Comparison between Neural Networks (NN) and Principal Component Analysis (PCA): Structure Activity Relationships of 1,4-Dihydropyridine Calcium Channel Antagonists (Nifedipine Analogues)

Yuji Takahata,*,† Maria Cristina Andreazza Costa,† and Anderson Coser Gaudio‡

Instituto de Química, Universidade Estadual de Campinas, CP 6154, Campinas, São Paulo, 13083-970, Brazil, and Departamento de Física, Centro de Ciências Exatas, Universidade Federal do Espírito Santo, Campus de Goiabeiras, Vitória, ES, 29.060-900, Brazil

Received November 5, 2001

The applicability of the neural network computer package PSDD (Perceptron Simulator for Drug Design/Perceptron-type Neural Network Simulator) in structure−activity relationship (SAR) studies was investigated. A group of 1,4-dihydropyridine derivatives was used in order to compare the PSDD results with those obtained previously with PCA. Calculated atomic and molecular descriptors using the semiempirical AM1 method were mainly used. It was shown that the predictive capability demonstrated by PSDD in SAR analysis were almost equivalent to that of PCA.

INTRODUCTION

The 1,4-dihydropyridine derivatives (DHPs; Figure 1) are an important class of drugs known as calcium antagonists. They have been used in the treatment of a number of cardiovascular disorders such as variant and exert ional angina, certain types of cardiac arrhythmias, hypertension, and others.[1] These drugs act directly on the voltage-dependent calcium channels, localized in the cell membrane, blocking the flux of calcium (Ca^{2+}) from the extra cellular medium to the cell cytoplasm.

A number of SAR[2−9] and QSAR[10−18] studies on DHP derivatives are available. Coburn and co-workers[12] determined the biological activity and performed a QSAR study with a set of 46 DHP derivatives (Table 1). Most of the QSAR analysis that appeared after the Coburn's paper made use of their biological activity data set. In our previous QSAR study of the DHP derivatives, we applied, first, multiple linear regression (MLR) analysis,[14] and then, principal component analysis.[17] Due to the different nature of the two methods employed, MLR enabled QSAR, while PCA enabled SAR.

The neural networks (NN) methodology has been applied in SAR and QSAR studies for the last 10 years.[19−27] The computer program PSDD (Perceptron Simulator for Drug Design/Perceptron-type Neural Network Simulator)[28] is one among many available to run NN. The original authors of PSDD showed its applicability in the area of SAR/QSAR.[19,20] Viswanadhan and co-workers[18] applied NN for the set of the DHP derivatives and found that NN demonstrated higher correlation between predicted biological activities with observed ones than did standard multiple linear regressions. They used information theoretic descriptors. The objective of the present work is to evaluate the applicability of the neural network package PSDD in SAR (*not* Quantitative SAR) studies, taking the Coburn DHP biological data set[12] and using quantum chemical and classical parameters instead of theoretical information descriptors. Particularly, we want to know how SAR results of NN compare with those obtained by PCA previously.[17]

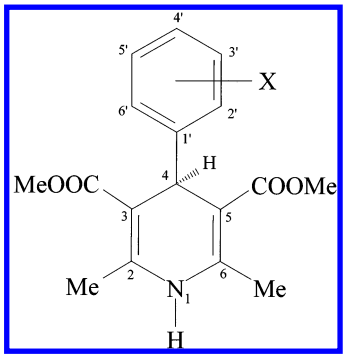

Figure 1. Structure and numbering of DHP derivatives, where X represents the phenyl ring substituent (see Table 1).

METHOD

The molecular descriptors that were selected for the PCA analysis are listed in Tables 1 and 2.[17] Table 3 defines seven different sets, set 1 to set 7, that were formed out of the compounds listed in Table 1. The first five sets, set 1 to set 5, are exactly identical, not only the number and the type of molecules but also the number and the type of descriptors, to those defined and used in the previous PCA work.[17] The set 1 consists of eight *para*-monosubstituted compounds. They are the compounds 13 and 29−35 (Table 1). The $\log(1/IC_{50})$ ranges between 7.55 (compound 13) and 4.00 (compound 33). We chose compounds 13 and 29 as high active, whereas the remaining six compounds were chosen as low active. The separation between the high active and low active groups is more than 1.39 units apart in the $\log(1/IC_{50})$ scale. There exist two clearly defined grouping. The set 2 consists

* Corresponding author phone: 55-19-3788-3103; fax: 55-19-3788-3023; e-mail: taka@iqm.unicamp.br.
† Universidade Estadual de Campinas.
‡ Universidade Federal do Espírito Santo.

10.1021/ci010117m CCC: $25.00 © 2003 American Chemical Society
Published on Web 02/21/2003

Table 1. Biological Data, Expressed as $\log(1/IC_{50})$, Molecular Polarizability[14] (α), Verlop Minimum Width[12] (B_1) and Length[12] (L) of the Substituent, and Rotational Barrier[14] (ϵ_{rot}) for the Set of DHP Derivatives[a]

no.	X	$\log(1/IC_{50})$	α	B_1	L	ϵ_{rot}
1	3′-Br	8.89	26.53	1.95	3.83	1.66
2	2′-CF_3	8.82	26.43	1.98	3.30	9.73
3	2′-Cl	8.66	25.75	1.80	3.52	4.73
4	3′-NO_2	8.40	27.45	1.70	3.44	2.05
5	2′-CH=CH_2	8.35	28.00	1.60	4.29	6.36
6	2′-NO_2	8.29	27.09	1.70	3.44	6.54
7	2′-Me	8.22	26.24	1.52	3.00	6.67
8	2′-Et	8.19	27.49	1.52	4.11	7.41
9	2′-Br	8.12	26.10	1.95	3.83	5.60
10	2′-CN	7.80	26.80	1.60	4.23	5.82
11	3′-Cl	7.80	26.06	1.80	3.52	1.58
12	3′-F	7.68	25.27	1.35	2.65	1.97
13	H	7.68	24.98	1.00	2.06	1.48
14	3′-CN	7.46	27.23	1.60	4.23	1.71
15	3′-I	7.38	26.99	2.15	4.23	1.65
16	2′-F	7.37	25.16	1.35	2.65	4.80
17	2′-I	7.33	26.51	2.15	4.23	6.20
18	2′-OMe	7.24	27.12	1.35	3.98	3.88
19	3′-CF_3	7.13	26.62	1.98	3.30	1.80
20	3′-Me	6.96	26.34	1.52	3.00	1.49
21	2′-OEt	6.96	28.47	1.35	4.92	5.27
22	3′-OMe	6.72	27.34	1.35	3.98	1.61
23	3′-NMe_2	6.05	29.25	1.50	3.53	1.67
24	3′-OH	6.00	25.72	1.35	2.74	1.41
25	3′-NH_2	5.70	26.39	1.50	2.93	1.54
26	3′-OAc	5.22	29.06	1.35	4.87	1.51
27	3′-OCOPh	5.20	36.18	1.70	8.15	1.57
28	2′-NH_2	4.40	26.24	1.50	2.93	5.82
29	4′-F	6.89	25.36	1.35	2.65	1.52
30	4′-Br	5.40	26.70	1.95	3.83	1.43
31	4′-I	4.64	27.16	2.15	4.23	1.40
32	4′-NO_2	5.50	27.68	1.70	3.44	1.40
33	4′-NMe_2	4.00	29.50	1.50	3.53	1.55
34	4′-CN	5.46	27.45	2.06	4.23	1.41
35	4′-Cl	5.09	26.19	1.80	3.52	1.47
36	2′,6′-Cl_2	8.72	26.65			9.24
37	F_5	8.36	26.71			8.52
38	2′-F,6′-Cl	8.12	26.04			9.44
39	2′,3′-Cl_2	7.72	26.87			4.99
40	2′-Cl, 5′-NO_2	7.52	28.39			5.25
41	3′,5′-Cl_2	7.03	27.13			1.63
42	2′-OH, 5′-NO_2	7.00	28.26			4.37
43	2′,5′-Me_2	7.00	27.63			6.66
44	2′,4′-Cl_2	6.40	27.04			4.71
45	2′,4′,5′-$(OMe)_3$	3.00	31.98			3.84

[a] IC_{50} is the molar concentration of the drug required to inhibit 50% of the contraction of guinea pig ileum induced by methylfurmethide.[12]

Table 2. Net Atomic Charge (δ), Frontier Electron ($F^{(e)}$) and Orbital ($F^{(o)}$) Densities, and Molecular Hardness (η) for the 45 DHP Set Calculated with AM1[a]

compd	$\delta_{3'}$	$\delta_{6'}$	$F_{4'}^{(e)}$	$F_{5'}^{(o)}$	$F_{12'}^{(o)}$	η
1	−0.1683	−0.1075	0.00041	0.00347	0.00930	4.173
2	−0.0955	−0.1275	0.00013	0.04050	0.06001	4.199
3	−0.1323	−0.1111	0.00023	0.01681	0.03488	4.179
4	−0.1338	−0.0799	0.00028	0.0897	0.02145	4.151
5	−0.1270	−0.1192	0.00020	0.02520	0.16782	4.167
6	−0.0869	−0.1296	0.00020	0.14475	0.14228	4.147
7	−0.1393	−0.1150	0.00145	0.00772	0.01810	4.160
8	−0.1394	−0.1154	0.00036	0.19349	0.21433	4.161
9	−0.1132	−0.1185	0.00028	0.00315	0.01132	4.184
10	−0.0986	−0.1258	0.00027	0.00308	0.01104	4.178
11	−0.0647	−0.1182	0.00025	0.01375	0.02800	4.169
12	+0.0841	−0.1352	0.00020	0.00987	0.02562	4.171
13	−0.1343	−0.1195	0.00015	0.03618	0.05323	4.164
14	−0.0228	−0.0961	0.00032	0.00443	0.01173	4.173
15	−0.2659	−0.1012	0.00028	0.01418	0.02820	4.169
16	−0.1743	−0.0952	0.00092	0.00411	0.01020	4.171
17	−0.1079	−0.1207	0.00030	0.00233	0.00805	4.198
18	−0.2109	−0.0879	0.00008	0.04871	0.20501	4.177
19	−0.1658	−0.0902	0.00019	0.00446	0.01043	4.177
20	−0.0756	−0.1223	0.00015	0.00411	0.01118	4.162
21	−0.2132	−0.0877	0.00020	0.00657	0.01875	4.177
22	+0.0719	−0.1517	0.00024	0.01593	0.03053	4.163
23	+0.0442	−0.1457	0.00025	0.00357	0.01220	3.972
24	+0.0735	−0.1551	0.00012	0.00566	0.01268	4.161
25	+0.0460	−0.1597	0.00010	0.00789	0.01632	3.979
26	+0.0544	−0.1273	0.00009	0.02339	0.2173	4.166
27	+0.554	−0.1252	0.00022	0.00345	0.01169	4.168
28	−0.2026	−0.0825	0.00035	0.00269	0.00952	3.903
29	−0.1689	−0.0962	0.00025	0.00328	0.01162	4.170
30	−0.1073	−0.1213	0.00028	0.00372	0.01019	4.167
31	−0.1013	−0.1237	0.01180	0.00417	0.01170	4.166
32	−0.0706	−0.1339	0.00007	0.01824	0.03209	4.121
33	−0.1553	−0.0970	0.20599	0.00156	0.00583	3.951
34	−0.0949	−0.1258	0.04055	0.00326	0.00924	4.167
35	−0.1290	−0.1112	0.27441	0.00268	0.00825	4.170
36	−0.1346	−0.0358	0.00014	0.15594	0.04693	4.134
37	+0.0205	+0.1041	0.00006	0.01948	0.01969	4.077
38	−0.1767	−0.0214	0.00026	0.00567	0.01092	4.131
39	−0.0674	−0.1110	0.00045	0.00337	0.01036	4.180
40	−0.1450	−0.0343	0.00049	0.00070	0.00745	3.980
41	−0.0578	−0.1148	0.00006	0.00509	0.01200	4.171
42	−0.2353	−0.0127	0.00018	0.00207	0.00739	4.027
43	−0.1357	−0.1136	0.06596	0.00150	0.00893	4.160
44	−0.1293	−0.1036	0.27320	0.00194	0.00782	4.177
45	−0.2580	−0.1327	0.26425	0.00217	0.01189	3.916

[a] Data taken from ref 14.

of 13 ortho derivatives. The 13 derivatives were arbitrarily divided into two training groups: high and low active ones. Seven derivatives (2, 3, 5, 6, 7, 8, and 9 in Table 1) were grouped as high active, whereas six derivatives (10, 13, 16, 17, 18, and 21) were chosen as low active. The set 3 consists of ortho and para monoderivatives. Almost all ortho derivatives have $\log(1/IC_{50})$ values greater than 7.00. The compound 28 is the only ortho derivative that has as low value as 4.40. On the other hand, almost all para derivatives have $\log(1/IC_{50})$ values less than 5.50. The compound 29 is the only para derivatives that has a corresponding $\log(1/IC_{50})$ value as high as 6.89. Those derivatives whose $\log(1/IC_{50})$ are greater than 6.89 were grouped as high active, whereas those less than 5.50 were grouped as low active. The set 4 consists of all of the 45 compounds in Table 1. The compounds whose $\log(1/IC_{50})$ are greater than 6.72 were grouped as high active. They are the compounds 1−22, 29, and 36−43. On the other hand, those compounds whose $\log(1/IC_{50})$ are less than 6.40 were grouped as low active. They are the compounds 23−28, 30−35, 44, and 45. The set 5 consists total of 35 compounds. The high active group consists of 14 compounds: 1−11 and 36−38 which have $\log(1/IC_{50})$ values between 8.89 and 7.80. The low active group consists of 13 compounds: 23−28, 30−35, and 45 which have $\log(1/IC_{50})$ values between 6.05 and 3.0. The high active group and low active one in the set 5 are well separated. The definitions and the numerical values of the descriptors are listed in Tables 1 and 2. With the selection of the identical sets, it is possible to compare the capability of NN with that of PCA in studying SAR. In the first five sets, the compounds are divided into the two categories: high activity and low activity. The set 6 and the set 7 are newly added ones. The set 6 consists of a total of 35 compounds that are divided into three categories: high activity (+3), medium activity (+2), and low activity (+1). The category (+3) consists of those compounds that have $\log(1/IC_{50})$

Table 3. Definition of the Seven Test Sets[a]

Classification in Two Categories (High Activity and Low Activity)

set	descriptors	total no. of compds	high activity	low activity
set 1	α, B_1, L	8	13, 29	30−35
set 2	$\delta_{6'}$, $\delta_{3'}$, ϵ_{rot}	13	2, 3, 5−9	10, 13, 16−18, 21
set 3	η, $F^{(o)}{}_{2'}$, ϵ_{rot}	21	2,3,5−10, 13 16−18, 21, 29	28, 30−35
set 4	η, $F^{(e)}{}_{4'}$, ϵ_{rot}	45	1−22,29 36−43	23−28, 30−35 44, 45
set 5	η, $F^{(o)}{}_{5'}$, ϵ_{rot}	27	1−11, 36−38	23−28, 30−35, 45

Classification in Three Categories

set	descriptors	total no. of compds	activity +3	activity +2	activity +1
set 6	η, $F^{(e)}{}_{4'}$, ϵ_{rot}, α, $\delta_{6'}$, $\delta_{3'}$	35	1−9	10−22, 29	23−28, 30−35

Classification in Four Categories

set	descriptors	total no. of compds	activity +4	activity +3	activity +2	activity +1
set 7	η, $F^{(e)}{}_{4'}$, ϵ_{rot}, α, $\delta_{6'}$, $\delta_{3'}$	35	1−9	10−22, 29	23−27, 30, 32, 34, 35	28, 31, 33

[a] The descriptors used in each set are listed in the second column, and the nunbers in the last columns with "activity" are the compound numbers defined in Table 1.

values greater or equal to 8.12. The category (+2) consists of those whose $\log(1/IC_{50})$ values are between 7.80 and 6.72. The category (+1) consists of those compounds that have $\log(1/IC_{50})$ value less than or equal to 6.05. The set 7 consists the same 35 compounds as the set 6. But they are divided into four categories (+4, +3, +2, +1). The compounds of category (+4) and (+3) in the set 7 are identical to those of category (+3) and (+2) of the set 6, respectively. The category (+2) consists of those whose $\log(1/IC_{50})$ values are between 6.05 and 5.09. The category (+1) consists of those compounds that have $\log(1/IC_{50})$ value less than or equal to 4.64. The PSDD is shown to be capable of multiple classifications[19] for the cases such as set 6 and set 7. It is of interest to test such a capability of PSDD using our descriptors. The neural network consists of three layers. For the most of the cases, the number of neurons in the first layer coincides with the number of the descriptors adopted in the each set plus one, which is bias. The value of a neuron in each layer is expressed by a sigmoid function. The back-propagation method is used in PSDD. The process of calculation was carried out in a supervised manner.

RESULTS AND DISCUSSION

The NN structure used for the NN calculations for the seven sets are summarized in Table 4. ρ is the nonlinear parameter of the sigmoid functions, and θ is a threshold value for a neuron (see eq 1 of ref 19). We take the set 4 as an example to discuss the results of calculations. It is the largest set among the seven set studied. It contains the whole 45 compounds of Table 1. The first layer of the set 4 consists of three neurons. The second layer consists of 10 neurons. There are two neurons in the third layer to form two different patterns to represent the two categories, low activity and high activity. Table 5 lists the results of two types of NN calculations, RECALL and LONE (*leave-one-out)*. In RECALL, NN is asked to *recall* the data used for training NN with the training pattern. In LONE, NN is asked to *predict* the pattern of each member of the data set, which is left out, one after another, during the training NN. Five compounds were

Table 4. Structure and Parameters of Neural Networks for the Seven Sets[a]

layer	no. of neurons	ρ	θ	layer	no. of neurons	ρ	θ
set 1				set 2			
1	4			1	4		
2	5	1.0	0.0	2	5	4.0	0.0
3	2	1.0	0.0	3	2	4.0	0.0
set 3				set 4			
1	4			1	3		
2	10	1.0	0.0	2	10	1.0	0.0
3	2	1.0	0.0	3	2	1.0	0.0
set 5				set 6			
1	3			1	7		
2	10	1.0	0.0	2	12	2.5	0.0
3	2	1.0	0.0	3	3	5.0	0.0
set 7				none			
1	7						
2	12	2.5	0.0				
3	4	5.0	0.0				

[a] ρ is the nonlinear parameter of the sigmoid functions and θ is a threshold value for a neuron. (See eq 1 of ref 19. The letter alpha instead of ρ is used in ref 19.)

incorrectly classified both in RECALL and LONE. They are the compounds 13, 20, 26, 27, and 29. The calculated RECALL and LONE patterns of these compounds disagree with their respective TRAINING patterns. The wrong patterns are marked with parentheses in Table 5. Five out of 45 compounds were incorrectly classified. The percent of correct classification was 89% for both RECALL and LONE. This value is slightly better than 82% that is the percentage of correct classification obtained by PCA for the same set (Table 6). The third and fourth columns in Table 6 summarize the resultant percentage (%) of correct prediction of RECALL and LONE for the different set studied. The last column of the table lists the percentage of correct prediction (%) obtained with PCA for only two category classification. No PCA values are available for more than three category classifications.

The percentages of correct prediction by the process of RECALL were almost always close to or equal to 100%.

Table 5. Two Category Classification of the 45 Molecules of Set 4 Using the Three Descriptors, η, $F^{(o)}{}_{2'}$, and ϵ_{rot}[a]

molecule	category	TRAINING pattern		RECALL pattern		LONE pattern	
1	1	1	0	0.8	0.2	0.8	0.2
2	1	1	0	1.0	0.0	1.0	0.0
3	1	1	0	1.0	0.0	1.0	0.0
4	1	1	0	1.0	0.0	1.0	0.0
5	1	1	0	1.0	0.0	1.0	0.0
6	1	1	0	1.0	0.0	1.0	0.0
7	1	1	0	1.0	0.0	1.0	0.0
8	1	1	0	1.0	0.0	1.0	0.0
9	1	1	0	1.0	0.0	1.0	0.0
10	1	1	0	1.0	0.0	1.0	0.0
11	1	1	0	0.7	0.3	0.7	0.3
12	1	1	0	1.0	0.0	1.0	0.0
13	1	1	0	(0.4	0.6)	(0.4	0.6)
14	1	1	0	0.9	0.1	0.9	0.1
15	1	1	0	0.8	0.2	0.8	0.2
16	1	1	0	1.0	0.0	1.0	0.0
17	1	1	0	1.0	0.0	1.0	0.0
18	1	1	0	1.0	0.0	1.0	0.0
19	1	1	0	0.9	0.1	0.9	0.1
20	1	1	0	(0.4	0.6)	(0.5	0.5)
21	1	1	0	1.0	0.0	1.0	0.0
22	1	1	0	0.7	0.3	0.7	0.3
23	2	0	1	0.0	1.0	0.0	1.0
24	2	0	1	0.2	0.8	0.3	0.7
25	2	0	1	0.0	1.0	0.0	1.0
26	2	0	1	(0.5	0.5)	(0.7	0.3)
27	2	0	1	(0.6	0.4)	(0.6	0.4)
28	2	0	1	0.0	1.0	1.0	0.0
29	1	1	0	(0.5	0.5)	(0.5	0.5)
30	2	0	1	0.3	0.7	0.4	0.6
31	2	0	1	0.0	1.0	0.0	1.0
32	2	0	1	0.1	0.9	0.3	0.7
33	2	0	1	0.0	1.0	0.0	1.0
34	2	0	1	0.0	1.0	0.0	1.0
35	2	0	1	0.0	1.0	0.0	1.0
36	1	1	0	1.0	0.0	1.0	0.0
37	1	1	0	1.0	0.0	1.0	0.0
38	1	1	0	1.0	0.0	1.0	0.0
39	1	1	0	1.0	0.0	1.0	0.0
40	1	1	0	1.0	0.0	0.8	0.2
41	1	1	0	0.8	0.2	0.8	0.2
42	1	1	0	1.0	0.0	1.0	0.0
43	1	1	0	1.0	0.0	1.0	0.0
44	2	0	1	0.0	1.0	0.0	1.0
45	2	0	1	0.0	1.0	0.0	1.0
%age of correct classification				89%		89%	

[a] The RECALL pattern is the one calculated with the same input data as those used for the initial training of NN. LONE pattern refers to the results of the *leave-one-out* experiment. Those patterns in parentheses are incorrect ones.

The percentage of correct prediction by LONE experiment for the first five sets, sets 1−5, are almost always approximately above 80%. The first five sets consist of the class of two categories. The percentages of LONE are very close to those percentages of correct classification by PCA. This is an implication that predictive power of NN is about the same as PCA. Table 7 lists the results of three category classification of the 35 molecules of set 6 using the six descriptors, η, $F^{(e)}{}_{4'}$, $\epsilon_{rot,}$ α, $\delta_{6'}$, $\delta_{3'}$. The percentage of correct classification of RECALL is 100%. In LONE, the category of 22 molecules out of 35 was correctly predicted, while that of 13 was incorrectly predicted. The 13 incorrectly predicted patterns are marked with parentheses in the last column headed by "LONE pattern". The percentages of correct prediction by LONE for the set 6 is 60%. This is not satisfactory. Corresponding value for set 7 is 37%. The set 6 is of classification in three categories, while the set 7 is of that in four categories. When the number of category increases, the percentage of correct prediction of LONE decreases substantially. With those descriptors listed in Table 6 for set 6 and set 7, no good values of LONE percentages were obtained.

Table 6. Comparison of Percentages (%) of Correct Prediction between RECALL and LONE of NN and PCA for the Seven Sets Defined in Table 1

set	employed descriptors	NN (%) RECALL	NN (%) LONE	PCA (%)
Classification in Two Categories (High and Low Activity)				
set 1	α, B_1, L	100	100	100
set 2	$\delta_{6'}$, $\delta_{3'}$, ϵ_{rot}	100	77	85
set 3	η, $F^{(o)}{}_{2'}$, ϵ_{rot}	100	90	90
set 4	η, $F^{(e)}{}_{4'}$, ϵ_{rot}	89	89	82
set5	η, $F^{(o)}{}_{5'}$, ϵ_{rot}	96	93	89
Classification in Three Categories				
set 6	η, $F^{(e)}{}_{4'}$, ϵ_{rot}, α, $\delta_{6'}$, $\delta_{3'}$	100	60	
Classification in Four Categories				
set 7	η, $F^{(e)}{}_{4'}$, ϵ_{rot}, α, $\delta_{6'}$, $\delta_{3'}$	100	37	

Table 7. Three Category Classification of the 35 Molecules of Set 6 Using the Six Descriptors, η, $F^{(e)}{}_{4'}$, $\epsilon_{Rot,}$ α, $\delta_{6'}$, and δ_3[a]

molecule	category	TRAINING pattern			RECALL pattern			LONE pattern		
1	3	0	0	1	0.0	0.0	1.0	(0.0	1.0	0.0)
2	3	0	0	1	0.0	0.0	1.0	0.0	0.0	1.0
3	3	0	0	1	0.0	0.0	1.0	(0.0	0.0	1.0)
4	3	0	0	1	0.0	0.0	1.0	0.0	1.0	0.0
5	3	0	0	1	0.0	0.0	1.0	0.0	0.0	1.0
6	3	0	0	1	0.0	0.0	1.0	(0.0	1.0	0.0)
7	3	0	0	1	0.0	0.0	1.0	0.0	0.0	1.0
8	3	0	0	1	0.0	0.0	1.0	0.0	0.0	1.0
9	3	0	0	1	0.0	0.0	1.0	(0.0	1.0	0.0)
10	2	0	1	0	0.0	1.0	0.0	0.0	1.0	0.0
11	2	0	1	0	0.0	1.0	0.0	0.0	1.0	0.0
12	2	0	1	0	0.0	1.0	0.0	0.0	1.0	0.0
13	2	0	1	0	0.0	1.0	0.0	(0.0	0.1	0.9)
14	2	0	1	0	0.0	1.0	0.0	0.0	1.0	0.0
15	2	0	1	0	0.0	1.0	0.0	(0.0	0.0	0.5)
16	2	0	1	0	0.0	1.0	0.0	0.0	1.0	0.0
17	2	0	1	0	0.0	1.0	0.0	(0.0	0.1	0.9)
18	2	0	1	0	0.0	1.0	0.0	0.0	1.0	0.0
19	2	0	1	0	0.0	1.0	0.0	0.0	0.8	0.3
20	2	0	1	0	0.0	1.0	0.0	0.3	0.9	0.0
21	2	0	1	0	0.0	1.0	0.0	(0.0	0.0	0.5)
22	2	0	1	0	0.0	1.0	0.0	1.0	0.0	0.0
23	1	1	0	0	1.0	0.0	0.0	1.0	0.0	0.0
24	1	1	0	0	1.0	0.0	0.0	(0.0	1.0	0.0)
25	1	1	0	0	1.0	0.0	0.0	1.0	0.0	0.0
26	1	1	0	0	1.0	0.0	0.0	(0.3	0.7	0.0)
27	1	1	0	0	1.0	0.0	0.0	1.0	0.0	0.0
28	1	1	0	0	1.0	0.0	0.0	(0.5	0.0	1.0)
29	2	0	1	0	0.0	1.0	0.0	0.0	1.0	0.0
30	1	1	0	0	1.0	0.0	0.0	(0.1	0.6	0.0)
31	1	1	0	0	1.0	0.0	0.0	1.0	0.0	0.0
32	1	1	0	0	1.0	0.0	0.0	1.0	0.0	0.0
33	1	1	0	0	1.0	0.0	0.0	1.0	0.0	0.0
34	1	1	0	0	1.0	0.0	0.0	1.0	0.0	0.0
35	1	1	0	0	1.0	0.0	0.0	1.0	0.0	0.0
%age of correct classification					100%			63%		

[a] The RECALL pattern is the one calculated with the same input data as those used for the initial training of NN. LONE pattern refers to the results of the *leave-one-out* experiment. Those patterns in parentheses are incorrect ones.

To compare the efficiency of classical descriptors and those calculated with the AM1 semiempirical method in RECALL and LONE calculations with the NN, we used set 2 and selected the three classical descriptors, π, σ_m, B_1 and the three calculated ones, $\delta_{6'}$, $\delta_{3'}$, $\epsilon_{rot.}$ The percentage of correct prediction in RECALL was 100% with both the classical descriptors and the calculated ones. While the percentages of correct prediction in LONE were 77% with both the classical parameters and the calculated ones. The calculated descriptors demonstrated equal efficiency to the well-established classical descriptors, such as π, σ_m, B_1, in SAR/QSAR. This is an implication that calculated descriptors can be conveniently used to SAR study with NN. Calculated descriptors can be obtained without difficulty for almost all type of molecules.

CONCLUSION

PSDD results showed that the percentage of correct classification of RECALL was more than 90%, while the percentage of correct prediction by Leave-One-Out experiment was more than 80% for two category classifications of most cases studied. It was shown that the predictive capability demonstrated by PSDD in the SAR analysis was almost equivalent to that of PCA for two category classifications of the five different sets tested using mainly atomic and molecular descriptors calculated with the semiempirical AM1 method. The percentages of LONE correct prediction for three and four categories were low. The calculated descriptors demonstrated equal efficiency to the well-established classical descriptors.

REFERENCES AND NOTES

(1) Janis, R. A.; Silver P. J.; Triggle, D. J. Drug action and cellular calcium regulation. *Adv. Drug. Res.* **1987**, *16*, 309−591.
(2) Rodenkirchen, R.; Bayer, R.; Steiner, R.; Bossert, F.; Meyer, H.; Möller, E. Structure−Activity Studies on Nifedipine in Isolated Cardiac Muscle. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **1979**, *310*, 69−78.
(3) Joslyn, A. F. Structure−Activity Studies on 1,4-Dihydropyridine Calcium Channel Antagonists and Activators. *Dissert. Abstr. Int.* **1989**, *49*, 5244B−5245B.
(4) Triggle, D. J. Calcium Channel Drugs: Structure−Function Relationships and Selectivity of Action. *J. Cardiovasc. Pharmacol.* **1991**, *18*, S1−S6.
(5) Langs, D. A.; Kwon, Y. W.; Strong, P. D.; Triggle, D. J. Molecular-Level Model for the Agonist Antagonist Selectivity of the 1,4-Dihydropyridine Calcium Channel Receptor. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 95−106.
(6) Höltje, H.-D. Molecular Modelling Studies on 1,4-Dihydropyridines Acting at Ca−Channel. *Quant. Struct.-Act. Relat.* **1992**, *11*, 224−227.
(7) Iqbal, N.; Triggle, C. R.; Knaus, E. E. Novel Hantzsch 1,4-Dihydropyridines to Study the Structure−Function Relationships of Calcium Channels and Photoinduced Relaxation. *Drug Dev. Res.* **1997**, *42*, 120−130.
(8) Natale, N. R.; Rogers, M. E.; Staples, R.; Triggle, D. J.; Rutledge, A. Lipophilic 4-Isoxazolyl-1,4-dihydropyridines: Synthesis and Structure−Activity Relationships. *J. Med. Chem.* **1999**, *42*, 3087−3093.
(9) Nguyen, J. T.; McEwen, C. A.; Knaus, E. E. Hantzsch 1,4-Dihydropyridines Containing a Nitrooxyalkyl Ester Moiety to Study Calcium Channel Antagonist Structure−Activity Relationships and Nitric Oxide Release. *Drug Dev. Res.* **2000**, *51*, 233−243.
(10) Seidel, W.; Meyer, H.; Born, L.; Kazda, S.; Dompert, W. *Rigid Calcium Antagonists of the Nifedipine-Type: Geometric Requirements for the Dihydropyridine Receptor*, In: *QSAR and Strategies in the Design of Bioactive Compounds*; Seydel, J. K., Ed.; 1984.
(11) Mahmoudian, M.; Richards, W. G. QSAR of Binding of Dihydropyridine-type Calcium Antagonists to Their Receptor on Ileal Smooth Muscle Preparations. *J. Pharm. Pharmacol.* **1986**, *38*, 272−276.
(12) Coburn, R. A.; Wierzba, M.; Suto, M. J.; Solo, A. J.; Triggle, A. M.; Triggle, D. J. 1,4-Dihydropyridine Antagonist Activities at the Calcium Channel: A Quantitative Structure−Activity Relationship Approach. *J. Med. Chem.* **1988**, *31*, 2103−2107.
(13) Höltje, H.-D.; Marrer, S. Qualitative and Quantitative Structure−Activity Relationships for Calcium Channel Modulating 1,4-Dihydropyridine Derivatives: A Hypothetical Molecular Receptor Model. *Quant. Struct.-Act. Relat.* **1988**, *7*, 174−178.
(14) Gaudio, A. C.; Korolkovas, A.; Takahata, Y. Quantitative Structure−Activity Relationships for 1,4-Dihydropyridine Calcium Channel Antagonists (Nifedipine Analogues): A Quantum Chemical/Classical Approach. *J. Pharm. Sci.* **1994**, *83*, 1110−1115.
(15) Hadjipavlou-Litina, D. Review, Reevaluation, and New Results in Quantitative Structure−Activity Studies of Anticonvulsants. *Med. Res. Rev.* **1998**, *18*, 91−119.
(16) Mahmoudian, M.; Richards, W. G. A Quantum Chemical Study of Structure−Activity Relationships of Dihydropyridine Calcium Antagonists. *J. Sci. I. R. Iran* **1990**, *1*, 261−265.
(17) Costa, M. C. A.; Gaudio, A. C.; Takahata, Y. A Comparative Study of Principal Component and Linear Multiple Regression Analysis in SAR and QSAR Applied to 1,4-Dihydropyridine Calcium Channel Antagonists (Nifedipine Analogues). *J. Mol. Struct. (THEOCHEM)* **1997**, *394*, 291−300.
(18) 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 QSAR Approaches: Application to 1,4-Dihydropyridine-based Calcium Channel Antagonists. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 505−515.
(19) Aoyama, T.; Suzuki, Y.; Ichikawa, H. Neural Networks Applied to Structure−Activity Relationships. *J. Med. Chem.* **1990**, *33*, 905−908.
(20) Aoyama, T.; Suzuki, Y.; Ichikawa, H. Neural Networks Applied to Quantitative Structure−Activity Relationship Analysis. *J. Med. Chem.* **1990**, *33*, 2583−2590.
(21) So, S.-S.; Richards, W. G. Application of Neural Networks: Quantitative Structure−Activity Relationships of the Derivatives of 2,4-Diamino-5-(substituted-benzyl)pyrimidines as DHFR Inhibitors. *J. Med. Chem.* **1992**, *35*, 3201−3207.
(22) Manallack, D. T.; Livingstone, D. J. Limitations of Functional-Link Nets as Applied to QSAR Data Analysis. *Quant. Struct.-Act. Relat.* **1994**, *13*, 18−21.
(23) Tang, Y.; Jiang, H. L.; Chen, K. X.; Ji, R. Y. QSAR Study of Artemisinin (Qinghaosu) Derivatives using Neural Network Method. *Indian J. Chem.* **1996**, *35B*, 325−332.
(24) Maddalena, D. J. Applications of Artificial Neural Networks to Quantitative Structure−Activity Relationships. *Exp. Opin. Ther. Patents* **1996**, *6*, 239−251.
(25) Novic, M.; Nikolovska-Coleska, Z.; Solmajer, T. Quantitative Structure−Activity Relationship of Flavonoid p56lck Protein Tyrosine Kinase Inhibitors. A Neural Network Approach. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 990−998.
(26) Montanari, C. A.; Tute, M. S. A Neural Network Analysis for Antileishmaniasis Compounds. *Quant. Struct.-Act. Relat.* **1997**, *16*, 480−481.
(27) Schneider, G.; Wrede, P. Artificial Neural Networks for Computer-Based Molecular Design. *Prog. Biophys. Mol. Biol.* **1998**, *70*, 175−222.
(28) PSDD: Perceptron-type Neural Network Simulator, Release QCPE 615, Indiana University.

CI010117M