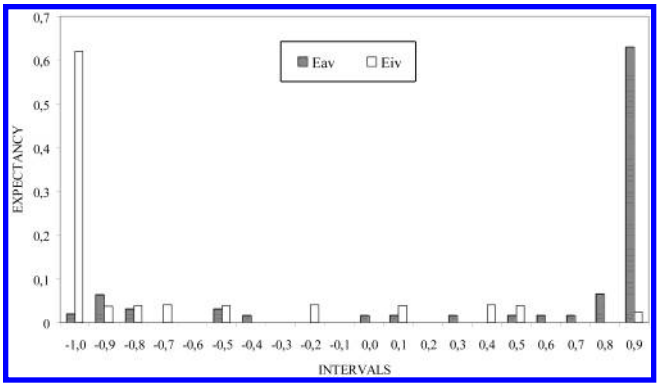

Figure 4. Pharmacological distribution diagram of MLP output values for pharmacological activity. (E_{av} and E_{iv}, activity and inactivity expectancy of validation group, respectively).

categories (except for the antiinflammatory one), accuracy is more than 80% and is actually greater than 90% for six categories.

As per Figure 5, the graphic representation of the accuracy percentage versus the average output value of the network obtained for each therapeutic category shows that, for the molecules in each therapeutic group, the network output

Table 6. Average of Accuracy and Output Values Obtained in Neural Network for Different Therapeutic Categories

therapeutic category	average of accuracy (%)	average of output values in neural network
analgesic	84.4	0.911
antibacterial	92.9	0.932
antidepresant	87.5	0.965
antidiabetic	100	0.991
antifungal	96.0	0.952
antihypertensive	93.6	0.927
antihistaminic	96.7	0.963
antiinflammatory	71.4	0.955
diuretic	92.3	0.980
antihyperlipoproteinemic	83.3	0.967
sedative	82.4	0.970

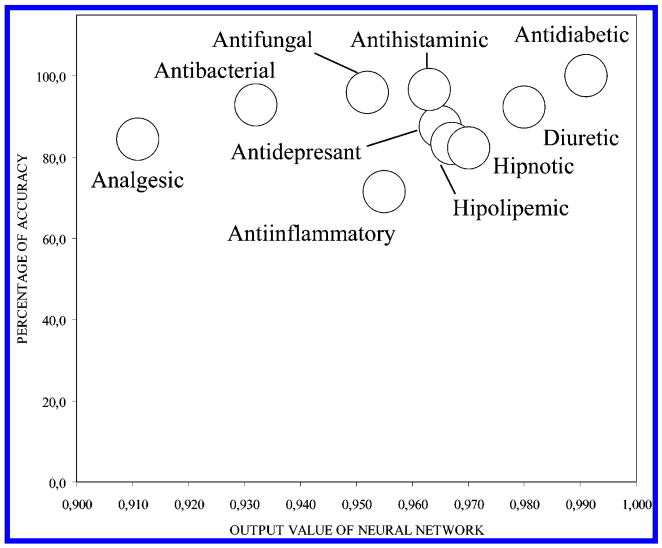

Figure 5. Accuracy percentage versus average output value of the network for each therapeutic category.

values concentrate around specific average values that allow molecules defined as active by the trained neural network to be therapeutically classified.

4. CONCLUSIONS

On the basis of the obtained results, we could state that the described topological−structural indices enable us to recognize general pharmacological activity. Combined with the use of artificial neural networks, molecular topology could become a useful tool in discriminating general pharmacological activity and so be used in large chemical compound databases, given its simplicity and quick calculation process.

Out of the highly heterogeneous compounds it was trained with, the designed neural network managed to find a pharmacological activity topological pattern that can also distinguish the therapeutic category of active-classed compounds. This aspect can be extremely useful in drug design.

REFERENCES AND NOTES

(1) Madan Babu, M. *Integrating Bioinformatics, Medical Sciences and Drug Discovery*. National Conference on Medical Informatics, Vijayawada, November 2000.
(2) Ajay; Walters, W. P.; Murcko, M. A. Can We Learn To Distinguish between "Drug-like" and "Nondrug-like" Molecules? *J. Med. Chem.* **1998**, *41*, 3314−3324.
(3) Sadowski, J.; Kubinyi, H. A Scoring Scheme for Discriminating between Drugs and Nondrugs. *J. Med. Chem.* **1998**, *41*, 3325−3329.
(4) Frimurer, T. M.; Bywater, R.; Nærum, L.; Lauritsen, L. N.; Brunak, S. Improving the Odds in Discriminating "Drug-like" from "Non Drug-like" Compounds. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1315−1324.
(5) Gálvez, J.; de Julián-Ortiz, J. V.; García-Domenech, R. General topological patterns of known drugs. *J. Mol. Graphics Modell.* **2001**, *20*, 84−94.
(6) Balaban, A. T.; Motoc, I.; Bonchev, D.; Mekenyan, O. Topological Indices for Structure−Activity Correlations. *Top. Curr. Chem.* **1983**, *114*, 21−55.
(7) Estrada, E.; Uriarte, E. Recent advances on the role of topological indices in drug discovery research. *Curr. Med. Chem.* **2001**, *8*, 1573−1588.
(8) Stankevich, M. I.; Stankevich, I. V.; Zefirov, N. S. Topological Indexes in Organic Chemistry. *Usp. Khimii* **1988**, *57*, 337−366.
(9) Rouvray, D. H. Should We Have Designs on Topological Indexes? *Stud. Phys. Theor. Chem.* **1983**, *28*, 159−177.
(10) Balaban, A. Chemical Graphs. XXXIV. Five New Topological Indices for the Branching of Tree-like Graphs. *Theor. Chim. Acta* **1979**, *53*, 355−375.
(11) Seybold, P. G.; May, M.; Bagal, U. A. Molecular Structure−Property Relationships. *J. Chem. Educ.* **1987**, *64*, 575−581.
(12) Jaen-Oltra, J.; Salabert-Salvador, M×bb 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.
(13) 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. (Teochem)* **2000**, *504*, 249−259.
(14) Murcia-Soler, M.; Pérez-Giménez, F.; García-March, F. J.; Salabert-Salvador, M. T.; Díaz-Villanueva, W.; Medina-Casamayor, P. Discrimination and selection of new potential antibacterial compounds using simple topological descriptors. *J. Mol. Graphics Modell.* **2003**, *21*, 375−390.
(15) Valls Sánchez, J. M. Diseño de fármacos mediante topología molecular: desarrollo de la interfaz gráfica, Proyecto fin de Carrera, Licenciatura de Ingeniería Informática, Universidad de Valencia, Spain, 2001.
(16) Randić, M.; Kleiner, A. F.; DeAlba, L. M. Distance/Distance Matrixes. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 277−286.
(17) Budavari, S. *The Merck Index*, 12th ed. [CD-ROM]; Merck & Co., Inc.: Whitehouse Station, NJ, 1996.
(18) Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. Learning internal representations by error propagation. In *PDP: Computational Models of Cognition and Perception I*. Rumelhart, D. E., McClelland, J. L., Ed.; MIT Press: Cambridge, MA, 1986.
(19) Bishop, C. M. *Neural Networks for Pattern Recognition*. Oxford University Press: New York, 1995.
(20) Zell, A.; et al. SNNS Stuttgart Neural Network Simulator V 4.2, Institute for Parallel and Distributed High Performance Systems (IPVR), University of Stuttgart, Germany, 1995.
(21) 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 Modell.* **1996**, *14*, 272−276.

CI0302862