Journal of Computer-Aided Molecular Design, **15:** 67–80, 2001.
KLUWER/ESCOM

Structure-toxicity relationships of polycyclic aromatic hydrocarbons using molecular quantum similarity

Ana Gallegos, David Robert, Xavier Gironés & Ramon Carbó-Dorca*
Institute of Computational Chemistry, University of Girona, Campus Montilivi, E-17071 Girona, Catalonia, Spain

Received 27 March 2000; Accepted 7 July 2000

Key words: atomic shell approximation (ASA) density functions, carcinogenic activity, molecular quantum similarity measures (MQSM), percutaneous absorption, polycyclic aromatic hydrocarbons (PAHs), quantitative structure-activity relationships (QSAR)

Summary

The establishment of quantitative structure-activity relationship (QSAR) models for the toxicity of polycylic aromatic hydrocarbons (PAHs) is described. Two properties, *in vitro* percutaneous absorption in rat skin and discrete levels of carcinogenic activity, are examined using molecular quantum similarity measures (MQSM). The results show that MQSM produces comparable, or even better, results than other approaches using physicochemical, topological and quantum-chemical molecular descriptors. Furthermore, a careful analysis puts into evidence that most of the information characterized by the original descriptors is in fact contained in the molecular density functions, the basis of MQSM. The present paper, together with several other reported by our laboratory, proves that MQSM might be appropriate theoretical tools for QSAR and computer-aided drug design, comparable to other highly predictive QSAR methodologies.

Introduction

Polycyclic aromatic hydrocarbons (PAHs) are present throughout the environment in air, water and soil [1]. PAHs are produced in different incomplete combustion processes: waste incineration [2], coal tar [3], biomass burning [4], automobile exhaust [5] and cigarette smoke. There are diverse routes of exposure for humans and animals: orally, by inhalation and through skin contact. There exists early documentation of appearance of scrotal cancer diseases in chimney sweeps, supposed to be related to PAHs [6]. Later, animal studies under controlled conditions provided the majority of the evidence for PAHs mutagenic and carcinogenic properties [7]. As a result, PAHs have become an important problem for human health, and all the current risk assessment programs include information about these compounds [8].

*To whom correspondence should be addressed. E-mail: director@iqc.udg.es

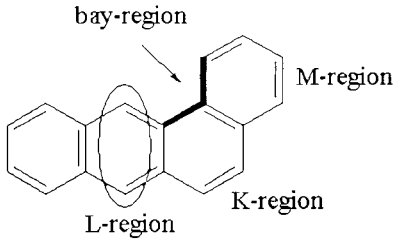

Figure 1. The K, L, M and bay regions of a polycyclic aromatic hydrocarbon: benzo[*a*]anthracene.

Although PAHs structure is rather homogeneous, their carcinogenic effect ranges from highly active to inactive [9]. Dipple [10] proposed a classification into four classes according to their power: inactive, slight, moderate and high. A great effort was made to understand the reaction mechanism and connect the molecular structure to its toxic effect. A consistent explanation for the carcinogenic action of many PAHs was given by the so-called K-L-M-'bay region' theory. Figure 1 shows a graphical example of these region locations for a particular PAH. In 1939, Schmidt [11]

suggested that PAHs carcinogenicity was related to the distribution of π-electrons. Within Schmidt's hypothesis, PAHs with high π-electron density in the L-region are strongly carcinogenic. Pullman and Pullman [12] pointed out the coupling of a high π-electron density in the K-region and low density in the L-region to be determinant for a high carcinogenicity. Later, Gayoso and Kimri [13] suggested that the M region could replace the K region for an improvement of the structure-carcinogenic activity relationships. This theory decisively evolved by the contribution of Jerina [14], who found stabilization of the 'bay region' cation to be relevant for carcinogenicity. Thus, the diol epoxide formed at the 'bay region' was established as the main factor responsible for the toxic action, since it is believed to directly react with a nucleic base and give rise to cancer [15]. Other different approaches made use of the 'bay region' considerations [16–18].

Once the decisive molecular features for activity were identified, several quantities were proposed to characterize PAHs. Quantitative structure-activity relationships (QSAR) were then constructed to estimate PAHs carcinogenic power, using different empirical and theoretical methods [19–24]. In the present work, an alternative perspective is provided, consisting of the use of the *molecular quantum similarity* theory [25]. The basis for an effective application of any molecular similarity technique to QSAR rests on the underlying assumption that *similar* molecules will possess *similar* properties. Quantum similarity provides a suitable quantification of the resemblance between two molecular structures, based on quantum-mechanical principles. Carbó et al. [26] demonstrated the formal connection between quantum similarity and QSAR via the discretization of the expectation value rule. Since then, this theory has been successfully applied to construct QSAR models, both in pharmacological [27] and toxicological [28] environments. Together with QSAR, quantum similarity has been applied to a great number of problems: determination of molecular alignment [29], chemical reactivity [30], detection of chirality [31], comparison of quantum-chemical basis sets [32] and comparison of computational methodologies [33]. An up-to-date review of the theory can be found in different papers [34].

Methods

Molecular quantum similarity measures (MQSM)

Quantum similarity attempts to give a quantitative similarity measure between two molecular structures, based on Quantum Mechanical ideas. Concretely, a *molecular quantum similarity measure* (MQSM) [25] can be defined as the scalar product between the first-order molecular density functions of the two compared molecules, weighted by a non-differential positive definite operator:

$$Z_{AB} = \iint \rho_A(\mathbf{r}_1)\Omega(\mathbf{r}_1,\mathbf{r}_2)\rho_B(\mathbf{r}_2)\, d\mathbf{r}_1\, d\mathbf{r}_2 \quad (1)$$

According to the form of this operator, different types of MQSM can be defined. In particular, two kinds of MQSM have been used in the present study: the so-called *overlap MQSM* [25], in which the Ω operator is a Dirac's delta distribution:

$$Z_{AB} = \iint \rho_A(\mathbf{r}_1)\delta(\mathbf{r}_1-\mathbf{r}_2)\rho_B(\mathbf{r}_2)\, d\mathbf{r}_1\, d\mathbf{r}_2 = \int \rho_A(\mathbf{r})\rho_B(\mathbf{r})\, d\mathbf{r} \quad (2)$$

On the other hand, if the Coulomb operator $\Omega = |\mathbf{r}_1 - \mathbf{r}_2|^{-1}$ is employed, this yields the *Coulomb MQSM* [35]:

$$Z_{AB} = \iint \rho_A(\mathbf{r}_1)\frac{1}{|\mathbf{r}_1-\mathbf{r}_2|}\rho_B(\mathbf{r}_2)\, d\mathbf{r}_1\, d\mathbf{r}_2 \quad (3)$$

The overall set of pairwise MQSM for a given molecular family can be expressed in matrix form: $\mathbf{Z}=\{ZAB\}$. Quantum similarity matrices will be used as a source of QSAR descriptors.

Computation of the molecular density functions and determination of the molecular alignment

In order to avoid computationally expensive calculations, fitted first-order density functions will be used instead of ab initio ones. The *Promolecular Atomic Shell Approximation* (ASA) algorithm [36] has been employed to obtain accurate fitted molecular density functions. This approach considers the molecular electron density as a sum of atomic density contributions. The atomic densities are taken as a linear combination of 1*S* Gaussian functions, which are fitted to an atomic ab initio calculation with a 3-21G basis set [36a]. Once the fitted atomic densities are computed, they are stored in a database, and the molecular densities are then built simply by conveniently adding these

elementary pieces. Coefficients and exponents for the 3-21G promolecular ASA fitting can be found on the IQC Website [37]. The differences in MQSM built ab initio and ASA densities are less than 2%, so the use of this procedure is clearly justified. Molecular geometries have been optimized at the semiempirical AM1 level [38], using the AMPAC 6.55 software [39]. The particularly rigid structure of the examined compounds, made of series of joined benzene rings, made further conformational analyses unnecessary.

As can be deduced from Equations 1–3, MQSM depends on the relative orientation of the molecules compared. Thus, as in other 3D QSAR methodologies, an alignment criterion needs to be chosen in order to overlay the structures appropriately. In the present work, two molecules are considered to be optimally aligned if the MQSM integral value is maximal. Different algorithms have been proposed to maximize MQSM [40], and the one proposed by Constans et al. [40a] has been employed here. The MOLSIMIL-97 software [41], developed in our laboratory, was used to perform the density fitting, the matching procedure and the MQSM computations.

Treatment of quantum similarity matrices

Once the quantum similarity matrices are computed, a transformation is necessary to reduce the dimensionality of the data. The method chosen here was *classical scaling* [42], which considers the molecules as points in a low-dimensional Euclidean space and then finds coordinates for these points in such a way that the interpoint distances match the original similarities as good as possible. More details on the algorithm can be found in Reference 42. Molecular coordinates in the new multidimensional space are known as the *principal coordinates* (PCs) of the system. Classical scaling produces a reorganization of the information encoded by the similarity matrices, in which the data variance explained by each PC is associated to the corresponding eigenvalue. In order to avoid the participation of PCs related to the transformation background noise, all those PCs accounting for less than 1% variance were neglected, and therefore the matrix dimensionality is effectively reduced.

From all the surviving PCs, those have to be chosen which best correlate with the external data. To do this, the *most predictive variables method* (MPVM) [43] was used. This variable selection technique defines a correlation coefficient between each PC and the molecular property, and arranges the PCs according to this index. Finally, the connection between the most descriptive PCs and the property analyzed is carried out by means of a multilinear regression. The statistical parameters computed to assess the model accuracy are the conventional correlation coefficient r^2 and the leave-one-out cross-validation coefficient q^2.

In order to detect possible chance correlations or excess of parameters, the so-called *randomization test* was adopted [44]. In this statistical technique, the property data are randomly permuted in their positions. Then, new predictive models are built for the altered responses. If a real structure-activity relationship exists, only the correctly ordered variables should led to significant results. Otherwise, if the QSAR equations are able to correlate randomized data, the model is suspected of overparametrization. The statistical treatment was carried out using the TQSAR-SIM software [45], developed in our laboratory.

QSAR for estimating the percutaneous absortion of PAHs

Human exposure to PAHs, present in mineral oils, coal-tars and derivatives, mainly comes from skin contact. For this reason, evaluation of dermal penetration of these materials is of great importance in hazard assessment programs. Several *in vivo* and *in vitro* experiments have been performed, designing specific measurement methods for these products [46]. The possibility of establishing simple and accurate mathematical relationships between the PAHs structure and their percutaneous absorption can avoid expensive biological assays and become an alternative to animal testing. In this section, QSAR equations built from MQSM for a PAHs series are presented as a suitable option, with comparable results to classical physicochemical descriptors.

Molecular set and experimental data

The molecular set used in the first application example consists of 60 commercially available PAHs, with a structure of 3 up to 7 fused rings. Percutaneous absorption was measured *in vitro* on rat skin sections of about 350 μm. Property values are expressed as the percentage of applied dose (PADA) penetrating the skin 24 h after the application of the dosing solutions (40 nmol per cm^2 skin surface). One of the advantages of *in vitro* analytical methods consists of avoiding the use of radiotracers, which are expensive and of limited availability. More details on the technical aspects

Table 1. Structures and dermal penetration (PADA) values (%) for 60 PAHs

No.	Compound	PADA (%)	No.	Compound	PADA (%)
1	Coronene	0.7	31	3-Ethylfluoranthene	20
2	Dibenzo[*a*,*l*]pyrene	2	32	Triphenylene	20
3	9,10-Diphenylanthracene	6	33	7,8,9,10-Tetrahydroacephenanthrene	20
4	Perylene	7	34	2,3-Benztriphenylene	20
5	Dibenzo[*a*,*i*]pyrene	8	35	Benzo[*c*]phenanthrene	20
6	3-Methylcholanthrene	8	36	1-Methylpyrene	22
7	9-Benzylidenefluorene	8	37	3,9-Dimethylbenz[*a*]anthracene	24
8	7,10-Dimethylbenzo(*a*)pyrene	8.3	38	2,3-Benzofluorene	25
9	Indeno(1,2,3-*cd*)pyrene	9	39	1,2-Benzofluorene	25
10	Dibenz[*a*,*h*]anthracene	9.4	40	9-Benzylfluorene	26
11	Benzo[*e*]pyrene	10	41	9-m-Tolylfluorene	29
12	Benzo[*g*,*h*,*i*]perylene	10	42	Pyrene	30
13	9-p-Tolylfluorene	10	43	2-Ethylanthracene	30
14	6-Ethylchrysene	10	44	10-Methylbenzo[*a*]pyrene	32
15	9-Cynnamylfluorene	11	45	1-Methylanthracene	33
16	6-Methylbenz[*a*]anthracene	14	46	2-Methylfluoranthene	33
17	Benzo[*k*]fluoranthene	14	47	3,6-Dimethylphenanthrene	33
18	Benzo[*a*]pyrene	15	48	Benzo[*a*]anthracene	35
19	3-Ethylpyrene	18	49	Fluorene	36
20	1-Methyl-7-isopropylphenanthrene	20	50	2-Methylphenanthrene	38
21	2-(tert-Butyl)anthracene	20	51	9-Ethylfluorene	38
22	9-Phenylanthracene	20	52	1-Methylphenanthrene	40
23	3-Methylcholanthrene	20	53	9,10-Dihydrophenanthrene	40
24	10-Methylbenz[*a*]anthracene	20	54	9-Vinylanthracene	40
25	5-Methylbenz[*a*]anthracene	20	55	Anthracene	42
26	9,10-Dihydroanthracene	20	56	Fluoranthene	42
27	9-Phenylfluorene	20	57	1-Methylfluorene	49
28	1,2,3,6,7,8-Hexahydropyrene	20	58	2-Methylanthracene	50
29	n-Butylpyrene	20	59	4H-Cyclopenta(*d*,*e*,*f*)phenanthrene	50
30	5,6-Dihydro-4H-dibenz[*a*,*k*,*l*]anthracene	20	60	Phenanthrene	50

of the measurement experiments can be found in the work of Roy et al. [47]. PAHs with 1 or 2 aromatic rings were not considered because of the difficulty in measuring their dermal penetration, due to their volatility and loss from the skin surface during the biological assays. For obvious reasons, interest was primarily focused on the carcinogenic PAHs, mainly comprising 4–6 ring structures. The concrete molecular set and the activity values are given in Table 1. It must be noted that some errors in the chemicals' nomenclature were noticed and amended: see original tables for comparison. The structures and molecular coordinates can be seen and downloaded from the IQC Website [48].

Results and discussion

Following the protocol exposed in the previous section, the most predictive classical scaling solution of the quantum similarity matrices was correlated with PAHs activity. First, calculations employing both overlap and Coulomb MQSM were performed on this molecular set. When analyzing the results, it was observed that the quality of the models derived with Coulomb MQSM were notably better than those built employing overlap MQSM, which did not achieve statistically significant results. A plausible hypothesis to explain this situation is connected to the particular structure of the studied PAHs. The presence of 5-

Table 2. QSAR models for the dermal penetration of 60 PAHs. The optimal model (3 PCs) has been marked in bold face

No. PCs	Selected PCs	r^2	q^2
1	1	0.589	0.560
2	1, 2	0.659	0.615
3	**1, 2, 13**	**0.684**	**0.634**
4	1, 2, 13, 6	0.693	0.637
5	1, 2, 13, 6, 5	0.699	0.632

membered rings and internal single bonds within the molecular structures hinders an exact intermolecular atom-atom matching. Overlap MQSM is highly sensitive to exact atom superpositions, and the fast decrease of density peaks makes the similarity contributions of close atoms almost negligible [34c]. On the other hand, the Coulomb operator acts as an attenuation factor for density distributions, and the contribution of not exactly superposed atoms to the total MQSM is more important. Subsequently, all models presented in this section correspond to Coulomb-like MQSM.

The Coulomb quantum similarity matrix for the 60 PAHs was transformed utilizing the previously exposed chemometric tools. The classical scaling solution evidenced that the distribution of the explained variance by the different PCs is quite smooth, and 13 PCs are necessary to include a 90% data variation. As a result, there are several PCs explaining a very small part of the total variance. In particular, only the first 17 PCs accounted for eigenvalues higher than 1%. Those axes were the only ones used in the variable selection analysis, discarding the remaining 43.

The statistical coefficients for the resulting QSAR models are shown in Table 2. As can be seen, MQSM produce acceptable correlation and prediction results, with q^2 about 0.6 in all cases. Considering the evolution of r^2 and q^2 coefficients, the 3 PCs equation has been taken as the optimal QSAR model for this family, leading to $r^2 = 0.684$ and $q^2 = 0.634$. There are higher-order models with better predictive capabilities, but the improvement achieved, in terms of q^2, when adding more parameters to the regression does not justify the increase in the equation complexity.

As can be seen, the most descriptive PCs are not those accounting for the maximal variance. In particular, the third most predictive axis is the 13th, explaining only 1.8% of the total data variation. However, the variance explained by the 'best-fitting' three-

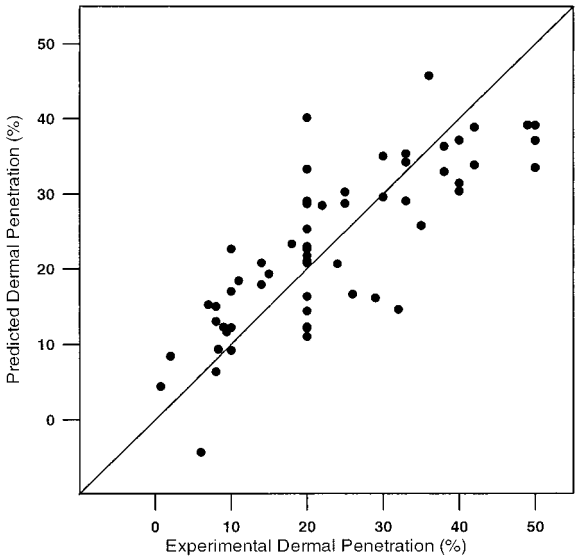

Figure 2. Cross-validated versus experimental percutaneous absorption values.

dimensional subspace, made up of PCs {1, 2, 3}, is 51.4%, differing only into 9.3% with regard of the {1, 2, 13} subspace: 42.1%. Thus, the great increase in the predictivity of the model is compatible with a variance of the same order, which justifies the use of variable selection techniques. In spite of the low variance explained by any three-dimensional subspace, it is enough to generate satisfactory prediction models.

The resulting equation obtained with the regression treatment for the 3 PCs model is:

$$\mathbf{y} = -1.263\mathbf{x}_1 + 0.489\mathbf{x}_2 + 0.914\mathbf{x}_{13} + 23.173 \quad (4)$$

Figure 2 shows a representation of the cross-validated versus the experimental percutaneous absorption values. As this plot evidences, a possible cause for the poor results in correlation and prediction can be attributed to the particular distribution of the experimental data. Except for three PAHs, the measurement of the dermal penetration is expressed by an integer number, precluding the appreciation of slight differences within the studied compound activity. For instance, there are 16 molecules (PAHs **20**–**35**) with the same PADA value: 20%. Most of these compounds possess the highest residuals in the model. Furthermore, there is a single molecule, 9,10-diphenylanthracene (**3**), which is predicted to have a negative dermal penetration, which is obviously a lack in the least-squares procedure used to solve the multilinear equation system.

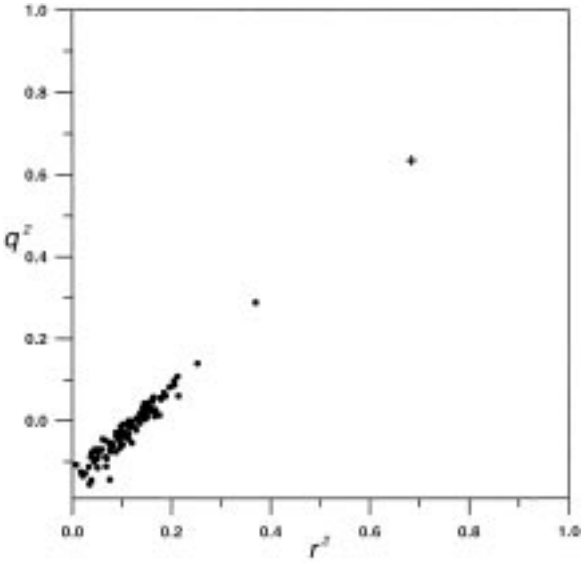

Figure 3. Randomization test for the optimal model. The randomized responses (100) have been marked with circles, and the correctly ordered activity has been marked with a cross.

As was discussed before, chance correlations were examined with the randomization test. The results of this test are given in Figure 3. As is clearly appreciated, a separation exists between the statistical coefficients belonging to real data and those associated with randomized properties. Even though correlation and prediction results are not very good, it is important to remark that those corresponding to altered data do not reach statistically significant levels, with $q^2<0.3$ for all cases. This fact ensures that a real structure-property relationship has been found, connecting the percutaneous absorption of the studied PAHs with changes in the electron distributions, quantified by means of MQSM.

This molecular data set was studied by other authors [47, 49]. Roy et al. [47], who experimentally measured the dermal penetration for this set, also reported a QSAR study using different kinds of molecular descriptors in a Hansch-type analysis [50]. Starting from a pool of 50 physicochemical, topological, geometric, electronic and chromatographic parameters, a study to detect collinearities was carried out using the ADAPT program [51], from which only 18 descriptors remained. The optimal model used only two parameters: the octanol-water partition coefficient, log P, and a shape descriptor defined as the two-dimensional projections of the molecule onto one of its planes normalized for its size. This QSAR equation yielded a $r^2 = 0.64$ value.

Later, Gute et al. [49] reported a similar study in which topostructural, topochemical, geometric, quantum-chemical and physicochemical indices were used to characterize the molecular features. A hierarchical approach was employed to select the most predictive descriptors, and the optimal model was found to be formed by a unique parameter, of topochemical type: the first-order bond path connectivity index $^1X^b$. This index accounts for general size and bonding patterns within the molecules. A $r^2 = 0.695$ value was achieved. Adding other indices to the QSAR equation did not significantly improve the description.

As can be appreciated by comparing r^2 values, the three different models are equally effective in correlating dermal penetration for PAHs. This fact is not surprising, keeping in mind that the basis of MQSM is the first-order density function, which is a physical observable characterizing the electron distribution of the molecules. The parameters found to be responsible for the activity differences quantify molecular size, shape and hydrophobicity. Both size and shape description are obviously contained in the molecular electron densities. In addition, it has been proved that for highly homogeneous series, log P can be accurately characterized by MQSM [27d].

Finally, it must be stated that other QSAR studies on dermal penetration of different series of PAHs have been reported, dealing with different molecular descriptors. For instance, Potts and Guy [52] proposed a model using molecular weight, molar volume and log P, yielding comparable results. This confidence level, about 60–70%, seems to be an upper bound for most of the QSAR analyses on skin penetration. Even though size and shape descriptors characterize this family reasonably well, some additional information, still missing, is required to exceed this limit.

QSAR for estimating the carcinogenic activity of PAHs

As has been commented in the Introduction, historically the derivation of molecular descriptors for structure-toxicity relationships in PAHs was strongly influenced by the K, L, M and 'bay region' concepts. Thus, the stabilization of the 'bay region' cation was characterized by several theoretical reactivity indices [53], and different descriptors provided information on the electronic structure of the relevant rings and regions, such as the ionization potential [54], the averaged ^{13}C chemical shifts [55], etc.

Recently, Barone et al. [56, 57] presented a novel methodology based on the concept of *local density of states* (LDOS), which employed the electron density of the ring with the highest bond order, plus some molecular energy levels computed using the Hückel method [58], to identify if a given PAH molecule would present a carcinogenic activity or not. This theory can be applied even to those PAHs that do not contain K, L, M or 'bay regions'. Vendrame et al. [59] utilized this theory to explore the structure-toxicity relationships for the same series of PAHs, treated with more sophisticated statistical tools.

Some parallelism can be established between LDOS theory and quantum similarity, because the latter is nothing but a *global* density-based approach, which makes use of the whole electron density to derive the QSAR parameters instead of local densities computed on specific molecular regions. This is in better agreement with Quantum Mechanics, in which the delimitation of molecular regions or fragments is not univocally defined. The results presented in this section will confirm that the relevant information of local density is in fact contained in the whole electron distribution, and can be retrieved using quantum similarity techniques and linear multivariate analysis tools.

Molecular set and experimental data

The molecular set studied in this case example is formed by 78 PAHs. Following structural criteria, it was divided into two subsets attending to the presence or absence of methyl substitutions. In this sense, a first class made up of 32 non-methylated PAHs and a second class of 46 methylated PAHs were analyzed. Afterwards, the whole set of 78 PAHs was also studied. This set of compounds has been taken from the work of Vendrame et al. [59], who reported a QSAR study using molecular descriptors based on LDOS theory, dealt with PCA and neural networks. Results from MQSM will be compared to those achieved by the mentioned authors. Due to the fact that no experimental data measured under the same conditions is available, the discrete classification in classes was used. Because of the difficulty in correlating PAHs according to Dipple's 4-classes classification [10], the carcinogenic activity was reduced here to only two classes: inactive (I) and active (A). Experimental data used to define activity classes come from the Iball [60] and Badger [61] experimental indices, as well as from the Cavalieri et al. scale [54]. A compound was considered to be noncarcinogenic if it was classified as inactive or slightly active with the Iball and Badger indices; otherwise, it was assumed to be carcinogenic. This division does not coincide with the one proposed by Villemin et al. [23], in which only the inactive PAHs were considered to be noncarcinogenic, hence some PAHs are differently classified in both approaches. Vendrame's classification criterion has been adopted here because of confidence reasons. Experimental carcinogenic data come from studies of two decades ago; at that time, experimental techniques were not reliable enough to ensure that a very low activity value was not caused by uncontrolled external effects, such as impurities in the solution. A very strict rule such as Villemin's might tend to overrate toxic effects of many PAHs.

The concrete molecular set and the carcinogenic activity are shown in Table 3. Some errors have been detected in this family as well, which have been amended in the calculations. Corrected structures and molecular coordinates can be found on the IQC Website [48].

Preliminary statistical considerations

A fundamental difference from the previous case must be stressed. The property values in this example do not correspond to real numbers which can be entered in a common regression analysis, but to two categories related to the molecular action: active-inactive. In order to perform a QSAR study for this property, real numbers {1, 0} were assigned to the active and inactive categories, respectively. Although these values have been arbitrarily chosen, it must be noted that any other pair of numbers could be expressed as a linear transformation of these ones, and subsequently, the quality of the model remains unchanged. It is well known that shifts or dilation of the whole set of data do not affect r^2 and q^2 coefficients. Attaching {1,0} conveniently to the experimental data, the multilinear regression is carried out exactly as in the previous case. However, the adjusted and cross-validated activities are dealt with in a different manner: they are assigned to a category according to a pre-established threshold. A value of 0.5, the arithmetic mean of the two reference values, has been taken here as this limit. Like any other choice of this kind, the adjustment process can present some unavoidable decision conflicts for those molecules placed close to the frontier threshold.

The quality of the models was quantified by discrete indices substituting the continuous r^2 and q^2

Table 3. Structures and carcinogenic activities for 78 PAHs. A: Active; I: Inactive; NA: not available

No.	Compound	Act	No.	Compound	Act
Non-methylated PAHs					
1	Dibenzo[3,4:9,10]pyrene	A	17	Benzo[1,2]pyrene	I
2	Benzo[3,4]pyrene	A	18	Phenanthrene	I
3	Dibenzo[3,4:8,9]pyrene	A	19	Triphenylene	I
4	Dibenzo[3,4:6,7]pyrene	A	20	Benzo[1,2]naphthacene	I
5	Dibenzo[1,2:3,4]pyrene	A	21	Dibenzo[3,4:5,6]phenanthrene	I
6	Naphtho[2,3:3,4]pyrene	A	22	Picene	I
7	Dibenzo[1,2:5,6]anthracene	A	23	Tribenzo[1,2:3,4:5,6]anthracene	I
8	Tribenzo[3,4:6,7:8,9]pyrene	A	24	Dibenzo[1,2:5,6]pyrene	I
9	Dibenzo[1,2:3,4]phenanthrene	A	25	Phenanthra[2,3:1,2]anthracene	I
10	Tribenzo[3,4:6,7:9,10]pyrene	A	26	Benzo[1,2]pentacene	I
11	Dibenzo[1,2:5,6]phenanthrene	I	27	Anthanthrene	I
12	Benzo[1,2]anthracene	I	28	Benzene	I
13	Chrysene	I	29	Naphthalene	I
14	Benzo[3,4]phenanthrene	I	30	Pyrene	I
15	Dibenzo[1,2:7,8]anthracene	I	31	Benzo[*ghi*]perylene	I
16	Dibenzo[1,2:3,4]anthracene	I	32	Coronene	I
Methylated PAHs					
33	7,12-Dimethylbenz[*a*]anthracene	A	58	3-Methylbenzo[*c*]phenanthrene	I
34	6,12-Dimethylbenz[*a*]anthracene	A	59	6-Methylbenzo[*c*]phenanthrene	I
35	6,8,12-Trimethylbenz[*a*]anthracene	A	60	6-Methylbenz[*a*]anthracene	I
36	2-Methylbenzo[*a*]pyrene	A	61	12-Methylbenz[*a*]anthracene	I
37	4-Methylbenzo[*a*]pyrene	A	62	6-Methylanthanthrene	I
38	11-Methylbenzo[*a*]pyrene	A	63	6,12-Dimethylanthanthrene	I
39	12-Methylbenzo[*a*]pyrene	A	64	1-Methylbenzo[*c*]phenanthrene	I
40	1-Methylbenzo[*a*]pyrene	A	65	2-Methylbenzo[*c*]phenanthrene	I
41	4,5-Dimethylbenzo[*a*]pyrene	A	66	10-Methylbenzo[*a*]pyrene	I
42	3-Methylbenzo[*a*]pyrene	A	67	6-Methylchrysene	I
43	1,2-Dimethylbenzo[*a*]pyrene	A	68	3-Methylbenz[*a*]anthracene	I
44	2,3-Dimethylbenzo[*a*]pyrene	A	69	1-Methylbenz[*a*]anthracene	I
45	3,12-Dimethylbenzo[*a*]pyrene	A	70	11-Methylbenz[*a*]anthracene	I
46	1,3-Dimethylbenzo[*a*]pyrene	A	71	9-Methylbenz[*a*]anthracene	I
47	1,4-Dimethylbenzo[*a*]pyrene	A	72	2-Methylbenz[*a*]anthracene	I
48	5-Methylbenzo[*c*]phenanthrene	A	73	5-Methylbenz[*a*]anthracene	I
49	5-Methylchrysene	A	74	8-Methylbenz[*a*]anthracene	I
50	6,8-Dimethylbenz[*a*]anthracene	A	75	2-Methylpyrene	I
51	7-Methylbenz[*a*]anthracene	A	76	4-Methylpyrene	I
52	5-Methylbenzo[*a*]pyrene	A	77	1-Methylpyrene	I
53	7-Methylbenzo[*a*]pyrene	A	78	7,10-Dimethylbenzo[*a*]pyrene	I
54	6-Methylbenzo[*a*]pyrene	A	79	6,10-Dimethylbenzo[*a*]pyrene	NA
55	1,6-Dimethylbenzo[*a*]pyrene	A	80	8-Methylbenzo[*a*]pyrene	NA
56	3,6-Dimethylbenzo[*a*]pyrene	A	81	9-Methylbenzo[*a*]pyrene	NA
57	4-Methylbenzo[*c*]phenanthrene	I			

coefficients, in particular, by the percentage of correct classifications, and by the percentage of correctly classified carcinogenic compounds. The latter coefficient is of major interest because it is very important

for the predictive model to be able to correctly assign the active molecules to their category, even more than considering an inactive chemical as carcinogenic. Finally, it must be commented that this pseudo-binary multilinear regression can be viewed as a kind of multidimensional linear discrimination analysis.

Results and discussion

Exactly as in the previous case, but taking into account the property discretization discussed above, the carcinogenic action of PAHs was analyzed using MQSM. Initially, calculations employing both overlap and Coulomb MQSM were performed on the three molecular sets. When analyzing the results, it was observed that the quality of the models was comparable, in contrast to the previous case example. Here, all the compounds are exclusively made of fused benzenes, without the presence of 5-membered rings or bonds connecting rings. This fact allows exact atom-atom superpositions for all the existing molecular pairs, and therefore overlap MQSM is also able to provide similarity terms with reliable values. By the way, overlap measures produced better adjustments in the major part of the cases, and subsequently, results presented in this section correspond to this kind of MQSM.

The first set analyzed with the MQSM approach consisted of the 32 non-methylated PAHs (compounds **1**–**32** in Table 3). The application of the previously exposed protocol yielded the results given in Table 4. As can be observed, the percentage of correct classifications for the different models is excellent. The 4-parameter model was considered the optimal one, and it is able to correctly classify 30 molecules out of a total of 32 (93.8%). The two misclassified PAHs correspond to molecules **7** and **22** of Table 3. If only carcinogenic compounds are considered, 9 out of 10 (90.0%) are included in the correct category. In order to test the predictive capacity of the model, the cross-validation process was carried out. In this procedure, one molecule is removed from the data set, and the remaining molecules serve as a training set. The QSAR equation obtained for the reduced set is then applied to predict the carcinogenicity of the extracted PAH. This process is repeated for all the molecules of the set. As before, goodness of the cross-validation predictions is quantified by means of the percentage of correct classifications. Thus, the 4 PCs model achieves 27 out of 32 (84.4%) if both active and inactive agents are considered, while 8 out of 10 (80.0%) are correctly classified if only the carcinogenic PAHs are taken into account. Models containing more parameters improve the description of the model, but the improvement is not significant enough (96.9% versus 93.8%) to justify the more complex derived equation.

As was previously explained, Villemin et al. [23] proposed a different class grouping for this family. The differences in the carcinogenic power assignments change for molecules **11**–**17** and **31**, which are considered to be active. QSAR models were reconstructed using Villemin's classification for non-methylated PAHs. The optimal model arises when using 4 PCs, leading to 81.3% correctly classified PAHs and 83.3% correctly classified carcinogenic PAHs. Results are not as good as for Vendrame's classification, but they are still significant.

Regarding the methylated set, made up of 46 PAHs, results using MQSM can be seen in Table 5. Again good results have been obtained, as the statistical coefficients show. In this case, the 5 PCs model has been taken as the optimal one, leading to a correct classification of 40 molecules out of 46 (87.0%), and 20 out of 24 (83.3%) for the active PAHs. Misclassifications correspond now to PAHs **33**, **48**, **50**, **51**, **61** and **67**. The same slight decrease in the percentages can be observed from the cross-validation results for almost all the models. In particular, for the optimal model, the number of correctly classified PAHs decreases in one unit in relation to the adjusted results: 39 out of 46 (84.8%) and 19 out of 24 (79.2%), respectively.

In this case, Villemin's grouping criterion is much stricter than Vendrame's, and the major part of the methylated PAHs (38 out of 46: 83%) would have to be considered carcinogenic. Concretely, molecules **57**–**67**, **70**, **71** and **74** are classified as active under Villemin's assumptions. Correlations derived from the different QSAR models could therefore be artificially overrated due to the disproportion between the size of both samples, and results will not be given.

Finally, both subsets are put together and examined with the MQSM methodology. Results for the 78-molecule set are gathered in Table 6. For this set, although slightly lower percentages are attained, the equation needs a smaller number of parameters to reach a maximum. For the optimal 3 PCs model, 64 out of 78 (82.1%) compounds and 24 out of 34 (70.6%) carcinogenic compounds are correctly assigned. The large difference between both values, a situation not found in the previous analyses, indicates that the QSAR equation for the full set concentrates the misclassifications in the active compounds, which is obviously less interesting than fitting the active

Table 4. QSAR results for the non-methylated PAHs subset. The optimal model has been marked in bold face

Number of PCs	Adjustment		Cross-validation	
	% Correct classification	% Correct classification for carcinogenic compounds	% Correct classification	% Correct classification for carcinogenic compounds
1	78.1	30.0	78.1	30.0
2	78.1	60.0	78.1	60.0
3	87.5	80.0	87.5	80.0
4	**93.8**	**90.0**	**84.4**	**80.0**
5	96.9	90.0	84.4	80.0
6	96.9	90.0	84.4	80.0

Table 5. QSAR results for the methylated PAHs subset. The optimal model has been marked in bold face

Number of PCs	Adjustment		Cross-validation	
	% Correct classification	% Correct classification for carcinogenic compounds	% Correct classification	% Correct classification for carcinogenic compounds
1	58.7	79.2	58.7	79.2
2	67.4	66.7	65.2	66.7
3	73.9	66.7	71.7	66.7
4	82.6	79.2	76.1	75.0
5	**87.0**	**83.3**	**84.8**	**79.2**
6	89.1	83.3	80.4	79.2

Table 6. QSAR results for the entire PAHs set. The optimal model has been marked in bold face

Number of PCs	Adjustment		Cross-validation	
	% Correct classification	% Correct classification for carcinogenic compounds	% Correct classification	% Correct classification for carcinogenic compounds
1	74.4	70.6	73.1	70.6
2	70.5	61.8	70.5	61.8
3	**82.1**	**70.6**	**80.8**	**70.6**
4	80.8	73.5	79.5	70.6
5	83.3	73.5	82.1	73.5
6	83.3	73.5	82.1	73.5

ones. The model fails on classifying compounds **4**, **7**, **13**, **22**, **25**, **33**, **34**, **35**, **38**, **48**, **49**, **50**, **51** and **67**. All those PAHs, except molecule **61**, which were incorrectly assigned in the separated subsets, are again misclassified. In addition, when the whole 78 PAHs set is analyzed, several molecules (**4**, **13**, **25**, **34**, **35**, **38** and **49**) which were satisfactorily predicted within their subset, are now assigned to an incorrect class. In relation to the predictive ability of the QSAR model, comparable values are obtained for prediction: 80.8% and 70.6%. The decrease in the predictivity of the models could be expected *a priori* because of the nature of the molecular descriptors employed. As any other similarity-based QSAR approach, the more structurally homogeneous the molecular set studied, the better the methodology applies.

For this family, no randomization test has been carried out. The fact that the activity values are just ones or zeros makes the generation of non-redundant randomized response sequences really difficult. Nevertheless, the presented results can be considered to be reliable because of the small number of descriptors employed and the linearity of the transformations involved.

The same family of 78 PAHs was examined by different authors [56, 57]. Barone et al. [56] reported a study in which the carcinogenicity of the first 26 non-methylated PAHs was correlated using the *electronic index methodology* (EIM) based on the LDOS theory. Later, the same approach was employed to analyze the remaining molecules: **27**–**32** non-methylated PAHs and **33**–**81** methylated ones [57]. In addition, Vendrame et al. [59] verified the importance of quantum chemical parameters by comparing them using three different methods: EIM, principal component analysis (PCA) and neural networks (NN). The optimal descriptors selected were: the highest occupied molecular orbital (HOMO) energy and its lower level (HOMO-1), the difference between these two energies, the HOMO, HOMO-1 contribution to the LDOS over the PAH ring containing the highest bond order and their relative difference. The study of the methylated subset separately found other predictive descriptors: the lowest unoccupied molecular orbital (LUMO) energy and the chemical hardness, defined as (LUMO−HOMO)/2.

Table 7 summarizes the percentages of correct classification for the different methodologies: EIM, PCA, NN and MQSM. As observed, MQSM improve the description provided by the other techniques, even if nonlinear statistical tools, such as NN, are used. This fact is especially relevant for the methylated subset case, for which the percentage of correct classification is almost 9% better than the PCA and NN methods. The methylated subset is particularly interesting because the presence of methyl groups breaks the electron-hole symmetry present in the non-methylated PAHs. This subset was carefully analyzed by Vendrame, who allowed the introduction of additional descriptors. The derived results, however, were poorer than those achieved for both the non-methylated and the full sets. The notable improvement when using MQSM shows that a global density approach might encode some relevant information for characterization of PAHs carcinogenicity that is not present in the local densities.

Table 7. Comparison between the percentages of correctly classified compounds using different QSAR methodologies. NM: non-methylated PAHs; M: methylated PAHs

Method	32 NM PAHs	46 M PAHs	Full set of 78 PAHs
EIM	84.4	73.9	78.2
PCA	84.4	78.3	80.8
NN	93.8	78.3	84.6
MQSM	93.8	87.0	82.1

When analyzing the misclassifications, it can be seen that at least half the PAHs incorrectly classified with the MQSM method coincide with the ones misclassified by PCA (**7**, **34**, **35**, **48**, **49**, **50** and **51**). However, this correspondence is not found with the NN method, where the wrong assignments differ in all cases except two (**48**, **49**). This situation could be explained by the fact that MQSM build molecular descriptors related to changes in the molecular electron distributions, a quantum-chemical magnitude, and the information contained in the similarity matrices is extracted using a linear method, namely classical scaling, which is equivalent to PCA for proximity data. This seems to indicate that information produced by MQSM is somehow similar to that included in the quantum-chemical parameters utilized by Vendrame, and therefore, a similar behavior is retrieved in terms of predictive capacity and objects misclassified. Alternatively, the nonlinear internal manipulations carried out by NN make this approach more different from MQSM treatment. In any case, some caution should be taken with molecules badly predicted by all the methods (**48**, **49**). It seems that none of the different molecular descriptors used to characterize the mole-

cular structure is able to shed light on the structure-activity relationships for these two PAHs. This may indicate two different things: either some decisive information is missing in these pictures, or the activity measurement for these two compounds might be doubtful.

An important remark on the predictivity of the MQSM QSAR models must be pointed out. The last three compounds in Table 3 (**79**, **80** and **81**) possess a carcinogenic activity not experimentally measured yet. MQSM, as well as the other QSAR approaches discussed, have formulated hypotheses about the carcinogenic action of these PAHs. MQSM found PAHs **80** and **81** to be unequivocally active, in agreement with predictions by EIM, PCA and NN. On the other hand, assignment for molecule **79** is close to the threshold between both categories, although the optimal model classifies it as inactive. This differs from the previous studies, which also classified this compound as active.

The success of the different proposed approaches confirms the Benigni et al. hypothesis [62] stating that, whereas hydrophobicity plays a central role in determining toxic potency, the electronic effects, those derived from LDOS and quantum similarity theories, are able to discern between active and inactive PAHs.

Conclusions

This paper has discussed the usefulness of quantum similarity for predicting toxicity-related properties of PAHs. Two different properties have been analyzed: the percutaneous absorption and the carcinogenic activity of two different series of PAHs. In both cases, QSAR models constructed using MQSM have produced comparable, or even better, results than the models found in the literature. A further look at the descriptors evidences that a connection between many of them and MQSM exists, since most of the information described by several quantum-chemical indices or even physicochemical ones is contained within the molecular electron density, and can be extracted from MQSM using simple linear transformations.

This study is included in a series of papers that tries to demonstrate that quantum similarity might be a valuable tool for QSAR and computer-aided drug design, producing comparable results to other highly predictive and widely established 3D-QSAR approaches. In contrast to other methodologies, MQSM produce unbiased descriptors not depending on the parameter database or on arbitrarily chosen grid features. Furthermore, the theoretical connection between MQSM and the molecular properties, via the quantum-mechanical expectation value formula in discrete form [26], ensures that the correlations found, after passing the convenient statistical tests, come from an underlying causal relationship.

Acknowledgements

This work has been partially supported by the European Commission contract ENV4-CT97-0508. Financial support from the *Fundació María Francisca de Roviralta* is also acknowledged. One of us (X.G.) acknowledges the University of Girona for a pre-doctoral fellowship. Finally, thanks to Ricardo Braga and to Richard H. Guy for their kindness and interesting comments.

References

1. Grimmer, G. (Ed.) Environmental Carcinogens, Polycyclic Aromatic Hydrocarbons, Chemistry, Occurrence, Biochemistry, Carcinogenicity, CRC Press, Boca Raton, FL, 1983.
2. Taylor, P., Dellinger, B. and Lee, C.C., Environ. Sci. Technol., 24 (1990) 316.
3. Grimmer, G., Naujac, K.W., Dettbarn, G., Brune, H., Deutsch-Wenzel, R. and Misfeld, J., In Cooke, M., Dennis, A.J. and Fisher, G.L. (Eds.) Polynucleic Aromatic Hydrocarbons: Physics, Biology, and Chemistry, 6th International Symposium, Bartelle Press, Columbus, OH, 1981.
4. Freeman, D.J. and Cattell, F.C.R., Environ. Sci. Technol., 24 (1990) 1581.
5. Takada, H., Onda, T. and Ogura, N., Environ. Sci. Technol., 24 (1990) 1179.
6. Pott, P., Surgical observations relative to the cancer of the scrotum. (1775). Reprinted in Natl. Cancer Inst. Monogr. 10 (1973) 7.
7. a. United States Environmental Protection Agency (EPA). Integrated Risk Information System (IRIS). Environmental Criteria and Assessment Office, Office of Health and Environmental Assessment, Cincinnati, OH, 1994.
   b. Agency for Toxic Substances and Disease Registry (ATSDR). Toxicological Profile for Polycyclic Aromatic Hydrocarbons. Acenaphthene, Acenaphthylene, Anthracene, Benzo(a)anthracene, Benzo(a)pyrene, Benzo(b)fluoranthene, Benzo(g,i,h)perylene, Benzo(k)fluoranthene, Chrysene, Dibenzo(a,h)anthracene, Fluoranthene, Fluorene, Indeno(1,2,3-c,d)pyrene, Phenanthrene, Pyrene. Prepared by Clement International Corporation, under Contract No. 205-88-0608. ATSDR/TP-90-20, 1990.
   c. International Agency for Research on Cancer (IARC). IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Polynuclear Aromatic Compounds. Part 1. Chemical, Environmental and Experimental Data, Vol. 32. World Health Organization, Lyon, 1983.

8. Govers, H.A.J., In Karcher, W. and Devillers, J. (Eds.) Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology, Kluwer, Dordrecht, 1990.
9. Harvey, R.G. and Geacintov, N.E., Acc. Chem. Res., 21 (1988) 66.
10. a. Dipple, A., In Searle, C.E. (Ed.) Polynuclear Aromatic Carcinogens, Chemical Carcinogens, American Chemical Society, Washington, DC, 1976.
    b. Dipple, A., Moschel, R.C. and Bigger, C.A.H., In Searle, C.E. (Ed.) Polynuclear Aromatic Carcinogens, Chemical Carcinogens, 2nd Ed., American Chemical Society, Washington, DC, 1984.
11. Schmidt, O.Z., Physik. Chem., B42 (1939) 83.
12. a. Pullman, A. and Pullman, B., Rev. Sci., 84 (1946) 145.
    b. Pullman, A. and Pullman, B., Adv. Cancer Res., 3 (1955) 117.
13. a. Gayoso, J. and Kimri, S., Int. J. Quantum Chem., 38 (1990) 461.
    b. Gayoso, J. and Kimri, S., Int. J. Quantum Chem., 38 (1990) 487.
14. Jerina, D.M., Lehr, R.E., Schaefer, R.M., Yagi, H., Karle, J.M., Thakker, D.R., Wood, A.W. and Conney, A.H., In Hiatt, H., Watson, J.D. and Winstin, I. (Eds.) Origins of Human Cancer, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1977.
15. a. Ulrich, V., Roots, I., Hildebrant, A.G. and Eatabrook, R.K.W., In Conney, A.H. (Ed.) Microsomes and Drug Oxidations, Pergamon Press, Oxford, 1977.
    b. Geacintov, N.E., In Cooke, M., Dennis, A.J. and Fisher, G.L. (Eds.) Polynuclear Aromatic Hydrocarbons: Physics, Biology, and Chemistry, 6th International Symposium, Bartelle Press, Columbus, OH, 1981.
    c. Hodgson, R.M., Cary, P.D., Grover, P.L. and Sims, P., Carcinogenesis (London), 4 (1983) 1153.
16. Silverman, J.P.B.D., Acc. Chem. Res., 17 (1984) 332.
17. Szentpály, L.V., J. Am. Chem. Soc., 106 (1984) 6021.
18. Kimri, S. and Gayoso, J., J. Mol. Struct. (THEOCHEM), 362 (1996) 141.
19. Hansch, C. and Fujita, T., J. Am. Chem. Soc., 86 (1964) 1616.
20. Nordén, U.E. and Wold, S., Acta Chem. Scand., B32 (1978) 602.
21. Klopman, G., J. Am. Chem. Soc., 106 (1984) 7315.
22. Lall, R.S., Match., 15 (1984) 251.
23. Villemin, B., Cherqaoui, D. and Mesbah, A., J. Chem. Inf. Comput. Sci., 34 (1994) 1288.
24. Isu, Y., Nagashima, U., Aoyama, T. and Haruo, H., J. Chem. Inf. Comput. Sci., 36 (1996) 286.
25. Carbó, R., Arnau, J. and Leyda, L., Int. J. Quantum Chem., 17 (1980) 1185.
26. Carbó, R., Besalú, E., Amat, L. and Fradera, X., J. Math. Chem., 18 (1995) 237.
27. a. Fradera, X., Amat, L., Besalú, E. and Carbó-Dorca, R., Quant. Struct.-Act. Relat., 16 (1997) 25.
    b. Lobato, M., Amat, L., Besalú, E. and Carbó-Dorca, R., Quant. Struct.-Act. Relat., 16 (1997) 465.
    c. Amat, L., Robert, D., Besalú, E. and Carbó-Dorca, R., J. Chem. Inf. Comput. Sci., 39 (1998) 624.
    d. Amat, L., Carbó-Dorca, R. and Ponec, R., J. Comput. Chem., 19 (1998) 1575.
    e. Robert, D., Amat, L. and Carbó-Dorca, R., J. Chem. Inf. Comput. Sci., 39 (1999) 333.
    f. Ponec, R., Amat, L. and Carbó-Dorca, R., J. Phys. Org. Chem., 12 (1999) 447.
    g. Ponec, R., Amat, L. and Carbó-Dorca, R., J. Comput.-Aided Mol. Design, 13 (1999) 259.
    h. Robert, D., Gironés, X. and Carbó-Dorca, R., J. Comput.-Aided Mol. Design, 13 (1999) 597.
    i. Amat, L., Carbó-Dorca, R. and Ponec, R., J. Med. Chem., 42 (1999) 5162.
    j. Robert, D., Gironés, X. and Carbó-Dorca, R., J. Chem. Inf. Comput. Sci., 40 (2000) 839.
    k. Gironés, X., Amat, L., Robert, D. and Carbó-Dorca, R., J. Comput.-Aided Mol. Design, 14 (2000) 477.
28. a. Robert, D., Gironés, X. and Carbó-Dorca, R., SAR QSAR Environ. Res., 10 (1999) 401.
    b. Gironés, X., Amat, L. and Carbó-Dorca, R., SAR QSAR Environ. Res., 10 (1999) 545.
29. Constans, P., Amat, L. and Carbó-Dorca, R., J. Comput. Chem., 18 (1997) 826.
30. Solà, M., Mestres, J., Carbó, R. and Duran, M., J. Am. Chem. Soc., 116 (1994) 5909.
31. Mezey, P.G., Ponec, R., Amat, L. and Carbó-Dorca, R., Enantiomeres, 4 (1999) 371.
32. Forés, M., Duran, M. and Solà, M., TMMEC, 1 (1997) 50.
33. Solà, M., Mestres, J., Carbó, R. and Duran, M., J. Chem. Phys., 104 (1996) 636.
34. a. Carbó-Dorca, R. and Besalú, E., J. Mol. Struct. (THEOCHEM), 451 (1998) 11.
    b. Carbó-Dorca, R., Amat, L., Besalú, E. and Lobato, M., In Carbó-Dorca, R. and Mezey, P.G. (Eds.), Advances in Molecular Similarity, Vol. 2, JAI Press, Greenwich, CT, 1998.
    c. Carbó-Dorca, R., Besalú, E., Amat, L. and Fradera, X., In Carbó-Dorca, R. and Mezey, P.G. (Eds.), Advances in Molecular Similarity, Vol. 1, JAI Press, Greenwich, CT, 1996.
    d. Carbó, R. and Besalú, E., In Carbó, R. (Ed.) Molecular similarity and reactivity: from quantum chemical to phenomenological approaches, Kluwer, Dordrecht, 1995.
    e. Carbó-Dorca, R., Amat, L., Besalú, E., Gironés, X. and Robert, D., In Carbó-Dorca, R. and Mezey, P.G. (Eds.) The Fundamentals of Molecular Similarity, Kluwer, Dordrecht, to be published.
    f. Carbó-Dorca, R., Amat, L., Besalú, E., Gironés, X. and Robert, D., J. Mol. Struct. (THEOCHEM), 504 (2000) 181.
35. Carbó, R. and Domingo, L., Int. J. Quantum Chem., 23 (1987) 517.
36. a. Amat, L. and Carbó-Dorca, R., J. Comput. Chem., 18 (1997) 2023.
    b. Amat, L. and Carbó-Dorca, R., J. Comput. Chem., 20 (1999) 911.
37. ASA coefficients and exponents for an assorted sample of atoms can be seen and downloaded from the website: http://iqc.udg.es/cat/similarity/ASA/funcset.html
38. Dewar, M.J.S., Zoebisch, E.G., Healy, E.F. and Stewart, J.J.P., J. Am. Chem. Soc., 107 (1985) 3902.
39. AMPAC 6.55, 1999 Semichem, 7128 Summit, Shawnee, KS 66216 D.A.
40. a. Constans, P., Amat, L. and Carbó-Dorca, R., J. Comput. Chem., 18 (1997) 826.
    b. McMahon, A.J. and King, P.M., J. Comput. Chem., 18 (1997) 151.
    c. Parretti, M.F., Kroemer, R.T., Rothman, J.H. and Richards, W.G., J. Comput. Chem., 18 (1997) 1344.
41. Amat, L., Constans, P., Besalú, E. and Carbó-Dorca, R., MOLSIMIL 97, Institute of Computational Chemistry, University of Girona, Spain, 1997.
42. Cox, T.F. and Cox, M.A.A., Multidimensional Scaling, Chapman & Hall, London, 1994.

43. Cuadras, C.M. and Arenas, C., Commun. Stat. Theor. Method., 19 (1990) 2261.
44. Wold, S. and Eriksson, L., In Van der Waterbeemd, H. (Ed.) Chemometric methods in molecular design, VCH, New York, NY, 1995.
45. Amat, L., Robert, D. and Besalú, E., TQSAR-SIM, Institute of Computational Chemistry, University of Girona, Spain, 1997.
46. a. Yang, J.J., Roy, T.A., Neil, W. and Krueger, A.J., Toxicol. Ind. Hlth., 3 (1987) 405.
    b. Yang, J.J., Roy, T.A. and Mackerer, C.R., Toxicol. Ind. Hlth., 2 (1986) 409.
    c. Yang, J.J., Roy, T.A. and Mackerer, C.R., Toxicol. Ind. Hlth., 2 (1986) 79.
47. Roy, T.A., Krueger, A.J., Mackerer, C.R., Neil, W., Arroyo, A.M. and Yang, J.J., SAR QSAR Environ. Res., 9 (1998) 171.
48. Structures and coordinates can be seen and downloaded from the website: http://iqc.udg.es/cat/similarity/QSAR/PAHs.html
49. Gute, B.D., Grundwald, G.D. and Basak, S.C., SAR QSAR Environ. Res., 10 (1999) 1.
50. Kubinyi, H. (Ed.) QSAR: Hansch Analysis and Related Approaches, VCH, Weinheim, 1993.
51. Stuper, A.J., Brugger, W.E. and Jurs, P.C., Computer Assisted Studies of Chemical Structure and Biological Function, Wiley, New York, NY, 1979.
52. Potts, R.O. and Guy, R.H., Pharm. Res., 9 (1992) 663.
53. a. Berger, D., Smith, A., Seybold, G. and Serve, M.P., Tetrahedron Lett., 3 (1978) 231.
    b. Sjentplay, L.V., J. Am. Chem. Soc., 106 (1984) 6021.
54. Cavalieri, E.L., Rogan, E.G., Roth, R.W., Saugier, R.K. and Hakan, A., Chem. Biol. Interact., 47 (1983) 87.
55. Sakamoto, Y. and Watanabe, S., Bull. Chem. Soc. Jpn., 59 (1986) 3033.
56. Barone, P.M.V.B., Camilo, A., Jr. and Galvão, D.S., Phys. Rev. Lett., 77 (1996) 1186.
57. Braga, R.S., Barone, P.M.V.B. and Galvao, D.S., J. Mol. Struct. (THEOCHEM), 464 (1999) 257.
58. Streitwieser, A., Molecular Orbital Theory, Wiley, New York, NY, 1961.
59. Vendrame, R., Braga, R.S., Takahata, Y. and Galvão, D.S., J. Chem. Inf. Comput. Sci., 39 (1999) 1094.
60. Iball, J., Am. J. Cancer, 35 (1939) 188.
61. Badger, G.M., Br. J. Cancer, 2 (1948) 309.
62. Benigni, R., Andreoli, C. and Giuliani, A., Environ. Mol. Mutagen., 24 (1994) 208.