Identifying Relevant Molecular Descriptors Related to Carcinogenic Activity of Polycyclic Aromatic Hydrocarbons (PAHs) Using Pattern Recognition Methods

V. R. Coluci,* R. Vendrame, R. S. Braga, and D. S. Galvão

Instituto de Física, UNICAMP, CP 6165, CEP 13083-970, Campinas - SP - Brazil

Received August 12, 2002

Polycyclic Aromatic Hydrocarbons (PAHs) constitute an important family of molecules capable of inducing chemical carcinogenesis. In this work we report structure−activity relationship (SAR) studies for 81 PAHs using the pattern-recognition methods Principal Component Analysis (PCA), Hierarchical Clustering Analysis (HCA) and Neural Networks (NN). The used molecular descriptors were obtained from the semiempirical Parametric Method 3 (PM3) calculations. We have developed a new procedure that is capable of identifying the PAHs' carcinogenic activity with an accuracy higher than 80%. PCA selected molecular descriptors that can be directly correlated with some models proposed to PAHs' metabolic activation mechanism leading to the formation of PAHs-DNA adducts. PCA, HCA and NN validate the energy separation between the highest occupied molecular orbital and its next lower level as a major descriptor defining the carcinogenic activity. This descriptor has been only recently discussed in the literature as one new possible universal parameter for defining the biological activity of several classes of compounds.

1. INTRODUCTION

Cancer is a disease of multicellular organisms involving multistep processes in which cells accumulate genetic alterations as they progress to a more malignant phenotype.[1] It is believed that many factors can be associated with cancer induction (such as virus, radiation, chemical agents, etc.), but the chemical component is the most important. Among the chemicals that are known to induce cancer PAHs constitute a very important class. Only mycotoxin mold metabolites are more carcinogenic than PAHs.[2] The PAHs' carcinogenic power varies from some of the strongest carcinogens known to inactive ones.[2]

The investigation of why some of these very similar molecules present carcinogenic activity and others do not started in the 1930s with the work of Cook and co-workers.[3] They reported a relationship between carcinogenic activity (malignant tumors in rats) and some molecular geometrical features. Pullman and Pullman[4] demonstrated that it is possible to correlate the PAHs' carcinogenic activity with some critical index values over specific molecular regions (later known as K and L regions (see inset of Figure 1)). These indices were obtained from quantum chemical calculations using simple Hückel theory.[5] Later similar theories evolved to include what is called the 'bay region'[6−8] (B region in the inset of Figure 1).

These and others theories using electronic and geometrical properties, statistical analysis, neural networks, and artificial intelligence methods have achieved only partial success.[9−12] 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[13,14] a new methodology called EIM (Electronic Indices Methodology) was proposed to identify PAH carcinogenic activity. Through very simple rules derived from Hückel calculations it was possible to group and identify active and inactive PAH compounds.[13,14] Further studies[15] using the well-known semiempirical method PM3 (Parametric Method 3)[16] imposing $\sigma-\pi$ separation (in order to provide a direct comparison with Hückel results) produced similar rules indicating that the general conclusions were not method dependent at this level. These results[13−15] were validated in the framework of more standard structure−activity relationship (SAR) studies[15,17] using the pattern-recognition methods Principal Component Analysis (PCA), Hierarchical Clustering Analysis (HCA) and Neural Networks (NN).[18−20]

In the present work we have expanded these studies carrying out PM3 fully geometrical optimizations on methylated and nonmethylated PAHs coupled to the EIM methodology. The obtained molecular descriptors are then used in SAR analysis (PCA, HCA and NN).

Our results showed that with the use of these combined methodologies we can group and classify (certainty higher than 80%) the PAH compounds in terms of their carcinogenic activity. We have also observed a direct relationship between some of electronic descriptors obtained from EIM and PCA analysis with proposed mechanism models for PAHs metabolic activation.

2. METHODOLOGY

We have investigated the 81 PAH molecules shown in Figures 1 and 2, in their neutral and ionic (cations and anions) forms, totalizing 243 analyzed structures. The methylated structures shown in Figure 2 are structurally related to the nonmethylated ones shown in Figure 1. These molecules

* Corresponding author phone: +55-19-3788-5369; fax: +55-19-3788-5376; e-mail: coluci@ifi.unicamp.br.

10.1021/ci025577+ CCC: $22.00 © 2002 American Chemical Society
Published on Web 10/25/2002

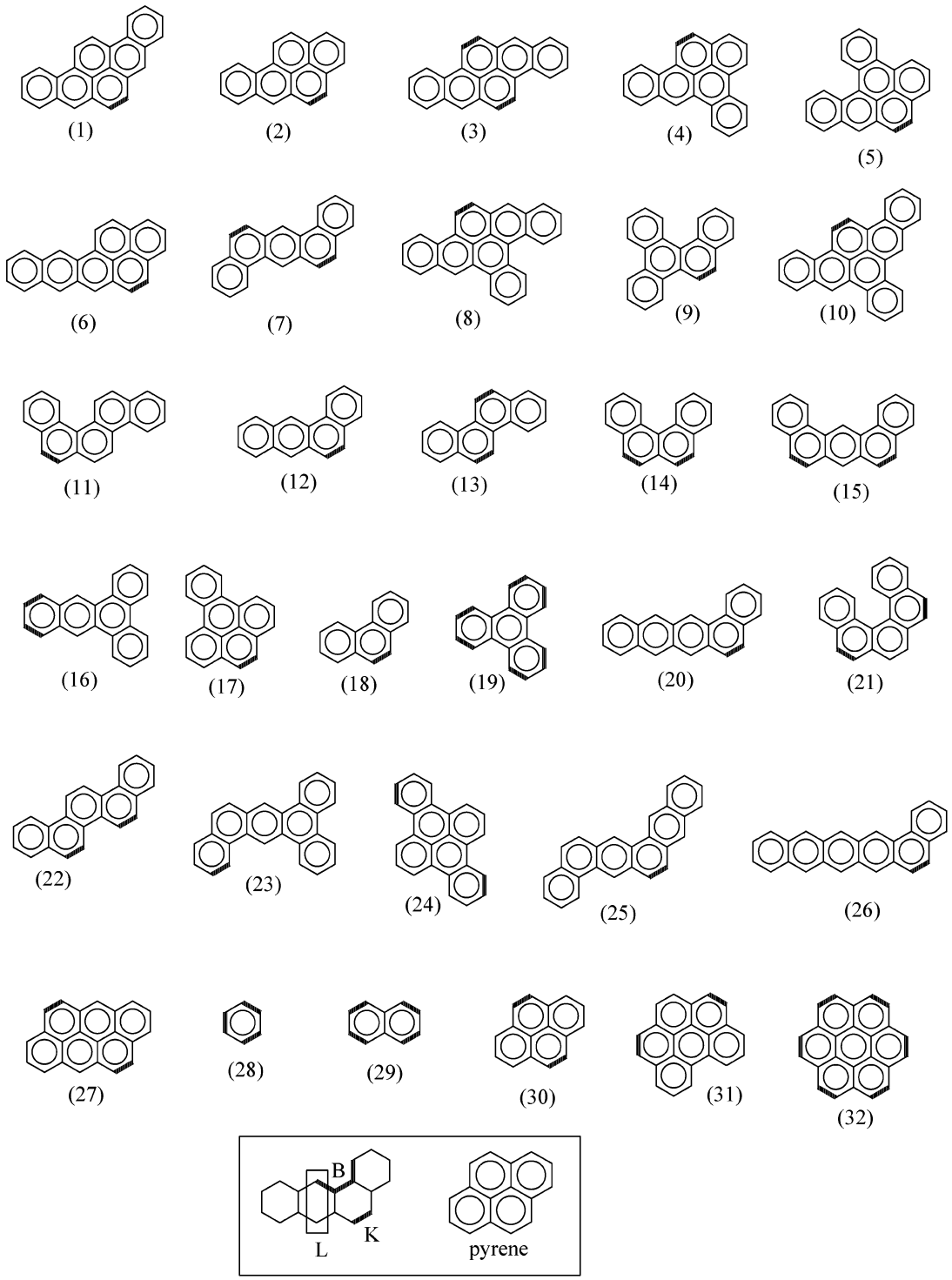

Figure 1. Molecular structure of 32 nonmethylated polycyclic aromatic hydrocarbon (PAH) molecules. The darker bonds indicate the calculated highest bond order. In the inset are shown the pyrene structure and typical L, K and Bay (B) regions for PAH molecules. Table 1 lists their IUPAC names.

were selected considering the availability of experimental data for their carcinogenic activity (see Tables 1 and 2). We used the same approach proposed by Villemin et al.[11] simply classifying the compounds into two classes: active (A) and inactive (I). The experimental data used here to define active and inactive compounds are from the Iball indices[9] and from the scale proposed by Cavaliere et al.[21]

The PM3 fully geometrical optimizations calculations were carried out using MOPAC6 package[22] contained in the Chem2Pac[23,24] program. ChemBats2Pac is a free software package developed at our group for the Windows operating system and used in the EIM analysis.

The following calculated physicochemical descriptors were chosen to be used in our SAR analysis:

- Ionization Potential (IP), which was approximated by taking the negative value of the HOMO (Highest Occupied Molecular Orbital) energy (ϵ_{HOMO}) (Koopmans' theorem).

- Electron Affinity (EA), which was approximated by taking the negative value of the LUMO (Lowest Unoccupied Molecular Orbital) energy (ϵ_{LUMO}) (Koopmans' theorem).

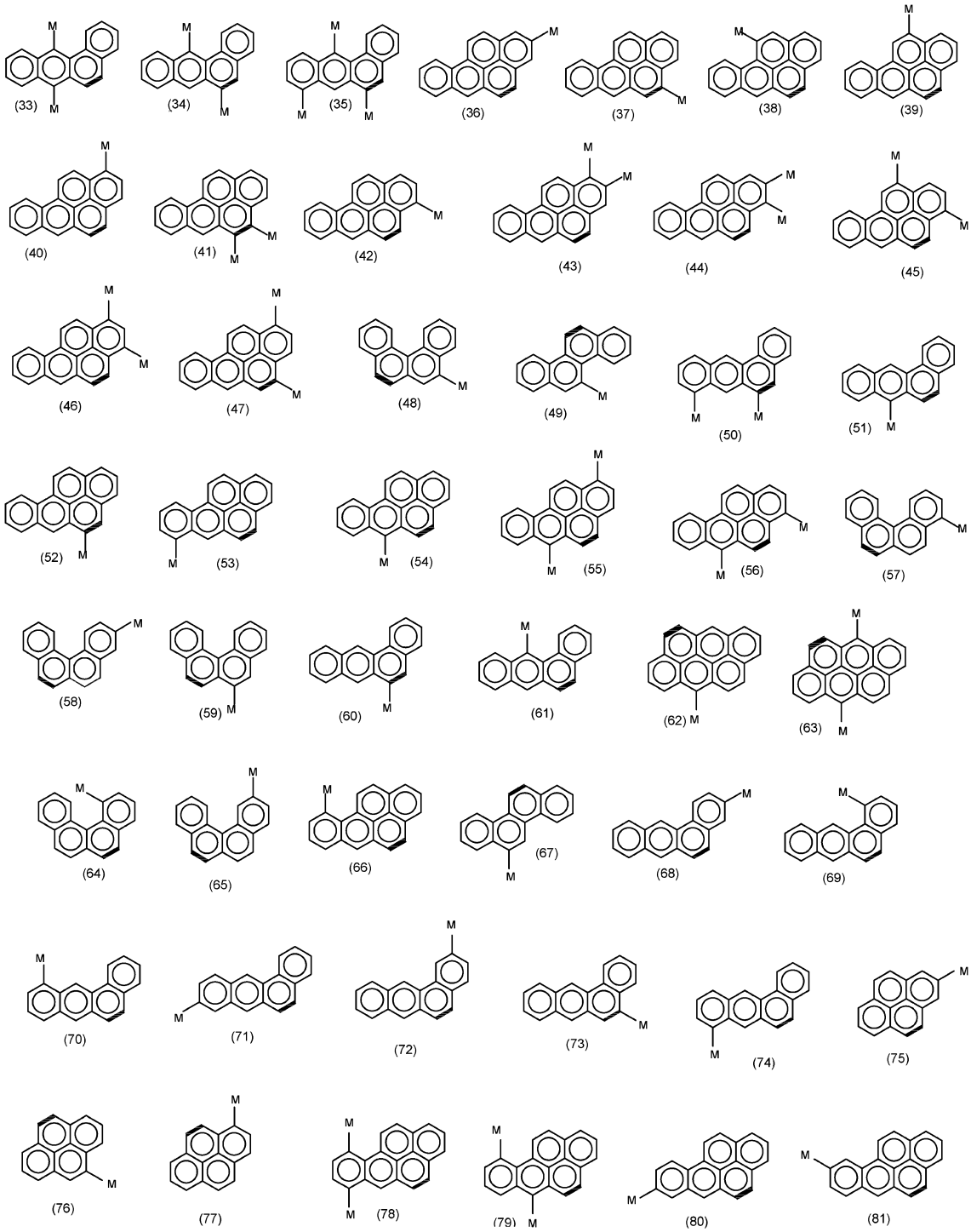

Figure 2. Molecular structure of 49 methylated polycyclic aromatic hydrocarbon (PAHs) molecules. The darker bonds indicate the calculated highest bond order. Table 2 lists their IUPAC names.

- Hardness,[25] approximated as HD = $(\epsilon_{LUMO} - \epsilon_{HOMO})/2$.

- Mulliken eletronegativity,[26] approximated as $\chi = -(\epsilon_{HOMO}+\epsilon_{LUMO})/2$.

- The stabilization energy to form PAH radical cations and anions, calculated from the difference in heats of formation values between neutral and ionic structures.

- Coefficient of molecular partition octanol−water (log P) and the empirical descriptor[27] that gives the molecular size suitable to carcinogenic activity: Nat = $(N - 20)^3$, where N is the number of C-atoms in the molecules. Log P was calculated using parameters of the substituents hydrophobicity.[28]

- The HOMO-1 energy, which is molecular orbital one level below the HOMO.

- The energy difference between HOMO and HOMO-1 (ΔH).

- The molecular orbital energy of LUMO+1, which is the molecular orbital one level above the LUMO.

- The energy difference between LUMO and LUMO+1 (ΔL).

- The HOMO and HOMO-1 contributions (CH and CH-1, respectively) to the local density of states (LDOS)[13,14] calculated over the PAH ring presenting the highest bond

Table 1. 32 Nonmethylated Polycyclic Aromatic Hydrocarbons (PAHs)[a]

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

[a] See Figure 1 for their molecular structures. The carcinogenic activity data are adapted from Iball index experiments[10,41,42] and from Cavaliere et al. scale.[21] **A** and **I** refer to active and inactive, respectively.

Table 2. 49 Methylated Polycyclic Aromatic Hydrocarbons (PAHs)[a]

molecule	CA[21]
(33) 7,12-dimethylbenz[*a*]anthracene	A
(34) 6,12-dimethylbenz[*a*]anthracene	A
(35) 6,8,12-trimethylbenz[*a*]anthracene	A
(36) 2-methylbenzo[*a*]pyrene	A
(37) 4-methylbenzo[*a*]pyrene	A
(38) 11-methylbenzo[*a*]pyrene	A
(39) 2-methylbenzo[*a*]pyrene	A
(40) 1-methylbenzo[*a*]pyrene	A
(41) 4,5-dimethylbenzo[*a*]pyrene	A
(42) 3-methylbenzo[*a*]pyrene	A
(43) 1,2-dimethylbenzo[*a*]pyrene	A
(44) 2,3-dimethylbenzo[*a*]pyrene	A
(45) 3,12-dimethylbenzo[*a*]pyrene	A
(46) 1,3-dimethylbenzo[*a*]pyrene	A
(47) 1,4-dimethylbenzo[*a*]pyrene	A
(48) 5- methylbenzo[*c*]phenanthrene	A
(49) 5- methylchrysene	A
(50) 6,8-dimethylbenz[*a*]anthracene	A
(51) 7-methylbenz[*a*]anthracene	A
(52) 5-methylbenzo[*a*]pyrene	A
(53) 7-methylbenzo[*a*]pyrene	A
(54) 6-methylbenzo[*a*]pyrene	A
(55) 1,6-dimethylbenzo[*a*]pyrene	A
(56) 3,6-dimethylbenzo[*a*]pyrene	A
(57) 4-methylbenzo[*c*]phenanthrene	I
(58) 3-methylbenzo[*c*]phenanthrene	I
(59) 6- methylbenzo[*c*]phenanthrene	I
(60) 6-methylbenz[*a*]anthracene	I
(61) 12-methylbenz[*a*]anthracene	I
(62) 6-methylanthanthrene	I
(63) 6,12-dimethylanthanthrene	I
(64) 1-methylbenzo[*c*]phenanthrene	I
(65) 2-methylbenzo[*c*]phenanthrene	I
(66) 10-methylbenzo[*a*]pyrene	I
(67) 6-methylchrysene	I
(68) 3-methylbenz[*a*]anthracene	I
(69) 1-methylbenz[*a*]anthracene	I
(70) 11-methylbenz[*a*]anthracene	I
(71) 9-methylbenz[*a*]anthracene	I
(72) 2-methylbenz[*a*]anthracene	I
(73) 5-methylbenz[*a*]anthracene	I
(74) 8-methylbenz[*a*]anthracene	I
(75) 2-methylpyrene	I
(76) 4-methylpyrene	I
(77) 1-methylpyrene	I
(78) 7,10-dimethylbenzo[*a*]pyrene	I
(79) 6,10-dimethylbenzo[*a*]pyrene	NA
(80) 8-methylbenzo[*a*]pyrene	NA
(81) 9-methylbenzo[*a*]pyrene	NA

[a] See Figure 2 for their molecular structures. The carcinogenic activity data are adapted from Cavaliere et al. scale[21]. **A** and **I** refer to active and inactive, respectively. The carcinogenic activity of the last three molecules is not available (NA).

order (RHBO) (most susceptible to electrophilic attack) and their difference $\eta H = (CH) - (CH\text{-}1)$

- The LUMO and LUMO+1 contributions (CL and CL+1, respectively) to the local density of states (LDOS) calculated over RHBO, and their difference $\eta L = (CL) - (CL+1)$.

The mentioned parameters of the first six items are usual parameters in SAR analysis, while the ones listed in the last six items are suggested to be relevant from the new EIM methodology. We use statistical methods to select the best set of descriptors correlating to the experimental carcinogenic activity.

For SAR studies we divided the set of 81 PAH molecules (sequentially numbered in Figures 1 and 2) into two groups: **A** and **B**. The **A** group (or category 1) consists of 34 active molecules (1−10, Figure 1; 33−56, Figure 2). The **B** group (or category 2) consists of 44 inactive molecules (11−32, Figure 1; 57−78, Figure 2). There are no experimental data available for the remaining three molecules (79−81, Figure 2), and they are used for prediction purposes.

PCA and HCA are methods widely used in pattern-recognition studies. PCA is an extremely useful explorative tool which maps samples through scores and individual variables by the loadings (see Figures 3 and 4) in a new vector space defined by the principal components (PC). Score plots allow sample identification, clarifying whether they are similar or dissimilar, typical or outliers. From loading plots the more important variables can be easily identified as well as the correlation patterns among them. The first PC is generated in a way that it has maximum correlation with all of the variables and accounts for a large portion of the total data variance. From the remaining data variance (after the removal of the first PC) a second PC is extracted which is completely uncorrelated (orthogonal) with the first one and accounts for the maximum possible remaining data set variance. The procedure is then repeated until all PCs are generated. This corresponds to a N-dimensional space, isomorph to the N-dimensional variable space (N is the total number of used variables). HCA, also an exploratory tool, is used to validate the grouping previously identified by PCA. The primary goal of HCA is to emphasize natural grouping of similar samples based on their proximity in the multidimensional space spanned by the used variables. The results,

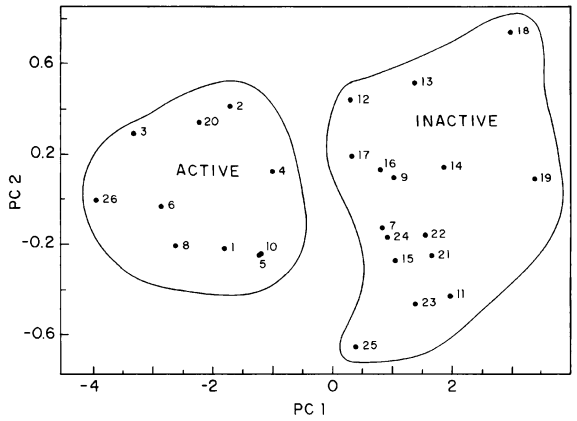

Figure 3. The score graph of the first two principal components (PC1 and PC2) for the set of 26 nonmethylated PAH molecules shown in Figure 1.

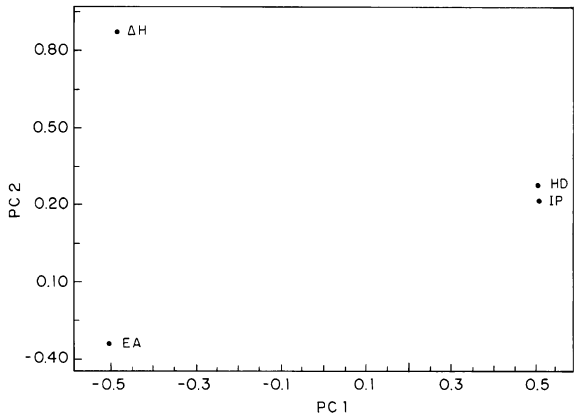

Figure 4. The loading graph for the 4 physicochemical descriptors selected by PCA for the set 26 nonmethylated PAH molecules (Figure 1): IP, EA, HD, and ΔH refer to ionization potential, the electron affinity, hardness, and the difference in energy between HOMO and HOMO-1, respectively.

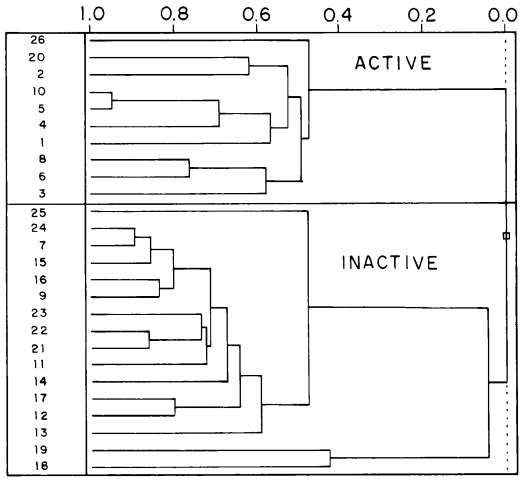

Figure 5. Hierarchical dendogram for the set of 26 nonmethylated PAH molecules. They were classified into two groups: active and inactive.

qualitative in nature, are presented in the form of a dendogram (see Figure 5), allowing the visualization of clusters and correlation among samples. In HCA, the distances (Euclidean metric) between the samples are calculated and transformed into a similarity matrix whose elements are the similarity indexes ranging from zero to one; a smaller distance means a larger index.[29] In principle we can expect that points representing carcinogenic compounds will cluster in one limited region of the space in the score graphic of the principal components, while the points representing noncarcinogenic compounds will cluster elsewhere.

The PCA study was carried out using the program package Pirouette[30] which contains the PCA and related methods.

In connection with PCA and HCA analysis we have also carried out Neural Networks studies (NN). We use the relevant descriptors selected by PCA to define the PAH carcinogenic activity to train the NN. After the learning process the NN are used to identify (by similarity) active and inactive PAH compounds. This constitutes an independent test for the statistical relevance of the descriptors obtained from PM3, EIM and PCA/HCA analysis.

NN are a well-known and widely used methodology in the field of pattern recognition.[31,32] NN is a computer-based system derived from the simplified concept of the brain in which a number of nodes, called processing elements or neurons, are interconnected in a netlike structure. The NN characteristics have been found to be nonlinear making them suitable for data processing in which the relationship between cause and results cannot be linearly defined. NN analysis offers great advantage over the conventional SAR methods for the cases where the targeted properties are not linearly dependent on the chosen parameters.[33−36] Three components constitute a NN: the processing elements, the topology of the connections between the nodes, and the learning rules.

Our NN analysis was carried out using the program package Perceptron-type Neural Network Simulator (PSDD) (Quantum Chemistry Exchange Program #615).[37] PSDD stands for Perceptron Simulator for Drug Design, which consists of a three-layer perceptron network. PSDD was originally developed for drug design, but it can be applied to many other problems as a general nonlinear classification method.

3. RESULTS AND DISCUSSION

The first molecular set studied with PCA was the 26 nonmethylated PAHs listed in Figure 1 (numbered from 1 to 26). The descriptors selected by PCA method that generate the best separation in active and inactive compounds are (see Table 3) as follows:

- Ionization Potential (IP);
- Electron Affinity (EA);
- Hardness (HD) and;
- the energy difference between HOMO and HOMO-1 (ΔH).

Figure 3 shows the scores of the first two principal components (PC1 and PC2) for the set of 26 nonmethylated PAHs. The molecules are distributed into two distinct regions in the figure. The active group is on the left side and the inactive one on the right side. The molecules 7 and 9 are incorrectly classified as inactive and the molecules 20 and 26 are incorrectly classified as active, respectively. 22 molecules out of the 26 are correctly classified (accuracy of 84.6%).

Table 3. Four Descriptors (IP, EA, HD and Δ*H*) Obtained from PM3 Calculations Used in PCA and NN Calculations for Nonmethylated (Figure 1) and Methylated (Figure 2) PAHs[a]

molecule	IP	EA	HD	Δ*H*	molecule	IP	EA	HD	Δ*H*	molecule	IP	EA	HD	Δ*H*
(1)	7.987	1.288	3.350	0.699	**(28)**	9.751	0.396	4.678	0.000	**(55)**	7.876	1.204	3.336	0.983
(2)	8.042	1.221	3.411	0.860	**(29)**	8.835	0.407	4.214	0.600	**(56)**	7.878	1.189	3.345	0.929
(3)	7.808	1.461	3.174	1.072	**(30)**	8.249	1.010	3.620	0.793	**(57)**	8.480	0.733	3.874	0.266
(4)	8.140	1.163	3.489	0.676	**(31)**	8.139	1.167	3.486	0.569	**(58)**	8.478	0.724	3.877	0.267
(5)	8.079	1.212	3.434	0.600	**(32)**	8.290	1.063	4.614	0.000	**(59)**	8.502	0.722	3.890	0.175
(6)	7.848	1.421	3.214	0.913	**(33)**	8.110	0.918	3.596	0.714	**(60)**	8.261	0.924	3.669	0.595
(7)	8.377	0.918	3.730	0.326	**(34)**	8.167	0.913	3.627	0.647	**(61)**	8.220	0.922	3.639	0.636
(8)	7.888	1.410	3.239	0.829	**(35)**	8.110	0.900	3.605	0.685	**(62)**	7.683	1.495	3.094	1.223
(9)	8.423	0.875	3.774	0.359	**(36)**	8.018	1.199	3.410	0.787	**(63)**	7.606	1.483	3.062	1.275
(10)	8.087	1.228	3.430	0.607	**(37)**	7.983	1.205	3.389	0.882	**(64)**	8.447	0.747	3.850	0.303
(11)	8.518	0.787	3.866	0.070	**(38)**	7.985	1.208	3.389	0.873	**(65)**	8.507	0.720	3.894	0.189
(12)	8.328	0.934	3.697	0.563	**(39)**	7.972	1.205	3.384	0.907	**(66)**	7.985	1.206	3.390	0.881
(13)	8.496	0.783	3.857	0.420	**(40)**	7.971	1.209	3.381	0.913	**(67)**	8.403	0.775	3.814	0.494
(14)	8.538	0.748	3.895	0.243	**(41)**	7.936	1.182	3.377	0.909	**(68)**	8.291	0.915	3.688	0.522
(15)	8.400	0.902	3.749	0.254	**(42)**	7.973	1.194	3.390	0.862	**(69)**	8.301	0.901	3.700	0.491
(16)	8.400	0.907	3.747	0.402	**(43)**	7.951	1.182	3.385	0.836	**(70)**	8.277	0.923	3.677	0.593
(17)	8.335	0.967	3.684	0.490	**(44)**	7.948	1.178	3.385	0.789	**(71)**	8.287	0.904	3.692	0.525
(18)	8.740	0.535	4.103	0.236	**(45)**	7.908	1.179	3.365	0.904	**(72)**	8.278	0.915	3.682	0.579
(19)	8.773	0.556	4.109	0.000	**(46)**	7.906	1.183	3.362	0.913	**(73)**	8.258	0.917	3.671	0.558
(20)	7.962	1.297	3.333	0.918	**(47)**	7.915	1.194	3.361	0.932	**(74)**	8.256	0.922	3.667	0.617
(21)	8.465	0.795	3.835	0.164	**(48)**	8.456	0.737	3.860	0.284	**(75)**	8.230	0.982	3.624	0.666
(22)	8.477	0.824	3.827	0.210	**(49)**	8.427	0.775	3.826	0.412	**(76)**	8.178	0.994	3.592	0.830
(23)	8.448	0.889	3.780	0.154	**(50)**	8.198	0.912	3.643	0.638	**(77)**	8.151	0.990	3.581	0.847
(24)	8.409	0.930	3.740	0.304	**(51)**	8.218	0.931	3.644	0.647	**(78)**	7.928	1.193	3.368	0.898
(25)	8.274	1.026	3.624	0.248	**(52)**	7.986	1.207	3.390	0.897	**(79)**	7.891	1.203	3.344	0.933
(26)	7.693	1.573	3.060	1.088	**(53)**	7.993	1.206	3.394	0.875	**(80)**	8.020	1.195	3.413	0.772
(27)	7.762	1.508	3.127	1.170	**(54)**	7.942	1.215	3.364	0.933	**(81)**	7.987	1.202	3.393	0.883

[a] See text for discussions. All values in electron-Volts (eV).

Table 4. Correlation Matrix of the 4 Physicochemical Descriptors (Table 3) for the First 26 Nonmethylated PAH Molecules (Figure 1)

	IP	EA	HD	Δ*H*
IP	1.0000	−0.9972	0.9993	−0.9340
EA		1.0000	−0.9993	0.9182
HD			1.0000	−0.9268
Δ*H*				1.0000

The two principal components (PC1 and PC2) are given by eqs 1 and 2:

$$\mathbf{PC1} = 0.51\,\mathbf{IP} - 0.50\,\mathbf{EA} + 0.50\,\mathbf{HD} - 0.49\,\Delta\mathbf{H} \quad (1)$$

$$\mathbf{PC2} = 0.21\,\mathbf{IP} - 0.35\,\mathbf{EA} + 0.28\,\mathbf{HD} + 0.87\,\Delta\mathbf{H} \quad (2)$$

The PC1 and PC2 respond to 97.2% and 2.7% of the variance, respectively. The remaining PCs respond to less than 0.1%. All the descriptors are equally important in PC1. The PC2 major descriptor is the difference in energy between HOMO and HOMO-1 (Δ*H*). The Δ*H* appears to be the most important variable, since that it is equally important to the other PC1 descriptors and dominates PC2. It is interesting to notice that Δ*H* is exactly one of the two major EIM descriptors.

The correlation matrix of four descriptors are given in Table 4. They are highly correlated. Figure 4 shows the loading graph for these four physicochemical descriptors. The descriptors are grouped in two regions: one at the left side of the figure, with the values of the PC1 axis approximately at −0.5 (descriptors Δ*H* and EA) and the other at right side with the values of the PC1 axis approximately at 0.5 (descriptors HD and IP). From Figures 3 and 4 we observe that two descriptors, Δ*H* and EA, are responsible for pulling the active molecules toward the left side in the score graph (Figure 3). The descriptors that are mostly responsible for pulling inactive molecules toward the right side in Figure 4 are HD and IP.

Figure 5 shows the hierarchical clustering diagram for the set of the first 26 PAHs (Figure 1). We can see that with the use of 4 descriptors used in PCA calculations two clusters are formed separated by the horizontal line between the #3 and #25 molecules. One group is mostly composed by active molecules (upper half of the figure) and the other by inactive molecules (lower half). These two clusters have zero similarity which demonstrates that active and inactive molecules are well separated in the four-dimensional PCA space. The molecules incorrectly classified using HCA are the same observed in the PCA analysis.

To study the predictive ability of the PCA method we applied it to the molecules 27−32 (Figure 1) and 33−78 (Figure 2), using exactly the same 4 descriptors used for the set of 26 PAHs (Table 3). The PCA score graph is illustrated in Figure 6. The active group is located on the left side of the figure and the inactive one on the right side. The molecules 27, 31, 62, 63, 66 and 78 are incorrectly classified as actives; also the molecules 33, 34, 35, 48, 49, 50 and 51 are incorrectly classified as inactives.

Considering the sets of 32 nonmethylated and 46 methylated PAHs, PCA correctly classifies 26 (81.3%) and 35 (76.1%) molecules, respectively. For the global set the accuracy is 61 out of 78 (78.2%).

When methylated PAHs were separately analyzed in our previous work,[17] using quantum chemical descriptors derived from the Hückel method, the same four descriptors (IP, EA, HD and Δ*H*) were selected as efficient parameters for the PAHs classification. In this case, the percentage of correct classification is 82.6%.[17] The PCA selected descriptors derived from PM3 method (with and without geometrical

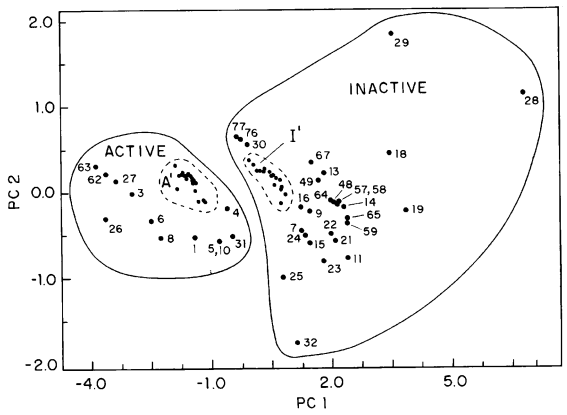

Figure 6. The score graph of the first two principal components (PC1 and PC2) for the global set of 81 PAH molecules. The cluster surrounded on the left side (A′) in the figure is composed by molecules #2, #20, #36, #37, #38, #39, #40, #41, #42, #43, #44, #45, #46, #47, #52, #53, #54, #55, #56, #66, #78, #79, #80 and #81 and the cluster surrounded on the right side (I′) is composed by molecules #12, #17, #33, #34, #35, #50, #51, #60, #61, #68, #69, #70, #71, #72, #73, #74 and #75.

constraints) are consistent with those derived from the Hückel calculations for molecules without electron−hole symmetry.

The correct PCA global predictions using descriptors derived from PM3 method with and without imposed $\sigma-\pi$ separation are exactly the same (78.2%). This is in very good agreement with the results obtained from Hückel calculations[17] (80.8%).

In general the results derived from Hückel and PM3 methods (with and without geometrical constraints) are very similar. On average PCA predicts carcinogenic activity with an accuracy of approximately 80% for the three different classes of compounds (nonmethylated, methylated and the overall results) using molecular descriptors derived from Hückel or PM3 methods. The selection of the same parameters producing the same degree of accuracy using quite different methods such as Hückel and PM3 supports the statistical relevance of these parameters in the classification of active and inactive PAHs.

To further investigate these aspects we have also carried out NN (perceptron-type three layers NN[32,33]) analysis using the same PCA descriptors.

The descriptors used as input data (unscaled) are listed in Table 3. We started studying the 26 nonmethylated PAH molecules used in PCA. Each neuron was set to have a value ranging from 0 to 1 (usual input data rescaling[32]). The NN training was carried out accordingly to the back-propagation algorithm[37] until the error function[28] attains convergence criterion (<0.03). The usual NN parameters (α, θ and ϵ)[28] are shown in Table 5. α is a parameter which expresses the nonlinearity of the neuron's operation. 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 (accordingly to the values in Table 5) until the error function attains convergence. θ is a threshold value for neuron in the second and third layers. It was set to its usual value (0.0). ϵ is a parameter which determines the shift for correction in recursive cycles.

The NN were trained with the first 26 nonmethylated PAHs to predict carcinogenic activity of methylated and nonmethylated PAHs. Table 6A shows the NN training

Table 5. Parameters Used (α, θ and ϵ) in the Neural Networks Calculations[a]

layer	neurons	α	θ	ϵ
Parameters Used in Table 6A				
1	4			
2	10	2.0	0.0	0.1
3	2	15.5	0.0	0.05

[a] α is the nonlinear parameter of the sigmoid functions, θ is a threshold value for a neuron and ϵ is a parameter which determines the shift for correction in recursive cycles.

results and Table 6B shows the prediction results for the methylated and nonmethylated 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 6A. In the initial phase of "NN learning", the weight matrix was calculated with the training pattern using the 4 parameters for each of the 26 molecules. The NN "learned" the training pattern with 100% of success. The four molecules, 7, 9, 20 and 26, incorrectly classified with PCA are now correctly described. The number of iterations was 53,101. Although the neurons at the first layer (input pattern) can take continuous values between **0** and **1**, those of the last layer (output pattern) are required to assume discrete values **0** or **1**. However, in general, the final output patterns do not show a complete set of discrete values, which shows the limit of the resolution ability of the network, but when, the output patterns are close to (**1 0**) or (**0 1**) this allows the network decisions, active or inactive.

Considering the sets of 32 nonmethylated and 46 methylated PAHs, the NN correctly classified 30 (93.8%) and 37 (80.4%) molecules, respectively. For the global set the accuracy is 67 out of 78 (85.9%).

Table 7 compares percentage of correct classification (%) of the two different methods (PCA and NN), for the three different classes of compounds (nonmethylated (NM), methylated (M) and overall results (NM+M)). On average the two methods have approximately the same margin of correct prediction (80%). However, the NN seems to give slightly better percentage than PCA. Both methods (PCA and NN) predict the molecules 79, 80 and 81 (experimental activity not available) as actives. The four descriptors selected by PCA analysis were validated in the NN framework. It is very intriguing that the complex phenomena of discriminating active and inactive PAHs could be attained with high precision using only few parameters and without any specific assumption about the involved biochemical processes. Based on that we decide to investigate the possible relationships among the descriptors selected by PCA and the proposed mechanisms of PAHs metabolic activation.

Recent studies[38−40] suggest that two major mechanisms are involved in the PAHs' metabolic activation to initiate cancer: (1) monooxigenation with formation of bay-region diol epoxides and (2) one-electron oxidation to produce radical cations. There are evidences that some PAHs are activated exclusively by one of these mechanisms and others are activated by a combination of both.

In Figure 7 we show the proposed mechanism of metabolic activation of benzo[*a*]pyrene by P450 mono-oxygenase enzymes, one of the most widely investigated study cases.[41,42] It shows that after initial oxidation at 7,8-position by the

Table 6. Results from the Neural Networks (NN) Calculations[a] and Prediction of Carcinogenic Activity for 42 PAHs[b]

A. Results from the Neural Networks (NN) Calculations

molecule	category	training pattern		output pattern		molecule	category	training pattern		output pattern	
(1)	1	1	0	0.988	0.012	**(14)**	2	0	1	0.000	1.000
(2)	1	1	0	0.987	0.023	**(15)**	2	0	1	0.010	0.989
(3)	1	1	0	0.954	0.046	**(16)**	2	0	1	0.032	0.968
(4)	1	1	0	1.000	0.000	**(17)**	2	0	1	0.029	0.972
(5)	1	1	0	0.999	0.001	**(18)**	2	0	1	0.000	1.000
(6)	1	1	0	0.987	0.013	**(19)**	2	0	1	0.000	1.000
(7)	1	1		0.974	0.026	**(20)**	2	0	1	0.383	0.617
(8)	1	1	0	0.987	0.013	**(21)**	2	0	1	0.000	1.000
(9)	1	1	0	0.960	0.039	**(22)**	2	0	1	0.000	1.000
(10)	1	1	0	0.999	0.001	**(23)**	2	0	1	0.000	1.000
(11)	2	0	1	0.000	1.000	**(24)**	2	0	1	0.075	0.925
(12)	2	0	1	0.029	0.971	**(25)**	2	0	1	0.014	0.986
(13)	2	0	1	0.007	0.993	**(26)**	2	0	1	0.008	0.992

B. Prediction of Carcinogenic Activity for 42 PAHs

molecule	category	predicted pattern		molecule	category	predicted pattern		molecule	category	predicted pattern	
(27)	2	0.243	0.756	**(46)**	1	0.778	0.225	**(64)**	2	0.000	1.000
(28)	2	0.000	1.000	**(47)**	1	0.781	0.223	**(65)**	2	0.000	1.000
(29)	2	0.007	0.993	**(48)**	2	0.000	1.000	**(66)**	1	0.933	0.069
(30)	1	0.947	0.054	**(49)**	2	0.004	0.996	**(67)**	2	0.008	0.992
(31)	1	1.000	0.000	**(50)**	2	0.047	0.952	**(68)**	2	0.029	0.972
(32)	2	0.000	1.000	**(51)**	2	0.050	0.949	**(69)**	2	0.028	0.972
(33)	1	0.918	0.083	**(52)**	1	0.945	0.057	**(70)**	2	0.030	0.970
(34)	2	0.086	0.911	**(53)**	1	0.948	0.054	**(71)**	2	0.029	0.972
(35)	1	0.586	0.408	**(54)**	1	0.762	0.241	**(72)**	2	0.029	0.971
(36)	1	0.960	0.042	**(55)**	1	0.561	0.441	**(73)**	2	0.030	0.970
(37)	1	0.934	0.068	**(56)**	1	0.585	0.417	**(74)**	2	0.032	0.968
(38)	1	0.922	0.081	**(57)**	2	0.000	1.000	**(75)**	2	0.165	0.830
(39)	1	0.925	0.077	**(58)**	2	0.000	1.000	**(76)**	1	0.995	0.006
(40)	1	0.911	0.092	**(59)**	2	0.000	1.000	**(77)**	1	0.996	0.004
(41)	1	0.918	0.084	**(60)**	2	0.031	0.969	**(78)**	1	0.796	0.207
(42)	1	0.930	0.072	**(61)**	2	0.042	0.958	**(79)**	1	0.557	0.444
(43)	1	0.883	0.120	**(62)**	2	0.000	1.000	**(80)**	1	0.965	0.037
(44)	1	0.829	0.174	**(63)**	2	0.000	1.000	**(81)**	1	0.953	0.049
(45)	1	0.801	0.202								

[a] The NN were trained with 26 PAHs, molecules 1−26 (Figure 1) using the parameters listed in Table 5. [b] Molecules 27−32 (Figure 1) and 33−81 (Figure 2), using the trained NN in part A.

Table 7. Percentage of Correct Classification and Prediction for the Two Different Methods (PCA and NN) for Methylated (M) and Nonmethylated (NM) Molecules Considered Separately and the Overall Results (M+NM)

percentage of correct classification (%)	nonmethylated (32 PAHs)	methylated (46 PAHs)	nonmethylated + methylated (78 PAHs)
PCA	81.3	76.1	78.2
NN	93.8	80.4	85.9

enzyme the molecule is transformed into a diol-epoxide. This epoxide is then transformed in the triol carbonium ion that it is believed to react with a nucleic base to form covalent adducts that lead to mutations. This ultimately could give rise to an event leading to cancer.[43−45]

PAHs radical cations obtained by one-electron oxidation are produced by removal of a π electron by the biological oxidants P450, peroxidases such as horseradish peroxidase (HRP) and prostaglandin H synthase (PHS).[38] The PAH ability to form a radical cation is related to its Ionization Potential (IP),[21] lower IP implies easier one-electron oxidation. Studies carried out by Cavaliere et al.[21] indicate that PAHs must have IP less than 7.3 eV (experimental determination) to be activated by one-electron oxidation. Results obtained by Jerina et al.[46] also indicate that PAH carcinogenicity is directly related to IP values.

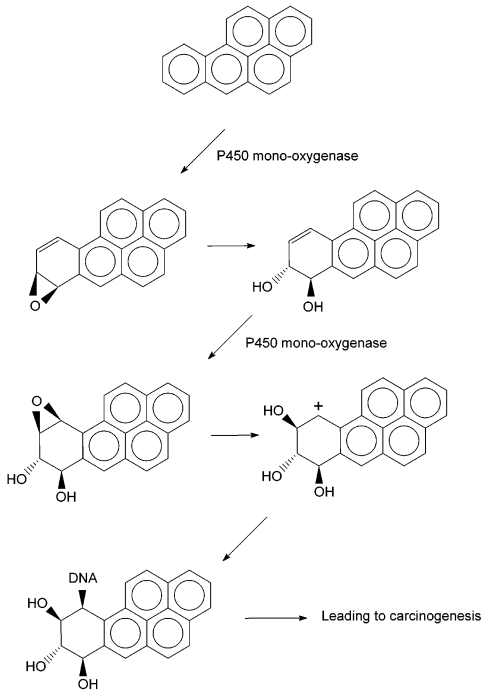

Figure 7. Proposed mechanism of metabolic activation leading to carcinogenesis by benzo[*a*]pyrene.[41,42]

PAHs activation by one-electron oxidation[38] requires a relatively low IP. In fact it is believed that some active molecules such as dibenzo[3,4;9,10]pyrene, benzo[3,4]pyrene (molecules 1 and 2, respectively, in Table 1) and 7,12-dimethylbenz[*a*]antracene (molecule 33, Table 2) are activated mainly by one-electron oxidation.[38] However this is not a sufficient condition since other features such as geometric configurations and charge localization effects can block this mechanism.

For both mechanisms (mono-oxygenation and one-electron oxidation), the ultimate carcinogen is an electrophilic agent that binds to nucleophilic sites in DNA through covalent binding. It is known that the electrophilic reactants have a relatively low LUMO energy[47] or high electron affinity. For this reason high electron affinities should be related to high carcinogenic activity, or in other words electrophilic PAHs should be related to high carcinogenic activity.

As mentioned above for both mechanisms the ultimate PAH metabolite binds to sites in DNA typically through covalent bonding, forming the so-called PAH-DNA adduct. For this reason we can expect that a small separation between HOMO and LUMO energy of PAH (or 'soft' molecules) should favor binding with DNA since "soft likes soft" in a chemical reaction.[25] Due to the proximity between the frontier orbitals of carcinogenic and receptor, the covalent interaction between them is the most probable.[25] Low hardness are then related to high activity.

In fact, in the present study low hardness values are related to high activity, or 'soft' molecules show greater carcinogenic activity than 'hard' ones.

From the discussions above we can summarize the relevance of these descriptors for the PAHs metabolic activation:

- IP for the initial oxidation processes (lower IP higher carcinogenic activity);

- EA for the ultimate carcinogen (higher EA higher carcinogenic activity);

- HD (which combines IP and EA values) for the final covalent binding between the ultimate carcinogen and DNA bases (lower HD (easier binding) higher carcinogenic activity).

In the present study the active molecules or those that are classified as active have theoretical IP values between 7.61 and 8.14 eV, while the inactive ones have IP values between 8.15 and 9.75 eV. There are two exceptions to this rule: the active molecules 33 and 34, both possess IP equal to 8.11 eV and they are classified as inactive.

With respect to EA the active molecules or those that are classified as active have EA values between 1.16 and 1.57 eV, while values for the inactive molecules are between 0.40 and 1.06 eV.

The hardness values (HD) also present distinct patterns for active and inactive molecules. The active molecules or those that are classified as active present HD values between 3.06 and 3.49 eV, while the inactive ones values between 3.58 and 4.68 eV.

These theoretical results, from the combined methodologies (PM3, EIM and pattern-recognition methods) not only identified these 3 descriptors (IP, EA, and HD) as among the most relevant to define the biological activity, as well as the obtained values for these magnitudes (see Table 3, and discussions above) follow exactly the experimental tendencies (higher/lower↔carcinogenic activity).

The three different employed methods for SAR analysis, PCA, HCA and NN, correctly classified PAH carcinogenic activity with high accuracy, using the four selected descriptors by PCA: IP, EA, HD and ΔH. These results suggest that the carcinogenic activity is highly correlated with these four descriptors as mentioned above. In fact, as discussed above, it was possible to correlate directly three of these descriptors with the mechanisms of PAH metabolic activation to initiate cancer proposed in the literature. This validates our methodology as an efficient and useful tool to classificatory SAR studies.

However, the major descriptor defining the carcinogenic activity, ΔH, is not clearly correlated with the mechanisms proposed in the literature.[38]

From our results, higher energy separation between the HOMO and HOMO-1 levels (ΔH) can be correlated to higher carcinogenic activity. The major part of active molecules possess $\Delta H > 0.600$ eV, while for the inactive ones $\Delta H < 0.600$ eV. There are some exceptions. It seems that a "clean" frontier orbital, i.e., large separation between HOMO and HOMO-1 energy, is a necessary but not sufficient condition for carcinogenic activity.

It remains to be elucidated why this descriptor is so important. It cannot be directly associated with any specific chemical or physical property. Also, it seems to have an 'universal' character since it appears as a major EIM descriptor in many distinct organic classes. These classes include compounds with very differentiated biological activity, such as the following: carcinogens,[13,14] antibiotics[49] (mitomycins), HIV protease inhibitors,[50] and progestational hormones.[51] Preliminaries studies with another organic classes such as talidomides[52] and some steroids with binding affinity to CBG (corticosteroid binding globulin)[53] also show similar results.

Preliminary calculations[48] for PAHs indicate that this descriptor can perhaps be understood as an indirect measure of the chemical reactivity and lifetime of excited states in the process of charge-transfer and/or covalent bond formations to PAH-DNA adducts. However these arguments do not apply to other classes or organic classes[48−53] where ΔH is also a major descriptor allowing to distinguish active and inactive compounds. The only aspect that all these classes have in common is that ΔH is obtained from the Schrödinger equation under the same approximations. Perhaps this ΔH 'universality' class is an evidence of the existence of common features (fractal/chaotic/scaling).[53−55] Further studies are needed to clarify these aspects.

4. SUMMARY AND CONCLUSIONS

Polycyclic Aromatic Hydrocarbons (PAHs) constitute an important class of molecules capable of inducing chemical carcinogenesis. It has been a challenge to explain why these very similar molecules present such variation in terms of carcinogenic activity (from some of strongest known carcinogens to inactive ones).

Recently[13,14] a new methodology (EIM-Electronic Indices Methodology) exploring the concepts of local density of states and critical values for energy separation involving

frontier orbitals was proposed to group and identify the PAH's carcinogenic activity.

In this work we expand these ideas in the framework of more conventional structure−activity studies using PCA (Principal Component Analysis) and NN (Neural Networks) methods with descriptors derived from fully geometrical PM3 (Parametric Method 3) calculations.

We have investigated 81 methylated and nonmethylated PAHs (Figures 1 and 2) in their neutral and ionic (anions and cations) states. It was possible to classify active and inactive compounds with high accuracy (80%).

Three out of four selected PCA descriptors (Ionization Potential, Electron Affinity, and Hardness) can be directly related to some proposed mechanisms[38−40] of PAHs metabolic activation leading to carcinogenic events.

The remaining descriptor ΔH (energy separation between the highest occupied molecular orbital and its lower level) cannot be related to these mechanisms. ΔH is one of the two major EIM descriptor and has been statistically validated by our present PCA and NN studies. ΔH seems to have an 'universal' character since it appears as a major descriptor to distinguish active and inactive compounds in distinct organic classes such as carcinogenics,[13,14] antibiotics,[49] HIV protease inhibitors,[50] etc.

Since it is not possible to associate ΔH with a specific chemical or physical property its importance remains to be elucidated.

ACKNOWLEDGMENT

This work was supported in part by the Brazilian Agencies FAPESP and CNPq. Computational support from CENAPAD − SP is acknowledged. The authors are in debt to Prof. P. M. V. B. Barone, Prof. M. A. Cotta, and M. Cyrillo for helpful discussions.

REFERENCES AND NOTES

(1) Sugimura, T. Multistep Carcinogenesis − A 1992 Perspective. *Science* **1992**, *258*, 603−607.

(2) Harvey, R. G.; Geacintov, N. E. Intercalation and Binding of Carcinogenic Hydrocarbon Metabolites to Nucleic-Acids. *Acc. Chem. Res.* **1988**, *21*, 66−73.

(3) Coulson, C. A. Electronic configuration and carcinogenesis. *Adv. Cancer Res.* **1953**, *1*, 1−56 and references therein.

(4) Pullman, A.; Pullman, B. Electronic structure and carcinogenic activity of aromatic molecules − new developments. *Adv. Cancer Res.* **1955**, *3*, 117−169 and references therein.

(5) Streitwieser, A. *Molecular Orbital Theory;* Wiley: New York, 1961.

(6) Lowe, J. P.; Silverman, B. D. Predicting Carcinogenicity of Polycyclic Aromatic-Hydrocarbons. *Acc. Chem. Res.* **1984**, *17*, 332−338.

(7) Jerina, D. M. et al. In *Carcinogenesis: Fundamental Mechanisms and Environmental Effects*; Pullman, B., Ts'O, P. O., Gelboin, H. et al., Ed.; D. Reidel Publishing Co.: Dordrecht, Holland, 1980.

(8) Szentpály, L. V. Carcinogenicity by polycyclic aromatic hydrocarbons: a multilinear regression on new type PMO indices. *J. Am. Chem. Soc.* **1984**, *106*, 6021−6028.

(9) Gayoso, J.; Kimri, S. Unification of quantum theories of polyacene carcinogenesis 0.1. M, L, and B regions theory. *Int. J. Quantum Chem.* **1990**, *38*, 461−486; Unification of quantum theories of polyacene carcinogenesis 0.2. Role of the K-region in metabolic activation process leading to the ultimate carcinogen − M, L, and BK regions theory. *Int. J. Quantum Chem.* **1990**, *38*, 487−495.

(10) Nordén, U. E.; Svante, W. Carcinogenicity of polycyclic aromatic hydrocarbons studied by SIMCA pattern recognition. *Acta Chem. Scand. B* **1978**, *32*, 602−608.

(11) Villemin, D.; Cherqaoui, D.; Mesbah, A. Predicting carcinogenicity of polycyclic aromatic-hydrocarbons from back-propagation neural-network. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1288−1293.

(12) Song, X.-H.; Xiao, M.; Yu, R.-Q. Artificial neural network applied to classification of mutagenic activity of nitrosubstituted polycyclic aromatic-hydrocarbons. *Comput. Chem.* **1994**, *18*, 391−396.

(13) Barone, P. M. V. B.; Camilo, A., Jr.; Galvão, D. S. Theoretical approach to identify carcinogenic activity of polycyclic aromatic hydrocarbons. *Phys. Rev. Lett.* **1996**, *77*, 1186−1189.

(14) Braga, R. S.; Barone, P. M. V. B.; Galvão, D. S. Identifying carcinogenic activity of methylated polycyclic aromatic hydrocarbons (PAHs). *J. Mol. Struct. (THEOCHEM)* **1999**, *464*, 257−266.

(15) Vendrame, R.; Braga, R. S.; Takahata, Y.; Galvão, D. S. Structure-carcinogenic activity relationship studies of polycyclic aromatic hydrocarbons (PAHs) with pattern-recognition methods. *J. Mol. Struct. (THEOCHEM)* **2001**, *539*, 253−265.

(16) Stewart, J. J. P. Optimization of parameters for semiempirical methods 0.1. method. *J. Comput. Chem.* **1989**, *10*, 209−220; Optimization of parameters for semiempirical methods. 2. applications. *J. Comput. Chem.* **1989**, *10*, 221−264.

(17) Vendrame, R.; Braga, R. S.; Takahata, Y.; Galvão, D. S. Structure−activity relationship studies of carcinogenic activity of polycyclic aromatic hydrocarbons using calculated molecular descriptors with principal component analysis and neural network methods. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1094−1104.

(18) Zhang, L. T.; Sannes, K.; Shusterman, A. J.; Hansch, C. The structure−activity of skin carcinogenicity of aromatic-hydrocarbons and heterocycles. *Chem. Biol. Interactions* **1992**, *81*, 149−180.

(19) Isu, Y.; Nagashima, U.; Aoyama, T.; Hosoya, H. Development of neural network simulator for structure−activity correlation of molecules (NECO). Prediction of endo/exo substitution of norbornane derivatives and of carcinogenic activity of PAHs from C-13-NMR shifts. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 286−293.

(20) Miyashita, Y.; Takahashi, Y.; Daiba, S.; Abe, H.; Sasaki, S. Computer-assisted structure-carcinogenicity studies of polynuclear aromatic-hydrocarbons by pattern-recognition methods. The role of the Bay and L-Regions. *Anal. Chim. Acta* **1982**, *143*, 35−44.

(21) Cavaliere, E. L.; Rogan, E. G.; Roth, R. W.; Saugier, R. K.; Hakam, A. The relationship between ionization-potential and horseradish-peroxidase hydrogen-peroxide catalyzed binding of aromatic-hydrocarbons to DNA. *Chem. Biol. Interact.* **1983**, *47*, 87−109.

(22) *Mopac Program*, version 6.0; Quantum Chemistry Exchange No. 455.

(23) Cyrillo M.; Galvão, D. S.; Chem2Pac: A computational chemistry integrator for Windows. *EPA Newsletter* **1999**, *67*, 31−34.

(24) *Chem2Pac: A Computational Chemistry Integrator;* [www http://www.ifi.unicamp.br/gsonm/chem2pac].

(25) Parr, R. G.; Pearson, R. G. Absolute hardness − companion parameter to absolute eletronegativity. *J. Am. Chem. Soc.* **1983**, *105*, 7512−7516.

(26) Mulliken, R. S. A new electroaffinity scale; Together with data on valence states and on valence ionization potentials and electron affinities. *J. Chem. Phys.* **1934**, *2*, 782−793.

(27) Govers, H. A. J. *Practical Applications of Quantitative Structure−Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W., Devilers, J.; Kluwer Academic publishers: Dordrecht, 1990; p 411.

(28) Nys, G. G.; Rekker, R. F. Concept of Hydrophobic fragmental constants (F-values). 2. Extension of its applicability to calculation of lipophilicities of aromatic and heteroaromatic structures. *Eur. J. Med. Chem. Chim. Ther.* **1974**, *9*, 361−375.

(29) Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. *Chemometrics: A Practical Guide*; John Wiley: 1998.

(30) *Pirouette Multivariate Data Analysis for IBM PC Systems*, Version 2.0; Infometrix: Seattle, WA, 1996.

(31) Aoyama, T.; Suzuki, Y.; Ichikawa, H. Neural networks applied to pharmaceutical problems. 3. Neural networks applied to quantitative structure activity relationship analysis. *J. Med. Chem.* **1990**, *33*, 2583−2590.

(32) Gassner, H.; Probst, M.; Lauenstein, A.; Hermansson, K. Representation of intermolecular potential functions by neural networks. *J. Phys. Chem. A* **1998**, *102*, 4596−4605.

(33) Aoyama, T.; Suzuki, Y.; Ichikawa, H. Neural networks applied to pharmaceutical problems. 1. Method and application to decision-making. *Chem. Pharm. Bull.* **1989**, *37*, 2558−2560.

(34) Aoyama, T.; Ichikawa, H. Neural networks applied to pharmaceutical problems. 5. Obtaining the correlation indexes between drug activity and structural parameters using a neural network. *Chem. Pharm. Bull.* **1991**, *39*, 372−378.

(35) Zupan, J.; Gasteiger, J. *J. Neural Networks for Chemists*; VCH Publishers: New York, 1993.

(36) Hertz, J.; Krogh, A.; Palmer, R. G. *Introduction to the Theory of Neural Computation*; Addison-Wesley: New York, 1990.

(37) Ichikawa, H. *PSDD: Perceptron-type Neural Network Simulator*, QCPE 615.

(38) Cavalieri, E. L.; Rogan, E. G. Central role of radical cations in metabolic-activation of polycyclic aromatic-hydrocarbons. *Xenobiotica* **1995**, *25*, 677−688;
(39) Harvey, R. G. Mechanisms of carcinogenesis of polycyclic aromatic hydrocarbons. *Polycycl. Aromat. Compounds* **1996**, *9*, 1−23.
(40) Jerina, D. M. From arene oxides to diol epoxides and DNA. *Polycycl. Aromat. Compounds* **2000**, *19*, 5−36.
(41) Ladik J.; Förner, W. *The Beginnings of Cancer in the Cell*; Springer-Verlag Berlin, Heidelberg, 1994; p. 55.
(42) Silverman, B. D. *Computer Simulation of Carcinogenic Processes*, CRC Press: Boca Raton, Florida, 1988; p 94.
(43) Ulrich, V.; Roots, I.; Hildebrant, A. G.; Eatabrook, R. K. W. *Microsomes and Drug Oxidations*; Conney, A. H.; Pergamon Press: Oxford, 1977.
(44) Geacintov, N. E. In *Reaction pathways of benzo*[*a*]pyrene diol epoxide and DNA. Polynuclear Aromatic Hydrocarbons: *Physics, Biology, and Chemistry*, 6th International Symposium; Cooke, M.; Dennis, A. J., Fisher, G. L., Eds.; Bartelle Press: Columbus, OH, 1981; p 311.
(45) Hodgson, R. M.; Cary, P. D.; Grover, P. L.; Sims, P. Metabolic-activation of Chrysene by Hamster-embryo cells − Evidence for the formation of a Bay-region diol-epoxide-N2-guanine adduct in RNA. *Carcinogenesis (London)* **1983**, *4*, 1153−1158.
(46) Jerina, D. M.; Lehr, R. E. *Microsomes and Drug Oxidations*; Conney, A. H.; Pergamon Press: Oxford, 1977; p 709.
(47) Flemming, I. *Frontier Orbitals and Organic Chemical Reactions*; Wiley: New York, 1976.
(48) Barone, P. M. V. B.; Braga, R. S.; Camilo, A., Jr.; Galvão, D. S. Electronic indices from semiempirical calculations to identify carcinogenic activity of polycyclic aromatic hydrocarbons. *J. Mol. Struct. (THEOCHEM)* **2000**, *505*, 55−66.
(49) Santo, L. L. E.; Galvão, D. S. Structure−activity study of indole-quinones bioreductive alkylating agents. *J. Mol. Struct. (THEOCHEM)* **1999**, *464*, 273−279.
(50) Cyrillo, M.; Galvão, D. S. Structure−activity relationship study of some inhibitors of HIV-1 integrase. *J. Mol. Struct. (THEOCHEM)* **1999**, *464*, 267−272.
(51) Braga, R. S.; Vendrame, R.; Galvão, D. S. Structure−activity relationship studies of substituted 17 alpha-acetoxyprogesterone hormones. *J. Chem. Inf. Comput. Sci*. **2000**, *40*, 1377−1385.
(52) Mazzali, L. L.; Coluci, V. R.; Galvão, D. S. Structure-Actrivity Relationship Study for some N-Phenylphthalimides Analogues − to be published.
(53) Vendrame, R.; Coluci, V. R.; Braga, R. S.; Galvão, D. S. Structure−Activity Relationship (SAR) Studies of the Tripos Benchmark Steroids. *J. Mol. Struct. (THEOCHEM)* − in press.
(54) Coluci, V. R.; Braga, S. F.; Oyarzun, G. G. C.; Galvão, D. S. − unpublished.
(55) Dewey, T. G. *Fractals in Molecular Biophysics*; Oxford University Press: Oxford, 1997.

CI025577+