Structure−Activity Relationship Studies of Carcinogenic Activity of Polycyclic Aromatic Hydrocarbons Using Calculated Molecular Descriptors with Principal Component Analysis and Neural Network Methods

R. Vendrame,[†] R. S. Braga,[†] Y. Takahata,[‡] and D. S. Galvão*,[†]

Instituto de Física Gleb Wataghin and Instituto de Química, UNICAMP, CP 6154, CEP 13083-970, Campinas, SP, Brasil

Received May 24, 1999

Recently a new methodology based on local density of state (LDOS) calculations using topological and semiempirical methods was proposed to identify the carcinogenic activity of polycyclic aromatic hydrocarbons (PAHs). In this work we perform a comparative study of this methodology with principal component analysis (PCA) and neural networks (NN). The PCA and NN results show that LDOS quantum chemical descriptors are relevant descriptors to identify the carcinogenic activity of methylated and non-methylated PAHs. Also, we show that the combination of these distinct methodologies can be an efficient and powerful tool in the structure−activity studies of PAHs compounds. We have studied 81 methylated and non-methylated PAHs, and our study shows that with the use of these methods it is possible to correctly predict the carcinogenic activity of PAHs with accuracy higher than 80%.

1. INTRODUCTION

In the past decade the presence of polycyclic aromatic hydrocarbons (PAHs) in the environment and their biological effects have received substantial attention. This is due in part to the carcinogenic activity of certain PAHs.[1] The PAH's carcinogenic power varies from the very strongest carcinogens known to inactive ones.[2] The theoretical investigation of why some of these very similar molecules present carcinogenic activity and others do not started with the work of Cook and co-workers.[3] They reported a relationship between carcinogenic activity (malignant tumors in rats) and some geometrical features of molecules. Later Pullman and Pullman[4] proposed the K, L region theory (see inset of Figure 1) based on quantum chemical calculations, using simple Hückel theory,[5] and expressed it in terms of critical index values over specific molecular regions. Other similar theories evolved to include what is called the "bay region"[6−9] (see inset of Figure 1).

These theories based on electronic indices and more recent ones using statistical analysis, neural networks, and artificial intelligence methods[10−13] have achieved only partial success. Some of them work well for a specific subset of compounds and fail for others and vice versa. On account of the increasing levels of PAHs present in urban air due to auto exhaust and in many common processed foods, the search for a theory that could predict, at least at a qualitative level, whether a specific PAH will be carcinogenic or not continues to be a challenge.

Recently[14,15] a new methodology was proposed to identify PAH carcinogenic activity. This new methodology, based on electronic indices (EIM), is able to analyze all the PAHs molecules, even those that do not contain K, L, or bay regions, since it does not explicitly involve such concepts. This study was carried out first on non-methylated molecules 1−26,[14] shown in the Figure 1, and later it was extended to the other 55 molecules[15] (non-methylated molecules numbered 27−32, Figure 1, and methylated molecules numbered 33−81, Figure 2). This methodology is based on the concepts of the local density of state[14] (LDOS) over the ring that contains the highest bond order (RHBO) and on critical values for the energy separation (ΔH) between HOMO (highest occupied molecular orbital) and HOMO-1, which is one level below to HOMO. The density of states[14,15] (DOS) is defined as the number of electronic states per energy unit. The LDOS is the DOS calculated over a specific molecular region. It gives us detailed information on the contributions of specific geometrical regions of the molecules to the chemical reactivity, optical response, etc.

Through a single simple rule it was possible to identify whether a specific PAH molecule (methylated and non-methylated) will (or will not) exhibit carcinogenic activity.[15] This rule is based on the values of relative HOMO and HOMO-1 contribution (ηH) to the LDOS over the RHBO and in a critical value for their energy separation (ΔH energy).

In this work we have investigated the carcinogenic activity of this same set of molecules using principal component analysis (PCA) and the neural network (NN) methods. Since the quantum molecular descriptors used in the ElM methodology (ηH and ΔH energy) were never used before in the literature, we decided to investigate their importance and validity in the framework of the more standard structure−activity relationship (SAR) studies.[16−18]

The studies of SAR can aid in understanding the nature of carcinogenic activity as well as the mechanism of interaction between carcinogenic molecules and their recep-

* Corresponding author. E-mail: galvao@ifi.unicamp.br.
[†] Instituto de Física Gleb Wataghin.
[‡] Instituto de Química.

10.1021/ci990326v CCC: $18.00
Published on Web 10/19/1999

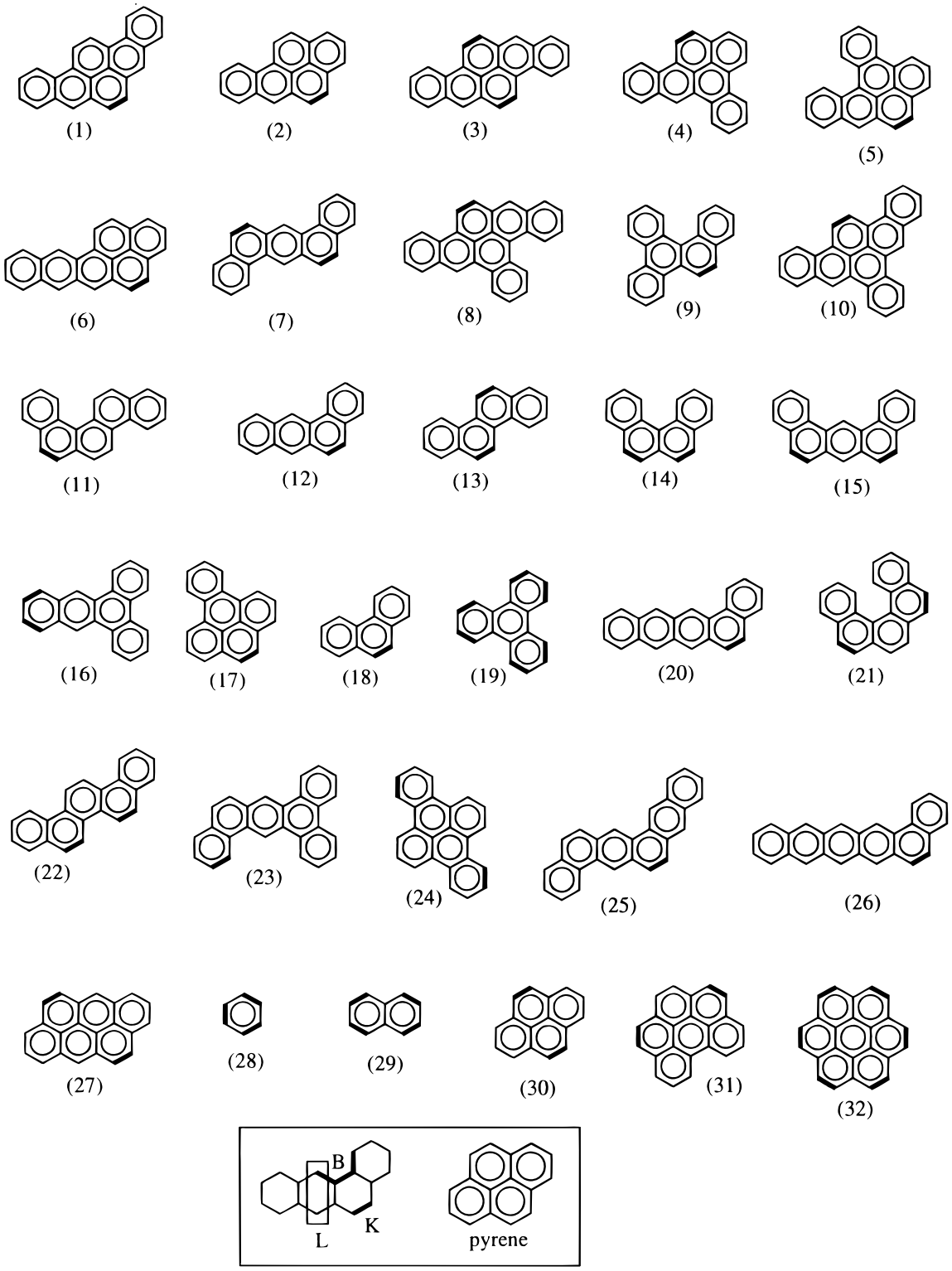

Figure 1. Molecular structures of the 32 non-methylated polycyclic aromatic hydrocarbon (PAH) molecules. The darker bonds indicate bonds with the highest bond orders.[14] In the inset are shown the pyrene structure and typical L, K, and bay (B) regions for PAH molecules. Table 1 lists their IUPAC names.

tors. Our objectives are to investigate the SAR of the compounds listed in Figures 1 and 2 using EIM quantum-chemical descriptors and some other calculated descriptors, and to compare the results of three entirely different methods (EIM, PCA, and NN methods), to investigate if they are consistent with the same physicochemical descriptors and whether these descriptors can be used to correlate to the carcinogenic activity.

2. METHODOLOGY

In the present work we have studied 81 PAH molecules shown in Figures 1 and 2. The methylated structures shown in Figure 2 are structurally related to the non-methylated ones shown in Figure 1, in order to provide a direct comparison. These molecules were selected by bearing in mind the criterion of the availability of experimental data for chemical

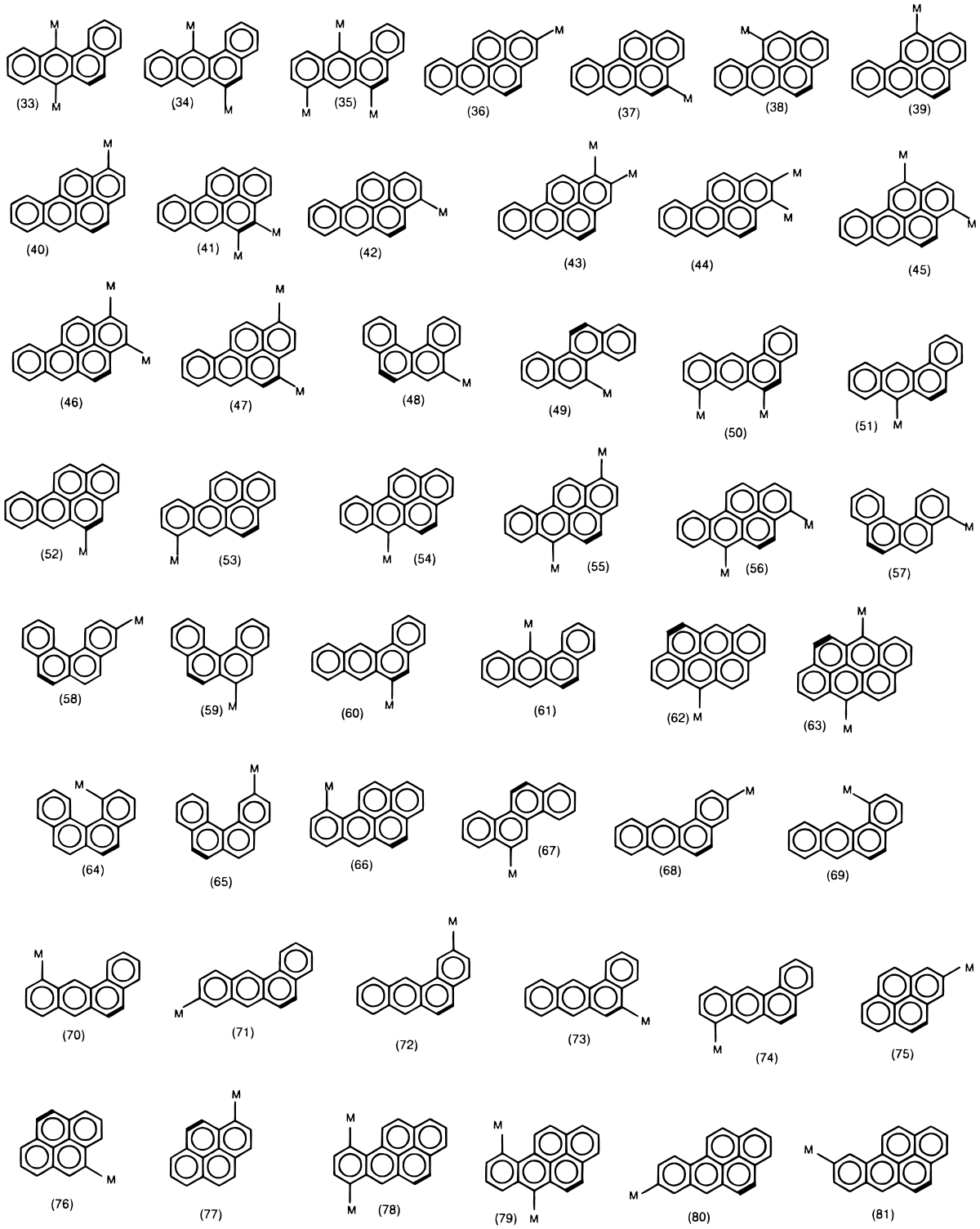

Figure 2. Molecular structures of 49 methylated polycyclic aromatic hydrocarbon (PAH) molecules. The darker bonds indicate bonds with the highest bond orders.[14] Table 2 lists their IUPAC names.

carcinogenesis. In some cases in the literature on these compounds have been classified into several categories because of the intrinsic dependence of the carcinogenic potency on the metabolic processes.[12] For instance some compounds that are classified as moderately active as injected intramuscularly into rats are classified as a slight carcinogen when painted on mouse skin. Because of that and since we do not have experimental data under the same conditions for all the compounds studied here, we have used the same approach proposed by Villemin et al.,[12] simply classifying the compounds into two classes: active (A) and inactive (I). Our experimental data to define active and inactive compounds are from the Iball indices[10] and from the scale proposed by Cavaliere et al.[19]

The set of 81 PAH molecules, sequentially numbered (Figures 1 and 2; Tables 1 and 2), was divided into two groups: **G1** and **G2**. Group G1, category 1, consists of 34 active molecules (1−10 in Figure 1, 33−56 in Figure 2). Group G2, category 2, consists of 44 inactive molecules (11−32 in Figure 1, 57−78 in Figure 2). The remaining three

Table 1. Carcinogenic Activity (ca) of the 32 Non-Methylated Polycyclic Aromatic Hydrocarbons (PAHs)[a]

molecule	ca[10,19]	molecule	ca[10,19]
(**1**) dibenzo[3,4;9,10]pyrene	A	(**17**) benzo[1,2]pyrene	I
(**2**) benzo[3,4]pyrene	A	(**18**) phenantrene	I
(**3**) dibenzo[3,4;8,9]pyrene	A	(**19**) triphenylene	I
(**4**) dibenzo[3,4;6,7]pyrene	A	(**20**) benzo[1,2]naphthacene	I
(**5**) dibenzo[1,2;3,4]pyrene	A	(**21**) dibenzo[3,4;5,6]phenantrene	I
(**6**) naphto[2,3;3,4]pyrene	A	(**22**) picene	I
(**7**) dibenzo [1,2;5,6]anthracene	A	(**23**) tribenzo[1,2;3,4;5,6]anthracene	I
(**8**) tribenzo[3,4;6,7;8,9]pyrene	A	(**24**) dibenzo[1,2;5,6]pyrene	I
(**9**) dibenzo[1,2;3,4]phenantrene	A	(**25**) phenanthra[2,3;1,2]anthracene	I
(**10**) tribenzo[3,4;6,7;8,9]pyrene	A	(**26**) benzo[1,2]pentacene	I
(**11**) dibenzo[1,2;5,6]phenantrene	I	(**27**) anthanthrene	I
(**12**) benzo[1,2]anthracene	I	(**28**) benzene	I
(**13**) chrysene	I	(**29**) naphthalene	I
(**14**) benzo[3,4]phenantrene	I	(**30**) pyrene	I
(**15**) dibenzo[1,2;7,8]anthracene	I	(**31**) benzo[*ghi*]preylene	I
(**16**) dibenzo[1,2;3,4]anthracene	I	(**32**) coronene	I

[a] See Figure 1 for the molecular structures. The carcinogenic activity data are adapted from Iball index experiments[10,27,28] and from the Cavaliere et al. scale:[19] A and I refer to active and inactive, respectively.

Table 2. Carcinogenic Activity (ca) of the 49 Methylated Polycyclic Aromatic Hydrocarbons (PAHs)[a]

molecule	ca[19]	molecule	ca[19]
(**33**) 7,12-dimethylbenz[a]antracene	A	(**58**) 3-methylbenzo[c]phenanthrene	I
(**34**) 6,12-dimethylbenz[a]antracene	A	(**59**) 6-methylbenzo[c]phenanthrene	I
(**35**) 6,8,12-trimethylbenz[a]antracene	A	(**60**) 6-methylbenz[a]antracene	I
(**36**) 2-methylbenzo[a]pyrene	A	(**61**) 12-methylbenz[a]antracene	I
(**37**) 4-methylbenzo[a]pyrene	A	(**62**) 6-methylanthanthrene	I
(**38**) 11-methylbenzo[a]pyrene	A	(**63**) 6,12-dimethylanthanthrene	I
(**39**) 2-methylbenzo[a]pyrene	A	(**64**) 1-methylbenzo[c]phenanthrene	I
(**40**) 1-methylbenzo[a]pyrene	A	(**65**) 2-methylbenzo[c]phenanthrene	I
(**41**) 4,5-dimethylbenzo[a]pyrene	A	(**66**) 10-methylbenzo[a]pyrene	I
(**42**) 3-methylbenzo[a]pyrene	A	(**67**) 6-methylchrysene	I
(**43**) 1,2-dimethylbenzo[a]pyrene	A	(**68**) 3-methylbenz[a]anthracene	I
(**44**) 2,3-dimethylbenzo[a]pyrene	A	(**69**) 1-methylbenz[a]antracene	I
(**45**) 3,12-dimethylbenzo[a]pyrene	A	(**70**) 11-methylbenz[a]antracene	I
(**46**) 1,3-dimethylbenzo[a]pyrene	A	(**71**) 9-methylbenz[a]antracene	I
(**47**) 1,4-dimethylbenzo[a]pyrene	A	(**72**) 2-methylbenz[a]antracene	I
(**48**) 5-methylbenzo[c]phenanthrene	A	(**73**) 5-methylbenz[a]antracene	I
(**49**) 5-methylchrysene	A	(**74**) 8-methylbenz[a]antracene	I
(**50**) 6,8-dimethylbenz[a]antracene	A	(**75**) 2-methylpyrene	I
(**51**) 7-methylbenz[a]antracene	A	(**76**) 4-methylpyrene	I
(**52**) 5-methylbenzo[a]pyrene	A	(**77**) 1-methylpyrene	I
(**53**) 7-methylbenzo[a]pyrene	A	(**78**) 7,10-dimethylbenzo[a]pyrene	I
(**54**) 6-methylbenzo[a]pyrene	A	(**79**) 6,10-dimethylbenzo[a]pyrene	NA
(**55**) 1,6-dimethylbenzo[a]pyrene	A	(**80**) 8-methylbenzo[a]pyrene	NA
(**56**) 3,6-dimethylbenzo[a]pyrene	A	(**81**) 9-methylbenzo[a]pyrene	NA
(**57**) 4-methylbenzo[c]phenanthrene	I		

[a] See Figure 2 for the molecular structures. The carcinogenic activity data are adapted from the Cavaliere et al. scale:[19] A and I refer to active and inactive, respectively. The carcinogenic activity of the last three molecules is not available (NA).

molecules (79−81 in Figure 2) were chosen for comparative purposes; there are no experimental data available for them.

We have carried out EIM calculations in the framework of tight-binding Hamiltonian[14,15] (simple Hückel-HMO); that is, a natural choice in terms of the vast literature of K, L, and bay region (KLB) models (see ref 15 and references cited therein). We were able to show that the right choice of the molecular descriptors can give much better results than that previously obtained with KLB works, with the advantage that EIM can be applied even to PAH structures without K, L, or B regions.

The PCA study was carried out using the program package Pirouette[20] that contains PCA and related methods. The calculated physicochemical descriptors used in the present work are as follows: the highest occupied molecular orbital (HOMO) energy and its lower level (HOMO-1); the energy difference between HOMO and HOMO-1 (ΔH energy); the lowest unoccupied molecular orbital (LUMO) energy and its upper level (LUMO+1); the energy difference between LUMO and LUMO-1 (ΔL energy); hardness,[21] approximated as HD = (LUMO − HOMO)/2; Mulliken eletronegativity, approximated as $\chi = -$(HOMO + LUMO)/2; the HOMO, HOMO-1 contribution (CH and CH-1, respectively) to the local density of states (LDOS) over the PAH ring with the highest bond order (RHBO) and their difference, $\eta H =$ (CH) − (CH-1); the LUMO, LUMO+1 contribution (CL and CL+1, respectively) to the local density of states (LDOS) over the PAH ring with the highest bond order (RHBO) and their difference, $\eta L =$ (CL) − (CL+1); frontier electron density[22] ($F^{(e)}_n$), frontier orbital density ($F^{(o)}_n$), and frontier radical density ($F^{(r)}_n$) at the atomic position where the highest bond order is located (see Figures 1 and 2); and coefficient

Table 3. Six Descriptors Used in PCA and NN Calculations for 32 Non-Methylated Molecules (Figure 1) and 46 Methylated Molecules (Figure 2), Obtained from Hückel Calculations[a]

molecule	*H*	H-1	Δ*H*	CH	CH-1	*ηH*	molecule	*H*	H-1	Δ*H*	CH	CH-1	*ηH*
1	−0.342	−0.682	0.340	0.457	0.131	0.327	**42**	−0.319	−0.765	0.446	0.487	0.184	0.302
2	−0.371	−0.802	0.431	0.441	0.231	0.209	**43**	−0.302	−0.743	0.441	0.453	0.429	0.024
3	−0.303	−0.793	0.490	0.379	0.283	0.096	**44**	−0.318	−0.685	0.367	0.493	0.252	0.241
4	−0.422	−0.742	0.320	0.443	0.288	0.155	**45**	−0.284	−0.760	0.476	0.447	0.241	0.206
5	−0.398	−0.669	0.271	0.460	0.272	0.188	**46**	−0.272	−0.765	0.493	0.519	0.184	0.335
6	−0.303	−0.648	0.345	0.356	0.186	0.170	**47**	−0.284	−0.787	0.503	0.528	0.391	0.137
7	−0.474	−0.684	0.210	0.548	0.403	0.146	**48**	−0.502	−0.639	0.137	0.541	0.872	−0.331
8	−0.338	−0.671	0.333	0.426	0.135	0.292	**49**	−0.485	−0.738	0.253	0.573	0.537	0.036
9	−0.532	−0.711	0.179	0.784	0.351	0.434	**50**	−0.362	−0.672	0.311	0.608	0.613	−0.005
10	−0.396	−0.680	0.284	0.340	0.648	−0.308	**51**	−0.360	−0.715	0.355	0.561	0.598	−0.037
11	−0.550	−0.603	0.053	0.308	0.790	−0.482	**52**	−0.331	−0.798	0.467	0.536	0.269	0.267
12	−0.452	−0.715	0.263	0.542	0.593	−0.051	**53**	−0.333	−0.759	0.426	0.416	0.171	0.245
13	−0.520	−0.792	0.272	0.710	0.695	0.016	**54**	−0.286	−0.799	0.513	0.476	0.224	0.252
14	−0.568	−0.662	0.094	0.649	0.716	−0.067	**55**	−0.225	−0.795	0.570	0.498	0.190	0.308
15	−0.492	−0.618	0.126	0.519	0.500	0.019	**56**	−0.228	−0.759	0.531	0.517	0.180	0.338
16	−0.499	−0.714	0.215	0.672	0.342	0.330	**57**	−0.510	−0.646	0.136	0.451	0.830	−0.379
17	−0.497	−0.718	0.221	0.541	0.308	0.233	**58**	−0.535	−0.658	0.124	0.618	0.786	−0.168
18	−0.605	−0.769	0.164	0.917	0.551	0.366	**59**	−0.541	−0.592	0.051	0.325	0.908	−0.583
19	−0.684	−0.684	0.000	0.570	0.763	−0.193	**60**	−0.403	−0.690	0.287	0.715	0.553	0.162
20	−0.327	−0.687	0.360	0.336	0.706	−0.370	**61**	−0.375	−0.697	0.323	0.479	0.706	−0.227
21	−0.536	−0.657	0.121	0.598	0.452	0.147	**62**	−0.226	−0.728	0.502	0.276	0.889	−0.612
22	−0.502	−0.680	0.178	0.564	0.393	0.171	**63**	−0.150	−0.705	0.555	0.338	0.783	−0.445
23	−0.522	−0.637	0.115	0.370	0.456	−0.087	**64**	−0.507	−0.659	0.152	0.540	0.754	−0.215
24	−0.555	−0.673	0.118	0.393	0.395	−0.001	**65**	−0.543	−0.627	0.084	0.494	0.789	−0.295
25	−0.429	−0.555	0.126	0.554	0.250	0.304	**66**	−0.353	−0.728	0.375	0.413	0.191	0.221
26	−0.244	−0.618	0.374	0.226	0.581	−0.356	**67**	−0.445	−0.792	0.347	0.725	0.679	0.046
27	−0.291	−0.750	0.459	0.299	0.802	−0.503	**68**	−0.446	−0.665	0.219	0.525	0.616	−0.092
28	−1.000	−1.000	0.000	2.000	2.000	0.000	**69**	−0.452	−0.617	0.165	0.541	0.490	0.051
29	−0.618	−1.000	0.382	1.000	1.333	−0.333	**70**	−0.403	−0.696	0.293	0.430	0.679	−0.249
30	−0.445	−0.879	0.434	0.457	0.654	−0.197	**71**	−0.430	−0.684	0.254	0.490	0.641	−0.151
31	−0.439	−0.684	0.245	0.420	0.492	−0.072	**72**	−0.428	−0.690	0.262	0.603	0.510	0.093
32	−0.539	−0.539	0.000	0.431	0.450	−0.019	**73**	−0.405	−0.675	0.270	0.671	0.603	0.068
33	−0.259	−0.697	0.438	0.512	0.704	−0.192	**74**	−0.397	−0.709	0.312	0.460	0.623	−0.164
34	−0.337	−0.661	0.324	0.607	0.717	−0.111	**75**	−0.445	−0.779	0.334	0.457	0.599	−0.142
35	−0.292	−0.652	0.360	0.536	0.750	−0.214	**76**	−0.398	−0.860	0.462	0.374	0.732	−0.358
36	−0.369	−0.749	0.380	0.432	0.355	0.077	**77**	−0.377	−0.862	0.485	0.470	0.702	−0.232
37	−0.331	−0.787	0.455	0.520	0.392	0.128	**78**	−0.295	−0.673	0.378	0.365	0.201	0.164
38	−0.352	−0.742	0.390	0.405	0.189	0.216	**79**	−0.268	−0.728	0.460	0.454	0.192	0.262
39	−0.327	−0.801	0.475	0.397	0.257	0.140	**80**	−0.367	−0.722	0.355	0.439	0.206	0.233
40	−0.314	−0.799	0.486	0.469	0.202	0.267	**81**	−0.342	−0.801	0.459	0.431	0.241	0.190
41	−0.258	−0.765	0.507	0.684	0.654	0.029							

[a] References 14 and 15. The values for the HOMO (H), HOMO-1 (H-1) energies, and their difference (Δ*H*), the HOMO (CH), and HOMO-1 (CH-1) contribution to the LDOS and their relative contribution to the LDOS (*ηH*) over the RHBO are presented. The descriptors 3 and 6, (Δ*H*) and (*ηH*), were also used in the electronic index methodology.[14,15] Energies are expressed in terms of the usual β units (≈2.4 eV), and the relative contributions are expressed in the normalized population charge values (from 0 up to 2).

of molecular partition octanol−water (log *P*) and the empirical descriptor[1] that gives the molecular size suitable to carcinogenic activity Nat = $(N - 20)^3$, where *N* is the number of C-atoms in the molecules.

The above quantum chemical descriptors were obtained directly from the HMO calculations. log *P* was calculated using parameters of the substituents' hydrophobicity.[23] We expect that some of these descriptors can be correlated to the carcinogenic activity. We used statistical methods to choose a set of descriptors that best correlate to the carcinogenic activity.

The NN calculations were carried out using the program Perceptron-type Neural Network Simulator for Drug Design[24] (PSDD) (Quantum Chemistry Program Exchange 615). The back-propagation method was used, and the calculation process was performed in a supervised way.

3. RESULTS AND DISCUSSION

The EIM methodology allowed us to identify and classify the carcinogenic activity of PAH molecules as active or inactive on the basis of one simple rule stated as follows:

If $\eta H > 0$ and $\Delta H > 0.17\beta$, the molecule will be active; otherwise, the molecule will be inactive. β is the usual resonance integral in HMO, and it is ≈2.4 eV.

This simple rule correctly predicts the carcinogenic activity of the compounds shown in Figures 1 and 2 with an accuracy of 78.2%.

The first molecular set studied with the PCA method was the first 26 non-methylated compounds shown in Figure 1. The basis of our choice for this set was not only because it contains an expressive number of representative molecules of the different classes but mainly because it is the same set used by the EIM methodology.[14] This choice allows a more direct comparison between the different methodologies. From the descriptors list mentioned in Methodology the best separation in active and inactive compounds was obtained using the following descriptors (Table 3): HOMO, HOMO-1, Δ*H*, CH, CH-1, and *ηH*.

In Figure 3 we show the scores of the first two principal components (PC1 and PC2) for the set of 26 non-methylated

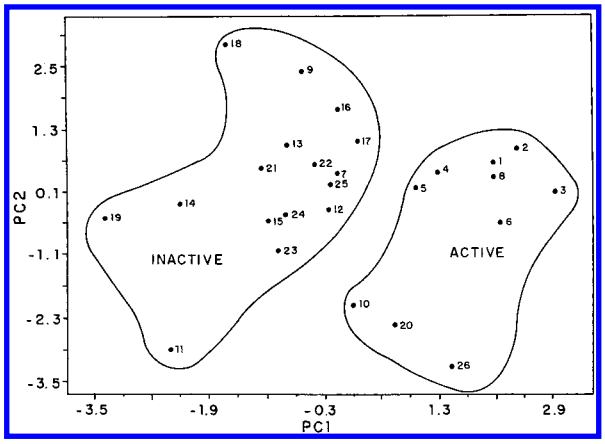

Figure 3. Score graph of the first two principal components (PC1 and PC2) for the set of 26 non-methylated PAH molecules shown in Figure 1.

PAHs. The molecules are distributed into two distinct regions in this figure. The active group is on the right side and the inactive one on the left side. The molecules 7, 9 and 20, 26 are incorrectly classified as inactive and active, respectively. Out of the 26 molecules, 22 are correctly classified, which corresponds to an accuracy of 84.6%.

PC1 and PC2 are given in eqs 1 and 2.

$$\mathrm{PC1} = 0.57(\mathrm{HOMO}) + 0.58(\Delta H) - 0.45(\mathrm{CH\text{-}1}) - 0.28(\mathrm{CH}) - 0.17(\mathrm{HOMO\text{-}1}) + 0.18(\eta H) \quad (1)$$

$$\mathrm{PC2} = -0.24(\mathrm{HOMO}) - 0.03(\Delta H) - 0.31(\mathrm{CH\text{-}1}) + 0.57(\mathrm{CH}) - 0.36(\mathrm{HOMO\text{-}1}) + 0.62(\eta H) \quad (2)$$

PC1 and PC2 respond to 42.6 and 35.2% of the variance, respectively. Equation 1 indicates that HOMO, the difference in energy between HOMO and HOMO-1 (ΔH), and the HOMO-1 contribution to the LDOS (CH-1) are the major descriptors of PC1. Major descriptors of PC2 are the relative ηH contribution to the LDOS (CH − CH-1) and the HOMO contribution to the LDOS (CH). It is interesting to note that the highest contributions for PC1 and PC2 are exactly the same most important variables derived from EIM methodology. The correlation matrix of the six descriptors is given in Table 4. Correlation between pairs of descriptors is generally less than 0.5, except in a few cases.

In Figure 4 we show loadings of the six physicochemical descriptors. The descriptors are grouped in two regions: at the left side of this figure, with the values of the PC1 axis less than −0.1 (descriptors HOMO-1, CH, and CH-1) and at the right side with the values of the PC1 axis greater than +0.1 (descriptors H, ΔH, and ηH). A comparison between Figures 3 and 4 shows that the three descriptors, HOMO-1, CH, and CH-1, are responsible for pulling the inactive molecules toward to the left side in the score graph (Figure 3). The descriptors that are mostly responsible for pulling active molecules toward to the right side in Figure 3 are H, ΔH, and ηH.

In order to study the predictive ability of the PCA method, we applied PCA to the molecules 27−32 (Figure 1) and 33−78 (Figure 2), using exactly the same six descriptors as those used for set of 26 PAHs (Table 3). The score graph of PCA is illustrated in Figure 5. The active group is located on the right side of this figure and the inactive one on the left side. Molecules 28 and 29 cannot be seen in Figure 5, because they are in the region with the value of the PC1 axis less than −3.0. Molecules 27, 62, 63, 66, and 78 are incorrectly classified as actives; also molecules 34, 35 and 48−51 are incorrectly classified as inactives. Out of the 52 newly added compounds, 41 molecules are correctly classified, which corresponds to an accuracy of 78.8%. If we consider only the set of 32 non-methylated PAHs, there are 27 molecules correctly classified (84.4%). If we consider only the set of 46 methylated compounds, a total of 36 molecules are correctly classified (78.3%). Out of the global set of 78 PAHs 63 molecules are correctly classified, which corresponds to 80.8% overall correct classification, slightly superior to EIM (78.2%).

The results for methylated compounds deserve special consideration. It is a well-known experimental fact that chemical substitution (methylation, for instance) at PAH molecules can drastically affect their carcinogenic activity,[6,9] depending on the site of substitution and on the number of substituted groups. Active molecules can become inactive or vice versa, or the carcinogenic power can be largely varied (increased or decreased). These facts have not been consistently explained in terms of K−L theories. Although the methylation process does not change the total number of π electrons, it produces perturbations on the π electronic density of states, such as changing the relative contribution of HOMO and HOMO-1 to the local density of states. If the EIM methodology is physically sound, we could expect methylation to induce a discontinuous transition in the carcinogenic activity; i.e., it could make active molecules inactive and vice versa. In fact, this has been confirmed in our previous work.[15] The good agreement between the EIM and PCA results, quite different approaches, substantiates the main features of the EIM methodology.

We have also separately analyzed the methylated set with PCA. The presence of methyl groups breaks the electron−hole symmetry present in non-methylated molecules, but the EIM methodology can treat them in the same way. On the other side the break of symmetry allows the introduction of new descriptors in PCA analysis that is not possible when methylated and non-methylated molecules are analyzed simultaneously. We would like to investigate if these new descriptors can improve the results.

This PCA analysis selected a new set of four descriptors: HOMO and LUMO energies, hardness (HD), and the difference in energy between HOMO and HOMO-1 (ΔH). From the original descriptors (eqs 1 and 2) two of them are present again: HOMO and ΔH. These descriptors are listed in Table 5.

Figure 6 shows the scores of the first two principal components (PC1 and PC2) for the set of 46 methylated PAHs. The molecules are grouped into three regions indicated in the figure as A, B (active), and C (inactive). Region A consists of the three most active compounds[19] (see Table 2). Molecules 48−51 are incorrectly classified as inactives. Molecules 62, 63, 66, and 78 are also incorrectly classified as actives. All the remaining molecules are correctly classified. A total of 38 molecules out of the 46 are correctly classified in Figure 6, which corresponds to an 82.6% correct classification, only a slight improve upon the joint analysis of methylated and non-methylated compounds (78.3%).

Table 4. Correlation Matrix of the Six Physicochemical Descriptors (Table 3) for the First 26 Non-Methylated PAH Molecules (Figure 1)

	H	H-1	ΔH	CH	CH-1	ηH
H	1.0000	0.0124	0.8717	-0.6292	-0.4621	-0.0297
H-1		1.0000	−0.4792	−0.3550	0.0877	−0.2982
ΔH			1.0000	−0.3783	−0.4486	0.1201
CH				1.0000	0.0624	0.5905
CH-1					1.0000	−0.7686
ηH						1.0000

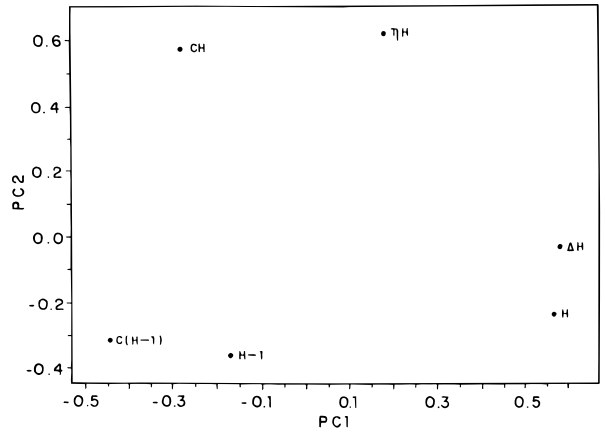

Figure 4. Loadings of the six physicochemical descriptors (described in Table 3) selected for PCA to the set of 26 non-methylated PAH molecules.

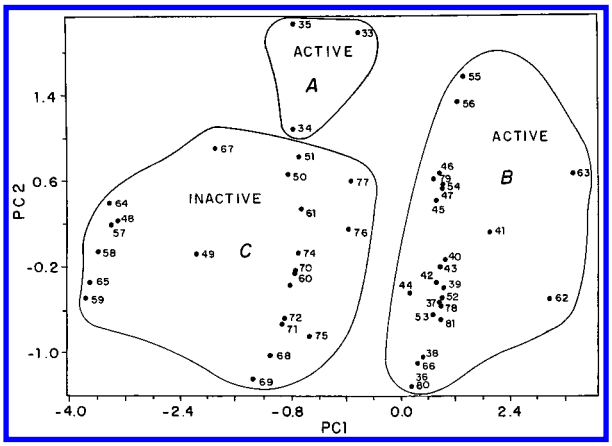

Figure 5. Score graph of the first two principal components (PC1 and PC2) for the set of 49 methylated PAH molecules (Figure 2).

The two principal components are given by

$$PC1 = 0.51(HOMO) - 0.42(LUMO) - 0.55(HD) + 0.51(\Delta H) \quad (3)$$

$$PC2 = 0.44(HOMO) + 0.81\,(LUMO) + 0.13(HD) + 0.36(\Delta H) \quad (4)$$

PC1 and PC2 respond to 81.5 and 15.9% of the variance, respectively. Equation 3 shows that the four variables are almost equally important, while in eq 4 LUMO dominates. It is interesting that the break of electron−hole symmetry induced by methylation has an important role in defining PC2.

The experimental carcinogenic activities of molecules 79−81 are not available, and they were used for comparison purposes. All three molecules are classified as active with the PCA method, as can be seen in Figures 3 and 6. These results are in agreement with previsions obtained by EIM methodology.[15]

Table 5. Selected Descriptors That Were Employed in the Statistical Analysis for the Set of 49 Methylated PAHs Obtained from Hückel Calculations[a]

molecule	*H*	*L*	HD	ΔH
33	−0.259	0.597	0.428	0.438
34	−0.337	0.573	0.455	0.324
35	−0.292	0.632	0.462	0.360
36	−0.369	0.372	0.371	0.380
37	−0.331	0.406	0.369	0.455
38	−0.352	0.385	0.369	0.390
39	−0.327	0.412	0.369	0.475
40	−0.314	0.427	0.371	0.486
41	−0.258	0.424	0.341	0.507
42	−0.319	0.419	0.369	0.446
43	−0.302	0.428	0.365	0.441
44	−0.318	0.429	0.374	0.367
45	−0.284	0.466	0.375	0.476
46	−0.272	0.478	0.375	0.493
47	−0.284	0.468	0.376	0.503
48	−0.502	0.598	0.549	0.137
49	−0.485	0.542	0.514	0.253
50	−0.362	0.549	0.456	0.311
51	−0.360	0.553	0.457	0.355
52	−0.331	0.406	0.369	0.467
53	−0.333	0.404	0.368	0.426
54	−0.286	0.469	0.378	0.513
55	−0.225	0.521	0.373	0.570
56	−0.228	0.511	0.369	0.531
57	−0.510	0.598	0.554	0.136
58	−0.535	0.589	0.562	0.124
59	−0.541	0.571	0.556	0.051
60	−0.403	0.491	0.447	0.287
61	−0.375	0.523	0.449	0.323
62	−0.226	0.362	0.294	0.502
63	−0.149	0.424	0.287	0.555
64	−0.507	0.611	0.559	0.152
65	−0.543	0.577	0.559	0.084
66	−0.353	0.384	0.369	0.375
67	−0.445	0.589	0.518	0.347
68	−0.446	0.456	0.451	0.219
69	−0.452	0.452	0.452	0.165
70	−0.403	0.492	0.447	0.293
71	−0.430	0.468	0.449	0.254
72	−0.428	0.469	0.449	0.262
73	−0.405	0.487	0.446	0.270
74	−0.397	0.499	0.449	0.312
75	−0.445	0.445	0.445	0.334
76	−0.398	0.484	0.441	0.462
77	−0.377	0.509	0.444	0.485
78	−0.295	0.408	0.352	0.378
79	−0.268	0.480	0.374	0.460
80	−0.367	0.375	0.371	0.355
81	−0.342	0.397	0.369	0.459

[a] References 14 and 15. Calculated HOMO and LUMO energies (*H*, *L*, respectively), hardness (HD), and energy difference between HOMO and (HOMO-1) levels (ΔH). Energies are expressed in terms of the usual β units (≈2.4 eV).

Figure 7 shows the hierarchical clustering dendogram. The hierarchical clustering dendogram (Hier) is a conceptually simple but effective clustering technique. Given a set of

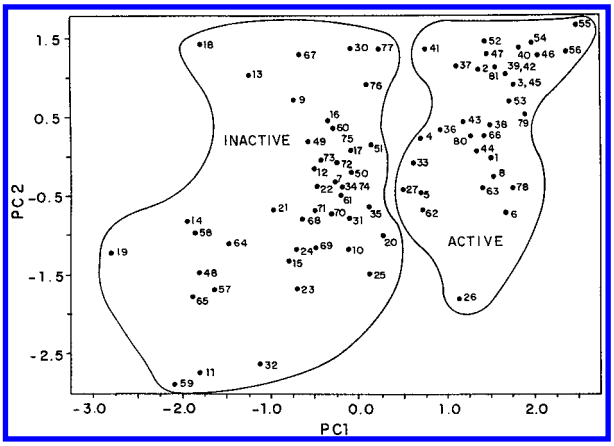

Figure 6. Score graph of the first two principal components (PC1 and PC2) for the global set of 81 PAH molecules. Molecules 28 and 29 cannot be seen, because they are in the region with a value of the PC1 axis less than −3.0.

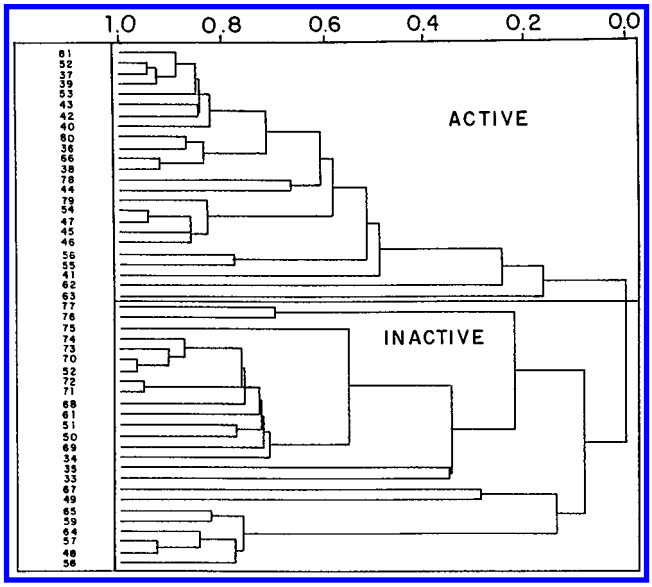

Figure 7. Dendogram obtained for the set of 49 methylated PAH molecules.

compounds and a list of descriptors, this technique clusters the compounds that share a common property (similarity). See refs 25 and 26 for details about dendogram techniques.

Hier uses the similarity measurements from eqs 1 and 2 below. It is assumed that the distance d_{ij} (in an N-dimensional space) between two points (molecules) is a good measure of their similarity. d_{ij} is defined by the expression

$$d_{ij} = [\sum_{k=1}^{N}(x_{ik} - x_{jk})^2]^{1/2} \quad (5)$$

where the summation is over all the descriptors.

Actually, d_{ij} is a reciprocal similarity measure because objects are more alike as d_{ij}, goes to zero. In order to remedy this, a similarity measure is defined as

$$S_{ij} = 1 - d_{ij}/\mathrm{MAX}(d_{ij}) \quad (6)$$

where $\mathrm{MAX}(d_{ij})$ is the largest interpoint distance. From this function the most unlike objects have $S = 0$ and identical ones $S = 1$. The procedure to build the dendogram is the following: (a) to calculate the distance matrix $[d_{ij}]$ for all the molecules, using eq 5; (b) to determine the shortest distance and the similarity (eq 6) and to register them in the dendogram; (c) to substitute the two points by a group (cluster) and to calculate the new distance matrix (eq 5). (this new matrix has a lower dimension ($N - 1$) than the initial one (N); the two points are substituted by a cluster in a halfway position between them); (d) to repeat steps b and c until all the points are included in the dendogram.

Table 6. Parameters Used (α and θ) in the Neural Networks Calculations[a]

layer	neurons	α	θ
1	6		
2	12	2.0	0.0
3	2	16.5	0.0

[a] α is the nonlinear parameter of the sigmoid functions, and θ is a threshold value for a neuron.[30] The NN have used 29 901 training epochs.

Table 7. Parameters Used (α and θ) in the Neural Network Calculations[a]

layer	neurons	α	θ
1	4		
2	15	3.0	0.0
3	2	12.0	0.0

[a] α is the nonlinear parameter of the sigmoid functions, and θ is a threshold value for a neuron.[30] The NN have used 29 980 training epochs.

From the dendogram (Figure 7) we can see that with the use of four descriptors used in the PCA calculations two clusters are formed (separated by the horizontal line between the no. 63 and no. 77 molecules). One group is mostly composed of active molecules (upper half of the figure) and the other by inactive molecules (lower half). These two clusters have zero similarity. This demonstrates that active and inactive molecules are well-separated in this four-dimensional space. The molecules incorrectly classified using Hier are the same as those observed with the PCA analysis.

In summary, the PCA analysis confirms the EIM descriptors as relevant in identifying the PAHs' carcinogenic activity. PCA combines other descriptors, resulting in an overall improvement of the molecular classifications: 78.3 and 73.9% for methylated compounds and 80.8 and 78.2% for global analysis (methylated and non-methylated), respectively, for PCA and EIM.

Next NN was employed to study the SAR of the PAHs. For these studies we adopted the following procedure. We started by studying the 26 non-methylated PAH molecules using six neurons in the first layer that are the same six physicochemical descriptors used in PCA in order to allow a direct comparison with the PCA results. The parameters (Table 3) used in NN calculations are listed in Table 6. The perceptron-type NN consists of three layers. The training of NN was carried out according to the back-propagation algorithm[29] until the error function[30] reaches the convergence criterion (≤0.000 01 in our case).

α is a parameter which expresses the nonlinearity of the neuron's operations.[30] It is a value of the sigmoid function of the neurons in the second and third layers. Its default value is 1.0. α was forced to change (according to the values in Tables 6 and 7) until the error function reaches the

Table 8

(a) Results from the Neural Network (NN) Calculations:

molecule	category	training pattern		output pattern		molecule	category	training pattern		output pattern	
1	1	1	0	0.998	0.002	**14**	2	0	1	0.000	1.000
2	1	1	0	0.999	0.002	**15**	2	0	1	0.004	0.996
3	1	1	0	0.999	0.002	**16**	2	0	1	0.009	0.991
4	1	1	0	0.999	0.001	**17**	2	0	1	0.009	0.991
5	1	1	0	0.999	0.001	**18**	2	0	1	0.001	0.999
6	1	1	0	0.999	0.002	**19**	2	0	1	0.000	1.000
7	1	1	0	0.990	0.010	**20**	2	0	1	0.004	0.996
8	1	1	0	0.999	0.002	**21**	2	0	1	0.000	1.000
9	1	1	0	0.990	0.010	**22**	2	0	1	0.004	0.996
10	1	1	0	0.990	0.010	**23**	2	0	1	0.004	0.996
11	2	0	1	0.000	1.000	**24**	2	0	1	0.005	0.995
12	2	0	1	0.004	0.997	**25**	2	0	1	0.000	1.000
13	2	0	1	0.002	0.999	**26**	2	0	1	0.007	0.993

(b) Prediction of the Carcinogenic Activity for 42 PAHs: Molecules **27**−**32** (Figure 1) and **33**−**81** (Figure 2), Using the NN Trained in a

molecule	category	predicted pattern		molecule	category	predicted pattern	
27	2	0.000	1.000	**55**	1	0.999	0.002
28	2	0.000	1.000	**56**	1	0.999	0.002
29	1	0.954	0.055	**57**	2	0.000	1.000
30	2	0.000	1.000	**58**	2	0.000	1.000
31	1	0.992	0.008	**59**	2	0.000	1.000
32	2	0.004	0.996	**60**	1	0.986	0.014
33	1	0.868	0.136	**61**	2	0.002	0.998
34	1	0.736	0.273	**62**	2	0.000	1.000
35	1	0.984	0.016	**63**	2	0.002	0.998
36	1	0.998	0.002	**64**	2	0.000	1.000
37	1	0.998	0.003	**65**	2	0.000	1.000
38	1	0.999	0.002	**66**	1	0.999	0.002
39	1	0.999	0.002	**67**	2	0.002	0.999
40	1	0.999	0.002	**68**	1	0.857	0.161
41	2	0.002	0.999	**69**	2	0.005	0.995
42	1	0.999	0.002	**70**	2	0.009	0.992
43	1	0.996	0.005	**71**	1	0.594	0.419
44	1	0.999	0.001	**72**	1	0.992	0.008
45	1	0.999	0.002	**73**	1	0.980	0.020
46	1	0.999	0.002	**74**	2	0.267	0.747
47	1	0.999	0.002	**75**	2	0.002	0.999
48	2	0.000	1.000	**76**	2	0.000	1.000
49	2	0.002	0.998	**77**	2	0.000	1.000
50	1	0.991	0.009	**78**	1	0.999	0.002
51	1	0.767	0.241	**79**	1	0.999	0.002
52	1	0.999	0.002	**80**	1	0.998	0.002
53	1	0.999	0.002	**81**	1	0.999	0.002
54	1	0.999	0.002				

Table 9. Percentage of Correct Classification for the Three Different Methodologies (EIM, PCA, and NN) for Methylated (M) and Non-Metylated (NM) Compounds Considerated Separately and the Overall Results (M + NM)

methodology	NM (32 PAHs)	M (46 PAHs)	M + NM (78 PAHs)
EIM	84.4%	73.9%	78.2%
PCA	84.4%	78.3%	80.8%
NN	93.8%	78.3%	84.6%

convergence. θ is a threshold value for neuron in the second and third layers.[30] It was set to its usual value (0.0).

We trained the NN with the first 26 non-methylated PAHs to predict carcinogenic activity of the methylated and non-methylated PAHs. Table 8a shows the results of NN training, and Table 8b shows the prediction results for the methylated and non-methylated PAH molecules. The active group belongs to category 1, and the inactive one belongs to category 2. The training pattern of category 1 is (1 0), whereas the training pattern of category 2 is (0 1), as seen in Table 8a. In the initial phase of "NN learning", the weight matrix was calculated with the training pattern using the six parameters for each of the 26 molecules. Although the neurons at the first layer can take continuous values between 0 and 1, those of the last layer (output patterns) are required to assume discrete values 0 or 1. However, in general, the final output does not present a complete set of discrete values. This is due to the limit of the resolution ability of the network, but it is expected that these values would be close to the discrete ones [(1 0) or (0 1)]. The classification criteria in these cases are based on its closer proximity to these limits [for instance (0.998 0.002) to (1 0) and (0.004 0.996) to (0 1)]. The NN "learned" the training pattern with success (100%). The four molecules, 7, 9, 20, and 26, incorrectly classified with PCA are now correctly described.

If we take the set of 32 non-methylated PAHs (see Figure 1), 30 molecules out of 32 are correctly classified (Table 8a,b), this corresponds to an accuracy of 93.8%. If we consider only the set of 46 methylated compounds, a total

Table 10

(a) Results of the NN Trained with the Same Parameters as Those in Table 7, for a Subset of 40 Methylated Compounds (Figure 2)

molecule	category	training pattern		output pattern		molecule	category	training pattern		output pattern	
36	1	1	0	0.998	0.002	**56**	1	1	0	1.000	0.000
37	1	1	0	1.000	0.000	**57**	2	0	1	0.001	0.999
38	1	1	0	0.990	0.010	**58**	2	0	1	0.000	1.000
39	1	1	0	1.000	0.000	**59**	2	0	1	0.000	1.000
40	1	1	0	1.000	0.000	**60**	2	0	1	0.000	1.000
41	1	1	0	0.995	0.005	**61**	2	0	1	0.002	0.998
42	1	1	0	1.000	0.000	**62**	2	0	1	0.001	0.999
43	1	1	0	1.000	0.000	**63**	2	0	1	0.002	0.998
44	1	1	0	0.992	0.008	**64**	2	0	1	0.008	0.992
45	1	1	0	1.000	0.000	**65**	2	0	1	0.000	1.000
46	1	1	0	1.000	0.000	**66**	2	0	1	0.011	0.989
47	1	1	0	1.000	0.000	**67**	2	0	1	0.006	0.994
48	1	1	0	0.992	0.009	**68**	2	0	1	0.000	1.000
49	1	1	0	0.993	0.006	**69**	2	0	1	0.000	1.000
50	1	1	0	0.997	0.003	**70**	2	0	1	0.002	0.998
51	1	1	0	1.000	0.000	**74**	2	0	1	0.007	0.994
52	1	1	0	1.000	0.000	**75**	2	0	1	0.002	0.998
53	1	1	0	1.000	0.000	**76**	2	0	1	0.002	0.998
54	1	1	0	1.000	0.000	**77**	2	0	1	0.005	0.995
55	1	1	0	1.000	0.000	**78**	2	0	1	0.007	0.993

(b) Prediction of Activity for Nine Methylated Molecules (**33**−**35**, **71**−**73**, **79**−**81**, Figure 2), Using the Trained Set of a

molecule	category	predicted pattern		molecule	category	predicted pattern	
33	1	0.998	0.002	**73**	2	0.002	0.998
34	1	0.998	0.002	**79**	1	1.000	0.000
35	1	0.998	0.002	**80**	2	0.000	1.000
71	2	0.002	0.998	**81**	1	1.000	0.000
72	2	0.002	0.998				

of 36 molecules out of 46 are correctly classified, which corresponds to 78.3%. A total of 66 molecules out of the 14 global set of 78 PAHs are correctly classified; that corresponds to 84.6% of correct global classification.

Table 9 compares the percentage of correct classification (%) of the three different methods (EIM, PCA, and NN), for the three different classes of compounds (non-methylated (NM), methylated (M), and the overall results (NM + M)). These results refer to PCA and NN data set using the same six descriptors mentioned above compared to EIM ones using only two of them. On average all of the three methods have approximately the same margin of correct prediction (80%). There are, however, slight differences in the percentages attained among the three different methods. The NN seems to give a slightly better percentage than PCA, which gives a slightly better percentage than EIM.

We have also separately analyzed the set of 49 methylated PAHS using the set of four physicochemical descriptors used on PCA analysis (eqs 3 and 4) by the same reasons mentioned above.

First, we divided the set of 49 methylated PAHs (Table 2) into two groups: GN1 (test) and GN2 (training). The GN1 group consists of nine molecules: no. 33, no. 34, and no. 35 (actives); no. 71, no. 72, and no. 73 (inactives); no. 79, no. 80, and no. 81 (unknown activity). Since we have three molecules with unknown activity, we have also selected three active and three inactive molecules randomly chosen from each group followed by two neighbors (similar potency). The GN2 group consists of the remaining 40 compounds. The parameters that are used in NN calculations to the group B molecules are listed in Table 7. There are four neurons in the first layer related to the physicochemical descriptors (parameters) listed in Tables 3 and 5. In the initial phase of "learning" in the NN, the weight matrix was calculated with the training pattern using the four parameters for each of the 40 molecules in the group B. The weight matrix was calculated in the initial phase of learning, and it was used to calculate the output pattern for each of the 40 molecules that were used to train the NN (Table 10a). A comparison between the training patterns and the output patterns shows a complete qualitative agreement between the two. The NN "learned" the training pattern with 100% success (Table 10a). All of the molecules are correctly classified. Molecules 33−35 are correctly classified as active; molecules 71−73 are also correctly classified as inactives. Molecules 79 and 81 are classified as actives, and molecule 80 is classified as inactive. The EIM and PCA methods predict these last three molecules as actives.

4. SUMMARY AND CONCLUSIONS

We have carried out a comparative study of three different methodologies in predicting the carcinogenic activity of polycyclic aromatic hydrocarbons (PAHs): electronic index methodology (EIM), principal component analysis (PCA), and neural networks (NN).

EIM[14,15] explores the concept of local density of states (LDOS) over a specific molecular region and critical values for the separation (in energy) between frontier molecular orbitals and their contribution to the LDOS. Our results show that these different methodologies have approximately the same predictive power (80%).

From the PCA analysis for the study of methylated and non-methylated PAH compounds six descriptors appear as efficient parameters for the classification (active or inactive). They are HOMO and HOMO-1 energies, their energy

difference, HOMO and HOMO-1 contribution, and their relative difference to the LDOS over the PAH ring containing the highest bond order (RHBO).

When methylated PAHs are separately analyzed, four descriptors are selected as efficient parameters for the classification: the HOMO and LUMO energies, hardness, and the difference in energy between HOMO and HOMO-1 levels.

Using the same sets of descriptors, the NN was successfully trained and can efficiently predict the PAHs activity (Table 9). The HOMO and HOMO-1 relative contribution to the LDOS over the RHBO and their difference in energy are the key descriptors in EIM and are again confirmed by PCA and NN as relevant descriptors for classifying the carcinogenic activity.

One of the major difficulties in applying EIM is to identify the variables and the specific molecular regions for deriving the classificatory rules. On the other hand PCA can be easily used to help in these issues. A combined use of these different methodologies could be an efficient tool for deriving faster and better rules.

ACKNOWLEDGMENT

This work was supported in part by the Brazilian agencies CNPq and FAPESP. Computational assistance from CENAPAD-SP is acknowledged. The authors wish to thank Prof. P. M. V. B. Barone for helpful discussions.

REFERENCES AND NOTES

(1) Govers, H. A. J. In *Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W., Devilers, J., Eds.; Kluwer Academic Publishers: Dordrecht, 1990; The Netherlands, 1990; p 411.
(2) Harvey, R. G.; Geacintov, N. E. *Acc. Chem. Res.* **1988**, *21*, 66.
(3) Coulson, C. A. *Adv. Cancer Res.* **1953**, 1, and references therein.
(4) Pullman, A.; Pullman, B. *Adv. Cancer Res.* **1955**, *3*, 17, and references cited therein.
(5) Streitwieser, A. *Molecular Orbital Theory*; John Wiley & Sons: New York, 1961.
(6) Silverman, J. P. B. D. *Acc. Chem. Res.* **1984**, *17*, 332.
(7) Jerina, D. M.; et al. In *Carcinogenesis: Fundamental Mechanisms and Environmental Effects*; Pullman, B., Ts'O, P. O., Gelboin, H., et al., Eds.; D. Reidel Publishing Co.: Dordrecht, The Netherlands, 1980.
(8) Szentpály, L. V. *J. Am. Chem. Soc.* **1984**, *106*, 6021.
(9) Kimri, S.; Gayoso, J. *J. Mol. Struct. (THEOCHEM)* **1996**, *362*, 141.
(10) Gayoso, J.; Kimri, S. *Int. J. Quantum Chem.* **1990**, *38*, 461; **1990**, *38*, 487.
(11) Nordén, U. E.; Svante, W. *Acta Chem. Scand.* **1978**, *B32*, 602.
(12) Villemin, D.; Cherqaoui, D.; Mesbah, A. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1288.
(13) Song, X.-H.; Xiao, M.; Yu, R.-Q. *Comput. Chem.* **1994**, *18*, 391.
(14) Barone, P. M. V. B.; Camilo, A., Jr.; Galvão, D. S. *Phys. Rev. Lett.* **1996**, *77*, 1186.
(15) Braga, R. S.; Barone, P. M. V. B.; Galvao, D. S. *J. Mol. Struct. (THEOCHEM)* **1999**, *464*, 257.
(16) Zhang, L.; Sannes, K.; Shusterman, A. J.; Hansch, C. *Chem.-Biol. Interact.* **1992**, *81*, 149.
(17) Isu, Y.; Nagashima, U.; Aoyama, T.; Hosoya, H. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 286.
(18) Miyashita, Y.; Takahashi, Y.; Daiba, S.; Abe, H.; Sasaki, S. *Anal. Chim. Acta* **1982**, *143*, 35.
(19) Cavaliere, E. L.; Rogan, E. G.; Roth, R. W.; Saugier, R. K.; Hakan, A. *Chem.-Biol. Interact.* **1983**, *47*, 87.
(20) *Pirouette Multivariate Data Analysis for IBM PC Systems*, Version 2.0, Infometrix: Seattle, WA, 1996.
(21) Parr, R. G.; Pearson, R. G. *J. Am. Chem. Soc.* **1983**, *105*, 7512.
(22) Fukui, K. In *Theory of orientation and stereo selection*; Hafner, K., Ed.; Reactivity and Structure Concepts in Organic Chemistry; Springer-Verlag: Berlin, 1975; Vol. 2, p 39.
(23) Nys, G. G.; Rekker, R. F. *Eur. J. Med. Chem. Chim. Ther.* **1974**, *9*, 361.
(24) Ichikawa, H. *PSDD: Perceptron-type Neural Network Simulator*, QCPE 615.
(25) Jyrol, R. C.; Bailey, D. E. *Cluster Analysis*; McGraw-Hill Book Co.: New York, 1970.
(26) Kowalski, B. R.; Bender, C. F. *J. Am. Chem. Soc.* **1972**, *94*, 5632.
(27) Daudel, P.; Daudel, R. *Chemical Carcinogenesis and Molecular Biology*; Wiley-Interscience: New York, 1966; pp 1−5.
(28) Herndon, W. C.; Szentpály, L. V. *J. Mol. Struct. (THEOCHEM)* **1986**, *148*, 141.
(29) Rumelhart, D. E.; McClelland, J. L. *Parallel Distributed Processing Exploration in Microstructure of Cognition*.
(30) Aoyama, T.; Suzuki, Y.; Ichikawa, H. *J. Med. Chem.* **1990**, *33*, 2583.

CI990326V