Braga, R.S.; Barone, P.M.V.B.; Galvão, D.S.

Identifying Carcinogenic Activity of Methylated and non-Methylated Polycyclic Aromatic Hydrocarbons (PAHs) Through Electronic and Topological Indices

Identifying Carcinogenic Activity of Methylated and non-Methylated Polycyclic Aromatic Hydrocarbons (PAHs) Through Electronic and Topological Indices

R.S. Braga[1*], P.M.V.B. Barone[2], and D.S. Galvão[1]

[1]*Instituto de Física, Universidade Estadual de Campinas - UNICAMP,*

Campinas-SP, Brasil, CP 6165, CEP 13081-970

[2]*Departamento de Física, Universidade Federal de Juiz de Fora - UFJF*

Juiz de Fora, MG, Brasil, CEP 36036-330

Polycyclic aromatic hydrocarbons (PAHs) are a class of planar molecules, abundant in urban environment, which can induce chemical carcinogenesis. Their carcinogenic power varies in a large range, from very strong carcinogens to inactive ones. In a previous study, we proposed a methodology to identify the PAHs carcinogenic activity exploring electronic and topological indices. In the present work, we show that it is possible to simplify that methodology and expand its applicability to include methylated PAHs compounds. Using very simple rules, we can predict their carcinogenic activity with high accuracy ($\approx 89\%$).

I 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]. In spite of many years of theoretical and experimental work, the details of the biochemical phenomena involved in the appearance of malignant tumors are not well-understood. It is believed today that although many factors can be associated with cancer induction, such as virus, radiation, chemical agents, etc., the chemical component is the most important. Among the chemicals that are known to induce cancer, the Polycyclic Aromatic Hydrocarbons (PAHs) are of special relevance. PAHs are a class of planar organic molecules (see Figs. 1 and 2) presenting carcinogenic power which varies from some of the strongest known carcinogens to inactive ones [2].

The reasons why some of these very similar molecules present carcinogenic activity, and others do not, have been the object of intense research since the thirties with the pioneer work of Cook and collaborators [3]. These first works tried to correlate the carcinogenic activity with some geometrical features of the molecules. These ideas were further developed by Pullman and Pullman [4] using quantum mechanical calculations (simple Hückel theory [5]) and were expressed in terms of critical index values over specific molecular regions named K and L (see inset of Fig. 1). Later, similar theories evolved to include one, which is called the bay region' [6-9] (see inset of Fig. 1). A semi-empirical study has been also reported [10] showing a close correlation between the electrophilic reactivity at specific carbon atoms of chrysene and methyl- chrysenes and their carcinogenic activity.

These theories (based on electronic indices) and more recent ones using statistical analysis, neural networks, and artificial intelligence methods [11-13] have achieved only partial success. Some of them work well for a specific subset of compounds and fail for others, and vice-versa. Due to the increasing levels of PAHs present in urban air (partly due to auto exhaust) and in many common processed foods, the search for a theory that could predict, at least at qualitative level, whether a specific PAH will be carcinogenic or not is a very important health challenge.

Recently [14] we proposed a new theoretical approach to identify PAH carcinogenic activity. This approach is based on the concepts of local density of electronic states and critical values for the energy separation between HOMO (highest occupied molecular orbital) and its next lower level HOMO-1. That study was carried out for the first 26 molecules shown in Fig. 1. With a few simple rules, we were able to group and identify their carcinogenic activity.

One interesting experimental fact associated with the PAHs is the role of attached chemical groups. It is a well-known experimental fact that chemical sub-

stitution (methylation, for instance) in PAH molecules can drastically affect their carcinogenic activity [15], 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 (i.e., 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 in 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 rules we have previously proposed [14] are correct, we could expect methylation to induce a discontinuous transition in the carcinogenic activity, i.e., it could make active molecules inactive and vice-versa. Thus, the study of methylated compounds presents itself as a very interesting test to our previously proposed methodology [14].

Figure. 2 Structural scheme of the 49 methylated polycyclic aromatic hydrocarbon (MPAHs) molecules studied in the present work. See Table 2 for their descriptive names. The darker bonds indicate the bonds with the calculated highest bond orders.

II Methodology

In the present work we have studied 81 PAH molecules (49 and 32 methylated and non-methylated PAHs, respectively). Their schematic structures are shown in Figs. 1 and 2. Most of these molecules were selected having in mind the criteria of availability of experimental data for chemical carcinogenesis. For the first 26 molecules shown in Fig. 1, the Iball index is available [9, 16, 17]. The Iball index is defined as the percentage of skin cancer in mice (skin painting experiments) divided by the average latent period in days for the affected animals multiplied by 100 [16, 17]. The remaining 6 molecules shown in Fig. 1 were chosen for comparison purposes; the Iball index is not available for them, so we have chosen the carcinogenic scale proposed by Cavalieri *et al.* [18]. For the methylated compounds, we have used the same scale [18], since the Iball indices are not available for all of them. The methylated structures shown in Fig. 2 are structurally related to the non-methylated molecules shown in Fig. 1, in order to provide a direct comparison.

The calculations analyzed in this work were carried out in the framework of simple Hückel theory [5]. As
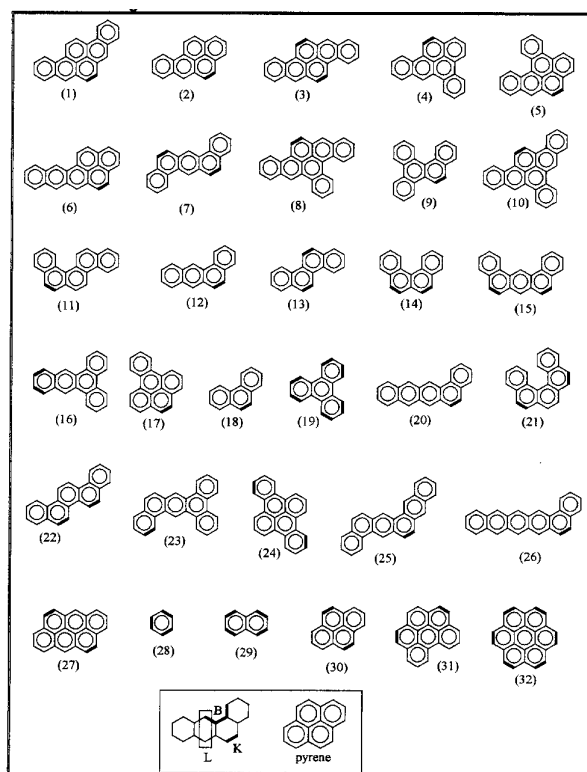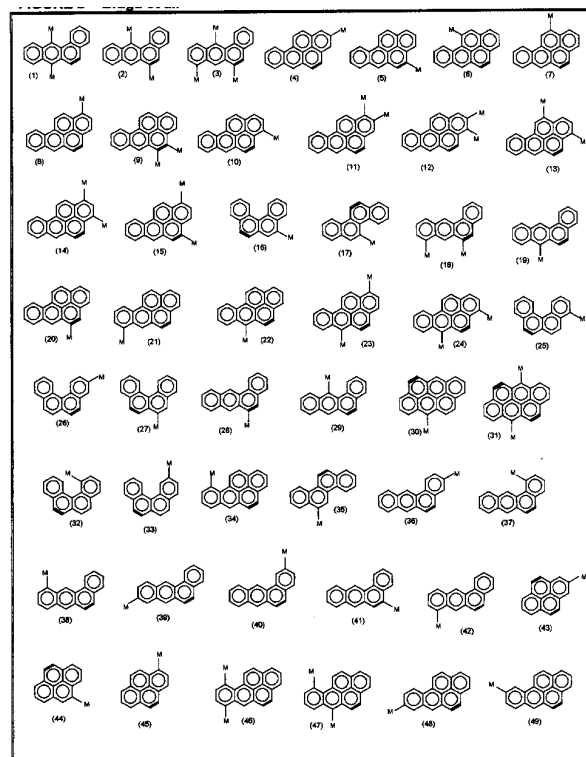

Figure. 1 Structural scheme of the 32 non-methylated polycyclic aromatic hydrocarbon (PAHs) molecules studied in the present work. See Table 1 for their descriptive names. The darker bonds indicate the bonds with the calculated highest bond orders. In the inset are shown the pyrene structure and also typical L, K and bay (B) regions for PAH molecules.

PAHs are planar molecules with a well-defined $\sigma - \pi$ separation the Hückel method is the simplest choice due to its simplicity and good qualitative power prediction. Also, it will allow us to compare our results with a large amount of theoretical studies carried out since the forties using Hückel models [3, 4, 12]. We have used the same method and parameters adopted by Pullman and Pullman [4] for their K-L theory to allow a direct comparison to their results and to our previous results [14, 19]. See Refs. [5] and [19] for details about the Hückel method.

In spite of its simplicity, the Hückel model and similar theories are still very useful in providing the relevant physical information for the qualitative analysis of the electronic and structural properties of organic compounds. For instance, Hückel models have been successfully used to investigate the electronic structure of conducting polymers and molecular crystals [20-24].

In the Hückel model, there is more than one way to treat methylated compounds. In this work, we have used the so-called inductive method, treating the methylation through an appropriate rescaling of the α parameter (α=-0.5) [5]. We have chosen this model (instead of the heteroatom or hyperconjugation models) because, in the present case, it is the best and simplest way to directly compare the electronic density of states (DOS) and the local density of states (LDOS) of methylated and non-methylated PAHs. Since the matrices will have the same dimensions, the summations are carried out over the same number of sites for the methylated and their structural parents.

The DOS is defined as the number of electronic states per energy unit. The related concept of LDOS, i.e., the DOS calculated over a specific molecular region, is introduced in order to also describe the spatial distribution of the states over the system under consideration. Due to the fact that we are carrying out molecular calculations, the eigenvalues form a discrete set and, in order to simulate a continuous set, this δ-function like spectrum has to be smoothed out using Gaussian or Lorentzian functions centered on the eigenvalues [19, 25]. For the LDOS calculations, the contribution of each carbon atom to an electronic level is weighted by the square of the (real) molecular orbital coefficient, i.e., by the probability density corresponding to the level in that site. In our previous works [14, 19] we have used a Lorentzian enveloping accordingly to the following expression:

$$LDOS(E) = \sum_{l=k}^{n_c} \frac{\gamma^2}{(E - E_l)^2 + \gamma^2} \sum_{m=n_1}^{n_f} |c_{ml}|^2. \quad (1)$$

Here, γ is the half-width of the Lorentzian peak ($\gamma = 0.01\beta$), and the spectra are generated varying the energy for a desired energy range (E). For the results shown here, we considered this range to be from -3.0 to 3.0 β (histograms with 500 points). E_1 refers to

the molecular energies and c_{ml} to the coefficients of the expansion of molecular orbitals expressed as a linear combination of atomic orbitals. The summation is carried out over the desired molecular region (initial (n_i) to final (n_f) carbon sites) including all the selected molecular energies ($l = k$ to n_c).

However, this procedure has some disadvantages. It has some intrinsic dependence on the chosen values for the Lorentzian enveloping and also on the initial energy values used to generate the simulated spectra. This precludes the direct comparison with LDOS generated with other methods. For example, if we use a method including all valence electrons, the half-width of the lorentzian-peak could produce artificial changes in the LDOS values through spurious overlap of the molecular levels that are very close. This does not happen with the Hückel method where only π-electrons are taken into account.

To solve these problems, we have rewritten eq. 1 to the following form:

$$LDOS(E_i) = 2 \sum_{m=n_i}^{n_f} |c_{mi}|^2. \quad (2)$$

Using the discrete modulation given by eq. 2 (instead of a continuous Lorentzian envelope) we avoid the problems involving eq. 1 and we are also able to directly compare DOS and LDOS calculated from any LCAO (Linear Combination of Atomic Orbitals) method.

The use of density of states (DOS) and local density of states (LDOS) concepts can give us detailed information on the contributions of specific geometrical regions of the molecules to the chemical reactivity, optical response, etc., and, consequently, to their biochemical behavior.

For the non-methylated PAHs molecules, it was shown [14] that the LDOS analysis over the K, L, and Bay regions (considered by some authors [4-10] to be the relevant molecular regions) did not provide patterns that could be correlated with the carcinogenic power. The same was observed for the LDOS involving terminal rings. However, when this analysis was carried out over the ring containing the highest bond-order (RHBO) in association with the difference in energy between the HOMO and HOMO-1 (Δ energy), a clear pattern appeared [14, 19]. Through very simple rules, it was possible to associate these electronic indices with the carcinogenic activity. For the present study of methylated compounds, we have analyzed these same molecular regions.

III   Results and discussions

The 81 PAH molecules we have studied here (49 methylated and 32 non- methylated ones) are indicated in Figs. 1 and 2. As can be seen from Fig. 2, the methylation does not change the location of the bonds with

the highest bond-orders in relation to the parent PAHs in Fig. 1.

In Tables 1 and 2, we show a summary of the Hückel results for the molecules shown in Figs. 1 and 2, respectively. The values for the HOMO (highest occupied molecular orbital), HOMO-1, their energy difference (Δ) and their LDOS relative contribution difference (η) are presented. The experimental carcinogenic activity is also indicated when available. From these tables we can notice that it is not possible to use any of these data separately as indicators for the carcinogenic activity. Our theoretical predictions are contrasted to the experimental data (when available) and with the results obtained with the methodology of previous work [14].

Table 1 - Summary of the Hückel results for the molecules numbered according to the scheme shown in Fig. 1. The highest occupied molecular orbital (H), the next lower occupied level (H-1), their energy difference value (Δ) and their relative contribution difference to the LDOS (η) are indicated. The theoretical results of the present work (this work) and a previous work (Ref. 14) are contrasted to the experimental data for carcinogenic activity (C.A.). A and D indicate agreement or disagreement, respectively. NA indicates the cases not analyzed in ref. 14. All the energy results are expressed in the usual Hückel resonance energy β (approximately 2.4 eV). See text for discussions.

Table 1 — molecule	H-1	H	Δ	η	C.A. [16-18]	This Work	Ref. [14]
(#1) db*[3,4;9,10]pyrene	-0.682	-0.342	0.340	0.327	74	A	A
(#2) b* [3,4]pyrene	-0.802	-0.371	0.431	0.209	72	A	A
(#3) db[3,4;8,9]pyrene	-0.793	-0.303	0.490	0.096	68	A	A
(#4) db[3,4;6,7]pyrene	-0.742	-0.422	0.320	0.155	50	A	A
(#5) db[1,2;3,4]pyrene	-0.669	-0.398	0.271	0.188	33	A	A
(#6) naphto[2,3;3,4]pyrene	-0.648	-0.303	0.345	0.170	27	A	A
(#7) db [1,2;5,6]anthracene	-0.684	-0.474	0.210	0.146	26	A	A
(#8) tb*[3,4;6,7;8,9]pyrene	-0.671	-0.338	0.333	0.292	20	A	A
(#9) db [1,2;3,4]phenantrene	-0.711	-0.532	0.179	0.434	18	A	A
(#10) tb [3,4;6,7;9,10]pyrene	-0.680	-0.396	0.284	-0.308	17	D	A
(#11) db [1,2;5,6]phenantrene	-0.603	-0.550	0.053	-0.482	10	A	A
(#12) b[1,2]anthracene	-0.715	-0.452	0.263	-0.051	07	A	A
(#13)chrysene	-0.792	-0.520	0.272	0.016	05	D	D
(#14) b[3,4]phenantrene	-0.662	-0.568	0.094	-0.067	04	A	A
(#15) db[1,2;7,8]anthracene	-0.618	-0.492	0.126	0.019	04	A	NA
(#16) db[1,2;3,4]anthracene	-0.714	-0.499	0.215	0.330	03	D	D
(#17) b[1,2]pyrene	-0.718	-0.497	0.221	0.233	02	D	A
(#18)phenantrene	0.769	-0.605	0.164	0.366	00	A	D
(#19)triphenylene	0.684	-0.684	0.000	-0.193	00	A	A
(#20) b[1,2]naphthacene	-0.687	-0.327	0.360	-0.370	00	A	A
(#21) db[3,4;5,6]phenantrene	-0.657	-0.536	0.121	0.147	00	A	NA
(#22) picene	-0.680	-0.502	0.178	0.171	00	D	D
(#23) tb[1,2;3,4;5,6]anthracene	-0.637	-0.522	0.115	-0.087	00	A	A
(#24) db[1,2;6,7]pyrene	-0.673	-0.555	0.118	-0.001	00	A	A
(#25) phenanthra[2,3;1,2]anthracene	-0.555	-0.429	0.126	0.304	00	A	NA
(#26) b[1,2]pentacene	-0.618	-0.244	0.374	-0.356	00	A	A
(#27) anthanthrene	-0.750	-0.291	0.459	-0.503	-	A	D
(#28) benzene	-1.000	-1.000	0	0	-	A	NA
(#29) naphthalene	-1.000	-0.618	0.382	-0.333	-	A	D
(#30) pyrene	-0.879	-0.445	0.434	-0.197	-	A	D
(#31) b[ghi]preylene	-0.684	-0.439	0.245	-0.072	-	A	A
(#32) coronene	-0.539	-0.539	0	-0.019	-	A	A

*b=benzo; db=dibenzo; tb=tribenzo

        Brazilian Journal of Physics, vol. 30, no. 3, September, 2000

Table 2 - Summary of the Hückel results for the methylated compounds, numbered according to the scheme shown in Fig. 2. The highest occupied molecular orbital (H), the next lower occupied level (H-1), their energy difference value (Δ), their relative contribution difference to the LDOS (η), and the experimental carcinogenic activity (C.A.) are indicated. The symbols (+++++), (++++), (+++), (++), (+), (±) and (-) mean extremely active, very active, active, moderately active, weakly active, very weakly active and inactive, respectively. The theoretical results of the present work (this work) and the ones obtained using the methodology of a previous work (Ref. 14) are contrasted to the experimental data for carcinogenic activity (C.A.). A and D indicate agreement or disagreement, respectively. All the energy results are expressed in the usual Hückel resonance energy β (approximately 2.4 eV). See text for discussions. Although for the last three molecules indicated in the table there are no experimental results available, the present approach predicted that they will be active.

Table 2 - molecule	H-1	H	Δ	η	C.A. [18]	This Work	Ref.[14]
(#1) 7,12-Dimethylbenz[a]antracene	-0.697	-0.259	0.438	-0.192	+ + + + +	D	D
(#2) 6,12-Dimethylbenz[a]antracene	-0.661	-0.337	0.324	-0.111	+ + + +	D	D
(#3) 6,8,12-Trimethylbenz[a]antracene	-0.652	-0.292	0.360	-0.214	+ + + +	D	D
(#4) 2-Methylbenzo[a]pyrene	-0.749	-0.369	0.380	0.077	+ + + +	A	A
(#5) 4-Methylbenzo[a]pyrene	-0.787	-0.331	0.456	0.128	+ + + +	A	A
(#6) 11-Methylbenzo[a]pyrene	-0.742	-0.352	0.390	0.216	+ + + +	A	A
(#7) 12-Methylbenzo[a]pyrene	-0.801	-0.326	0.475	0.140	+ + + +	A	A
(#8) 1-Methylbenzo[a]pyrene	-0.800	-0.314	0.486	0.267	+ + + +	A	A
(#9) 4,5-Dimethylbenzo[a]pyrene	-0.765	-0.258	0.507	0.029	+ + + +	A	A
(#10) 3-Methylbenzo[a]pyrene	-0.765	-0.319	0.446	0.302	+ + + +	A	A
(#11) 1,2-Dimethylbenzo[a]pyrene	-0.743	-0.302	0.441	0.024	+ + + +	A	A
(#12) 2,3-Dimethylbenzo[a]pyrene	-0.685	-0.318	0.367	0.241	+ + + +	A	A
(#13) 3,12-Dimethylbenzo[a]pyrene	-0.760	-0.284	0.476	0.206	+ + + +	A	A
(#14) 1,3-Dimethylbenzo[a]pyrene	-0.765	-0.272	0.493	0.335	+ + + +	A	A
(#15) 1,4-Dimethylbenzo[a]pyrene	-0.787	-0.284	0.503	0.137	+ + + +	A	A
(#16) 5- Methylbenzo[c]phenanthrene	-0.639	-0.502	0.137	-0.331	+ + +	D	D
(#17) 5- Methylchrysene	-0.738	-0.485	0.253	0.036	+ + +	A	A
(#18) 6,8-Dimethylbenz[a]antracene	-0.672	-0.362	0.310	-0.005	+ + +	D	D
(#19) 7-Methylbenz[a]antracene	-0.715	-0.360	0.355	-0.037	+ + +	D	D
(#20) 5-Methylbenzo[a]pyrene	-0.798	-0.331	0.467	0.267	+ + +	A	A
(#21) 7-Methylbenzo[a]pyrene	-0.759	-0.333	0.426	0.245	+ + +	A	A
(#22) 6-Methylbenzo[a]pyrene	-0.799	-0.286	0.513	0.252	+ + +	A	A
(#23) 1,6-Dimethylbenzo[a]pyrene	-0.795	-0.225	0.570	0.308	+ + +	A	A
(#24) 3,6-Dimethylbenzo[a]pyrene	-0.759	-0.228	0.531	0.338	+ + +	A	A
(#25) 4-Methylbenzo[c]phenanthrene	-0.646	-0.510	0.136	-0.379	+ +	A	A
(#26) 3-Methylbenzo[c]phenanthrene	-0.658	-0.535	0.123	-0.168	+ +	A	A
(#27) 6- Methylbenzo[c]phenanthrene	-0.592	-0.541	0.051	-0.583	+ +	A	A
(#28) 6-Methylbenz[a]antracene	-0.690	-0.403	0.287	0.162	+ +	D	D
(#29) 12-Methylbenz[a]antracene	-0.697	-0.375	0.322	-0.227	+ +	A	A
(#30) 6-Methylanthanthrene	-0.727	-0.223	0.504	-0.612	+ +	A	D
(#31) 6,12-Dimethylanthanthrene	-0.705	-0.150	0.555	-0.445	+ +	A	D
(#32) 1-Methylbenzo[c]phenanthrene	-0.659	-0.507	0.152	-0.215	+	A	A
(#33) 2- Methylbenzo[c]phenanthrene	-0.627	-0.543	0.084	-0.295	+	A	A
(#34) 10-Methylbenzo[a]pyrene	-0.728	-0.353	0.375	0.221	+	D	D
(#35) 6- Methylchrysene	-0.792	-0.445	0.347	0.046	±	D	D
(#36) 3-Methylbenz[a]anthracene	-0.665	-0.446	0.219	-0.092	-	A	A
(#37) 1-Methylbenz[a]antracene	-0.617	-0.452	0.165	0.051	-	A	D
(#38) 11-Methylbenz[a]antracene	-0.696	-0.403	0.293	-0.249	-	A	A
(#39) 9-Methylbenz[a]antracene	-0.684	-0.430	0.254	-0.151	-	A	A
(#40) 2-Methylbenz[a]antracene	-0.690	-0.428	0.262	0.093	-	D	D
(#41) 5-Methylbenz[a]antracene	-0.675	-0.405	0.270	0.068	-	D	D
(#42) 8-Methylbenz[a]antracene	-0.709	-0.397	0.312	-0.164	-	A	A
(#43) 2-Methylpyrene	-0.779	-0.445	0.334	-0.142	-	A	D
(#44) 4-Methylpyrene	-0.860	-0.398	0.462	-0.358	-	A	D
(#45) 1-Methylpyrene	-0.862	-0.377	0.485	-0.232	-	A	D
(#46) 7,10-Dimethylbenzo[a]pyrene	-0.673	-0.295	0.378	0.164	-	D	D
(#47) 6,10-Dimethylbenzo[a]pyrene	-0.728	-0.268	0.460	0.262	?	?	?
(#48) 8-Methylbenzo[a]pyrene	-0.722	-0.337	0.385	0.233	?	?	?
(#49) 9-Methylbenzo[a]pyrene	-0.801	-0.342	0.459	0.190	?	?	?

Barone *et al.* [14] have studied the first 26 molecules in Fig. 1 and proposed the following three simple rules to identify carcinogenic activity (based on the Δ energy values associated with the HOMO and HOMO-1 relative contribution to the LDOS over the RHBO):

Pyrenelike molecules.

(a) If the molecule contains a pyrenelike structure (see inset of Fig. 1) and Δ is greater than 0.25β ($\beta \approx 2.4$eV), it will be strongly carcinogenic. Otherwise, the molecule will be inactive.

Nonpyrene molecules

(b) If the HOMO is the highest (peak) contribution to the LDOS over RHBO, the molecule will be completely inactive.

(c) If the HOMO contribution to the LDOS over RHBO is greater than that of HOMO-1 (but not the highest peak) and $\Delta > 0.15\beta$, the molecule will present a strong or moderate carcinogenic activity. If the HOMO-1 contribution is greater than that of the HOMO, the molecule will present weak or no activity at all. Typical examples of these rules are shown in Fig. 3.
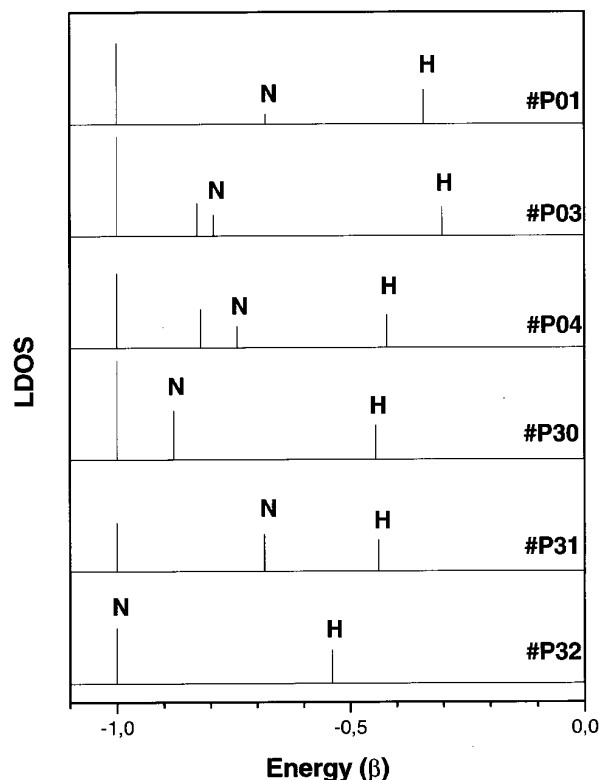

Figure. 3 Local density of states (LDOS) in arbitrary units (a.u.) over the ring that contains the highest bond order (RHBO) for typical active and inactive non-methylated molecules. For simplicity, only the valence states are displayed. H indicates the highest occupied molecular orbital (HOMO) and N is the next lower molecular orbital (HOMO-1).

This set of rule presents some limitations. If a compound have the HOMO contribution greater than HOMO-1 (positive η) and $\Delta < 0.15\beta$ (case not present in the Lorentzian analysis [14, 19]), these rules cannot be used to determine whether the compound is active or not.

However, this situation appears when we use the discrete representation of the spectra (see tables 1 and 2) and it needs to be considered. The present study including a larger number of compounds (methylated and non-methylated) and using discrete modulation to DOS and LDOS spectra allows us to treat this case. Besides that, the above set of three rules can be reduced to just one simpler and more encompassing rule:

- **If the $\eta > 0$ and $\Delta > 0.17\beta$, the molecule will present a strong or moderate carcinogenic activity. Otherwise, the molecule will present a weak or null activity.**

This rule explores the same concepts of the original set of rules, critical Δ values and relative LDOS contributions to HOMO and HOMO-1. In Fig. 3 we show typical results for active and inactive non-methylated compounds whose carcinogenic activity is correctly predicted by the above rule.

In what follows, we will discuss the results for methylated compounds. We would like to stress that the description of the methylation process in terms of simple perturbation of the α parameter for the carbon at which the chemical group is attached is a strong approximation. Even so, in this approach - as the results below will show - at least the qualitative behavior of the carcinogenic activity is appropriately described. This is a clear indication that the methodology we are proposing is physically sound.

We have examined 49 methylated molecules (with experimental data available for 46 of them) and the above new rule can correctly predict the absolute carcinogenic activity of 74% of them. This is an excellent result, considering the approximations we have used to treat the methylation process. If we use the above rule to analyze the tendency of changes in the Δ values under methylation (i.e., whether it increases or decreases) the agreement with the experimental data goes to 89%. Typical LDOS results for these molecules are shown in Fig. 4.

The most interesting cases are those for which, under methylation, the active molecules become inactive or vice-versa. This happens for 9 out of 46 compounds. In Table 3 we show the variation of Δ for these compounds. We can see from this table and from Fig. 5 that for the active compounds which become inactive, the Δ value decreases (#P2, #M34 and #M46) and for inactive compounds which become active, the Δ value increases (#P12, #M01 and #M03), following the tendency of the requirements for activity/inactivity of our single rule. These results are in agreement with the

experimental data for 8 out of the 9 compounds. For the remaining methylated structure, its activity is correctly predicted by our rule, which, however, fails in the prediction of its parent PAH structure.
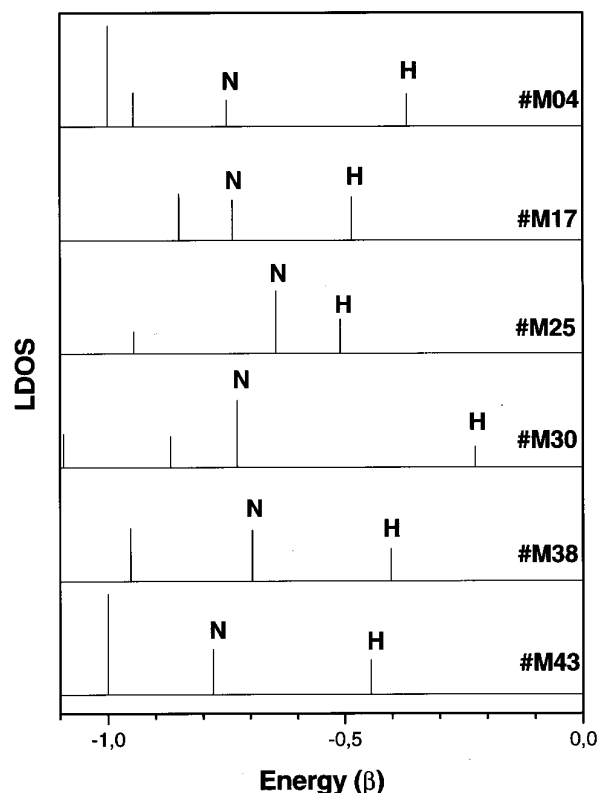

Figure. 4 Local density of states (LDOS) in arbitrary units (a.u.) over the ring that contains the highest bond order (RHBO) for methylated molecules representative of the rule stated in the text. For simplicity, only the valence states are displayed. H indicates the highest occupied molecular orbital (HOMO) and N is the next lower molecular orbital (HOMO-1). See text for discussion.

For the 78 molecules with available experimental data, our new single rule correctly describes the biological activity of 61 of them (78%). If we include the tendencies, we correctly describe 69 out of 78 (89%). Three of the methylated molecules shown in Fig. 2 (#M47, #M48 and #M49) are predicted by our rule to be carcinogenic but we do not have experimental data available for them.

The biochemical processes leading to chemical carcinogenesis are very complex phenomena, not well understood in all the details. It is very intriguing that, without assuming any biochemical mechanism and using a very simple rule based on the Hückel method, we are able to predict with high accuracy the carcinogenic activity of PAH molecules, methylated or not.
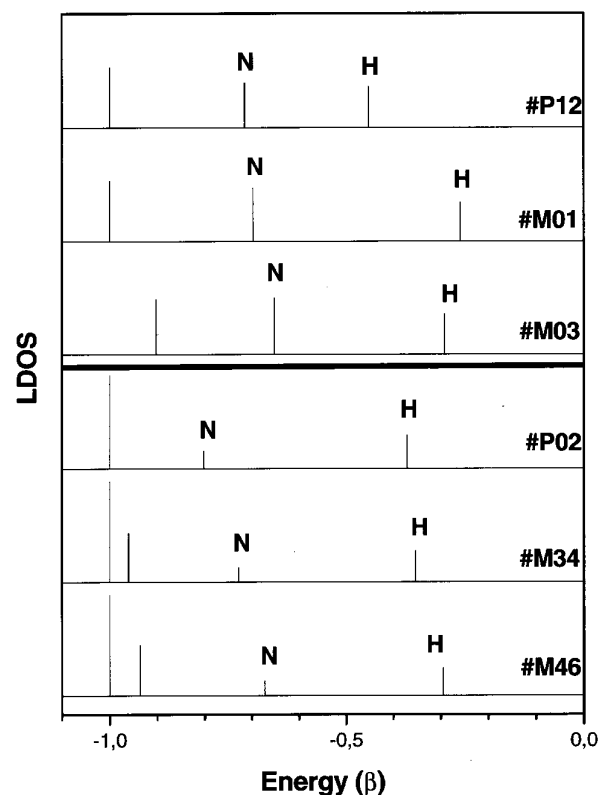

Figure. 5 Local density of states (LDOS) in arbitrary units (a.u.) over the ring that contains the highest bond order (RHBO) for typical molecules which under methylation present variation in their carcinogenic activity. For simplicity, only the valence states are displayed. H indicates the highest occupied molecular orbital (HOMO) and N is the next lower molecular orbital (HOMO-1). See text for discussion.

The K-L theories, as well the Bay theories, both assume the existence of a metabolic activation process inducing carcinogenic activity. These theories use energetic indices, which in fact represent activation energies [26]. One possible explanation of why the present methodology works is that the local density of states (over the ring that is the most susceptible to specific chemical reactions) measures these activation energies (believed to be directly correlated to the carcinogenic power [7]) more efficiently. But again, the existence of a minimum Δ value playing a decisive role in determining carcinogenic activity is a crucial feature. This was originally suggested by Barone *et al.* [14] and it might explain some of the K-L model failures. The physical meaning of the minimum Δ can be expressed in terms of frontier orbitals. It seems that a 'clean' frontier orbital, i. e., a HOMO well separated in energy from the HOMO-1, is a necessary but not sufficient condition for carcinogenic activity. It is the 'balance' between relative HOMO and HOMO-1 contributions (Δ values) and their energy separation (Δ values) that determines if a specific PAH molecule will be carcinogenic or not. In fact, preliminary results using more sophis-

ticated methods [27] (beyond Hartree-Fock level) indicate that active and inactive molecules have different patterns in terms of the mixing of configuration states. This suggests the existence of excited states with different lifetimes for active and inactive molecules and consequently different specific reactivities or activation energies. These aspects remain to be better elucidated.

Table 3 - Relative variations of the Δ values for the 9 methylated compounds (M) that change the carcinogenic activity of its related non-methylated parent structure (P). The numbering for the parent (P) and the related methylated compounds (M) is according to Figs. 1 (P) and 2 (M).

Table 3

Molecule	Δ	ΔRelative
P02	0.431	0
M34	0.375	-0.056
M46	0.378	-0.053
P12	0.263	0
M1	0.438	0.175
M2	0.324	0.061
M3	0.360	0.097
M18	0.311	0.048
M19	0.355	0.092
P13	0.272	0
M17	0.253	-0.019
P14	0.095	0
M16	0.137	0.042

We would like to stress that the present work, based on the electronic features of isolated molecules, can only be used to classify molecules as active or not, but it cannot be used to predict potency. Environmental aspects, such as hydrophobicity (not considered here), play a major role in determining the potency of the active compounds, while mainly electronic factors differentiate the active from the inactive ones. QSAR studies show no correlation between the electronic indices and the carcinogenic potency [28]. Also, the carcinogenic power of some PAHs compounds varies, depending on the way of application (subcutaneous injection or painted skin). Thus, the classification criteria for isolated molecules is better defined in terms of active or inactive [12].

In summary, we have presented new developments in the electronic indices methodology (EIM) to identify carcinogenic activity of methylated and non-methylated PAH molecules, which enlarge and generalize a previous methodology [14, 19]. This improvement allowed the construction of a single rule that explores the concept of relative HOMO and HOMO-1 contributions to the local density of states over the ring that contains the highest bond-order (RHBO) and on the separation in energy between these orbitals. We have used the simple Hückel method, but the methodology can be adapted to more sophisticated methods, semi-empirical or even good quality ab initio methods. In fact, it is an interesting question to know whether the rule is artificially produced by the Hückel parameterization. Preliminary calculations [27] using sophisticated semi-empirical methods for non-methylated molecules have produced very similar results and we believe that this can be extended also for the methylated compounds. That study is in progress.

References

[1] T. Sugimura, Science **258**, 603 (1992).

[2] R.G. Harvey, N.E. Geacintov, Acc. Chem. Res. **21**, 66 (1988).

[3] C.A. Coulson, Adv. Canc. Res. **1**, 1 (1953) and references therein.

[4] A. Pullman, B. Pullman, Adv. Canc. Res. **3**, 117 (1955) and references therein.

[5] A. Streitwieser, *Molecular Orbital Theory*, Wiley, N. York (1961).

[6] J.P. Lowe, B.D. Silverman, Acc. Chem. Res. **17, 332**, (1984).

[7] D.M. Jerina, *et al.*, in *Carcinogenesis: Fundamental Mechanisms and Environmental Effects*, B. Pullman, P. O. Ts'o and H. Gelboin (Eds.), D. Reidel Publishing Co., Holland, 1980, p. 1.

[8] L.V. Szentpály, J. Am. Chem. Soc. **106**, 6021 (1984).

[9] J. Gayoso, S. Kimri, Int. J. Quant. Chem. **38**, 461 (1990); **38**, 487 (1990).

[10] S. Kimri, J. Gayoso, J. Mol. Struc. (THEOCHEM) **362**, 141 (1996).

[11] U.E. Nordún, W. Svante, Acta Chem. Scand. B **32**, 602 (1978).

[12] D. Villemin, D. Cherqaoui, A. Mesbah, J. Chem. Inf. Comput. Sci. **34**, 1288 (1994).

[13] X.-H. Song, M. Xiao, R.-Q. Yu, Computers Chem.**18**, 391 (1994).

[14] P.M.V.B. Barone, A. Camilo Jr., D.S. Galvão, Phys. Rev. Lett. **77**, 1186 (1996).

[15] D.W. Jones, R.S. Matthews, in *Progress in Medical Chemistry*, G. Ellis and G.B. West (Eds.), North-Holland, **101**, 59 (1974).

[16] P. Daudel, R. Daudel, *Chemical Carcinogenesis and Molecular Biology,* Wiley- Interscience, New York, 1966, pp. 1-5.

[17] W.C. Herndon, L.V. Szentpály, J. Mol. Struct. (THEOCHEM) **148**, 141 (1986) and references therein.

[18] E.L. Cavalieri, E.G. Rogan, R.W. Roth, R.K. Saugier, A. Hakan, Chem. Biol. Interact. **47**, 87 (1983).

[19] R.S. Braga, P.M.V.B. Barone, D.S. Galvão, J. Mol. Struct. (THEOCHEM) **464**, 257 (1999).

[20] D.S. Galvão, D.A. Santos, B. Laks, C.P. Melo, M.J. Caldas, Phys. Rev. Lett. **63,** 786 (1989); **65,** 527 (1990).

[21] Z.G. Soos, S. Ramasesha, D.S. Galvão, Phys. Rev. Lett. **71**, 1609 (1993).

[22] F. Lavarda, M.C. Santos, D.S. Galvão, B. Laks, Phys. Rev. Lett. **73**, 1264 (1994).

[23] R.H. Baughman, D.S. Galvão, Nature **365**, 735 (1993).

[24] B. Laks, D.S. Galvão, Phys. Rev. B **56**, 967 (1997).

[25] L.E. Sansores, R.M. Valladares, J.A. Cogordan, A.A. Valladares, J. Non-Cryst. Solids. **143**, 232 (1992).

[26] R. Benigni, C. Andreoli, A. Giuliani, Environ. Mol. Mutagen. **24**, 208 (1994).

[27] P.M.V.B. Barone, R.S. Braga, A. Camilo Jr., D.S. Galvão, J. Mol. Struct. (THEOCHEM), **505**, 55 (2000).

[28] R. Vendrame, R.S. Braga, Y. Takahata, D.S. Galvão, J. Chem. Inf. Comp. Sci. **39**, 1094 (1999).