Journal of Computer-Aided Molecular Design, **14:** 477–485, 2000.
KLUWER/ESCOM

Use of electron-electron repulsion energy as a molecular descriptor in QSAR and QSPR studies

Xavier Gironés, Lluís Amat, David Robert & Ramon Carbó-Dorca
Institute of Computational Chemistry, University of Girona, Campus Montilivi, E-17071 Girona, Spain

Received 29 July 1999; Accepted 7 January 2000

Key words: electron-electron repulsion energy, molecular descriptor, molecular quantum similarity measures, QSAR, QSPR

Summary

Electron-electron repulsion energy ($\langle V_{ee} \rangle$) is presented as a new molecular descriptor to be employed in QSAR and QSPR studies. Here it is shown that this electronic energy parameter is connected to molecular quantum similarity measures (MQSM), and as a consequence can be considered as a complement to steric and electronic parameters in description of molecular properties and biological responses of organic compounds. The present strategy considers the molecule as a whole, thus there is no need to employ contributions of isolated fragments as in many calculations of molecular descriptors, like log *P* or the Free–Wilson analysis. The procedure has been tested in a widespread set of molecules: alcohols, alkanamides, indole derivatives and 1-alkylimidazoles. Molecular properties, as well as toxicity, are correlated using $\langle V_{ee} \rangle$ as a parameter, and extensions to the method are given for handling difficult systems. In almost all studied cases, satisfactory linear relationships were finally obtained.

Introduction

Due to continuous development of experimental chemistry, a great amount of new compounds are synthesised every year. The major part of these compounds are not tested for fundamental or relevant thermodynamic and physicochemical properties or biological activities, which are to remain unknown due to unavailability or costs of experimental methods. A procedure able to predict, within a reasonable error margin, molecular properties and biological activities for untested compounds is required to evaluate these molecular features in a fast and inexpensive way.

In order to provide a solution to the above mentioned problem, quantitative structure–activity relationships (QSAR) were defined. These models evaluate the dependence of any biological property under study with known physicochemical magnitudes, like in Hansch or the Free–Wilson analysis, based on purely molecular structural features [1–15]. Present applications are due to the linear free-energy model proposed by Hansch and Fujita [1] in 1964, based on the octanol/water partition coefficient, namely log *P*, or alternatively on the lipophilicity, π, parameters [1, 3, 6, 9]. Independently, Free and Wilson [2] developed a model based on addition of molecular substructures, and further improvements to QSAR came from combinations of both methods [3, 4, 9]. The major assumption in this kind of QSAR models is the capacity to relate biological activities, commonly on a logarithmic scale, to some addition of molecular features yielding to free-energy models. In this framework, several descriptors, briefly discussed below, satisfying this condition have been developed in addition to log *P*, π, and group contribution, such as topological [16–19], geometrical, polar [3–6, 9], electronic [20–22] or electrostatic [23] and steric [3–6, 9] parameters.

Topological parameters are intrinsically connected to molecular structures, since they are based on molecular connection matrices (see, for example, [16–19, 24–29]). Geometrical descriptors lead to the idea that molecular activity can be related to molecular

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

shape and size. Some of the most commonly used geometrical descriptors are molecular volume [30, 31], molecular surface area [32–37] or charged partial surface area [38, 39]. Polar parameters result from computing interactions that arise from dispersion forces, and usually are described by molar refractivity [3–6, 9]. Electronic parameters [20] are based on electronic interactions, as obtained from experimental data, semiempirical or quantum chemistry computations. Electronic descriptors, reflecting, for example, the electrophilic/nucleophilic molecular character (see [20–22] for a more extensive review) can be related to isolated molecules, like Hammett sigma (σ) [40–42]; or to intermolecular interactions, as HOMO/LUMO energy gaps, which are computed from molecular (MO) calculations [43]. Steric parameters can be used as a measure of molecular bulkiness, for example, to evidence the effect of variation of substituents on a common molecular skeleton, like Taft's E_s [44–46], or to describe the importance of substituent width or length in biological responses.

As a final review in QSAR/QSPR technique description, the recent appearance of 3D-QSAR models should be pointed out. These new methods, like CoMFA [47], MS-WHIM [48], E-State [49] and similarity methods, like Molecular Quantum Similarity Measures (MQSM) [50–62], include three-dimensional parameters which embody information about molecular conformation and stereochemistry.

MQSM can be used as an ideal computational tool to solve current chemical problems; several examples of these practical applications are related to the use of MQSM for QSAR and QSPR [63–71]. Such a proposal may be based on the recent description that QSAR/QSPR can be derived from MQSM theory [72]. From this reference, results can be inferred that MQSM provide QSAR models of a well defined theoretical foundation, and in this way, the usual statistical models obtained within MQSM represent QSAR associated to an approximate quantum mechanical relationship. The present work considers the possibility of characterising molecular properties using electron-electron repulsion energy ($\langle V_{ee} \rangle$) as a new kind of MQSM, allowing in this way the use of MQSM even if no specific software is available to compute regular MQSM.

Theoretical background

(A) *Molecular quantum similarity measures*

Arising from the theoretical foundation of MQSM [50], a general definition may be written as:

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

where $\rho_A(\mathbf{r}_1)$ and $\rho_B(\mathbf{r}_2)$ are density functions of molecules A and B, $\Omega(\mathbf{r}_1, \mathbf{r}_2)$ is a positive definite operator and $Z_{AB}(\Omega)$ is the resulting MQSM. Equation 1 may also be reordered:

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

for any non-differential operator. From here, a further step can be performed considering a molecular quantum self-similarity measure (MQS-SM), constructed as:

$$Z_{AA}(\Omega) = \int \Omega(\mathbf{r})\rho_A^2(\mathbf{r})\, \mathrm{d}\mathbf{r}, \quad (3)$$

which in turn, can be alternatively written in the form:

$$Z_{AA}^{(2)}(\omega) = \int \omega(\mathbf{R})\rho_A^2(\mathbf{R})\, \mathrm{d}\mathbf{R}, \quad (4)$$

where $\mathbf{R}$ represents a set of particle co-ordinates and $\omega(\mathbf{R})$ some function of them. On the other hand, definition (4) can be easily generated to a multiple MQS-SM:

$$Z_{AA}^{(n)}(\omega) = \int \omega(\mathbf{R})\rho_A^n(\mathbf{R})\, \mathrm{d}\mathbf{R}, \quad (5)$$

which in the most simple case, $n = 1$, becomes:

$$Z_{AA}^{(1)}(\omega) = \int \omega(\mathbf{R})\rho_A(\mathbf{R})\, \mathrm{d}\mathbf{R} \quad (6)$$

Equation 6 shows that quantum mechanical expectation values for non-differential operators can be considered as particular cases of MQS-SM, that is:

$$\langle \omega(\mathbf{R}) \rangle = Z_{AA}^{(1)}(\omega). \quad (7)$$

(B) *Coulomb and exchange operators expectation values: Electron-electron repulsion energy*

At the monoconfigurational ground closed shell discrete Hartree–Fock level, the Coulomb operator expectation value for an n-electron system can be written [73] as:

$$\langle V_{ee} \rangle = \sum_i \sum_j [2 \cdot (ii|jj) - (ij|ij)], \quad (8)$$

where $\{(ii|jj)\}$ and $\{(ij|ij)\}$ are the respective Coulomb and exchange integrals over the MO basis set.

Also, $\langle V_{ee}\rangle$ can be interpreted as a sum of similarity measures over the occupied MO set [51]. Starting from this previous point of view, $\langle V_{ee}\rangle$ can be considered an *n*-body self-similarity measure involving the positive definite Coulomb operator, $C(\mathbf{R})$, defined as follows:

$$C(\mathbf{R}) = \sum_{i>j} |\mathbf{r}_i - \mathbf{r}_j|^{-1} \wedge \mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n), \quad (9)$$

using a system density function, constructed in the usual way:

$$\rho_A(\mathbf{R}) = |\Psi_A(\mathbf{R})|^2, \quad (10)$$

where $\Psi_A(\mathbf{R})$ is the system's approximate wavefunction. So, an equivalent expression of Equation 8 appears to be written as:

$$\langle V_{ee}\rangle = \int C(\mathbf{R})\rho_A(\mathbf{R})\,\mathrm{d}\mathbf{R}. \quad (11)$$

Equation 11 has the same form as Equation 6, which, in turn, was deduced within an MQS-SM framework. In this manner, it can be said that the evaluation of Coulomb operator expectation values constitutes a calculation of a first-order MQS-SM class, and thus opens a new field for the potential use of this measure as a molecular descriptor.

Simple linear model using $\langle \mathbf{V_{ee}}\rangle$

Given a set of molecules, a molecular property can be associated to the expectation value of some unknown operator [72]. This general quantum mechanical definition can be approximated in a discrete framework constructing an $(n \times n)$ symmetric similarity matrix defined as: $\mathbf{Z}=\{Z_{AB}\}$, whose elements are built up using the MQSM between pairs of molecules as described in Equation 1. When homogeneous molecular series are studied, off-diagonal terms can be considered as constants, and a single vector that contains MQS-SM is used as the unique molecular descriptor [67, 68] for approximate molecular expectation values evaluation.

In this work, it is intended that $\langle V_{ee}\rangle$ be used as an alternative MQS-SM, and so to find a linear relationship between this parameter and the molecular properties, such as:

$$\pi_{\mathrm{I}} \cong a \cdot \langle V_{ee}\rangle_{\mathrm{I}} + b. \quad (12)$$

As presented in a previous work, MQS-SM have been proposed as an alternative to log *P* values in QSAR studies [68]. Here it is also suggested that MQS-SM based on $\langle V_{ee}\rangle$ can complement empirical parameters where steric and electronic effects play an important role. $\langle V_{ee}\rangle$ is strongly connected with the number of atoms and their type, hence $\langle V_{ee}\rangle$ is related to molecular volume. As it is based on quantum mechanical ideas, $\langle V_{ee}\rangle$ reflects the part of intramolecular interactions related to electron-electron repulsion energy.

Results and discussion

In this section it is intended that the usefulness of this new proposed methodology be proved. First of all, molecular modelling and the statistical considerations are explained. Then, an example that connects this new procedure with a previous use of MQS-SM in QSPR is presented. Finally, an assorted set of examples, involving molecular properties, biological activity and anomalous examples, is given.

(A) *Molecular modelling and statistical considerations*

The geometries of all compounds in this study have been considered in the gas phase and optimised according to the AM1 Hamiltonian using Ampac 6.01 [74]. $\langle V_{ee}\rangle$ values have been calculated with Gaussian 94 [75] within the HF/3-21G* computational level over the AM1 optimised geometry.

In addition, all values of $\langle V_{ee}\rangle$, $\{\langle V_{ee}\rangle_{\mathrm{I}}\}$ have been standardised in the statistical sense and then symbolised by: $\langle V_{ee}\rangle_{\mathrm{I}}^{(N)}$. That is, $\langle V_{ee}\rangle_{\mathrm{I}}^{(N)} = \left(\langle V_{ee}\rangle_{\mathrm{I}} - \langle \bar{V}_{ee}\rangle\right) \cdot s^{-1}$, where $\langle \bar{V}_{ee}\rangle$ is the arithmetic mean and *s* is the standard deviation of the various $\{\langle V_{ee}\rangle_{\mathrm{I}}\}$ values associated to the molecular set.

(B) *Connection with MQS-SM*

In order to illustrate the validity of the hypothesis formulated in Equation 12, a relationship between $\langle V_{ee}\rangle_{\mathrm{I}}^{(N)}$ and log *P* is presented for a set of amides. Compounds, $\{\langle V_{ee}\rangle_{\mathrm{I}}\}$ and log *P* are listed in Table 1. In addition, a cross-validation study is performed, whose results are also included in Table 1, to test the predictive capacity of the obtained relationship. In general terms, this process consists of leaving each element out of the model construction for later prediction of

Table 1. $\langle V_{ee}\rangle$ and log P values for a set of amides

Molecule	$\langle V_{ee}\rangle$	log P[a]	Cross-validated log P
Formamide	131.33	−1.51	−1.46
Acetamide	194.51	−1.26	−0.99
Propylamide	264.41	−0.66	−0.55
Butylamide	337.14	−0.21	−0.05
Pentylamide	413.94	0.33	0.49
Hexylamide	493.73	0.87	1.10
N-methylformamide	191.63	−0.97	−1.09
N-ethylformamide	258.85	−0.43	−0.63
N-propylformamide	330.24	0.11	−0.14
N-butylformamide	405.91	0.65	0.34

[a]From Reference 68.

its value [76]. The equation of the obtained model, as well as relevant statistical parameters: r^2 (coefficient of determination), s (root mean square for error between experimental and predicted values) and q^2 (predictive capacity parameter) are presented below.

$$\log\ P = 0.746 \cdot \langle V_{ee}\rangle^{(N)} - 0.308,$$

$$n = 10,\ \ r^2 = 0.955,\ \ s = 0.197,\ \ q^2 = 0.931$$

The good correlation observed, $r^2 = 0.955$, confirms the actual presence of a sound relationship between log P and $\langle V_{ee}\rangle$. This example also shows the close relation between $\langle V_{ee}\rangle$ and molecular size. As can be seen in Table 1, $\langle V_{ee}\rangle$ presents an additive behaviour, as log P does for homologous series, and thus increasing their respective values with the augmenting molecular size. In the previous study, a good correlation was also found using the overlap octanol-water MQS-SM, defined in [68], as a molecular descriptor, $r^2 = 0.954$, pointing out that both parameters contain almost the same information.

(C) *Molecular examples*

As mentioned above, three series of examples are presented. The first case consists of a relationship between $\langle V_{ee}\rangle$ and molecular properties. For this case, a set of 15 alcohols are tested for correlation with melting ($T_{m.p.}$) and boiling points ($T_{b.p.}$). The next example is related to toxicology, as a particular example of biological activity of large interest in the scientific community. Here, the relationship between $\langle V_{ee}\rangle$ and 50% inhibitory growth concentration ($-\log GC_{50}$) of *Tetrahymena pyriformis*, a ciliate often used to test ecotoxicology, is presented for a set of alkanamines.

Table 2. Values of $\langle V_{ee}\rangle$, $T^2_{m.p.}$ and $T^2_{b.p.}$ for a set of alcohols, all temperatures are given in Kelvin (K)

Alcohols	$\langle V_{ee}\rangle$	$T^2_{m.p.}$		$T^2_{b.p.}$	
		Exp.[a]	Pred.	Exp.[a]	Pred.
Methyl	81.07	30821.31	25459.32	114081.82	119433.95
Ethyl	135.21	25300.08	31081.85	123453.85	130076.48
Propyl	196.33	21626.64	36353.32	137166.53	141513.47
Butyl	263.38	33620.89	39480.91	152771.54	154125.36
Pentyl	335.28	37736.95	44289.67	168970.32	167862.25
Hexyl	411.32	52239.67	48417.68	185554.18	182536.32
Heptyl	490.92	57197.51	53868.74	202104.19	198006.25
Octyl	573.71	66388.68	59213.94	219267.43	214098.48
Nonyl	659.30	71909.79	65015.46	236643.33	230739.12
Decyl	747.48	78433.60	70883.90	254278.15	247862.33
Unidecyl	837.97	85357.47	76820.61	266421.15	265919.04
Dodecyl	930.62	88304.07	83268.95	283194.27	284113.39
Tetradecyl	1121.73	97756.28	96463.47	316023.87	322277.00
Hexadecyl	1319.77	103980.45	112079.30	368643.27	356904.69
Heptadecyl	1421.12	106902.84	121390.00	367429.95	385165.64

[a]From Reference 77.

Finally, a number of examples where no linear correlation can be found are handled. The first anomalous example consists of displacement of flunitrazepam from binding to bovine brain membrane. In this case the studied molecular set is formed by indole derivatives and it has been found that expectation value scaling solves the problem. The second difficult case is related to a set of 1-alkylimidazoles, where two biological activities are studied: aldrin epoxidation in armyworm gut ($-\log I_{50}$) and binding affinity to rat liver cytochrome P-450 ($-\log K_s$). A quadratic relationship, which is theoretically justified, is found to solve the problem.

(C.1) $\langle V_{ee}\rangle$ *vs. molecular properties*

The first example presented, in order to demonstrate the applicability of $\{\langle V_{ee}\rangle\}$ as a descriptor of molecular properties, consists of a series of 15 aliphatic alcohols. The values of $\langle V_{ee}\rangle$, $T^2_{b.p.}$ and $T^2_{m.p.}$ for the set of alcohols [77], including the predicted values for both properties, are summarised in Table 2 and the equation of the obtained model, including its statistical parameters, is presented below for $T^2_{m.p.}$:

$$T^2_{m.p.} = 27634 \cdot \langle V_{ee}\rangle^{(N)} - 63838,$$

$$n = 15,\ \ r^2 = 0.947,\ \ s = 7689.40,\ \ q^2 = 0.924$$

Table 3. Compounds, $\langle V_{ee}\rangle$, experimental and predicted $-\log GC_{50}$ for a set of amines

Molecule	$\langle V_{ee}\rangle$	$-\log GC_{50}$	
		Exp.[a]	Pred.
Isopropylamine	190.98	−0.88	−1.22
t-Butylamine	266.65	−0.90	−0.81
Propylamine	188.56	−0.71	−1.25
s-Butylamine	260.68	−0.67	−0.86
Butylamine	255.43	−0.57	−0.89
t-Amylamine	345.85	−0.70	−0.42
1-Ethyl-propylamine	336.33	−0.81	−0.46
1-Methyl-butylamine	333.63	−0.69	−0.48
Isoamylamine	333.42	−0.58	−0.49
Amylamine	324.66	−0.48	−0.54
2-Methylbutylamine	337.08	−0.48	−0.47
3-Phenyl-1-propylamine	616.32	0.28	−0.94
Hexylamine	402.83	−0.22	−0.15
4-Phenyl-1-butylamine	709.94	0.62	1.47
Heptylamine	485.12	0.21	0.25
Octylamine	568.59	0.61	0.65
Nonylamine	657.13	1.70	0.99
Decylamine	746.31	2.06	1.39
Unidecylamine	840.19	2.33	1.83
1,1-Dimethylpropargylamine	308.29	−0.91	−0.60
N-Methylpropylamine	256.91	−0.81	−0.87
N-Methylbutylamine	328.94	−0.68	−0.51
N-methylpropargylamine	224.79	−0.98	−1.03
N,N-dimethylethylamine	263.29	−0.91	−0.83

[a]From Reference 78.

and for $T^2_{b.p.}$:

$$T^2_{b.p.} = 79849 \cdot \langle V_{ee}\rangle^{(N)} - 226.40 \times 10^3,$$

$$n = 15,\ \ r^2 = 0.995,\ \ s = 6846.88,\ \ q^2 = 0.993$$

As can be seen from the results, in both cases a fairly good linear relationship is obtained when using $\langle V_{ee}\rangle$ as a single descriptor. The values of r^2 for the two properties are quite high, 0.947 and 0.995, indicating that there is a correlation present. In addition, predicted values are also quite close to experimental ones.

In this case, the relationship between $\langle V_{ee}\rangle$ and molecular size is also evidenced. As is well known, boiling and melting points for homologous series are very much influenced by the effect of molecular weight, and $\langle V_{ee}\rangle$ simulates that effect.

(C.2) $\langle V_{ee}\rangle$ *vs. toxicity*

In this example, the utility of $\langle V_{ee}\rangle$ as a molecular descriptor for toxicity is demonstrated. In this case, the studied molecular set consists of 24 alkanamides [78], from which 20 are primary, 3 are secondary and 1 tertiary. The values of $\langle V_{ee}\rangle$, experimental and predicted $-\log GC_{50}$ are presented in Table 3, and the correlation obtained for this system, as well as all relevant statistical parameters, are presented below:

$$-\log GC_{50} = 0.896 \cdot \langle V_{ee}\rangle^{(N)} - 0.174,$$

$$n = 24,\ \ r^2 = 0.885,\ \ s = 0.346,\ \ q^2 = 0.847$$

The model obtained presents an acceptable correlation between $\langle V_{ee}\rangle$ and $-\log GC_{50}$, and the predicted values are quite close to experimental ones. The previous study [78] yielded slightly improved results using $\log P$, $r^2 = 0.934$, however, it only involved a molecular subset of 20 primary amines.

This example also reveals the correspondence of $\langle V_{ee}\rangle$ with electronic effects. Several compounds form the studied molecular set with the same structural formula but different connectivity. If $\langle V_{ee}\rangle$ were only connected with molecular size, it would not be able to differentiate these compounds.

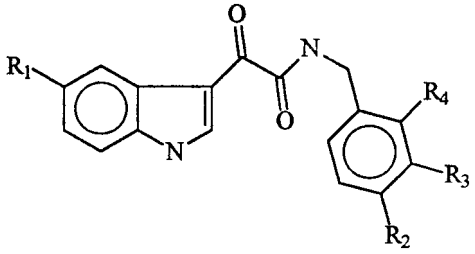
(C.3) *Extensions to the method*

In some cases, it is found that no correlation can be directly found between $\langle V_{ee}\rangle$ and any biological activity. This fact can be considered as common, since prediction of molecular activity with a single or small number of descriptors is still a hard task. However, some of these problems can be easily solved, as will be presented in this section.

(C.3.1) *Scaling of* $\langle V_{ee}\rangle$. When dealing with molecular sets presenting heavy atom substitutions, such as chlorines or bromines, the values of $\langle V_{ee}\rangle$ are highly increased with respect to the molecules that do not present such substitutions. In order to obtain a smooth distribution of $\langle V_{ee}\rangle$ values, a scaling factor is applied. This scaling factor has been chosen to be the square of the number of electrons (n_e^2). In this way, all values of $\langle V_{ee}\rangle$ are to be divided by n_e^2, and the resulting value $\langle V^{\gamma}_{ee}\rangle$, which is also standardised to $\langle V^{\gamma}_{ee}\rangle^{(N)}$, is tested for correlation.

The example presented to test this scaling procedure consists of a set of 20 indole derivatives, for which it had been found that no direct correlation between $\langle V_{ee}\rangle$ and $-\log K_{\mathrm{I}}$ is present, yielding $r^2 = 0.574$. When the normalised $\langle V^{\gamma}_{ee}\rangle$ descriptor is applied, the resulting model improves considerably, as can be seen from the statistical parameters below. All data concerning $\langle V_{ee}\rangle$, number of electrons, experimental [79] and predicted $-\log K_{\mathrm{I}}$ and the basic

Table 4. Basic molecular structure, compounds, $\langle V_{ee} \rangle$, experimental and predicted biological activity for a set of indole derivatives



Compound				$\langle V_{ee} \rangle$	n_{e^-}	$-\log K_1$	
R_1	R_2	R_3	R_4			Exp.[a]	Pred.
Cl	H	H	H	2177.44	162	6.21	5.96
NO_2	H	H	H	2209.08	168	6.93	6.72
H	OCH_3	H	H	2057.37	162	6.78	6.71
Cl	OCH_3	H	H	2465.27	178	6.68	6.81
NO_2	OCH_3	H	H	2500.32	184	7.27	7.44
H	H	OCH_3	H	2056.47	162	6.54	6.73
Cl	H	OCH_3	H	2464.26	178	6.79	6.81
NO_2	H	OCH_3	H	2499.60	184	7.42	7.42
H	OCH_3	OCH_3	H	2372.43	178	7.03	7.28
Cl	OCH_3	OCH_3	H	2790.32	194	7.52	7.35
NO_2	OCH_3	OCH_3	H	2829.30	200	7.96	7.88
H	H	H	Cl	2207.63	162	5.59	5.89
H	OH	H	H	1924.67	154	6.37	6.27
Cl	OH	H	H	2327.40	170	6.82	6.34
H	H	OH	H	1925.67	154	6.09	6.30
Cl	H	OH	H	2328.55	170	6.24	6.39
NO_2	H	OH	H	2362.13	176	7.19	7.03
H	OH	OH	H	2082.22	162	6.46	6.57
Cl	OH	OH	H	2490.17	178	6.75	6.68
NO_2	OH	OH	H	2525.53	184	7.32	7.30

[a]From Reference 79.

molecular structure are presented in Table 4.

$$-\log K_{\mathrm{I}} = -0.519 \cdot \langle V_{ee}^{\gamma} \rangle^{(N)} + 6.798,$$

$$n = 20, \;\; r^2 = 0.903, \;\; s = 0.179, \;\; q^2 = 0.879$$

The results indicate that, when $\langle V_{ee}^{\gamma} \rangle$ is applied, a fair correlation can be obtained. In this case, a correct relationship is obtained, $r^2 = 0.903$, with a considerable prediction capacity, as pointed out by the coefficient: $q^2 = 0.879$. However, the original set contained 23 molecules, obtaining a poor correlation, $r^2 = 0.603$, with a reasonable predictive capacity, $q^2 = 0.507$. The previous study [79], where Hammett's sigma was used as a single descriptor, obtained slightly poorer results; up to three molecules were also extracted from the full set, yielding $r^2 = 0.498$. Later addition of discrete descriptors depending on fragment presence increased to $r^2 = 0.728$ for one extra descriptor and $r^2 = 0.810$ for two extra descriptors, keeping a set of 20 molecules and providing no information about the predictive capacity.

In this example, the electronic effects are clearly evidenced, as the molecular size is not relevant in this biological process, but the electronic effects provided by the different substitutions at different locations are. $\langle V_{ee}^{\gamma} \rangle$ reflects these electronic effects and it is able to characterise this biological activity within the subset chosen.

(C.3.2) *Polynomial correlation using* $\langle V_{ee} \rangle$. In some cases, it can be seen by direct observation of the data that there exists a correlation between $\langle V_{ee} \rangle$ and some biological property. However, usage of a single or a small number of parameters is not enough to construct a valid linear model. The theoretical origin of this situation could be the fact that the ordinate at the origin, *b*, in Equation 12, can no longer be considered a constant. In this case, the proposed linear model can be extended to a general polynomial formulation, in order to take into account somehow the fluctuations in the off-diagonal MQSM, affecting the parameter *b*. Under these circumstances, the following equation can be used:

$$\pi_A \cong \sum_{p=1}^{N} a_p \cdot \langle V_{ee} \rangle_A^p + \beta. \tag{13}$$

This methodology is very close to the one proposed by Hansch and Fujita when they published the linear free-energy related model [1], using not only a single polynomial descriptor formulation, but many parameters, such as log *P* and Hammett's sigma, resulting in the so-called parabolic model. Further descriptor combinations and mathematical transformations, not applied here, can be performed (see, for example, [80] for assorted examples).

The example used to illustrate such a proposed situation consists of a set of 13 1-alkylimidazoles, in which two biological activities have been tested: $-\log I_{50}$ and $-\log K_S$. The results obtained for quadratic relationships are presented below with all relevant statistical parameters. All data concerning compounds, $\langle V_{ee} \rangle$, biological activities [81] and predicted values are presented in Table 5.

Table 5. Compounds, $\langle V_{ee} \rangle$, biological activities and predicted values for a set of 1-alkylimidazoles

Alkyl group	$\langle V_{ee} \rangle$	$-\log I_{50}$[a]		$-\log K_s$[b]	
		Exp.[c]	Pred.	Exp.[c]	Pred.
H	238.99	2.53	2.01	3.68	3.84
CH_3	314.30	2.90	3.20	3.70	4.25
CH_2CH_3	394.39	3.27	4.11	4.68	4.38
$(CH_2)_2CH_3$	474.88	4.85	4.71	4.94	4.68
$(CH_2)_3CH_3$	558.46	5.58	5.38	5.03	4.96
$(CH_2)_4CH_3$	643.95	6.41	5.94	5.39	5.15
$(CH_2)_5CH_3$	731.99	6.87	6.43	5.47	5.35
$(CH_2)_6CH_3$	821.99	6.94	6.86	5.56	5.49
$(CH_2)_7CH_3$	914.17	7.02	7.17	5.49	5.62
$(CH_2)_8CH_3$	1008.18	7.09	7.32	5.41	5.69
$(CH_2)_9CH_3$	1104.06	6.99	7.35	5.46	5.67
$(CH_2)_{11}CH_3$	1300.79	6.75	6.93	5.35	5.49
$(CH_2)_{13}CH_3$	1503.63	6.08	5.29	5.24	4.50

[a] Armyworm gut.
[b] Rat liver.
[c] From Reference 81.

$$-\log I_{50} = -1.034 \cdot \left(\langle V_{ee} \rangle^{(N)}\right)^2 + 1.624 \cdot \langle V_{ee} \rangle^{(N)} + 6.670,$$

$$n = 13, \;\; r^2 = 0.965, \;\; s = 0.231, \;\; q^2 = 0.931$$

$$-\log K_S = -0.399 \cdot \left(\langle V_{ee} \rangle^{(N)}\right)^2 + 0.588 \cdot \langle V_{ee} \rangle^{(N)} + 5.430,$$

$$n = 13, \;\; r^2 = 0.900, \;\; s = 0.156, \;\; q^2 = 0.745$$

In the armyworm case, it can be seen that the relationship found describes the system well, yielding $r^2 = 0.965$, and as can be observed in Table 5, the predicted values fit the experimental ones with low error, also evidenced by the predictive capacity, $q^2 = 0.931$. However, although an acceptable relationship results for the rat liver case, $r^2 = 0.900$, the obtained results are poorer, the error increases slightly in the predicted property, and as a result, the predictive capacity of the model is lower, yielding $q^2 = 0.745$. The previous study [81] obtained slightly better results, $r^2 = 0.966$ for the armyworm gut case and $r^2 = 0.923$ for the rat liver model using also a quadratic regression analysis with π as a descriptor.

As can be seen in this example, $\langle V_{ee} \rangle$ acts mainly as a descriptor of molecular size, even if the biological property at first increases and later decreases. This fact can be easily monitored by $\langle V_{ee} \rangle$ because of the negative coefficient of the quadratic term present in the correlation equation. At the same time, it can be seen that π and $\langle V_{ee} \rangle$ are fairly equivalent in this molecular set, as they achieve comparable results. Thus one can conclude that π can be safely substituted, in this case, by $\langle V_{ee} \rangle$.

Conclusions

Molecular sets composed of alcohols, alkanamides, indole derivatives and 1-alkylimidazoles have been analysed for a correlation search between experimental properties, biological activities, toxicology data and $\langle V_{ee} \rangle$, considered as an MQS-SM. Acceptable correlations have been obtained for all studied series, including those which have been handled by extensions of the method. The results indicate that MQSM, a general method based on quantum mechanical electron densities instead of the conventional atomic/fragment contribution approaches, may constitute a suitable alternative way of complementing or quantifying QSPR/QSAR in general.

Acknowledgements

This research has been partially supported by the Fundació Maria Francisca de Roviralta and an EU contract: #ENV4-CT97-0508. One of us, X.G., wishes to acknowledge the University of Girona for a predoctoral fellowship. The authors would also like to thank Ms. M. Crous for her bibliographical help supplies. Some comments by the evaluators have been helpful to ameliorate the structure of this paper, hence the authors wish to express their gratitude to them.

References

1. Hansch, C. and Fujita, T., J. Am. Chem. Soc., 86 (1964) 1616.
2. Free, S.M., Jr and Wilson, J.W., J. Med. Chem., 7 (1964) 395.
3. Ramsden, C.A. (Ed.), Comprehensive Medicinal Chemistry, Vol. 4, Pergamon, Oxford, 1990.
4. Kubinyi, H., QSAR: Hansch Analysis and Related Approaches, VCH, Weinheim, 1993.
5. Hansch, C. and Leo, A., Exploring QSAR. Fundamentals and Applications in Chemistry and Biology, American Chemical Society, Washington, DC, 1995.
6. Hansch, C., Leo, A. and Hoekman, D., Exploring QSAR. Hydrophobic, Electronic and Steric Constants, American Chemical Society, Washington, DC, 1995.
7. Van de Waterbeemd, H. (Ed.), Chemometric Methods in Molecular Design, VCH, Weinheim, 1995.
8. Van de Waterbeemd, H. (Ed.), Advanced Computer-Assisted Techniques in Drug Discovery, VCH, Weinheim, 1995.

9. Kubinyi, H., In Wolff, M.E. (Ed.) Burger's Medicinal Chemistry, 5th edn., Vol I, Wiley, New York, NY, 1995.
10. Sanz, F., Giraldo, J. and Manaut, F. (Eds), QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications, Proc. 10^{th} European Symp. on Quantitative Structure-Activity Relationships, Barcelona, 1994, Prous Science, Barcelona, 1995.
11. Pliska, V., Testa, B. and Van de Waterbeemd, H. (Eds), Lipophilicity in Drug Action and Toxicology, VCH, Weinheim, 1996.
12. Van de Waterbeemd, H. (Ed.), Structure-Activity Correlations in Drug Research, Academic, R.G. Landes Company, Austin, TX, 1996.
13. Van de Waterbeemd, H., In Wermuth, C.G. (Ed.) The Practice of Medicinal Chemistry, Academic, London, 1996.
14. Böhm, H.-J., Klebe, G. and Kubinyi, H., Wirkstoffdesign, Spektrum Akademischer Verlag, Heidelberg, 1996.
15. Van de Waterbeemd, H., Testa, B. and Folkers, G. (Eds), Computer-Assisted Lead Finding and Optimization, Proc. 11th European Symp. on Quantitative Structure-Activity Relationships, Lausanne, 1996, Verlag Helvetica Chimica Acta, Basel, and VCH, Weinheim, 1997.
16. Wiener, H., J. Am. Chem. Soc., 69 (1947) 17.
17. Hosoya, H., Bull. Chem. Soc. Jpn., 44 (1971) 2332.
18. Kier, L.B. and Hall, L.H., Molecular Connectivity in Chemistry and Drug Research, Academic, New York, NY, 1976.
19. Kier, L.B. and Hall, L.H., Molecular Connectivity in Structure-Activity Analysis, Research Studies Press, Letchworth, 1986.
20. Van de Waterbeemd, H. and Testa, B., Adv. Drug. Res., 16 (1987) 85.
21. Purcell, W.P., Bass, G.E. and Clayton, J.M., Strategy of Drug Design, Wiley, New York, NY, 1973.
22. Karelson, M., Lobanov, V.S. and Katritzky, A.R., Chem. Rev., 96 (1996) 1027.
23. Carbó, R., Martin, M. and Pons, V., Afinidad, 34 (1977) 348.
24. Brown, R.D. and Martin, Y.C., J. Chem. Inf. Comput. Sci., 36 (1996) 572.
25. Brown, R.D. and Martin, Y.C., J. Chem. Inf. Comput. Sci., 37 (1997) 1.
26. Doman, T.N., Cisulskis, J.M., Cisulskis, M.J., McCray, P.D. and Spangler, D.P., J. Chem. Inf. Comput. Sci., 36 (1996) 1195.
27. Klopman, G., Balthasar, D.M. and Rosenkranz, H.S., Environ. Toxicol. Chem., 12 (1990) 231.
28. Rosenkranz, H.S. and Klopman, G., Mutation Res., 228 (1990) 105.
29. Randic, M., J. Am. Chem. Soc., 97 (1975) 6609.
30. Bondi, A., J. Phys. Chem., 68 (1964) 441.
31. Pearlman, R.S., In Yalkowsky, S.H., Sikula, A.A. and Valvani, S.C. (Eds) Physical Chemical Properties of Drugs, Vol. 10, Marcel Dekker, New York, NY, 1980.
32. Pearlman, R.S., Quantum Chem. Prog. Exchange Bull., 1 (1981) 15.
33. Pearlman, R.S., In Dunn, W.J., Block J.H. and Pearlman, R.S. (Eds) Partition Coefficient Determination and Estimation, Pergamon, New York, NY, 1986.
34. Camilleri, P., Watts, A. and Boraston, J.A., J. Chem. Soc., Perkin. Trans. 2, (1988) 1699.
35. Lee, B. and Richards, F.M., J. Mol. Biol., 55 (1971) 379.
36. Hermann, R.B., J. Phys. Chem., 76 (1972) 2754.
37. Pearlman, R., SAREA, Quantum Chemistry Program Exchange, University of Indiana, Bloomington, IN, Program Number 432.
38. Grigoras, S., J. Comput. Chem., 11 (1990) 493.
39. Howel, J., Rossi, A., Wallace, D., Hiraki, K. and Hoffman, R., FORTICON 8, Quantum Chemistry Program Exchange, University of Indiana, Bloomington, IN, Program Number 469.
40. Hammett, L.P., J. Am. Chem. Soc., 59 (1937) 96.
41. Hammett, L.P., Physical Organic Chemistry, McGraw-Hill, New York, NY, 1940.
42. Unger, S.H. and Hansch, C., J. Med. Chem., 16 (1973) 745.
43. Streitweiser, A., Molecular Orbital Theory for Organic Chemists, Wiley, New York, NY, 1961.
44. Taft, R.W., J. Am. Chem. Soc., 86 (1968) 5175.
45. Taft, R.W., In Newman, M.S. (Ed.) Steric Effects in Organic Chemistry, Wiley, New York, NY, 1956.
46. Swain, C.G. and Lupton Jr., E.C., J. Am. Chem. Soc., 90 (1968) 4328.
47. Cramer III, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
48. Bravi, G., Gancia, E., Mascagni, P., Pegna, M., Todeschini, R. and Zalianni, A., J. Comput.-Aided Mol. Design, 11 (1997) 79.
49. Kellogg, G.E., Kier, L.B., Gaillard, P. and Hall, L.H., J. Comput.-Aided Mol. Design, 10 (1996) 513.
50. Carbó, R., Leyda, L. and Arnau, M., Int. J. Quant. Chem., 17 (1980) 1185.
51. Carbó, R. and Domingo, Ll., Int. J. Quant. Chem., 23 (1987) 517.
52. Besalú, E., Carbó, R., Mestres, J. and Solà, M., Top. Curr. Chem., 173 (1995) 31.
53. Carbó, R., Calabuig, B., Vera, L. and Besalú, E., Adv. Quant. Chem., 25 (1994) 253.
54. Carbó, R. (Ed.), Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches, Kluwer, Amsterdam, 1995.
55. Carbó-Dorca, R. and Mezey, P.G. (Eds), Advances in Molecular Similarity, Vol. 1, JAI Press, Greenwich, CT, 1996.
56. Carbó-Dorca, R. and Mezey, P.G. (Eds), Advances in Molecular Similarity, Vol. 2, JAI Press, Greenwich, CT, 1998.
57. Cioslowski, J. and Fleischmann, E.D., J. Am. Chem. Soc., 113 (1991) 64.
58. Burt, C., Richards, W.G. and Huxley, P., J. Comput. Chem., 10 (1990) 1139.
59. Mezey, P.G., Top. Curr. Chem., 173 (1995) 63.
60. Allan, N.L. and Cooper, D.L., Top. Curr. Chem., 173 (1995) 85.
61. Ponec, R., Top. Curr. Chem., 174 (1995) 1.
62. Carbó-Dorca, R. and Besalú, E., J. Mol. Struct., 451 (1998) 11.
63. Fradera, X., Amat, L., Besalú, E. and Carbó-Dorca, R., Quant. Struct.-Act. Relat., 16 (1997) 25.
64. Lobato, M., Amat, L., Besalú, E. and Carbó-Dorca, R., Quant. Struct.-Act. Relat., 16 (1997) 465.
65. Amat, L., Robert, D., Besalú, E. and Carbó-Dorca, R., J. Chem. Inf. Comput. Sci., 38 (1998) 624.
66. Robert, D., Amat, L. and Carbó-Dorca, R., J. Chem. Inf. Comput. Sci., 39 (1999) 333.
67. Ponec, R., Amat, L. and Carbó-Dorca, R., J. Comput.-Aided Mol. Design, 13 (1999) 259.
68. Amat, L., Carbó-Dorca, R. and Ponec, R., J. Comput. Chem., 19 (1998) 1575.
69. Ponec, R., Amat, L. and Carbó-Dorca, R., J. Phys. Org. Chem., 12 (1999) 447.

70. Amat, L., Carbó-Dorca, R. and Ponec, R., Simple Linear QSAR Models based on Quantum Similarity Measures, J. Med. Chem., 42 (1999) 5169.
71. Robert, D. and Carbó-Dorca, R., SAR QSAR Environ. Res., 10 (1999) 401.
72. Carbó, R., Besalú, E., Amat, L. and Fradera, X., J. Math. Chem., 18 (1995) 237.
73. Roothaan, C.C.J., Rev. Mod. Phys., 23 (1951) 69.
74. Ampac 6.01, 1994, Semichem, Schawnee, KS.
75. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Gill, P.M.W., Johnson, B.G., Robb, M.A., Cheeseman, J.R., Keith, T., 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, J.S., Defrees, D.J., Baker, J., Stewart, J.P., Head-Gordon, M., Gonzalez, C. and Pople, J.A., Gaussian-94 (Revision E.2), Gaussian, Inc., Pittsburgh, PA, 1995.
76. Allen, D.M., Technometrics, 16 (1974) 125.
77. Lide, D.R., Handbook of Chemistry and Physics, 76th Edition, CRC Press, Boca Raton, FL, 1995.
78. Sinks, G.D., Carver, T.A. and Schultz, W., SAR QSAR Envir. Res., 9 (1998) 217.
79. Hadjipavlou-Litina, D. and Hansch, C., Chem. Rev., 94 (1994) 1483.
80. Hansch, C., Kim, D., Leo, A.J., Novellino, E., Silipo, C. and Vittoria, A., CRC Crit. Rev. Toxicol., 19 (1989) 185.
81. Wilkinson, WC.F., Hetnarsky, K., Cantwell, P. and di Carlo, F. J., Biochem. Pharmacol., 23 (1974) 2377.