Journal of Computer-Aided Molecular Design, **13:** 259–270, 1999.
KLUWER/ESCOM

Molecular basis of quantitative structure-properties relationships (QSPR): A quantum similarity approach

Robert Ponec[a], Lluís Amat[b] & Ramon Carbó-Dorca[b,*]
[a]*Institute of Chemical Process Fundamentals, Czech Academy of Sciences, Prague 6, Suchdol 2, 165 02, Czech Republic;* [b]*Institute of Computational Chemistry, University of Girona, 17071 Girona, Catalonia, Spain*

Received 2 July 1998; Accepted 3 September 1998

Key words: log P, molecular quantum similarity measures, quantitative structure-properties relationships (QSPR), substituent effect

Summary

Since the dawn of quantitative structure-properties relationships (QSPR), empirical parameters related to structural, electronic and hydrophobic molecular properties have been used as molecular descriptors to determine such relationships. Among all these parameters, Hammett σ constants and the logarithm of the octanol-water partition coefficient, log P, have been massively employed in QSPR studies. In the present paper, a new molecular descriptor, based on quantum similarity measures (QSM), is proposed as a general substitute of these empirical parameters. This work continues previous analyses related to the use of QSM to QSPR, introducing molecular quantum self-similarity measures (MQS-SM) as a single working parameter in some cases. The use of MQS-SM as a molecular descriptor is first confirmed from the correlation with the aforementioned empirical parameters. The Hammett equation has been examined using MQS-SM for a series of substituted carboxylic acids. Then, for a series of aliphatic alcohols and acetic acid esters, log P values have been correlated with the self-similarity measure between density functions in water and octanol of a given molecule. And finally, some examples and applications of MQS-SM to determine QSAR are presented. In all studied cases MQS-SM appeared to be excellent molecular descriptors usable in general QSPR applications of chemical interest.

Introduction

One of the main goals of natural sciences is the formulation of simple models and concepts in terms of which the observed phenomena can be classified, understood and, finally, described. The sophistication of these models depends on the complexity of the phenomena studied and, also, on the degree of theoretical development of a given science. In chemistry, one of the most complex fields resisting rigorous mathematical description is constituted by the relationships between the structure and (re)activity or properties of molecular structures, the so-called QSAR or QSPR. Due to the difficulties associated to QSPR analysis, an important place in the problem formulation still belongs to various empirical relationships. The well-known Hammett or Taft equation [1–3] can serve as an example of this situation. Another important field, where empirical QSPR still play the dominant role, is the design of new materials or biologically active compounds [4–6]. Because of the immense practical importance of QSPR, considerable attention has been devoted to the elucidation of the factors responsible for the existence and validity of the empirical equations. In spite of the undeniable progress which these studies have brought to the solution of some particular problems (for instance, the theoretical rationalization of the Hammett ρ constants [7–9]), the understanding of the factors responsible for the existence of QSPR is still rather fragmentary. The main problem with these QSPR is that they cannot generally be derived from theoretically well-founded thermodynamic models, and this is why they are also called extra-thermodynamic relationships. Instead, they rely only on the intuitive

*To whom correspondence should be addressed.

empirical idea that a similar structural change, induced in a given series of structurally related compounds by a systematic variation of substitution, will also have a similar effect on the properties or reactivity of the corresponding molecules.

Because of the great impact which this intuitively understood idea of similarity has had on chemical reasoning, it is not surprising that in recent years a lot of effort has been devoted to formulating the qualitative concept of similarity within a well defined theoretical basis. Especially promising in this respect seems to be the introduction of the quantitative similarity measures and indices based on quantum mechanical ideas [10–22]. In terms of this approach a number of qualitative chemical concepts could indeed be justified. The aim of the present work is to demonstrate that in addition to previously reported applications [23–25], quantum similarity measures (QSM) also provide an elegant and universal framework for the formulation of QSPR, whether in linear-free energy relationships (LFER) or in a more general QSAR form.

In pursuing this goal, a brief review of the basic ideas of QSM and their connections to QSPR will be presented. After introducing this general methodology, various examples to show some practical applications of the approach will be given. The provided examples involve the use of QSM as theoretical descriptors, replacing empirical parameters like the Hammett σ constants or octanol-water partition coefficients (log P), in QSPR of both chemical and pharmacological interest.

Theoretical framework

Molecular quantum similarity measures (MQSM)

The philosophy underlying the introduction of MQSM as molecular descriptors arises from the simple idea that properties of molecules are determined by their electronic structure. As a consequence, any similarity in molecular properties has to be reflected in the similarity of electronic structure. This suggests that if one wants to look for theoretical descriptors of molecular similarity it is natural to look for similarity in electron distribution, which, from a quantum mechanical point of view, contains all the information associated to a given microscopic system [26, 27]. MQSM represent an attempt to characterize quantitatively the similarity in electronic structure. On the other hand, the simplest descriptor characterizing the molecular electron structure is the first order electron density $\rho(\mathbf{r})$, so it is quite natural to introduce the most easily computed MQSM just on this quantity. From the quantum mechanical point of view an MQSM involving two molecules *A* and *B* with associated first order densities $\rho_A(\mathbf{r})$ and $\rho_B(\mathbf{r})$, may be defined by means of the following integral:

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

where $\Omega(\mathbf{r}_1, \mathbf{r}_2)$ is a positive definite operator depending on the coordinates of two electrons. When both compared molecules are the same, the corresponding MQSM, $Z_{AA}(\Omega)$ is denoted as a molecular quantum self-similarity measure (MQS-SM). In this work only the so-called overlap-like MQSM is used, which corresponds to employing the Dirac delta function $\Omega(\mathbf{r}_1, \mathbf{r}_2) = \delta(\mathbf{r}_1 - \mathbf{r}_2)$ as a weighting operator. Then, in this case, Equation (1) simplifies to the form:

$$Z_{AB} = \int \rho_A(\mathbf{r})\rho_B(\mathbf{r})\,\mathrm{d}\mathbf{r}. \quad (2)$$

Calculation of integral (2) can be very time consuming, since a large number of four-center integrals has to be computed within the common LCAO approximation. Moreover, the value defined by any MQSM is positionally dependent, so that optimization of mutual position of both molecules to yield the maximum measure is also required [28]. In order to reduce the above computational problems, a simplification was proposed some time ago [29, 30], known as the ASA approximation, which fits the electronic first order density functions to a linear expression composed of spherical functions. Using ASA, the MQSM computational costs are indeed substantially reduced, while still preserving a reasonable accuracy. Further reduction of computational costs can be achieved using the so-called *promolecular* approximation [30]. This approximation is based on a discrete representation of the molecular electron density function, defined as a convex superposition of the atomic electron densities in a given molecule *A*:

$$\rho_A^{ASA}(\mathbf{r}) = \sum_{a \in A} P_a \rho_a^{ASA}(\mathbf{r}). \quad (3)$$

In all the examples studied within this work, the coefficients P_a have been defined as the number of *valence* electrons minus the effective charge located on each atom *a*. In the *promolecular* form of Equation (3), every ASA atomic density function, $\rho_a^{ASA}(\mathbf{r})$, is

constructed using a linear combination of normalized 1S-type Gaussian functions centered on the *a*-th atom:

$$\rho_a^{ASA}(\mathbf{r}) = \sum_{i \in a} w_i \, |S_i(\mathbf{r} - \mathbf{R}_a; \zeta_i)|^2 \tag{4}$$

where the coefficients w_i are forced to fulfill the convex constraints:

$$\{w_i > 0, \forall i\} \wedge \left\{ \sum_i w_i = 1 \right\}, \tag{5}$$

to preserve the statistical distribution probability meaning of the approximate density function [31]. In order to keep Equation (5) conditions, the convex coefficient set w_i has been optimized using an elementary Jacobi rotation technique [30]. The 1S-GTO exponents $\{\zeta_i\}$ have also been optimized simultaneously using a Newton method. Both sets of parameters have been fitted to a density function obtained using a 3–21 G basis sets for atoms H through Kr. Optimal coefficients and exponents can be downloaded from a WWW site [32]. Furthermore, the *promolecular* densities used in this work have been constructed using the following rule: one function for H, three functions for C, N, O, and F, four functions for S and Cl and five functions for Br.

Using the ASA formalism described above, the overlap-like MQSM between two molecules, as in Equation (2), is reduced to the final simple quadratic form:

$$Z_{AB} = \sum_{a \in A} \sum_{b \in B} P_a P_b Z_{ab}, \tag{6}$$

where P_a and P_b correspond to the same corrected atomic charges defined in the same way as in Equation (3), and the atomic QSM contributions Z_{ab} are calculated as the integrals:

$$Z_{AB} = \sum_{i \in a} \sum_{j \in b} w_i w_j \int |S_i(\mathbf{r} - \mathbf{R}_a)|^2 \left|S_j(\mathbf{r} - \mathbf{R}_b)\right|^2 \mathrm{d}\mathbf{r}, \tag{7}$$

which correspond to a simple and well-known overlap between two 1S-GTO functions [33].

MQSM and QSPR

Once the theoretical concept of MQSM and their computational implementations have been set, one can proceed to the application of the above abstract concepts to practical chemical problems. The most appealing is the recently proposed formulation of the general theory, where the MQSMs open the way for a plausible theoretical foundation of QSPR. Since the detailed and strict mathematical development of the procedure can be found in a previous study [23], only a brief resume will be given.

Suppose known a molecular set $\mathbf{M}=\{m_I\}$ formed by *n* molecules, and an attached set of density functions $\mathbf{D}=\{\rho_I\}$ in one-to-one correspondence with **M**. Now, having chosen a basis set of density functions, suppose computed an $(n \times n)$ similarity matrix $\mathbf{Z}=\{Z_{JI}\}$, whose elements correspond to some MQSM involving all pair of functions in **D**. This matrix can also be viewed as a row vector whose elements are their columns: $\mathbf{Z}=\{\mathbf{z}_I\}$. Every column $\mathbf{z}_I$ is in one-to-one correspondence with the functions of **D** and hence of **M**. The set $\mathbf{Z}=\{\mathbf{z}_I\}$ can be interpreted as an *n*-dimensional discrete representation of the functions of **D**. Since, according to quantum theory, the density functions can be regarded as the source of all the molecular information [26, 27], then any observable molecular property π_I, associated to some hermitian operator $\Omega(\mathbf{r})$ and attached to molecule m_I may be calculated using:

$$\pi_I = \int \Omega(\mathbf{r}) \rho_I(\mathbf{r}) \, d\mathbf{r}, \tag{8}$$

which may be interpreted as a scalar product, and can be formally written as:

$$\pi_I = \langle \Omega | \rho_I \rangle \, . \tag{9}$$

If it is taken into account that an *n*-dimensional discrete representation of the density function, ρ_I, is known in terms of the columns of the similarity matrix $\mathbf{Z}=\{\mathbf{z}_I\}$, the above continuous representation (8) of π_I can be rewritten as:

$$\pi_I \approx \mathbf{a}^\top \mathbf{z}_I \tag{10}$$

where **a** is an *n*-dimensional vector, representing the operator $\Omega(\mathbf{r})$ in the *n*-dimensional discrete space of MQSM vectors $\mathbf{z}_I$. The operator-vector **a** is generally unknown, but if the pairs $\{\pi_I, \mathbf{z}_I\}$ are well defined, the elements of **a** can be obtained by any least-squares technique. This computational procedure constitutes in this way a theoretical basis of QSPR. While Equation (10) represents a general theoretical form of QSPR, in some cases it is also useful to convert it to another, simpler form, which, could be closely related to the well-known LFER. In order to introduce this

simplification, Equation (10) can be rewritten in the form below, in which the self-similarity part, corresponding to the diagonal of the similarity matrix, is separated from the rest:

$$\pi_I = a_I Z_{II} + \sum_{J \neq I} a_J Z_{JI}. \quad (11)$$

Now, if the last term of the right hand side of Equation (11) is denoted as b_I and if it is further possible to admit that in a given series of molecules the terms $\{a_I, b_I\}$ could be roughly constant, Equation (11) transforms into:

$$\pi_I \approx a Z_{II} + b, \quad (12)$$

which expresses a simple linear correlation between the property π_I of molecule m_I and the MQS-SM: Z_{II}, of the same molecule. Such a form of relationship is frequently found empirically. For example, within the Hammett equation, where the molecular properties (like pK or log *k*) are correlated with the substituent constants σ in a series of substituted compounds. The following section will present several examples of the use of self-similarity measures as descriptors of substituent effects.

Results and discussion

Quantum self-similarity measures as descriptors of the substituent effect

In order to show the usefulness of MQSM as molecular descriptors, the applications of these measures to the description of substituent effect is reported as a first example. This field is traditionally represented by the broad class of the so-called LFER as, for example the Hammett and Taft equation. The intention of the present study is to show that appropriately selected MQS-SM do indeed describe the variation of substituent effect in a given reaction series, so that they can be regarded as equivalent to the usual substituent constants. In order to shed light on this equivalence, a series of dissociation equilibrium of several substituted carboxylic acids have been analyzed. As will be shown below, good correlations of MQS-SMs with Hammett σ constants, which can be regarded as a theoretical counterpart of empirical pK vs σ correlations, have been observed. Even if in principle the MQS-SM corresponding to whole molecules can be used in the correlations with σ, it has often been found useful not to characterize the whole molecule by the corresponding Z_{II} values, but only with certain fragments, which can be identified, in turn, with the molecular part actively participating in a given process: the reaction center. The philosophy underlying this replacement is based on the well-known fact that the reaction center is usually the part of the molecule which is the most strongly involved in the process. In this sense, neglecting any contaminating interactions with the remaining part of the molecule may result in the increased sensitivity of the corresponding descriptor to external effects. Such a specific limitation to a particular molecular fragment is, of course, possible only when the reaction center can unambiguously be determined. In the present case, where the substituent effect on the dissociation equilibrium of carboxylic acids has been analyzed, the active part of the molecule can clearly be identified with the COOH group. As a consequence, MQS-SM Z_{II}, calculated for the active COOH fragment, should be the appropriate descriptor in the present framework. In this case the theoretical counterpart of the usual Hammett pK vs σ correlation should be written as:

$$pK_X^R = a_R Z_{XX,R}^{COOH} + b_R, \quad (13)$$

or, using the equivalent form:

$$Z_{XX,R}^{COOH} = \alpha_R \sigma_X + \beta_R. \quad (14)$$

In these equations the index X runs over the set of substituents and R denotes the particular reaction series. In the present case, the analyzed reaction is constituted by the classical series of p-substituted benzoic acids (**I**), 5-substituted thiophen (**II**) and furan 2-carboxylic acids (**III**), p-substituted trans-cinnamic (**IV**) and phenylacetic (**V**) acids (see Scheme 1) with substituents involving both electron donor and electron acceptor groups. The calculations for all molecules have been performed using the semiempirical AM1 [34] method included within the MOPAC [35] package. In all cases, the structures of substituted acids have been completely optimized and the resulting geometries and charge distributions have been used as the input data for the subsequent calculations of MQS-SM. The observed correlations between the calculated MQS-SM for the COOH group and the Hammett substituent constants (Equation (14)) are depicted in Figures 1–5 and the resulting statistical parameters are summarized in Table 1.

The correlations are in all cases quite satisfactory, so that the assumption that MQS-SMs represent good

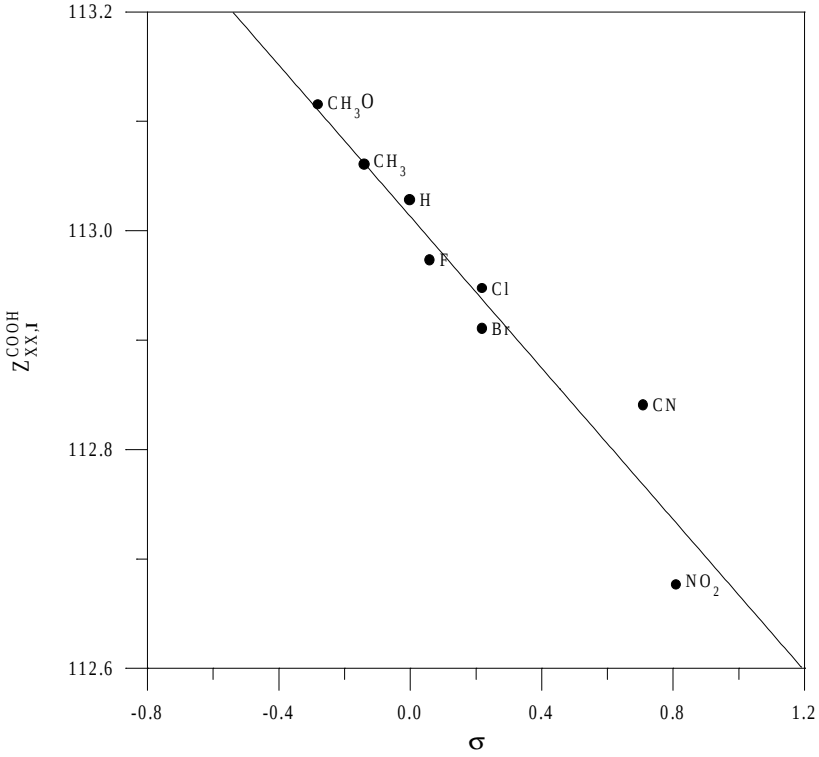


Figure 1. Dependence of calculated MQS-SM $Z^{COOH}_{XX,\mathbf{I}}$ for a series of substituted benzoic acids on the Hammett σ constants.

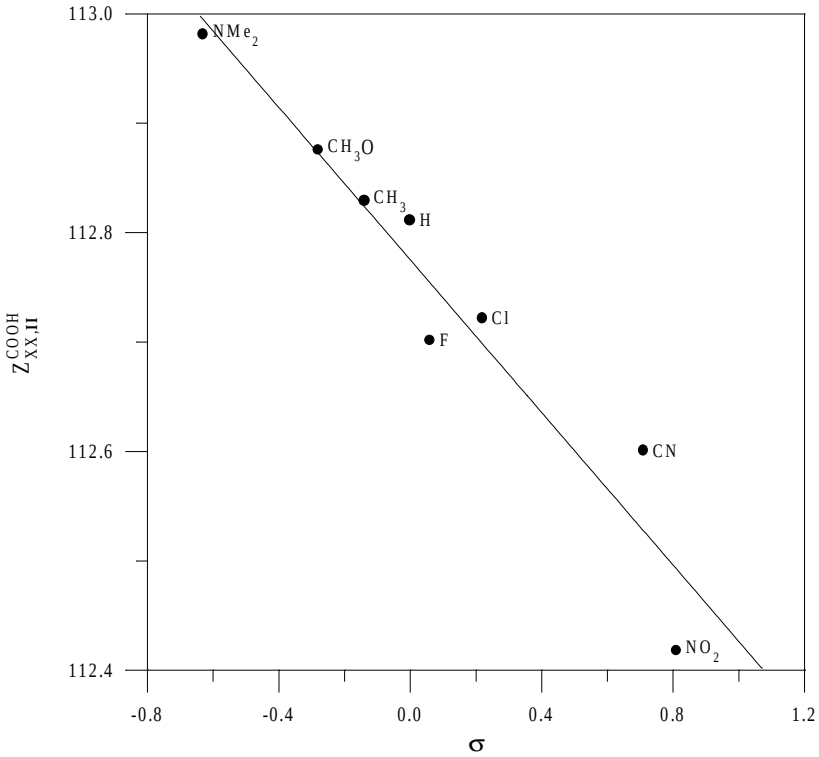


Figure 2. Dependence of calculated MQS-SM $Z^{COOH}_{XX,\mathbf{II}}$ for a series of substituted thiophen 2-carboxylic acids on the Hammett σ constants.

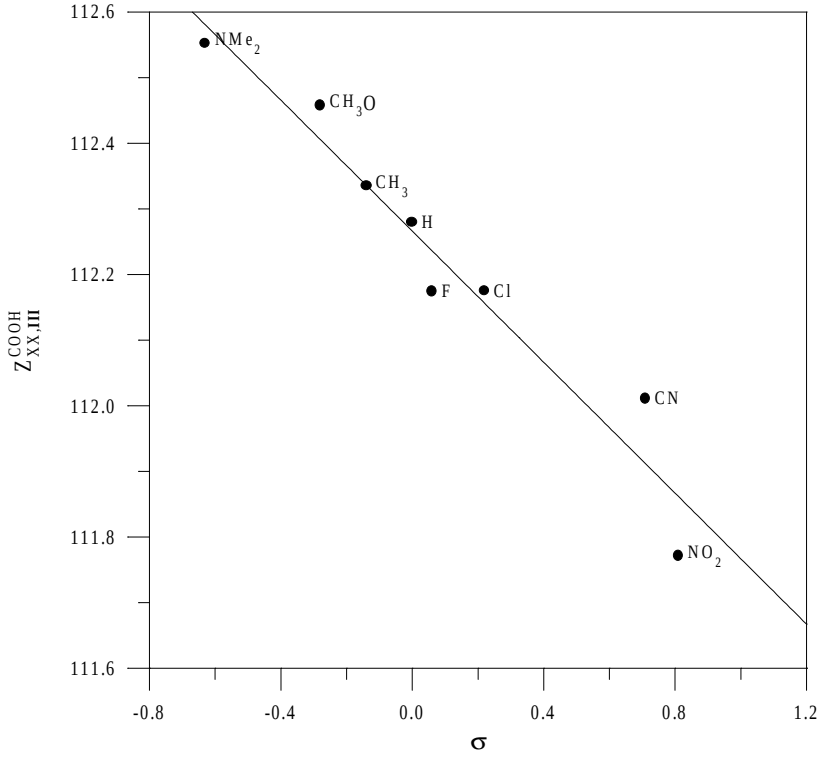
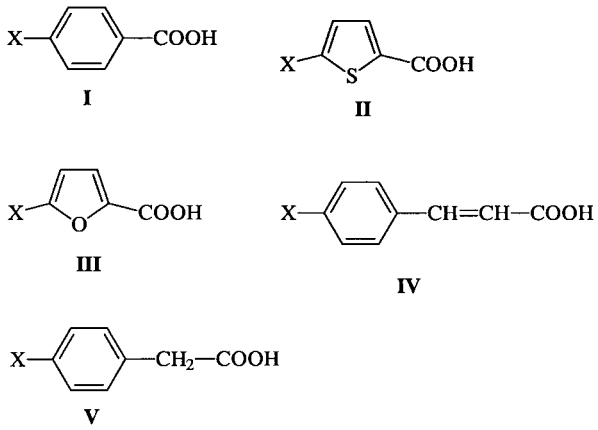


Figure 3. Dependence of calculated MQS-SM $Z^{COOH}_{XX,\mathbf{III}}$ for a series of substituted furan 2-carboxylic acids on the Hammett σ constants.

Table 1. Calculated statistical parameters of the correlations of MQS-SMs against Hammett substituent constants

Reaction series	α_R	β_R	n[a]	r[b]
I	−0.34	113.01	8	0.969
II	−0.37	112.77	8	0.966
III	−0.50	112.27	8	0.974
IV	−0.22	113.15	7	0.967
V	−0.14	113.17	8	0.961

[a] Number of molecules.
[b] Regression coefficient.



Scheme 1. Molecular structures for benzoic acid (**I**), thiophen 2-carboxylic acid (**II**), furan 2-carboxylic acid (**III**), trans-cinnamic acid (**IV**) and phenylacetic acid (**V**).

theoretical descriptors of substituent effects is indeed justified. In addition to this primary result there are also some other conclusions which can be deduced from the observed correlations. The most interesting of them all, corresponds to the possibility of using the slopes of the reported theoretical relationships for the estimation of the relative sensitivity of a given skeleton to the transmission of the substituent effect, measured by the Hammett ρ constant. Thus, if the slope α_o of the correlation (14) for the substituted benzoic acids is taken as an arbitrary unit (corresponding to ρ=1 for this process), then the ratios α_R/α_o of the corresponding slopes for the remaining reaction series can be expected, if everything is correct, to characterize the relative sensitivity of other skeletons to the substituent effect. That is: it will be connected with the experimental ρ constants. As can be seen in Table 2,

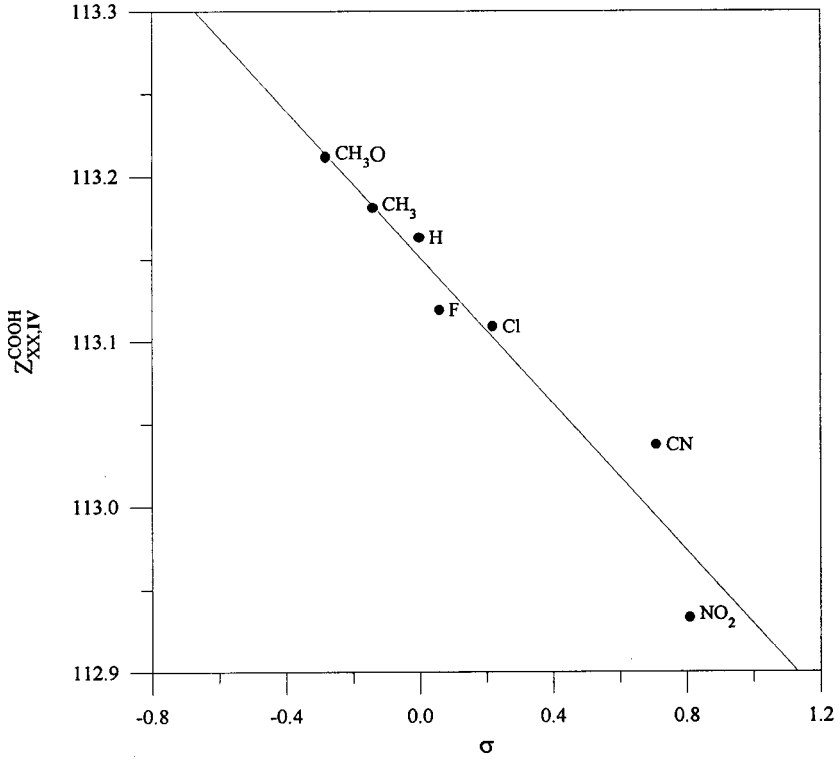


Figure 4. Dependence of calculated MQS-SM $Z^{COOH}_{XX,\mathbf{IV}}$ for a series of substituted trans-cinnamic acids on the Hammett σ constants.

where these ratios are summarized, the agreement between the experimental and 'theoretical' ρ constants is good. This suggests that the above reported linear relationships can be used advantageously as a means of calculating the experimental ρ constants. Here it is fair to say that similar estimations had been already reported on the basis of perturbation theory some time ago [7–9], but such calculations have been restricted only to aromatic conjugated systems. The above reported approach is, however, free of any limitation and can be generally used.

Another interesting and nontrivial conclusion, resulting from the above similarity approach, is represented by the observed correlations of pK_a of nonsubstituted acids (**I**-**V**) with the corresponding MQS-SM $Z^{COOH}_{HH,R}$. An example which surpasses the range of previously reported correlations is depicted in Figure 6. This is an especially interesting case, since it represents the example of the general relation (12) for which there is no known LFER counterpart.

In connection with the above reported relationships it is, of course, necessary to mention another important aspect. This aspect concerns the fact that while the experimental ρ constants have been obtained in water solution, the quantum chemical descriptors used correspond to the gas phase, so that in correlations like Equation (14), the data from two different phases are in fact compared. This, of course, is not theoretically perfect, but since the experimental data have in all cases been obtained under the same conditions (water solutions at 25 °C) and, moreover, the studied systems form a series of closely structurally related molecules, it is possible to expect that the comparison of gas phase and solution data will only be affected by a systematic shift. The relative comparisons between

Table 2. Comparison of experimental and calculated ρ constants for the dissociation of carboxylic acids **I**-**V** in water

Reaction series	ρ_{exp}	α_R/α_o
I	1	1.00
II	1.13[a]–1.20[b]	1.09
III	1.40	1.47
IV	0.46	0.64
V	0.56	0.42

[a] From ref. 36.
[b] From ref. 37.

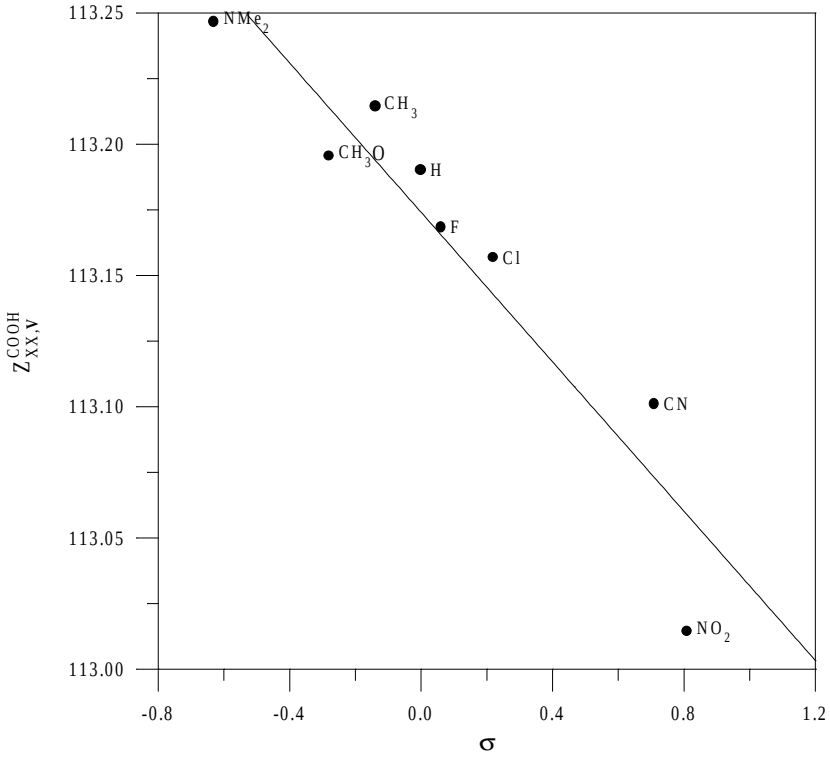


Figure 5. Dependence of calculated MQS-SM $Z^{COOH}_{XX,\mathbf{V}}$ for a series of substituted phenylacetic acids on the Hammett σ constants.

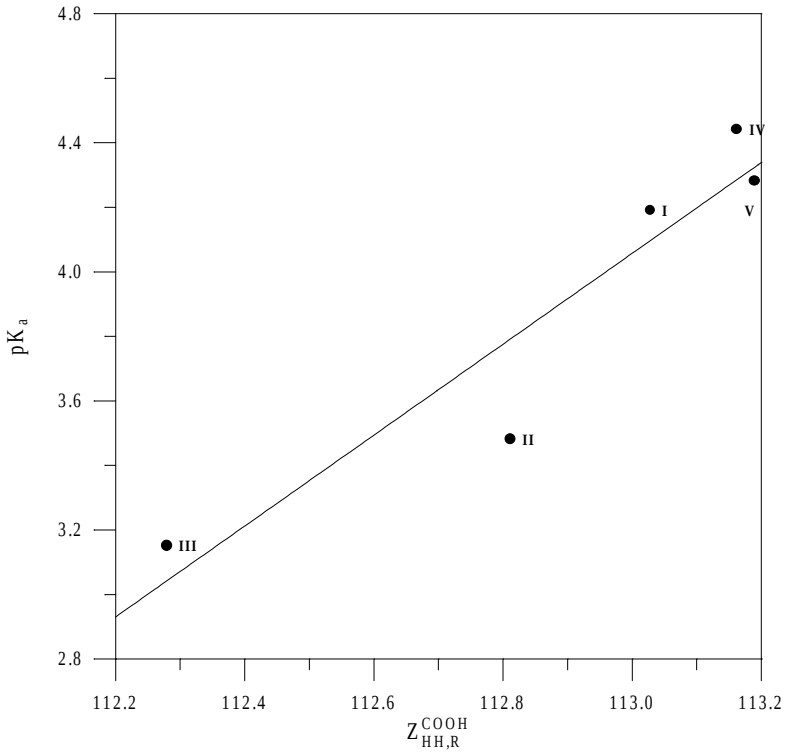


Figure 6. Dependence of calculated MQS-SM $Z^{COOH}_{HH,R}$ for the nonsubstituted acids (**I**-**V**) on the pK_a constants.

the different series in this context may still remain meaningful. There is, however, no problem to extract the necessary theoretical descriptors from the quantum chemical calculations including the solvent effect [39–41], so that empirical correlations like Equation (14) can be made more realistic. An example of correlations where the direct inclusion of the solvent effect is of crucial importance is represented here by the theoretical prediction of the partition coefficient between water and octanol, which is frequently used as a descriptor in QSAR calculations. In the following part an example of the procedure allowing the theoretical calculation of log *P* in a series of structurally related skeletons will be reported.

Self-similarity measures as descriptors of log P

Partition coefficients between water and octanol are empirical parameters which have long been used as a molecular descriptor of molecular hydrophobicity. Because of this descriptive importance, various empirical additivity schemes allowing its calculation have been proposed [6]. The present work pretends to show that a new theoretical scheme, producing an excellent prediction of log *P* in a series of structurally related molecules, can be formulated using MQS-SMs. This procedure is based on the direct inclusion of solvents (water and octanol in this case) into the quantum chemical calculations of electronic distribution in a molecule. A series of primary and secondary aliphatic alcohols and acetic acid esters have been used in order to test this approach. The calculations have been performed at ab initio HF level of theory using a 3-21 G* basis set within the Gaussian 94 program [38]. After determining the optimized gas phase structures for each molecule, the solvent effect was introduced using the polarized continuum model (PCM) [40, 41] incorporated in the mentioned program code. Within this approach, the molecules are placed in a cavity surrounded by a medium where the dielectric constant ϵ, the mutual polarization of this medium and the solute molecule are taken into account. Here, the values of dielectric constants ϵ=80.4 and ϵ=10.3 have been used for water and octanol respectively. Based on the calculations with included solvents, two *promolecular* ASA densities characterizing the modification of electron distributions in water $\rho_{A^W}^{\mathrm{ASA}}(\mathbf{r})$ and octanol $\rho_{A^o}^{\mathrm{ASA}}(\mathbf{r})$ have been determined. The theoretical descriptors have been constructed as an overlap-based self-similarity measure between ASA densities in water and octanol for each individual molecule. The comparison of calculated MQS-SMs and experimental log *P* values is summarized in Table 3 and presented graphically in Figure 7. As can be observed, the correlation splits into two separate lines for alcohols and esters, and a small systematic shift is in fact also observed between primary and secondary alcohols. As a result, within each class of molecules the correlation is indeed excellent. This situation is very interesting since it opens the possibility of using the above similarity approach as a new theoretical scheme for the calculation of log *P* in classes of structurally related molecules. As a consequence, the calculated MQS-SMs can directly be used in QSAR instead of log *P* themselves. An example of such a use for the correlation with biological data is presented below.

Table 3. Comparison of calculated MQS-SM with experimental log P values for a series of selected molecules

Molecule	log P[a]	$Z_{A^wA^o}$[b]
Methanol	−0.77	73.69
Ethanol	−0.31	91.23
1-Propanol	0.25	108.49
1-Butanol	0.88	125.50
1-Pentanol	1.56	142.30
1-Hexanol	2.03	159.68
2-Propanol	0.05	108.30
2-Butanol	0.61	125.60
2-Pentanol	1.19	142.79
3-Pentanol	1.21	142.51
2-Hexanol	1.76	159.48
3-Hexanol	1.65	159.49
Acetic acid (AcH)	−0.17	141.36
AcH methyl ester	0.18	157.71
AcH ethyl ester	0.73	175.18
AcH propyl ester	1.24	192.41
AcH butyl ester	1.78	209.31

[a] From Hansch et al. [42].
[b] See Results and discussion.

MQS-SM and correlations with biological data

The correlation of biological data with various molecular descriptors certainly constitutes an important and widely used field of the QSAR application. Because of the practical importance of such correlations for the rational drug design and recent work related to MQSM applied to QSAR [24, 25], it seems interesting to check whether the MQS-SMs could be of any help in this effort. Generally it is possible to expect that because of greater complexity of factors responsible for the biological activity, the finding of the

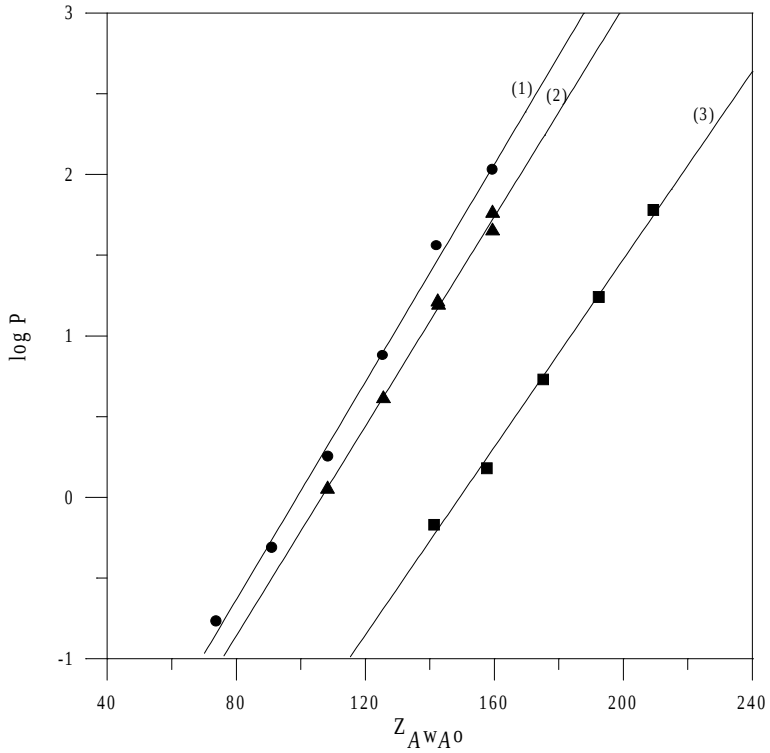


Figure 7. Dependence of calculated MQS-SM $Z_{A^wA^o}$ on log P for a series of primary (1) and secondary (2) aliphatic alcohols, and acetic acid esters (3).

appropriate molecular descriptor will be much more difficult and multiparameter correlations are often observed to be necessary for QSAR calculations. Among the descriptors which are used in such multiparameter correlations, the Hammett σ constants and log *P* are usually found. The previous demonstration that MQS-SM can successfully be used as descriptors of both the substituent effect and log *P*, has led to attempting the application of these new theoretical descriptors in QSAR as well. As a first example, the reported study of antibacterial and antifungal activity of substituted phenyl-isothiocyanates on *Escherichia coli* has been chosen. For this system the correlation of log ED_{50} with Hammett σ constants was early reported [43]. This has been interpreted as indicating that the toxicity of these molecules is apparently due to the reaction of the active compound with the target in the cells. A plausible candidate for such a process is the nucleophilic addition of the reactive groups in the protein to the isothiocyanate group [43]. The quantum molecular similarity analysis of this process can start from the fact that the carrier of the biological activity is in this case the isothiocyanate (NCS) group. In view of what has been previously discussed, this suggests that the appropriate theoretical descriptor could be the self-similarity measure Z_{XX}^{NCS} calculated in a series of substituted phenyl-isothiocyanates just for the active NCS group. The quantum chemical calculations, serving to generate the necessary density matrices, have been again performed by the semiempirical AM1 method and the resulting correlation of the MQS-SM with the Hammett substituent constants is depicted in Figure 8. As can be seen, the correlation is again very satisfactory so that the ability of the MQS-SMs to act as descriptors of the substituent effect is also confirmed in this case.

Having demonstrated the ability of MQS-SMs to act as descriptors of the substituent effect, in order to complete this study, one can try to find an example of the MQS-SMs application as a descriptor of log *P* in biological QSAR. The next example is related to the reported data on the biological activity of a broad class of molecules on the narcosis of tadpoles [44], which has been shown to be determined primarily by the log *P* [6]. The class of the studied molecules included the series of aliphatic alcohols and acetic acid

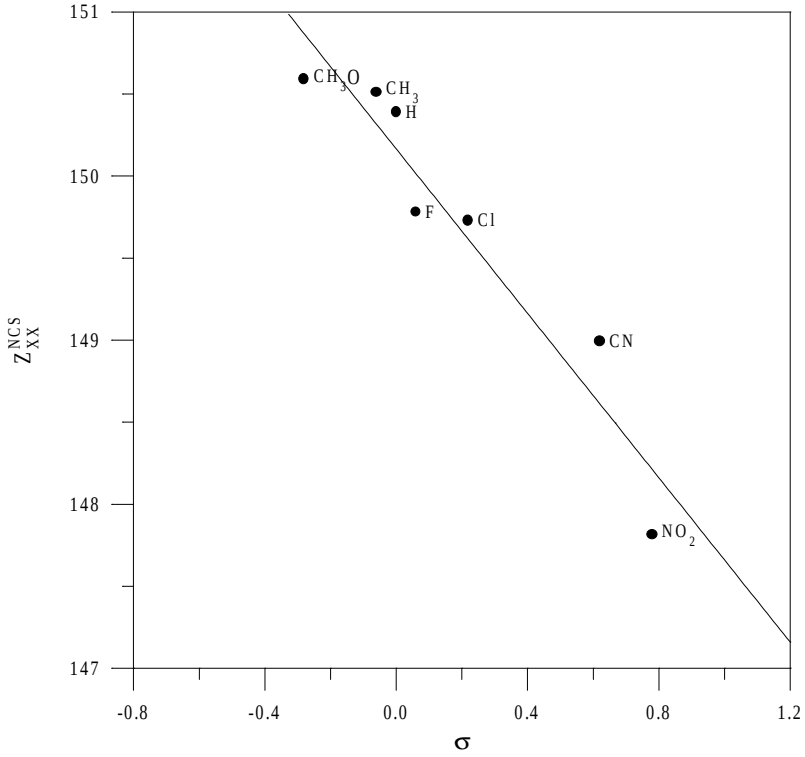


Figure 8. Dependence of calculated MQS-SM Z^{NCS}_{XX} for a series of substituted phenyl-isothiocyanate acids on the Hammett σ constants.

esters which have also been studied in the correlation of log *P* values. In this way, the previously calculated MQS-SMs can also be re-used for the correlation with the biological data on narcosis. A comparison of experimental and calculated biological activities is summarized in Table 4. It is possible to observe in this case how the agreement between the experimental activities and MQS-SMs as theoretical descriptors of log *P* again appears to be excellent.

Table 4. Comparison of experimental and calculated biological activities of aliphatic alcohols and acetic acid esters for the narcosis of tadpoles

Molecule	log 1/C	
	Obs[a]	Calc[b]
Methanol	0.24	0.19
Ethanol	0.54	0.58
1-Propanol	0.96	0.97
1-Butanol	1.42	1.35
2-Propanol	0.89	0.96
AcH methyl ester	1.10	1.11
AcH ethyl ester	1.52	1.52
AcH propyl ester	1.96	1.93
AcH butyl ester	2.30	2.32

[a] From ref. 44.
[b] Calculated using the equations: log 1/C = 0.0225 $Z_{A^wA^o}$ − 1.4713 for aliphatic alcohols, and log 1/C = 0.0235 $Z_{A^wA^o}$ − 2.5942 for acetic acid esters.

Conclusions

The present work complements previous studies of MQSM applied to determine QSPR, and confirms that this procedure appears as a useful tool for this purpose. The results in this paper encourage the consideration that in appropriate cases MQS-SM are able to provide an ab initio model to determine sound QSPR, as well as a plausible quantum mechanical justification background for such structure-properties linear relationships. In addition, the substitution of Hammett σ constants or log *P* values by MQS-SM, due to the universal computational structure of these new molecular descriptors, opens broad horizons in QSPR studies, and permits other kinds of relationships to be attained, just as the presented correlation example related to pK_a shows. Also, according to the obtained calculation pattern, significant correlations may be estimated

using MQS-SM for a vast number of biological series of compounds.

Acknowledgements

This work has been performed during the stay of one of us (R. Ponec) at the University of Girona, which has been made possible by a NATO grant. Thanks are also due to the Centre de Supercomputació de Catalunya (CESCA) and the Centre Europeu de Paral·lelisme de Barcelona (CEPBA) for a generous amount of computation time. This research has been partially supported by a CICYT grant: SAF 96-0158 and by the Fundació Maria Francisca de Roviralta. The authors gratefully acknowledge all these sources of support.

References

1. Hammett, L. P., Trans. Faraday Soc., 34 (1938) 96.
2. Taft, R. W., J. Am. Chem. Soc., 75 (1953) 4231.
3. Shorter, J., Chem. Brit., 5 (1969) 269.
4. Kubinyi, H. (Ed.) 3D QSAR in Drug Design: Theory Methods and Applications, ESCOM, Leiden, 1993.
5. Dean, P. M. (Ed.) Molecular Similarity in Drug Design, Blackie Academic & Professional, London, 1995.
6. Hansch, C. and Leo, A., Exploring QSAR. Fundamentals and Applications in Chemistry and Biology, ACS Professional Reference Book, Washington, DC, 1995.
7. Ponec, R. and Chvalovský, V., Collect. Czech. Chem. Commun., 39 (1974) 3091.
8. Ponec, R., Collect. Czech. Chem. Commun., 45 (1980) 1646.
9. Krygowski, T. M. and Perjessy, A., Bull. Acad. Sci. Polon., 22 (1974) 437.
10. Carbó, R., Leyda, L. and Arnau, M., Int. J. Quantum Chem., 17 (1980) 1185.
11. Carbó, R. (Ed.) Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches, Kluwer, Amsterdam, 1995.
12. Carbó-Dorca, R. and Mezey, P. G. (Eds.) Advances in Molecular Similarity, JAI Press Inc., Greenwich, CT, 1996, Vol. 1.
13. Cooper, D. L. and Allan, N. L., J. Comput.-Aided Mol. Design, 3 (1989) 253.
14. Cooper, D. L. and Allan, N. L., J. Am. Chem. Soc., 114 (1992) 4773.
15. Cioslowski, J. and Fleischmann, E. D., J. Am. Chem. Soc., 113 (1991) 64.
16. Cioslowski, J. and Nanayakkara, A., J. Am. Chem. Soc., 115 (1993) 11213.
17. Burt, C., Richards, W. G. and Huxley, P., J. Comput. Chem., 10 (1990) 1139.
18. Good, A. C., So, S. S. and Richards, W. G., J. Med. Chem., 36 (1993) 433.
19. Ponec, R. and Strnad, M., Int. J. Quantum Chem., 42 (1992) 501.
20. Ponec, R., J. Chem. Inf. Comput. Sci., 33 (1993) 805.
21. Mezey, P. G., J. Chem. Inf. Comput. Sci., 32 (1992) 650.
22. Luo, X. and Mezey, P. G., Int. J. Quantum Chem., 41 (1992) 557.
23. Carbó, R., Besalú, E., Amat, L. and Fradera, X., J. Math. Chem., 18 (1995) 237.
24. Fradera, X., Amat, L., Besalú, E. and Carbó-Dorca, R., Quant. Struct.-Act. Relat., 16 (1997) 25.
25. Lobato, M., Amat, L., Besalú, E. and Carbó-Dorca, R., Quant. Struct.-Act. Relat., 16 (1997) 465.
26. von Neumann, J., Mathematical Foundations of Quantum Mechanics, Princeton University Press, Princeton, NJ, 1955.
27. Bohm, D., Quantum Theory, Dover Pub. Inc., New York, NY, 1989.
28. Constans, P., Amat, L. and Carbó-Dorca, R., J. Comput. Chem., 18 (1997) 826.
29. Constans, P. and Carbó, R., J. Chem. Inf. Comput. Sci., 35 (1995) 1046.
30. Amat, L. and Carbó-Dorca, R., J. Comput. Chem., 18 (1997) 2023.
31. Carbó-Dorca, R., J. Mat. Chem., 22 (1997) 143; 23 (1998) 353; 23 (1998) 365.
32. ASA coefficients and exponents can be seen and downloaded from the WWW site: http://iqc.udg.es/cat/similarity/ASA/funcset.html
33. See for example: Saunders, V. R., Computational Techniques in Quantum Chemistry and Molecular Physics, D. Reidel Publ. Co., Dordrecht, Holland, 1975, pp. 347–424.
34. Dewar, M. J. S., Zoebisch, E. G., Healy, E. F. and Stewart, J. J. P., J. Am. Chem. Soc., 107 (1985) 3902.
35. Stewart, J. J. P. MOPAC6, QCPE 455, Indiana University, Bloomington, IN, 1993.
36. Exner, O. and Simon, W., Collect. Czech. Chem. Commun., 29 (1964) 2016.
37. Fringuelli, F., Mario, G. and Taticchi, A., J. Chem. Soc. Perkin Trans. II (1972) 1738.
38. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Gill, P. M. W., Johnson, B. G., Robb, M. A., Cheeseman, J. R., Keith, T. A., Petersson, G. A., Montgomery, J. A., Raghavachari, K., Al-Laham, M. A., Zakrzewski, V. G., Ortiz, J. V., Foresman, J. B., Cioslowski, J., Stefanov, B. B., Nanayakkara, A., Challacombe, M., Peng, C. Y., Ayala, P. Y., Chen, W., Wong, M. W., Andres, J. L., Replogle, E. S., Gomperts, R., Martin, R. L., Fox, D. J., Binkley, H. S., Defrees, D. J., Baker, H., Stewart, J. J. P., Head-Gordon, M., Gonzalez, C. and Pople, J. A., GAUSSIAN 94, Revision A.1, Gaussian, Inc.: Pittsburgh, PA, 1995.
39. Kirkwood, J. G., J. Chem. Phys., 2 (1934) 351.
40. Miertus, S., Scrocco, E. and Tomasi, J., Chem. Phys., 55 (1981) 117.
41. Miertus, S. and Tomasi, J., Chem. Phys., 65 (1982) 239.
42. Hansch, C., Leo, A. and Hoekman, D., Exploring QSAR. Hydrophobic, Electronic, and Steric Constants, ACS Professional Reference Book, Washington, DC, 1995.
43. Vlachová, D. and Drobnica, L., Collect. Czech. Chem. Commun., 31 (1966) 997.
44. Lipnick, R.L. In Suter II, G. W. and Lewis, M. A. (Eds.) Aquatic Toxicology and Environmental Fate, vol. II, ASTM, 1989, p. 468.