CROATICA CHEMICA ACTA
CCACAA **77** (1–2) 203–211 (2004)

ISSN-0011-1643
*CCA*-2918
*Original Scientific Paper*

# On the Usefulness of Graph-theoretic Descriptors in Predicting Theoretical Parameters. Phototoxicity of Polycyclic Aromatic Hydrocarbons (PAHs)*

**Ernesto Estrada** and Grace Patlewicz**

*Safety and Environmental Assurance Centre (SEAC), Unilever Colworth, Colworth House, Sharnbrook, Bedford MK44 1LQ, U.K.*

RECEIVED MAY 5, 2003; REVISED JULY 7, 2003; ACCEPTED AUGUST 6, 2003

Two main aspects on the use of theoretical approaches in predicting biological endpoints are addressed. The first concerns the appropriateness of theoretical methods and the misbelief that more sophisticated approaches produce better results. We demonstrate that the simple graph theoretic HMO approach produces results as good as »high level« *ab initio* calculations using only $10^{-4}$ to $10^{-7}$ of the time. Secondly, we investigate the feasibility of using *a priori* »mechanistic insights« in order to select the variables to be included in a QSAR model. As the majority of biological response measures are not specific for one particular mechanism, the use of this approach may well produce unrealistic results. The black box approach with its many descriptors and *a posteriori* interpretation of the results is much more appropriate in such cases where the biological response is the result of several mechanisms involving distribution and metabolism. All these aspects are analysed on the basis of the toxicity of PAHs after photoactivation by UV radiation.

*Key words*
• graph-theoretic descriptors
• graph theoretic HMO approach
• prediction of biological parameters
• QSAR models
• mechanistic interpretation of models
• polycyclic aromatic hydrocarbons (PAHs)

## INTRODUCTION

There has been great scientific, political and social interest in recent years about the contamination produced by polycyclic aromatic hydrocarbons (PAHs).[1] These chemicals are major products of combustion processes, and as a consequence they may be found in different environmental scenarios.[2] Amongst the intensive research dedicated to studying PAHs, modelling and prediction of their toxicity to different species has attracted a lot of attention.[3–9] Many PAHs are known to be carcinogenic to humans and have been found to be toxic to aquatic organisms when activated by ultraviolet (UV) light.[2]

This interest has justified the use of several theoretical tools to try to understand the physicochemical processes involved in the distribution, accumulation and toxicity of these chemicals in different environments.[3–9] In 1994, Mekenyan *et al*. published an interesting result where the acute lethality of PAHs to *Daphnia magna* was related to the energy gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO).[10] Due to the complex nature of the photo-induced toxicity, which is the result of competing processes such as stability and light absorbance, a multilinear relationship between toxicity and chemical structure was observed in this work. In fact, in

* Dedicated to Professor Nenad Trinajstić on the occasion of his 65th birthday and for his many contributions to theoretical chemistry.

** Author to whom correspondence should be addressed (E-mail: estrada66@yahoo.com).

order to obtain good correlations with the HOMO-LUMO gap, Mekenyan *et al*.[10] separated the PAHs into several groups. The theoretical methods used in this work were the semi-empirical molecular orbital (MO) methods AM1 (Austin model 1)[11] and PM3 (parameterisation method 3).[12] The results obtained by these authors are both theoretically sound and mechanistically interpretable in spite of the differences obtained between the computed energy gap and the experimentally determined excited state energies for the singlet and triplet states.

More recently, Betowski *et al*.[13] published »high level *ab initio* calculations« for a subset of the series of PAHs analysed by Mekenyan *et al.*[10] In this work, the authors used post-Hartree-Fock calculations using Configuration Interaction (CI) approach. The basis set analysed included CIS/6–311G(d,p); CISD/3–21G and UHF-RHF/6–311G(d,p), the first was used as a compromise between time consumption and precision of the results compared with those determined experimentally. The time consumption is not a minor problem in such calculations. Using a Cray C94 supercomputer these authors reported a computing time of 14 h 7 min 50.2 s for their calculations on perylene using CIS/6–311G(d,p) though the time extends to 24 days 32 min 39.2 s if CISD/3–21G is preferred. It has been this factor, namely the monstrous consumption of computer time in search of precision, which has motivated us to carry out the current research. Firstly we wish to make some comments about the use of quantum mechanical molecular models in chemistry and particularly in QSAR (quantitative structure-activity relationships).

## THEORETICAL METHODS IN CHEMISTRY

The main objectives of using theoretical approaches in chemistry are:

(i) to explain the chemical behaviour of molecules expressed through physicochemical or biological processes/properties and

(ii) to make predictions of such processes/properties for other molecules. The dream of predicting chemical behaviour from »first principles« started in the 1920s when Dirac claimed that the development of quantum mechanics opened up the possibility of predicting them *a priori*. As recalled by Dewar almost 20 years ago,[14] the Schrödinger equation is not exact, it is only an approximation where electron spin is incorporated in the results only as an artefact. Consequently, neither the Schrödinger nor the Dirac equation have been solved rigorously for any multielectron system. That is, the error is unknown.[14]

This fact makes the current *ab initio* methods of no higher chemical value in an *a priori* sense, though this does not mean they are not useful at »a less exalted level«. The main criticism made by Dewar to quantum theory in chemistry is described in the following paragraph extracted from the afore mentioned paper.[14] »There is no question of its (quantum theory) leading to solutions of chemical problems 'from first principles'. Being purely empirical, its justification, like that of any other empirical method, lies solely in its practical value.« As the same author stated that the only way to validate these results is through consideration of their chemical value. Consequently, we can obtain the same chemical information at different levels of »precision« compared with the experiment, which in fact means at different levels of theory.

Chemical graph theory[15,16] and graph-theoretic molecular descriptors (topological indices),[17,18] have been criticised by several authors due to their »over-simplification« of the molecular structure as well as their lack of physical meaning. It is known that most of the models created by using this type of theory or descriptors are based on the correlation between them and experimental properties, which clearly identifies the phenomenological nature of such approach. However, what happens with quantum theory in this respect? Is there any substantial difference? The Roothaan-Hall (RH) self-consistent field linear combination of molecular orbitals (SCF LCAO) is »a curve-fitting exercise in which the parameters in an arbitrary parametric function ($\phi$) are adjusted to make it fit a certain wave function ($\psi$). Since $\psi$ is unknown, there is no way to determine how good the fit is.«[14] In fact, the only criterion to evaluate the quality of this fit is by using an empirical comparison with experimental properties in the same way as any other semi-empirical approach, such as the proper graph theory. There are several reports where *ab initio* results are presented without any experimental comparison in the belief that they are based on »first principles«.

Concerning the topic of over-simplifications it is worth saying that chemical graph theory is not the only example of using them successfully in science. The use of a liquid-drop model to represent the atoms of super-heavy elements could be considered an over-simplification. However, this model developed by Bohr and Wheeler[19] allowed precise predictions on the fission of uranium to be made, and we recall again that prediction is the only reason for the existence of such theoretical models. Presently there are no doubts about the predictability of models created using graph-theoretical ideas in chemistry, such as QSPR and QSAR models. At this point the question is: Why should not such simple models be used to obtain important chemical insights in a fast and effective way? It is known that the Hückel molecular orbital (HMO) method[20] describes the chemical behaviour of PAHs very effectively. The topological nature of such an approach has been proved and it is extensively documented in the literature.[16] As we will show here, this approach is also useful in describing the excited states energies of PAHs in a comparable way to those »high level *ab initio*« methods.

## THEORETICAL APPROACH

Here we will use the Hückel molecular orbital theory (HMO) in order to make the calculation of the electronic structure of PAHs.[20] This theory begins with two structural assumptions:

The electrons of interest initially occupy a system of carbon 2p orbitals having a common nodal plane; that is, with their long axes parallel; they interact to form π-type molecular orbitals (MOs).

The rest of the electrons in the molecule occupy an s-orbital framework (σ MOs) that is orthogonal to the 2p orbitals and therefore does not interact with them.

The linear combination of the atomic orbitals to produce molecular orbitals (LCAO-MO) of the $C_{2p}$ orbitals is given as:

$$\psi = \sum_i c_i \varphi(i) \tag{1}$$

where $\varphi(i)$is a $C_{2p}$ orbital on atom $i$. The optimum coefficients and energies are found by solving the secular determinant:

$$\begin{vmatrix} H_{11}-\varepsilon S_{11} & H_{12}-\varepsilon S_{12} & \cdots & H_{1n}-\varepsilon S_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ H_{11}-\varepsilon S_{11} & H_{n1}-\varepsilon S_{n1} & \cdots & H_{nn}-\varepsilon S_{nn} \end{vmatrix} = 0 \tag{2}$$

where: $H_{ij} = \int \varphi(i) H \varphi(j) d\tau$ and $S_{ij} = \int \varphi(i)\, \varphi(j) d\tau$ are the Coulomb and resonance integrals, respectively.

This secular determinant can be represented as a graph. A graph G is a finite set of dots called vertices connected by lines called edges.[15,16] More formally a simple graph G = (V, E) is a (usually finite) set of vertices V and set of unordered pairs of distinct elements of V called edges, E. In this particular case we are not dealing with a simple graph but with a so-called pseudo-graph. Informally, a pseudograph is a graph with multiple edges or loops between the same vertices or the same vertex. Formally: a pseudograph is a set V of vertices along a set E of edges, and a function $f$ from E to $\{\{u,v\}|u,v$ in V$\}$. (The function $f$ shows which vertices are connected by which edge.) An edge is a loop if $f(e) = \{u\}$ for some vertex $u$ in V.

The vertex $i$ in the pseudograph is weighted by $H_{ii} - \varepsilon S_{ii}$ and the edge $(i, j)$ by the weight $H_{ij} - \varepsilon S_{ij}$. If these weights are elements of the sets $\xi_V$ and $\xi_E$, respectively, the weighted graph is formally defined as follows:

*Definition*. Let $\xi_V$ and $\xi_E$, be finite sets of vertices and edge weights, respectively. Let V be a finite nonempty set of vertices, $l$ a total function $l: V \rightarrow \xi_V$, E a set of unordered pairs of distinct vertices (called edges), and $\chi$ a total function $\chi: E \rightarrow \xi_E$. G = (V, $l$, E, $\chi$) is a weighted graph.

The HMO theory makes the following assumptions to reduce the complexity of the secular determinant:[20]

(i) All Coulomb integrals are equal to $\alpha$, regardless of the molecular environment of the particular carbon atom.
(ii) The integrals of the form $H_{ij}$ are equal to zero for non-adjacent atoms and equal to $\beta$ for adjacent ones.
(iii) The overlap integrals $S_{ij}$ are set equal to 1 if $i = j$ and equal to zero otherwise.

These assumptions transform the graph to a pseudograph with loops associated to each vertex. This implies that the secular determinant of the HMO approach can be expressed in terms of the adjacency matrix of this graph as follows:

$$(\alpha - \varepsilon)\boldsymbol{I} + \beta \boldsymbol{A} = 0 \tag{3}$$

where $\boldsymbol{I}$ is an unit matrix of order $n$ and the elements of the adjacency matrix $\boldsymbol{A}$ are defined as follows:[16]

$$A_{ij} = \begin{cases} 1 \text{ if } i \text{ and } j \text{ are adjacent} \\ 0 \text{ othervise} \end{cases}$$

If we divide the expression (3) by $\beta$ and substitute for $\gamma = (\alpha - \varepsilon)/\beta$, the problem of finding the energy of the molecules is reduced to the calculation of the eigenvalues of the adjacency matrix of the corresponding graph. These calculations can be carried out using a simple program in MATHCAD or using programs which are accessible on the web: www.chem.ucalgary.ca/SHMO/.

## EXCITATION ENERGIES OF PAHs

According to the mechanism of PAHs toxicity, these compounds absorb UV energy producing excited triplet states, which then transfer the energy to molecular oxygen producing singlet molecular oxygen or other free radicals. These radicals can react with cellular components and macromolecules thus producing damage.[21] Hence, the study of the excited state energies of PAHs constitutes an important factor in understanding their phototoxicity. Figure 1 depicts the molecular structures of the PAHs studied by Betowski *et al.*[13] (the first nine structures) which are included in the current study. We extend this data set to other nine structures to a total of 18 taken from the work of Mekenyan *et al.*[10] Fluorenes were not included in the current study on account of the presence of a $CH_2$ group which breaks the conjugation of the molecule and makes it unavailable for the simple HMO treatment. It is known that this group can be treated as a heteroatom in the HMO and the energy can be calculated after the inclusion of a parameter for this group.[20] However, for the sake of simplicity we excluded this compound from the current analysis as our objective could be fulfilled without its consideration.

Using the HMO approach we computed the HOMO--LUMO difference for these compounds that are given

in Table I together with those calculated by Betowski *et al*.[13] using *ab initio* CIS/6–311G(d,p) and those calculated by Mekenyan *et al*.[10] by the semiempirical AM1 and PM3 methods. In the last row of this table we give the correlation coefficient for the linear regressions between the experimental excitation energy and those calculated by the methods analysed for the reduced data set of Betowski *et al.* and for the extended one of 18 compounds.

As can be seen from the correlation coefficients of the linear regressions with the experimental triplet excitation energies, in the reduced data set the simple HMO values produce results as good as the high level *ab initio* CIS/6–311G(p,d) and better results than those obtained using the semiempirical quantum chemical methods AM1 and PM3. These differences with the semiempirical methods increases dramatically for the extended data set where the HMO method gives much better models than AM1 and PM3 methods. The linear correlation between the HMO HOMO-LUMO gap and the experimental excitation energies of the triplet state is illustrated in Figure 2.

At this point, the conclusion of the current work is very straightforward: for the analysis of the excitation energy of PAHs, which can be involved in the phototoxicity of these chemicals, the HMO theory produces results as good as high level *ab initio* calculations but in about $10^{-4}$ the time consumed by the last calculations at the CIS/6–311G(p,d) level or in $10^{-7}$ the time consumed by CISD/3–21G level. Other savings are the disk usage

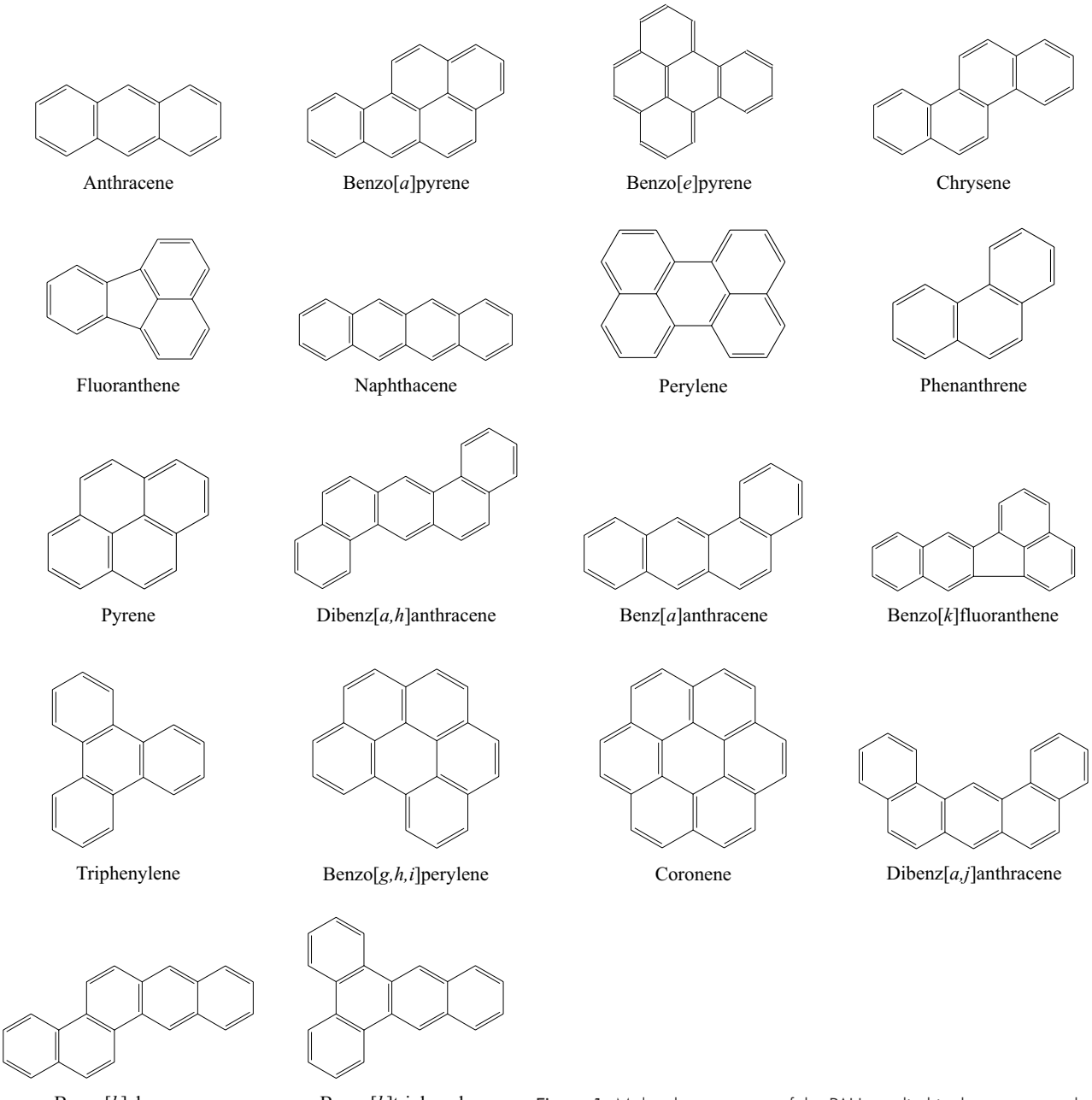



Figure 1. Molecular structures of the PAHs studied in the current work.

Table I. Experimental and calculated (by four different theoretical approaches) values of the triplet excitation energies of PAHs ($E$/eV)

| No. | Compound | Exp. | CIS/6–311G(d,p) | AM1 | PM3 | HMO ($\beta$) |
|---|---|---|---|---|---|---|
| 1 | Anthracene | 1.84 | 2.08 | 7.284 | 7.279 | 0.828 |
| 2 | Benzo[*a*]pyrene | 1.83 | 2.12 | 6.812 | 6.817 | 0.742 |
| 3 | Benzo[*e*]pyrene | 2.29 | 2.48 | 7.362 | 7.365 | 0.994 |
| 4 | Chrysene | 2.48 | 2.61 | 7.693 | 7.713 | 1.04 |
| 5 | Fluoranthene | 2.29 | 2.58 | 7.701 | 7.680 | 0.989 |
| 6 | Naphthacene | 1.27 | 1.54 | 6.517 | 6.517 | 0.590 |
| 7 | Perylene | 1.53 | 1.90 | 6.700 | 6.712 | 0.694 |
| 8 | Phenanthrene | 2.68 | 2.75 | 8.207 | 8.202 | 1.210 |
| 9 | Pyrene | 2.10 | 2.35 | 7.239 | 7.239 | 0.890 |
| 10 | Dibenz[*a,h*]anthracene | 2.25 | – | 7.452 | 7.458 | 0.946 |
| 11 | Benz[*a*]anthracene | 2.07 | – | 7.392 | 7.395 | 0.904 |
| 12 | Benzo[*k*]fluoranthene | 2.18 | – | 7.389 | 7.382 | 0.860 |
| 13 | Triphenylene | 2.92 | – | 8.215 | 8.204 | 1.368 |
| 14 | Benzo[*g,h,i*]perylene | 2.00 | – | 6.957 | 6.971 | 0.878 |
| 15 | Coronene | 2.40 | – | 6.967 | 6.907 | 1.078 |
| 16 | Dibenz[*a,j*]anthracene | 2.28 | – | 7.119 | 7.035 | 0.984 |
| 17 | Benzo[*b*]chrysene | 1.96 | – | 6.06 | 6.519 | 0.810 |
| 18 | Benzo[*b*]triphenylene | 2.20 | – | 7.163 | 7.088 | 0.998 |
| $R^2$ [(a)] | | | 0.978 | 0.899 | 0.904 | 0.968 |
| $R^2$ [(b)] | | | | 0.612 | 0.669 | 0.943 |

[(a)] Correlation coefficient for the first 9 compounds. [(b)] Correlation coefficient for the whole data set.

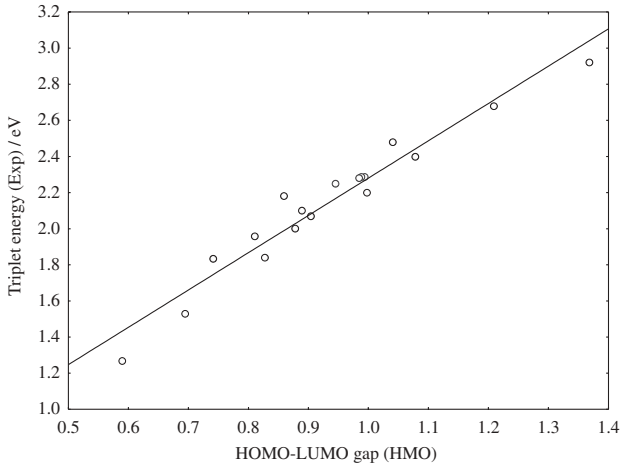



Figure 2. Linear plot of calculated (HMO) *vs.* experimental triplet excitation energies of the 18 PAHs.

that for CIS calculations is of 223 MB and for CISD/3–21G is 8034 MB. We want to call attention that we are not criticising the necessity of high level theoretical calculations in understanding chemical behaviour but the indiscriminate use of such calculations when the simplest ones produce the same results. As described earlier, Dewar[14] stated that the usefulness of any theoretical approach lies in its ability to reproduce experiments. Here it is obvious that the simple (graph-theoretical) HMO theory is able to reproduce the experimental excitation energies of PAHs to a high level of precision.

## THE PROBLEM OF THE TOXICITY OF PAHs

The main motivation in modelling the excitation energies of PAHs was the phototoxicity of these chemicals. Whilst it is known that most PAHs exert only narcotic toxicity at environmental concentrations, the toxicity of some of them is significantly increased by UV-photoactivation.[21–26] In analysing the phototoxicity of PAHs, Mekenyan *et al.*[10] and Betowski *et al.*[13] used the data produced by Newsted and Giesy in 1987.[27] In reality there is no linear correlation between the toxicity of these chemicals after photochemical activation and the energy of the singlet or triplet states. We have called here »toxicity after photoactivation« instead of phototoxicity with all the intention because when Newsted and Giesy[27] measured the toxicity of these chemicals they measured a global property of the biological system after the UV radiation was applied to it. However, this measurement (as many biological measurements) is not specific for the toxicity produced by the action of UV radiation but as a consequence of several possible mechanisms including the phototoxic one.[28]

If we plot the values of toxicity *versus* the HOMO--LUMO gap obtained by any of the theoretical methods used by Mekenyan *et al*.,[10] Betowski *et al*.[13] or the current work we obtain a graphic like that illustrated in Figure 3.

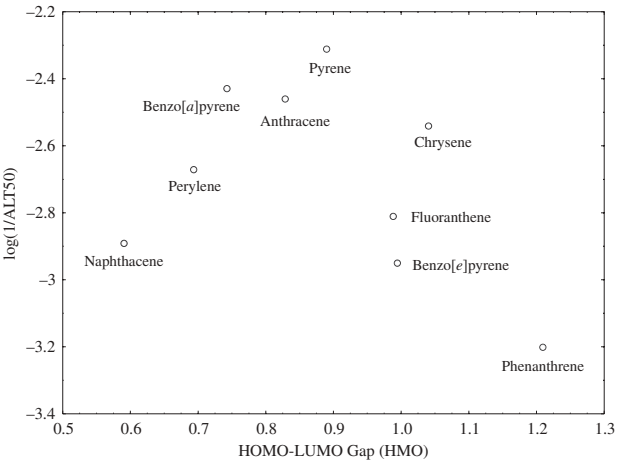
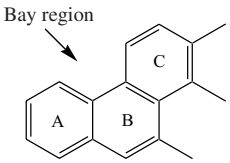



Figure 3. Typical plot of HOMO–LUMO gap versus toxicity after photoactivation of PAHs.

Some of the compounds in this limited data set, such as perylene, benz[a]pyrene, fluoranthene, benz[e]pyrene, phenanthrene and chrysene have a bay region as illustrated in the following scheme:




Scheme 1.

The importance of the presence of the bay region in some of these compounds can be understood by the role that this region plays in the toxicity of PAHs. Adducts are preferentially formed at this region since it provides an area of steric hindrance for detoxifying enzymes, whilst allowing oxidation to occur easily (see for instance Ref. 28, pp. 204–208).

Accordingly, it is probable that in measuring the toxicity of PAHs after photoactivation by UV radiation at least two competitive mechanisms are involved (we will go again to the problem of mechanisms later in this work):

*Mechanism 1.* Activation of the PAH molecule by UV light to the triplet state and the reaction of it with molecular oxygen giving rise to radical species which can react with macromolecules or tissues to produce the toxic effect.

*Mechanism 2.* Metabolic activation of PAH molecule by enzymes (mainly CYP1A1) to form electrophilic epoxides on the A ring, which are then transformed into diols. Formation of an epoxide in the bond closest to the bay region that can react with nucleic acid purines, *e.g.* guanine (see Figure 4), formation of diol-epoxides may result in carcinogenicity.

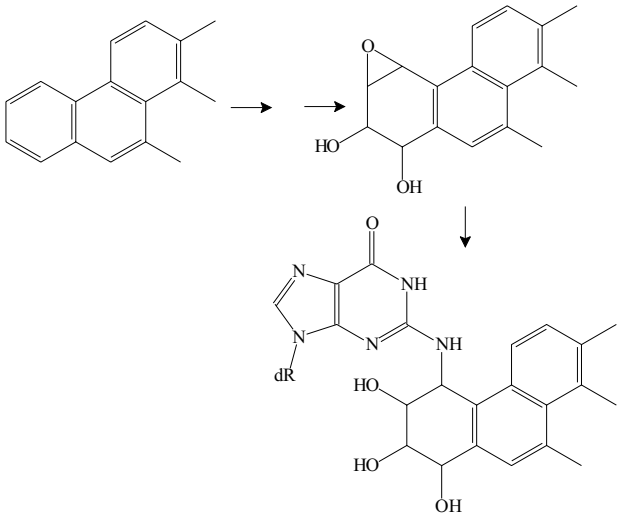



Figure 4. Mechanisms of DNA damage of PAHs after metabolic activation with participation of the bay region.

If this hypothesis is valid then it is impossible to expect a relationship between the toxicity and the HOMO--LUMO gap as the exclusive parameter. The relationship found by Newsted and Giesy[27] using the energy of the triplet state is non-linear and Mekenyan *et at*.[10] tried to solve this non-linearity by dividing the data set into two sub-sets. However, at the moment there is no clear evidence to confirm that only the photoactivation of the PAHs to the triplet state and the further transfer of this energy to oxygen generating radicals is responsible for the toxicity observed for these compounds. Consequently, the exclusion of other possible mechanisms such as the formation of epoxides is not justified at all, and the use of only one descriptor namely the HOMO--LUMO gap energy in the QSAR model appears not fully justified.

## QSPR/QSAR MODELS, OCCAM'S RAZOR AND »MECHANISTIC INTERPRETATIONS«

One of the current authors has previously stated that for the development of QSAR/QSPR models it is desired: [29]

»to have as many molecular descriptors as possible at your disposition but to include as few as possible into the model«.

The number of descriptors will guarantee coverage of the molecular structure space more efficiently than if limited to only a few descriptors, *e.g.* log *P*. The second part of this requisite is a sort of Occam's razor for maintaining the simplicity of the model under certain limits.

In its original form, the Occam's razor states that »*Numquam ponenda est pluritas sin necesitate*«, which can be translated as »Entities should not be multiplied beyond necessity«. In this case simplicity is loosely equated with the number of parameters in the model. If we understand predictive error to be the error rate for unseen examples, the Occam's razor can be stated for the selection of QSAR/QSPR models as:

*QSAR/QSPR Occam's Razor.* Given two QSAR/QSPR models with the same predictive error, the simpler one should be preferred because simplicity is desirable in itself.

This statement is domain specific and it does not deny the utility of the more complex models in other particular scenarios. For every domain where a simpler model is more accurate than a more complex one, there exists a domain where the reverse is true.[30] This rejects any argument about which model is preferable in general. Consequently, this denies the argument about which theoretical approach is preferable in general: empirical, semiempirical or *ab initio*. In the particular example of PAHs we have proved that the simpler, HMO approach produces similar error to the more complicated *ab initio* calculations and much better results than the semiempirical ones that use a large number of »fitting« parameters. According to the Occam's razor philosophy in this particular scenario it is of greater preference to use the HMO method than the other theoretical approaches. This kind of philosophy can be extended to the use of descriptors instead of approaches giving a general space for the use of the different types of molecular descriptors now in use in QSAR/QSPR.

The second point that requires some clarification is with respect to the »mechanistic interpretation« of the models developed. Some authors prefer some molecular descriptors or approaches to others because they claim that the first are more mechanistic than the second ones. Amongst those »mechanistic« descriptors that some authors prefer, the n-octanol/water partition coefficient (log *P*) is the star. There is no doubt about the importance that lipophilicity plays in the development of the biological activity as well as there is no doubt about the importance of electronic or steric effects. However, the use of the term mechanistic for these descriptors is not justified at all and it could reflect a lack of understanding of the proper term »mechanistic«. Firstly we will make an approach as to what we understand by mechanism.

*Mechanism.* A system of correlated parts working reciprocally together to give a final response.

In the particular case of a physicochemical process, *e.g.* partition between two phases, the final response is the physicochemical property, such as the log *P*, and the correlated parts are the set of inter- and intra-molecular interactions and physicochemical changes that take place during partition. Even in the simplest cases of physicochemical properties, the mechanisms are quite complex and their understanding is only possible at certain levels of approximation. When we refer to biological mechanisms we have to make the distinction between specific biological properties and global responses of organisms. The first are well exemplified by the chemical-protein interactions, such as enzyme inhibition. In this case, the understanding of the mechanism is reduced to the knowledge of the interactions between the chemical and the active site of the protein, as well as the changes that take place during these interactions. The global biological activity is referred here as those biological properties that are measured as the global response of an organism without considering the particular interactions of the xenobiotic with particular biological receptors. This is the case for the phototoxicity analysed here, other examples include skin sensitisation, carcinogenicity, and ecotoxicological endpoints such as $LD_{50}$ in different species, *etc*. In these cases the mechanism must include the penetration of the xenobiotics into organisms, their distribution, metabolism, general and specific interactions with organs, cells, receptors, *etc*. As it is obvious, a highly complex combination of processes.

Now, we can revisit the problem of designing some molecular descriptors or approaches as mechanistic ones. Using the example of log *P* and a global biological response that we will designate generically as Act. Suppose that we generated a QSAR model for Act as a function of log *P*: Act = $a + b \log P$. As we have assumed *a priori* that log *P* is a »mechanistic« variable we believe that we have a mechanistic interpretation of the biological response. However, if we try seriously in giving such interpretation we follow no simple difficulties. For instance, we have more than one possible mechanistic interpretation for the presence of log *P* in this QSAR model. Here are only three of them:

(i) the biological activity (Act) depends on the capacity of the xenobiotic to penetrate through lipophilic barriers, like the cell walls. The more lipophilic the compound is (because *b* is positive) the greater such capacity and higher the biological activity or

(ii) the biological activity (Act) depends on the capacity of the xenobiotic to interact with a hydrophobic pocket in a particular protein. The more lipophilic the compound is (because *b* is positive) the greater such capacity and higher the biological activity or

(iii) the biological activity (Act) depends on (i) and (ii).

As we can see at this point we have not gained any new insight about the system of correlated parts working reciprocally together to give the final biological response. Hence, there is no mechanistic interpretation for such model. The only thing that we can say is that the biological activity increases when the log *P* increases. The same can be said if the model were created through using molecular connectivity: the biological activity in-

creases (decreases) when molecular connectivity increases (decreases). In closing, log *P* is not more or less »mechanistic« that the connectivity index. *Ab initio* CIS/6–311G(d,p) HOMO-LUMO gap is not more or less mechanistic than the HMO one for the PAH compounds.

If there are no differences in terms of mechanistic interpretation of models between one and another descriptor or approach why do we read in the literature statements like »Topological indices, however, are difficult to interpret mechanistically and the use of other descriptors in QSAR is preferable«.[31] This statement from Cronin *et al.*[31] is not only false but lacks any fundamental basis. It is the consequence of the wrong application of the Occam's razor philosophy. The same that can conduce to deny general relativity because it makes more assumptions that Newton's gravitational law, and it is far more complex. The practioners of this wrong application of the Occam's razor can find quite illustrative that the model of a flat earth is preferable to that of a spherical one because the first is a linear model, while the second is quadratic and no better at explaining everyday observations in the Middle Ages.

## CONCLUSIONS

There are two straightforward conclusions from the current work. The first is related to the use of theoretical approaches in Chemistry. The selection of one theoretical approach or another should be carried out only on the basis of the utility of the model developed using it. Considering that all approaches used in chemistry are empirical, their unique value resides in their practical correspondence with the chemical reality. Hence, the selection of one approach over another should be based on the predictability and simplicity of the model developed. Sometimes simple approaches give results better or as good as more sophisticated ones, saving time and effort and producing the same results more simply.

The second lesson from the current work concerns the »mechanistic QSAR« approaches. The biological response of one organism to a xenobiotic is a very complex function that depends on several mechanisms of distribution, metabolism, and interactions of this chemical in the biological system. The assumption of one particular mechanism should be based on a careful selection of the biological experiment carried out in order to guarantee that this particular mechanism is the only one determining the biological response. In many cases what is being measured experimentally is the result of an amalgamation of different biological processes condensed in one numerical value that we call the biological response (activity, toxicity, *etc*.). In this context, the selection of one physicochemical variable *a priori* when dealing with one particular mechanism is very risky as we can ignore important variables that influence the other mechanisms participating in such a response. Consequently, considering the biological system as a black box and using several physicochemical, molecular and structural descriptors to derive the QSAR is a better and less risky choice in this case. The interpretation of such model *a posteriori* by considering several of the possible mechanisms can give rise to some important insights about the way in which the biological response is produced. However, this is a difficult task as in many cases not all the mechanisms are known. On the other hand, in most cases the same variables in the QSAR are describing different mechanisms simultaneously, making it difficult to interpret their role in the model.

*Acknowledgements*. – The authors want to thank the anonymous referees for their comments specially calling our attention to the Occam's razor and »mechanistic interpretations« in QSAR/QSPR models.

## REFERENCES

1. W. Karcher, J. Devillers, P. H. Garrigues, and J. Jacob, *Spectral Atlas of Polycyclic Aromatic Compounds*, Kluwer Academic Publishers, Dordrecht, 1991.
2. J. M. Neff, *Polycyclic Aromatic Compounds in the Aquatic Environment*, Applied Science Publishers, London, 1979.
3. Z. G. Zhou, Q. H. Dai, and T. Gui, *J. Chem. Inf. Comput. Sci.* **43** (2003) 615–621.
4. J. Devillers, *Polycyclic. Aromat. Comp.* **18** (2000) 231–241.
5. A. Gallegos, D. Robert, X. Girones and R. Carbo-Dorca, *J. Comput. Aided Mol. Des.* **15** (2001) 67–80.
6. B. D. Gute, G. D. Grunwald, and S. C. Basak, *SAR. QSAR. Environ. Res.* **10** (1999) 1–15.
7. M. M. C. Ferreira, *Chemosphere* **44** (2001) 125–146.
8. S. Markovic, Z. Markovic, J. P. Engelbrecht, and R. I. Mc Crindle, *J. Chem. Inf. Comput. Sci.* **42** (2002) 82–86.
9. A. Kitti, M. Harju, M. Tysklind, and B. van Bavel, *Chemosphere* **50** (2003) 627–637.
10. O. G. Mekenyan, G. T. Ankley, G. D. Veith, and D. J. Call, *Chemosphere*, **28** (1994) 567–582.
11. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, *J. Am. Chem. Soc*. **107** (1985) 3902–3909.
12. J. J. P. Stewart, *J. Comput. Chem*. **10** (1989) 209–220.
13. L. D. Betowski, M. Enslow, and L. Riddick, *Comput. Chem.* **26** (2002) 371–377.
14. M. J. S. Dewar, *J. Phys. Chem.* **89** (1985) 2145–2150 (and related cited references).
15. N. Trinajstić, *Chemical Graph Theory*, CRC Press, Boca Raton, 2nd ed., 1992.
16. I. Gutman and O. E. Polansky. *Mathematical Concepts in Organic Chemistry*, Springer-Verlag, Berlin, 1986.
17. E. Estrada and E. Uriate, *Curr. Med. Chem.* **8** (2001) 1573–1588.
18. J. Devillers and A. T. Balaban, *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon and Breach, Amsterdam, 1999.
19. N. Bohr and J. A. Wheeler, *Phys. Rev.* **56** (1939) 426–50.

20. A. Streitwiesen Jr, *Molecular Orbital Theory for Organic Chemists*, Wiley, New York, 1961.
21. G. T. Ankley, R. J. Erickson, B. R. Sheedy, P. A. Kosian, V. R. Mattson, and J. S Cox, *Aquat. Toxicol.* **37** (1997) 37–50.
22. B. L. Boese, J. O. Lamberson, R. C. Swartz, and R. J. Ozretich, *Arch. Environ. Contam. Toxicol.* **32** (1997) 389–393.
23. X-D. Huang, D. G. Dixon, and B. M. Greenberg, *Ecotoxicol. Environ. Saf.* **32** (1995) 194–200.
24. P. A. Kosian, E. A. Makynen, P. D. Monson, D. R. Mount, A. Spacie, O. G. Mekenyan, and G. T. Ankley, *Environ. Tox. Chem*. **17** (1998) 1021–1033.
25. M. C. Pelletier, K. T. Burgess, A. Ho, R. Kuhn, A. McKinney, and S. A. Ryba, *Environ. Toxicol. Chem*. **16** (1997) 2190–2199.
26. R. L. Spehar, S. Poucher, L. T. Brooke, D. J. Hansen, D. Champlin, and D. A. Cox. *Arch. Environ. Contam. Toxicol.* **37** (1999) 496–502.
27. J. L. Newsted and J. P Giesey, *Environ. Toxicol. Chem*. **6** (1987) 445–461.
28. U. A Boelsterli, *Mechanistic Toxicology, the Molecular Basis of How Chemicals Disrupt Biological Targets*, Taylor and Francis, London and New York, 2003.
29. E. Estrada, in: J. Devillers and A. T. Balaban (Eds.), *Topological Indices and Related Descriptors in QSAR and QSPR*. Gordon and Breach, Amsterdam, 1999, pp. 403–453.
30. P. Domingos, *Data Min. Know. Discov.* **3** (1999) 409–425.
31. G. P. Moss, J. C. Dearden, H. Patel and M. T. D. Cronin, *Tox. in Vitro* **16** (2002) 299–317.

## SAŽETAK

### O uporabivosti graf teorijskih opisivača u predviđanju teorijskih parametara. Fototoksičnost policikličkih aromatičkih ugljikovodika

**Ernesto Estrada i Grace Patlewicz**

Razmatrana su dva glavna aspekta teorijskoga pristupa predviđanju biologijskih svojstava molekula. Prvi se aspekt odnosi na prikladnost teorijskih metoda i na krivi nazor da sofisticiranije metode daju bolje rezultate. Autori su pokazali da jednostavna HMO metoda u graf teorijskoj formulaciji daje rezultate koji su jednako dobri kao *ab initio* računi visoke razine, ali sa znantno manjim utroškom vremena. Drugi se aspekt odnosi na ispitivanje izvedivosti uporabe *a priori* mehanističkih ideja u pokušaju selektiranja varijabli za uporabu u QSAR modeliranju. Međutim, kako većina biologijskih svojstava nije specifična za neki pojedini mehanizam, onda ovaj drugi pristup može rezultirati u neprihvatljivim predviđanjima. U takvim slučajevima, kada je biologijsko svojstvo odraz više mehanizama koji uključuju distribuciju i metabolizam, prihvatljivije je rabiti pristup crne kutije i *a posteriori* interpretirati dobivene rezultate. Svi su ovi pristupi analizirani pomoću toksičnosti policikličkih aromatičkih ugljikovodika nakon fotoaktivacije s UV zračenjem.